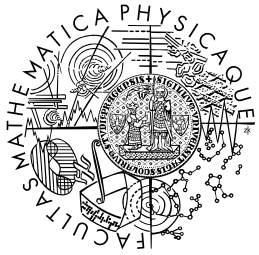

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Daragh Meehan

Learning Multiple Languages Through Parallel Audiovisual Input

Institute of Formal and Applied Linguistics

Supervisors of the master thesis:	RNDr. Jiří Hana, Ph.D. Prof. Dr. M.C. (Marije) Michel
Study programme:	Computer Science
Study branch:	Language Technologies and Computational Linguistics

Prague 2024

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

'I managed to get four copies of the Bible in different languages. So I read the Gospels, verse by verse, with the help of the little knowledge of German and French that I had acquired in school, and side by side with this a parallel reading in English and Italian. In a few months, I made excellent progress in this way.'

(Trotsky, 1930)

'Learning material of the kind you seek – parallel structures for the lessons of multiple languages – would greatly facilitate the learning of multiple languages. Not only would this suit people like you who seek to study various languages at once, but anyone who learned one language using such a method [would] have a great advantage thereafter learning another [...] since he or she would already know the content of the course. This is so obvious that one would imagine publishing houses would produce such courses, if not to actively encourage polyglottery, then simply to boost sales as they could truthfully market the ability to learn "two for one." Alas, this is not the case. There are no current methods that I know of that do this.'

(Arguelles, 2022)

Acknowledgements

First of all, I would like to thank my supervisors.

Thank you to Marije, for your patience and guidance through challenging times, your expertise, insightful feedback, kindness and enthusiasm.

Thank you to Jirka, for your passion for NLP, belief in my abilities, and your always excellent teaching.

I am grateful to the LCT Erasmus Mundus programme and the EU for providing me with such an unforgettable opportunity.

Special thanks goes to the always kind and helpful LCT coordinators Markéta Lopatková and Vladislav Kuboň at Charles University, and Tommaso Caselli at the University of Groningen, and all the LCT staff.

Thank you to Caroline for your support throughout my time in Groningen.

My research would not have been possible without my friends, new and old, for taking part in the experiment. Thank you.

Special thanks goes to Jente and Valentina for writing the subtitles of the Dutch and Italian episodes of Peppa Pig.

Special thanks also goes to Tom and Magnus for proofreading my work and providing helpful advice and criticism.

To the language learning enthusiasts on the Internet, especially the creators of the incredible sentence mining tools which served as inspiration for this project, thank you for helping me on my own language journey.

A special thank you goes to Xavier Nègre, the creator and curator of the Lexilogos website. I hope you get a chance to try out my learning tool!

Thank you Anne Marie for the regular visits and energy balls that saw me over the line.

Finally, my deepest thanks go to my friends and family, especially my parents and sister, for supporting me throughout my time as a student. You will be glad to see me finish this off.

Title: Learning Multiple Languages Through Parallel Audiovisual Input

Author: Daragh Meehan

Institute: Institute of Formal and Applied Linguistics

Supervisors: RNDr. Jiří Hana, Ph.D., Institute of Formal and Applied Linguistics
Prof. Dr. M.C. (Marije) Michel, Faculty of Arts, University of Groningen

Abstract: TV shows and movies, when dubbed and subtitled in multiple languages, offer a unique and untapped multilingual language learning opportunity. This thesis explores the use of audiovisual input (AVI) for learning multiple languages in parallel from two angles: first, a multilingual study tool which breaks down audiovisual input with multiple language audio tracks and matching subtitle files, into an interactive, multilingual text; and second, an experiment measuring the effectiveness of learning two new languages at once from the same audiovisual input.

The study tool presents the dialogue of subtitled videos in multiple languages in parallel. Learners can listen to specific subtitles, look up words and translate phrases, create multimedia flashcards to learn vocabulary, and export condensed audio files that can interleave spoken dialogues in different languages for multilingual listening practice.

The experiment employed three novel techniques: previewing a target video in learners' L1; repeating a video to learn two languages at once; dividing a video into short segments that repeat in multiple languages. 35 participants watched a short kid's TV show episode in English (the participants' first language), and two new languages, Dutch and Italian. Two viewing methods – watching the whole episode through in each language versus watching in short segments – both led to significant vocabulary gains, measured by form recognition and meaning recall, when compared to a control group.

Keywords: Audiovisual input, Multilingualism, Computer-assisted language learning, Parallel texts, Sentence mining

Contents

List of Tables

List of Figures

1. Introduction

Engaging with familiar content in a second language, whether it's revisiting a cherished childhood book series or a favourite TV show, is an enjoyable and effective way to learn another language. When a story, characters, and setting are familiar, it's easier to infer the meaning of words and phrases (Richards, 2017); the language used is much more comprehensible, leading to more effective language acquisition.

This approach is even more effective when learners can simultaneously view the same text in a language they're fluent in. It can act as a learning aid, a reference to understand new vocabulary and phrases. It allows the learner to focus on how the same ideas are expressed in a different cultural and linguistic context, while enjoying full comprehensibility of the material. For reading *Harry Potter*, this is entirely possible provided we have the same book in our first language (L1) and target language (L2) and can read them side-by-side. For watching *Friends* however, it's much harder to compare the dialogue from two versions of the same episode to learn unfamiliar vocabulary or grammar unless we are intimately familiar with the episode.

The practice of using parallel material in both a known language and a new language is not a new idea – it's been a cornerstone of language learning for millennia. In fact, some of the oldest instances of writing in history are bilingual texts that were used by translation apprentices to quickly gain proficiency in other languages (Bellos, 2012, p. 94). Romans used bilingual dialogues to learn Greek (Butzkamm and Caldwell, 2009, p. 144; Dickey and Ferri, 2012), while 'the self-translated, bilingual text was commonplace in the multilingual world of mediaeval and early modern Europe' (Hokenson and Munson, 2007). Today, this tradition continues with parallel text series like the renowned *Loeb Classical Library* and *Clay Sanskrit Library* for learners of classical languages, and modern bilingual book series intended for language learners like *Penguin's Parallel Text Series* or *Folio's Bilingue* books. Yet, despite extensive historical use, modern language learning research on bilingual texts remains scarce (Zhang and Webb, 2019).

Figure 1.1 shows a bilingual edition of *Le Petit Prince* intended for Japanese learners of French. If languages can be effectively learned with bilingual texts, why not use the same material to learn multiple languages simultaneously? A method to read *Harry Potter* in two new languages at once is shown in Figure 1.2. Multilingual texts have also been around for millennia, from ancient multilingual inscriptions (Payne, 2006) and 'polyglot' printings of the Bible which placed multiple language versions side-by-side (Giustiniani, 1516a), to *Orbis Pictus*, the first children's picture book which was reprinted in trilingual and quadrilingual editions (see Section 2.5.1). Conversation manuals and pocket dictionaries in up to 8 languages were even developed in Europe starting from the late 15th century, intended for the self-study of many languages at once by itinerant merchants (Franceschini, 2012; Hüllen, 2005). Yet today, multilingual texts are almost entirely overlooked in language learning.

Figure 1.1: A French/Japanese bilingual edition of *Le Petit Prince* ('The Little Prince'). Source: IBC対訳ライブラリー (ICB Taiyaku Library)

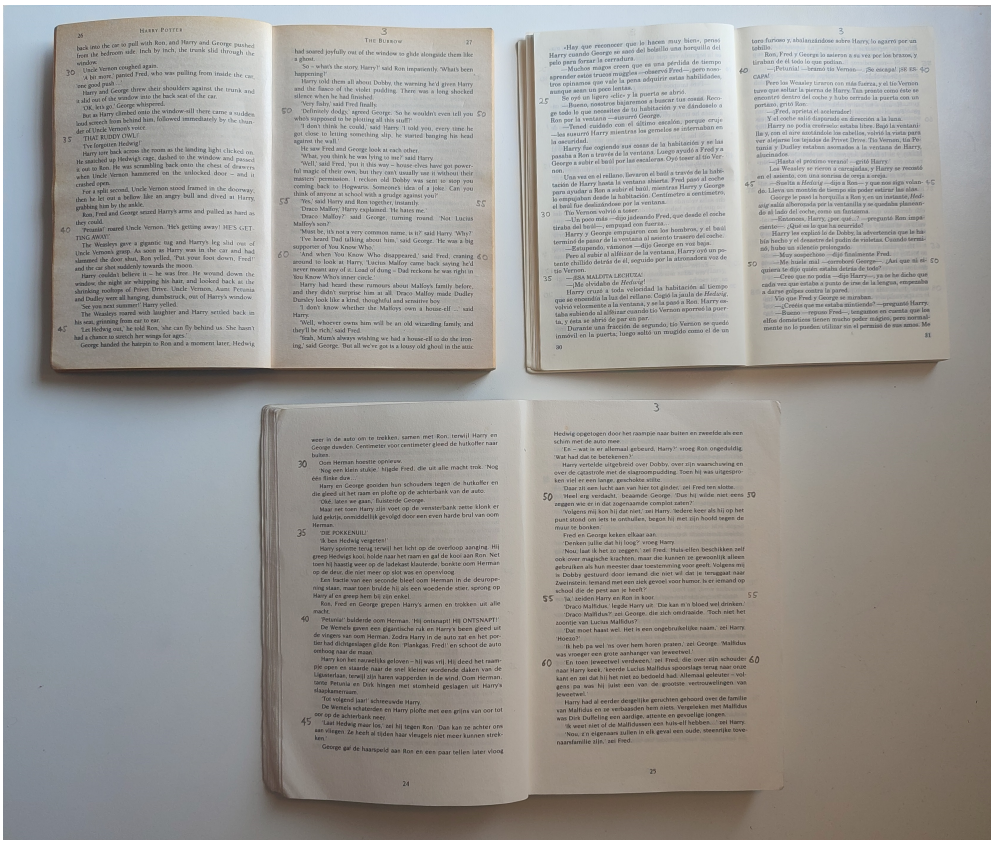

Figure 1.2: Reading *Harry Potter and the Chamber of Secrets* in three languages – English (top left), Spanish (top right), and Dutch (bottom). The custom section markings at every fifth paragraph aid the reader to match the paragraphs across the three languages. Credit: *Bloomsbury Publishing*, *Penguin Random House*, *Uitgeverij De Harmonie*

In comparison to the relatively limited collections of printed bilingual texts intended for language learners, there exists a massive multilingual set of language learning resources. These are the TV shows and movies available on streaming services like Netflix and Prime Video, which offer endless videos of entertaining native conversations, recorded in several languages with matching transcriptions.

Netflix, the industry-leading over-the-top (OTT) streaming platform, currently offers dubbed content in over 40 languages[1], with series and movies often dubbed in 4 or 5 languages. Some of Netflix's most popular films and TV series have around 20 language audio tracks[2], while newer movies and series can have over 30[3]. These multilingual resources were completely impossible to imagine before modern times. With thousands of TV shows and movies, or 'audiovisual input' (AVI), streaming platforms offer effectively unlimited language learning content—there has never been a better time for learning a second language.

Researchers in the field of second language acquisition (SLA) widely accept that input—language that learners are exposed to—is crucial to the speed and success of acquiring a second language (Li, 2017, p. 29). Recent review papers consistently demonstrate the clear benefits of watching audiovisual input for learning a second language. The additional effectiveness of L2 'captions'—on-screen transcriptions of dialogue in the target language—is also well established (Caruana, 2021; Muñoz, 2022; Montero Perez, 2022; Reynolds et al., 2022; Wei and Fan, 2022).

With endless content, easy access, and enhanced accessibility from captions, AVI meets the key criteria to take independent, informal language learners to high levels of proficiency (Vanderplank, 2019, p. 197). Long-term exposure to significant amounts of AVI can be a 'prime mover' for learning a language—it can lead to substantial language acquisition even in the absence of formal instruction (Caruana, 2021).

Language learning websites like *Lingopie*, *FluentU*, *LingQ*, and *Yabla* use AVI as their primary learning source. These platforms offer immersion in authentic and engaging language content through videos, TV shows, and movies, with support for learners through in-built dictionaries and explanations for difficult vocabulary and grammar. Alongside guided immersion in AVI, these websites incorporate digital flashcards that test learners' knowledge of words and phrases from videos they watch, systematically building and reinforcing vocabulary knowledge (Altiner, 2011; Seibert Hanson and Brown, 2020).

Vocabulary flashcards, backed by research in psychology and the study of memory (see Section 2.6.2), are also popular among self-taught language learning communities online. The lesser-known term 'sentence mining' refers to extracting sentences from authentic native materials that introduce new vocabulary or grammar, and reviewing them with a 'spaced repetition system' (SRS), a sys-

[1] as of April 2024, with Netflix accessed in Ireland

[2] *Stranger Things* has 20 dubbed languages, *Squid Game* 22, *Bridgerton* 17, *Wednesday* 19, and *Money Heist* 18

[3] *Concierge Pokémon* has 31, the new *Legend of Aang* has 35, and new *Pinocchio* movie has 36

tem for managing digital flashcards (Antimoon, 2002). Detailed in Appendix A, an online community of predominantly Japanese learners has developed sophisticated 'sentence mining' tools to quickly create multimedia flashcards from audiovisual input, thereby accelerating the language learning process.

Despite the support of captioned video in language learning research, and the proliferation of AVI-based language learning platforms, the potential of captioned AVI with multiple L2 language tracks has been overlooked. From a computational perspective, captioned AVI provides a unique opportunity to parallelise captioned dialogues across multiple languages. The timing metadata in subtitle files tells a media player when to display captions, thus linking spoken dialogue with a written transcription. By comparing subtitle timings across different subtitle files, dialogue can be matched across different language versions of the same video. Despite the rich historical tradition of parallel texts, the integration of parallel language practices into digital language tools and research with audiovisual input remains surprisingly limited.

Modern language learning platforms and tools have made language learning accessible and broadened its appeal, but none have offered the possibility to explicitly learn multiple languages in parallel from the same material. The SLA research community too has neglected the potential of multilingual AVI—from having learners preview AVI in their native language to quickly grasp content and instantly boost comprehension, like when reading bilingual texts, to reusing video content for learning several languages simultaneously.

What's missing is a multilingual approach in both research and practice. This thesis embraces this opportunity, combining the centuries-old idea of 'polyglot' parallel texts with modern software-based approaches to accelerating the process of learning a language from foreign media, like multimedia flashcards and condensed audio files for extensive listening practice, to develop an all-in-one study paradigm for learning multiple languages simultaneously using audiovisual input.

1.1 Research Questions

RQ0 – How can audiovisual content be used to study multiple languages at the same time?

This thesis is an exploratory study into the untapped potential of multilingual learning with AVI. **RQ0** serves as the core research question, which will be investigated in 3 ways – the first is theoretical (**RQ1**–**RQ2**), the second experimental (**RQ3**–**RQ6**), and the third technological (**RQ7**–**RQ9**).

1.1.1 Theoretical

RQ1 – Does language learning research support the idea of using audiovisual materials for learning multiple languages simultaneously?

RQ2 – What is sentence mining and how does it enhance learning from AVI?

1.1.2 Experimental

RQ3 – What kinds of language learning experiments are possible with multilingual AVI?

RQ4 – Is it possible to acquire vocabulary in two languages at once using captioned AVI repeated in multiple languages? If so, is it more effective to watch AVI the whole way through in each language being studied, or in short segments that repeat in each language?

RQ5 – Do the same factors that influence learning vocabulary from captioned AVI apply when learning two languages at once? In particular, are words with higher frequency of occurrence learned better, and does more prior vocabulary knowledge lead to greater vocabulary learning?

RQ6 – What are learners' reaction to viewing AVI in multiple languages?

1.1.3 Technological

RQ7 – What kinds of parallel learning materials can be created programmatically using AVI with multiple language audio tracks and corresponding subtitle files?

RQ8 – How can a language learning computer program facilitate the parallel study of multiple languages using audiovisual input? How can AVI tools and sentence mining practices be expanded to multilingual study?

RQ9 – What kinds of multimedia flashcards are possible to create from multilingual AVI?

1.2 Thesis Outline

RQ1 will be answered with a review of second-language acquisition (SLA), language learning with audiovisual input (AVI), computer-assisted language learning (CALL), multilingualism, parallel texts, and a look at the concepts of spaced repetition and interleaving (see Section 2).

RQ2 addresses the lack of literature investigating the field of 'sentence mining', a recent method for learning languages which focuses on extensive AVI input, and reviewing sentences just beyond the comprehension of the learner taken from authentic native materials using cutting-edge language learning tools. Sentence mining is introduced in Section 2.3.2.

RQ3 explores the fundamental experiments that emerge when videos can be viewed multiple times in both learners' L1 and more than one L2, as no multilingual AVI experiments of this nature have been conducted previously. These experiments were considered for this study but not performed. They are detailed in Section 3.1.

RQ4–**RQ6** will be addressed by an exploratory experiment in Section 3, which measures the incidental vocabulary acquisition—learning words as a by-product of engaging in an activity not explicitly focused on vocabulary learning—of complete

beginners to Dutch and Italian from watching the same audiovisual input, an episode of Peppa Pig, repeated in multiple languages. The study compares two viewing methods – watching the whole episode through, first in English (the learners' L1), then in Dutch, and finally Italian, versus watching in short segments that repeat in each language (see Figure 1.3). Both methods led to significant vocabulary gains, measured by form recognition (the ability to recognise the written form of a word) and meaning recall (the ability to remember the meaning of a word), in both languages when compared to a control group who only took the vocabulary tests. Always showing new material first in English as a way to increase comprehensibility is akin to the technique of pre-reading target texts in learners' first languages explored in bilingual text research (Zhang and Webb, 2019). Use of this technique is a first in AVI research.

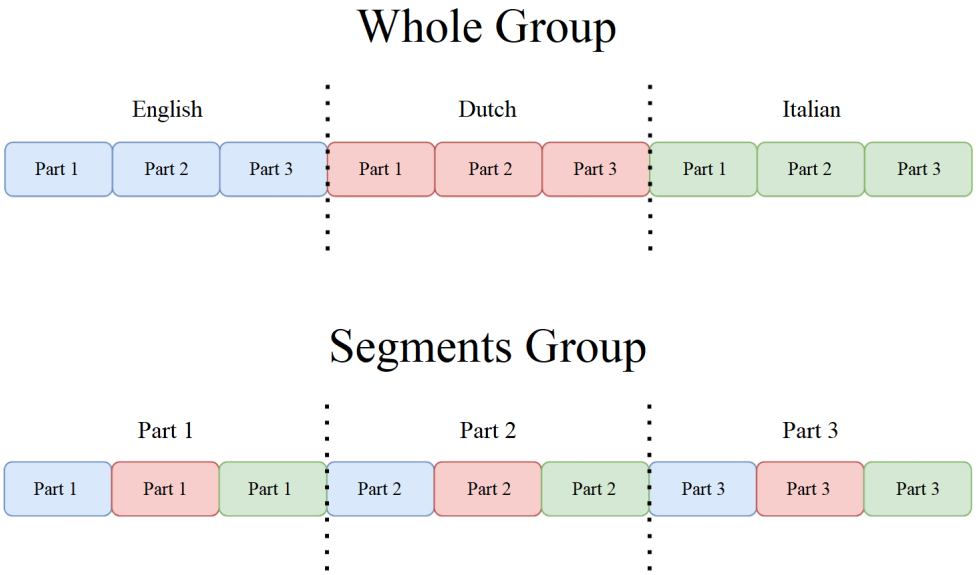

Figure 1.3: The 'Whole Group' of the language learning experiment watches the whole episode through in English, then Dutch, then Italian. The 'Segments Group' watches short segments of the show repeated in each language. This is a simplified diagram – there were 16 total segments in the video shown to the participants.

Finally, **RQ7**–**RQ9** investigate the development of a new tool for learning multiple languages at the same time using audiovisual content (see Section 4). An overview of the system is shown in Figure 1.4. The system processes videos with multiple audio tracks and matching subtitle files in several languages and creates an interactive multilingual text by aligning the captions of each language, building a parallel text in 2, 3, or more languages. While reading, learners can relisten to sections of the dialogue, and can easily check the meaning of words and phrases with in-app access to machine translation and a range of dictionaries. Additionally, inspired by sentence mining tools, users can create custom multimedia flashcards for long-term retention of vocabulary and phrases, with screenshots, audio, and text taken from any point in the video. Finally, learners can generate custom audio files from AVI for later listening practice, which condense the input by removing all non-speaking audio using subtitle timings, with the option of interleaving different language versions of the same dialogue

segments like the video shown to the 'Segments Group' in the experiment. We will examine the design considerations and computational challenges in developing the system, which constitutes a complete language learning suite of tools for the simultaneous study of several languages from any AVI input.

Figure 1.4: An overview of the language learning computer program. AVI in multiple languages is broken down into a multilingual text, and users can create condensed audio files and multimedia flashcards.

2. Theoretical Background

To answer **RQ1**, this section summarises and synthesises the research relevant to extending AVI-based language learning to a multilingual setting. This section also demonstrates the existing gaps in language research and language learning tools today, motivating the design of both the language learning experiment and computer program.

RQ2 will be answered with an introduction to sentence mining and flashcards (Section 2.3.2), a summary of advanced sentence mining tools (Appendix A), and an exploration of the evidence behind spaced repetition (Section 2.6.2).

2.1 Second Language Acquisition

The field of second language acquisition (SLA) studies how people learn a second language, or L2, after they have acquired their native language(s). Learning how a second language is acquired can shed light on how the brain works, and inform best practices in language education. Generally regarded as a discipline in the field of applied linguistics, SLA can draw on and influence fields such as education, psychology, cognitive science, neuroscience, and sociology.

The majority of research in SLA is driven by a number of fundamental questions (VanPatten and Smith, 2019):

- Is L2 (second language) acquisition different from L1 (first language) acquisition? This question arguably lies at the core of SLA research.
- What is the nature of development of a second language?
- What roles do input and output play?
- Can L2 learners achieve a native-like proficiency?
- What difference does instruction make?
- How do individual differences, like working memory and aptitude, affect acquisition?

2.1.1 Vocabulary

Building vocabulary is a key part of learning a language. There are many aspects to learning a word such as form, meaning, and use (Nation, 1998). Vocabulary learning can be incidental, where words are learned 'as a by-product of a task' (Ellis, 1999) like reading or listening, or intentional, where learners focus on deliberately learning words (Webb and Nation, 2017). Vocabulary knowledge can be receptive, knowing what a word means when listening or reading, or productive, being able to use a word when speaking or writing (Schmitt, 2014).

Another key question of SLA is how to measure acquisition. A wide variety of techniques exist for collecting data that measure learning (see VanPatten and Smith, 2019). Given that learning and retaining new words are essential goals

for language learners (Hao et al., 2021), measuring vocabulary acquisition is a standard measure of learning in SLA research.

Vocabulary tests often measure receptive and productive knowledge of form and meaning (Montero Perez, 2022; Webb et al., 2023):

- **Form recognition** (receptive) – e.g. do you recognise this word?
- **Form recall** (productive) – e.g. provide an L2 translation of an L1 word.
- **Meaning recognition** (receptive) – e.g. match L2 forms to L1 translations.
- **Meaning recall** (productive) – e.g. provide an L1 translation of an L2 word.

The language learning experiment in Section 3 measures incidental vocabulary acquisition by testing form recognition and meaning recall.

The most effective approaches to vocabulary learning 'include both an explicit, intentional learning component and a component based around maximising exposure and incidental learning' (Schmitt, 2008). In SLA, exposure to language is called input.

2.1.2 Role of Input

Input is of crucial importance in language acquisition: 'If learners do not receive exposure to the target language they cannot acquire it. In general, the more exposure they receive, the more and the faster they will learn' (Ellis, 2005, p. 38). In particular, exposure to 'authentic input that focuses on meaning' is arguably 'the most influential factor contributing to the speed and success of language learning' (Li, 2017, p. 29).

The five hypotheses of SLA pioneer Stephen Krashen's 'Monitor Model' have had great influence on the field of SLA. The key hypothesis, the 'Input Hypothesis', put forward the idea that language acquisition happens when a learner is exposed to 'comprehensible input', input which is just above the learner's current level, which he described as 'i + 1', where 'i' is the learner's current linguistic competence (Krashen, 1982). Today, 'many teachers and researchers believe comprehensible input is the primary determinant in SLA' (Li, 2017, p. 39).

Learning vocabulary gradually from input leads to a deeper understanding of meaning and use than vocabulary learning in isolation (Godwin-Jones, 2010). However, deliberate vocabulary learning alongside incidental learning from L2 input is more effective than relying solely on incidental learning (Nation, 2013, p. 437). The practice of 'sentence mining' combines extensive L2 input with intentional vocabulary learning using 'flashcards', or 'word cards', an easy, highly effective vocabulary learning technique (further discussed in Section 2.3.2). Sentence mining uses the context in which new vocabulary is encountered, in particular 'i + 1' sentences, for learning words intentionally (Sharp, 2010; Skritter, 2012; Zuo, 2021), while further exposure to studied vocabulary during 'immersion' in L2 input leads to deeper vocabulary knowledge and overall language acquisition.

2.2 Audiovisual Input

Audiovisual input (AVI), which refers to TV shows, movies, internet videos, is an excellent form of input for language learners (Caruana, 2021; Montero Perez and Rodgers, 2019). Watching L2 TV shows and movies is an enjoyable language learning activity (Wang, 2012) that provides large amounts of spoken input (Webb, 2015) and 'present[s] language in its cultural context' (Garrett, 1991). The 'richness, range, and variety of language available in television and films' is something 'no teacher, no classroom, no textbook can provide' (Vanderplank, 2019). It is close to the 'optimal' input for language acquisition, as it can be 'comprehensible, authentic, limitless, engaging, adaptable and relevant' (Muñoz, 2022).

Captions are on-screen text that provide a verbatim transcription of the language being spoken in L2 videos. Captions can be differentiated from subtitles, which provide an L1 translation of L2 audio (Montero Perez, 2022) – captions can be described as intralingual and subtitles as interlingual. These terms are often used interchangeably. There also exist a wide variety of additional caption types in AVI research like keyword, glossed, annotated, and enhanced captions (see Montero Perez, 2022; Wei and Fan, 2022).

Captions were identified in academia as an 'untapped resource' to improve the effectiveness of language learning using AVI over four decades ago (Price, 1983), although their potential was recognised before that (Lambert et al., 1981). Captions provide an 'orthographic, phonological and semantic foothold' into entertaining L2 material (Bird, 2005). The combination of listening to native dialogue while reading a written transcription helps learners develop both reading and listening skills (Lindgren and Muñoz, 2013). Watching captioned AVI has been shown to develop L2 vocabulary, grammar, listening comprehension, and pronunciation (Montero Perez, 2022). For learning vocabulary, captioned AVI is more effective than AVI with subtitles or without captions (Reynolds et al., 2022; Vanderplank, 2019).

Learning languages through captioned AVI can be considered a form of 'multimedia learning'—learning from 'both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video)' (Mayer, 2014). The principle of multimedia learning is that 'people learn better from words and pictures than from words alone' (Mayer, 2020), which builds off of Paivio's dual-coding theory (Paivio, 1986), a theory of cognition that suggests humans process information through two 'channels' – verbal and nonverbal. Multimedia learning posits that learning can be enhanced when both visual and verbal 'channels' are activated at the same time – 'the redundancy triggered by the matching auditory and visual stimuli functions as a catalyst to learning' (Maranzana, 2022). This helps explain why adding captions to L2 videos is so effective, and supports the use of images and audio in the multimedia flashcards produced by the computer program in Section 4.

Figure 2.1 shows a snippet from Karen Price's pioneering 1983 study on language learning using captioned video (Price, 1983). She noted that increasing the comprehensibility of a video should facilitate language acquisition. Many studies since

then have investigated how to improve learners' comprehension of L2 videos to boost language acquisition, as higher comprehension should lead to more learning (see Montero Perez, 2022). Surprisingly, no AVI studies have tested the effect of first showing a video in learners' L1 to guarantee full comprehension before presenting it in a target language.

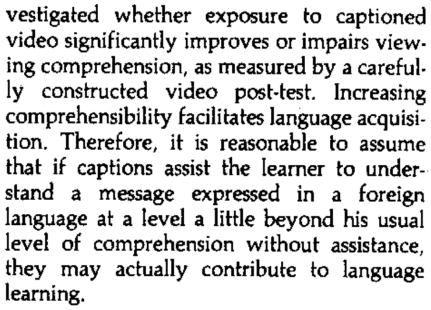
vestigated whether exposure to captioned video significantly improves or impairs viewing comprehension, as measured by a carefully constructed video post-test. Increasing comprehensibility facilitates language acquisition. Therefore, it is reasonable to assume that if captions assist the learner to understand a message expressed in a foreign language at a level a little beyond his usual level of comprehension without assistance, they may actually contribute to language learning.

Figure 2.1: An extract from Karen Price's 1983 study on the potential of captions for learning from L2 videos.

The use of audiovisual content for learning multiple languages simultaneously has also not been explored. Prior knowledge of the content of L2 input leads to improved incidental vocabulary acquisition (Huckin and Coady, 1999). Interesting and familiar content improves L2 vocabulary gains from watching TV shows (Wang, 2012), while the repeated viewing of episodes can improve comprehension and thus enjoyment and motivation from watching AVI (Webb, 2015). With TV shows and movies available in multiple languages with matching subtitles, it seems natural to both preview content in one's L1 for full comprehension, and reuse content for learning multiple languages. This study addresses this gap by exploring how these techniques can facilitate language acquisition across multiple languages.

In captioned AVI research, the most common measures of learning are comprehension and vocabulary, with some work investigating grammar, listening, and pronunciation (Montero Perez, 2022). In the study presented in Section 3, having learners preview the material in their L1 means testing comprehension doesn't make sense – they already have complete comprehension of the material. Instead we measure vocabulary acquisition for two main reasons: one, vocabulary acquisition has been studied relatively extensively in AVI research (see Montero Perez, 2022) and is a central tenet of language learning; two, the learning of vocabulary/phrases using flashcards is a key theme of this thesis and the 'sentence mining' approach to language learning (Antimoon, 2002), and one of two outputs from the language learning program of Section 4.

2.3 Computer-Assisted Language Learning

The field of computer-assisted language learning (CALL) can be defined as 'the search for and study of applications of the computer in language teaching and learning' (Levy, 1997), or 'any process in which a learner uses a computer and, as a result, improves [their] language' (Beatty, 2010). Important related terms

are MALL (mobile-assisted language learning) and TELL (technology-enhanced language learning).

In a meta-analysis of technology-supported language teaching methods from 1970 to 2006, it was found that technology-supported pedagogy was at least as effective as teaching without technology, and was often more effective (Grgurović et al., 2013). A lot has changed since 2006—the development of smartphones, social media, language learning apps, video platforms, and advances in AI and language technology. Language learning technologies have moved from a supplementary role to being central in language learning, enabling authentic language learning experiences anywhere, anytime (Otto, 2017).

2.3.1 Informal Language Learning and Extensive Language Input

Opportunities for learning through the Internet have 'transformed the scene for informal language learning, enabling significant numbers of independent, informal learners in foreign language contexts to achieve high levels of proficiency' (Cole and Vanderplank, 2016). The Internet, with its ample learning resources and potential for interaction with native speakers, supports self-directed language learners in creating 'highly personalised learning contexts', making language learning accessible at any time (Nishioka, 2023).

Smartphones in particular epitomise this new age of portable and accessible language learning. They offer 'on-the-move learning' (Hao et al., 2021) and are 'ideal for individualised informal learning' (Godwin-Jones, 2011); they can boost motivation and make learning more effective and entertaining (Huang et al., 2016; Jeong, 2022). Smartphones offer personalised, convenient, portable, flexible, self-directed, spontaneous, and informal language learning experiences (Jeong, 2022; Miangah, 2012).

One of the greatest advantages of the Internet age is the possibility for 'informal exposure to massive amounts of personally relevant comprehensible input' (Cole and Vanderplank, 2016). Today language learners can, for example, watch L2 videos about any topic they are interested in, any time of day. As long as L2 learners get 'quality input in sufficient quantities', they can reach a proficiency where they are rated as native-like by native speakers (VanPatten and Smith, 2019). Evidence has shown that, while classroom learners plateau at an upper-intermediate level, fully autonomous self-instructed learners (FASILs) can advance to 'native-like level of knowledge and use' (Cole, 2015).

2.3.2 Flashcards and 'Sentence Mining'

One of the greatest challenges in L2 acquisition is developing an extensive vocabulary (Webb, 2015). Classroom learners often struggle to acquire words incidentally due to limited input, making intentional vocabulary learning crucial (Nation, 1998). Effective vocabulary learning strategies combine explicit, intentional learning with exposure to extensive input (Schmitt, 2008). CALL offers informal language learners powerful digital tools to seamlessly integrate intentional vocabulary learning with exposure to extensive input. This is epitomised

by the 'sentence mining' method which prioritises extensive input and the use of vocabulary flashcards taken from context (Skritter, 2012; Zuo, 2021).

Paper flashcards, or word cards, are a simple technique to revise new vocabulary in a foreign language: write new words/phrases on one side of a small piece of paper, the flashcard, and the translation on the back. Revision then involves looking at the word and trying to remember its translation. Digital flashcards are a simple electronic extension of this idea - the 'front' of the flashcard is presented to learners and clicking a button reveals the 'back'.

We call a flashcard an 'understanding' or 'recognition' card when we see text in our target language and try to remember the meaning – see Figure 2.2 for examples. 'Understanding' flashcards test receptive vocabulary knowledge. 'Generation' or 'production' cards instead test productive vocabulary knowledge, where we see language in our L1 and try to produce it in the L2 – see Figure 2.3. Vocabulary flashcards strengthen words' 'form/meaning connection' which determines 'how readily the learner can retrieve the meaning when seeing or hearing the word form, and retrieve the word form when wishing to express the meaning.' (Nation, 2013, p. 73).

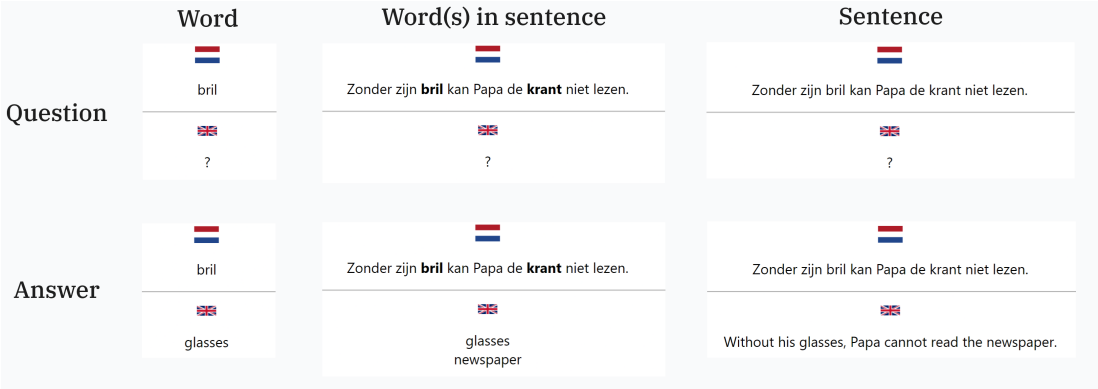

Figure 2.2: Some examples of digital 'understanding' flashcards.

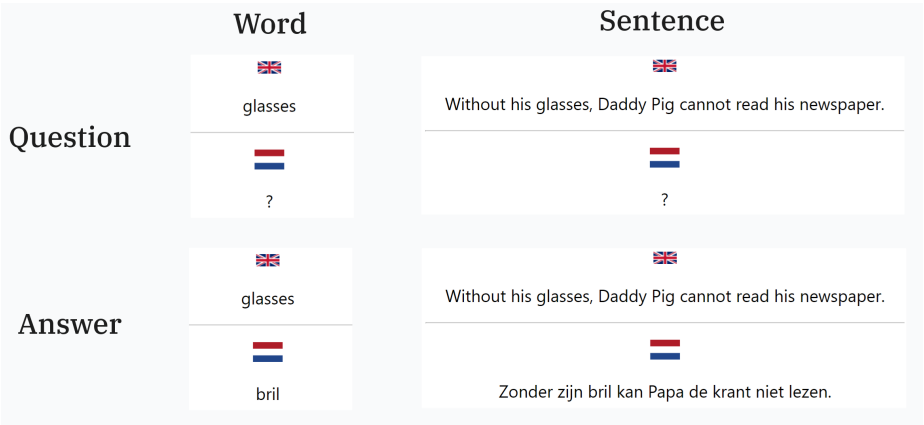

Figure 2.3: Some examples of digital 'generation' flashcards.

Flashcards are 'focused, efficient, and certain'. On their own, they are a very effective activity for vocabulary growth (Nation, 2013)—just three minutes a day can

lead to a significant increase in long-term vocabulary retention rate (Chukharev-Hudilainen and Klepikova, 2016). Even words learned in isolation, without the context of a sentence, get integrated into learners' 'existing knowledge systems' and contribute to overall language proficiency (Nation, 2013, p. 441).

However, flashcards and extensive input are best used in combination—they are 'complementary ways of learning which are partly overlapping and reinforcing' (Nation, 2013, p. 441). Learning vocabulary through flashcards yields a shallow understanding that facilitates comprehension while immersing in L2 input; extensive input then provides the repeated exposure to vocabulary over time required to develop comprehensive knowledge of form, meaning, and use (Ardasheva et al., 2019; Philip, 2010). Alternatively, unknown words encountered through extensive input can be captured in a flashcard and become the focus of deliberate study later on (Webb and Nation, 2017).

'Sentence mining' is a niche term that refers to the practice of creating flashcards from sentences found while engaging with target language input. 'Mined' sentences should contain at least one word or phrase the learner finds challenging, a 'target', in a sentence that is otherwise comprehensible (Antimoon, 2002; Xamuel, 2009). This 'sentence target' is typically a piece of vocabulary, but anything is possible: verb conjugations, idioms, characters (in Chinese/Japanese), cultural references. When learners find something new they wish to learn, they can 'mine' it, typically saving instructive sentences as digital flashcards inside a 'spaced repetition system' (SRS) (Sharp, 2010; Xamuel, 2009). SRSs, like *Anki*, *Quizlet*, and *SuperMemo*, are electronic systems for managing flashcards that schedule flashcard reviews to test learners just before they are likely to forget the answer, and are proven methods for retaining large amounts of vocabulary with minimum effort (Altiner, 2011; Seibert Hanson and Brown, 2020). Sections 2.6.1 and 2.6.2 discuss the psychological foundations of learning and memory with the use of flashcards and SRSs, such as active recall and the spacing effect.

Sentence mining is free, easy to implement, customisable, matches a learner's level, and improves reading and listening skills (Grigg, 2020). Examples are all taken from engaging source materials – whatever the learner wants to watch or read – and are contextualised, giving a rich sense of words' use and meaning. Pictures and audio taken automatically from L2 input can further enhance learning (cf. 'multimedia learning' in Section 2.2). Flashcard apps like *AnkiDroid* for Android offer efficient and convenient vocabulary learning (Hao et al., 2021). Language acquisition happens with long-term immersion in native content, with daily reviews of mined sentences a powerful supplementary activity to reinforce key vocabulary and grammar.

Originally developed on the Antimoon forum for the study of English (Antimoon, 2002), sentence mining was picked up by online Japanese language learning communities around 2007/2008 (Khatzumoto, 2007; Taurus, 2008). The concept is almost unexplored in the language learning literature, with the term only mentioned in one academic work (Kingman Rosero, 2022). It seems to offer a guaranteed path for success for motivated self-learners:

- First learn the basics of a target language's writing system, if necessary, and study the fundamentals of the grammar from a textbook or other resource.

- Perhaps learn the most frequent 2,000–3,000 words of the language – knowing 3,000 word families means recognising 95% of words in AVI (Montero Perez, 2022; Webb and Nation, 2017). This can be optimised with flashcards and an SRS.
- Repeat: Immerse, sentence mine i+1 sentences, review mined sentences every day.

A number of sentence mining tools, and tools developed by linguistics researchers for learning languages from sources of extensive input, are described in Appendix A.

2.3.3 Extensive Listening

The potential for extensive input from the Internet also extends to listening activities. The rise of podcasts for example, a powerful language learning tool (Hasan and Hoon, 2013), and the availability of online radio in foreign languages across the globe have afforded language learners engaging opportunities to engage in extensive listening practice, useful for developing an 'ear' for a language (Flowerdew and Miller, 2005) and improving both listening and speaking skills.

One form of extensive listening, 'passive immersion', involves listening to a language while engaging in other activities, harnessing the 'capacity for the brain to assimilate, sort, and store information' (Nida, 1982). It's been encouraged by sentence mining practitioners as a low-stakes way to build up crucial hours of L2 input (Ren, 2023). Ideally, passive immersion is best utilised by relistening to previously watched L2 content – comprehension is likely to be higher and learners can reinforce vocabulary knowledge studied from flashcards.

Language researchers have recommended extracting audio from audiovisual input for repeated listening practice. Listening, even passively, to the same L2 films can help learners to quickly expand passive vocabulary and acquire phrases and idioms effortlessly (Butzkamm and Caldwell, 2009, p. 191). Repeated listening can make sound–word links more automatic (Field, 2009), and help learners 'become familiar with the content, vocabulary, and structure' of L2 input (Goh and Vandergrift, 2011, p. 201).

Extensive listening can be paired with 'condensed audio', where non-speaking gaps in audio are removed, to maximise the density of listening materials. For AVI, this can be achieved using the timing information embedded in subtitle files—someone is only speaking whenever a subtitle is shown on screen. Creating condensed audio files programmatically from AVI is the focus of Section 4.4.2.

Many of the core principles of extensive listening set out by Gilliland (2014) are met by the audio output of the computer program:

- Access to a wide variety of authentic listening texts.
- Choice of listening content based on individual interests.
- Opportunities to listen repeatedly to selected texts.

Moreover, users of the system can create condensed audio that facilitates passive or active listening immersion in multiple languages simultaneously. This is yet to be explored in linguistic research.

2.4 Multilingualism

Multilingualism can be difficult to define. We use the following definition from Cenoz and Gorter (2015): 'Multilingualism at the individual level (also referred to as plurilingualism) brings together the process of acquiring second or foreign languages and the use of these languages.' We will use 'multilinguals' to refer to people who can speak more than one language. Historically, and today, multilingualism is 'the norm rather than the exception' (Richards and Rodgers, 2014). Herdina and Jessner (2002) argue: 'As the majority of the world's population is multilingual, research on linguistics should be centred on the multilingual speaker as the norm, not on the monolingual individual.'

Why have there been no multilingual AVI studies before? AVI with multiple audio tracks and subtitles has been hailed as 'the jewel in the crown of FLT [foreign language teaching] at school or at home' (Butzkamm and Caldwell, 2009, pp. 191–192). Researchers and educators have overlooked the opportunity to leverage L1 versions of videos to give full comprehension of AVI content, and the potential to reuse videos for learning multiple L2s.

There has historically been a 'monolingual bias' in SLA research and language teaching (Cenoz and Gorter, 2011). Buendgens-Kosten (2020) analysed the 'monolingual problem' in the CALL research field: they analysed 'CALL-related papers to demonstrate how far CALL is impacted by a monolingual bias that it inherited from language learning pedagogy', with terms like 'plurilingual', 'third language', 'L3', and 'translanguaging' vastly underrepresented. Ortega (2017) could not find 'a single published CALL article with "multilingual(ism)" or "bilingual(ism)" in the title', while these concepts do not appear in many historical overviews of CALL (Buendgens-Kosten, 2020).

Multilingualism was prevalent in societies the world over, from ancient to mediaeval times, which continues to be the case in many parts of the world today (Franceschini, 2012; King, 2018). Multilingual education is hundreds of years old (Baker and Lewis, 2015) while bilingual education is thousands of years old (García and Baetens Beardsmore, 2009, p. 13). What happened? David Crystal said it best: 'Speaking two or more languages is the natural way of life for three-quarters of the human race. [This] principle ... has been obscured in parts of Europe as a consequence of colonial history' (Crystal, 2005, p. 409). The French Revolution and development of European nationalism in the 19th and 20th centuries led to the idea of 'one nation, one language' and the rise of monolingual practices (King, 2018).

The monolingual bias in language research means that both prior knowledge and the interdependence of proficiency across languages are ignored (Cummins, 2008). Compared to monolinguals, multilinguals have 'an expanded and deepened language awareness', and 'tend to exploit the full array of their multilingual resources in language learning and language use' (Singleton and Aronin, 2007).

The more languages a learner knows, the more likely they are to exhibit positive transfer and metalinguistic awareness when learning a new language, especially when that language is similar to ones they already know (Festman, 2021). The closer such languages are to each other, the more resources language learners have at their disposal to help learning, like shared scripts, phonology, vocabulary, grammar, and aspects of pragmatics and discourse (Cenoz and Gorter, 2015, p. 2; Cenoz and Gorter, 2022). Students learning multiple languages side-by-side 'can use their resources cross linguistically and become more efficient language learners than when languages are taught separately' (Cenoz and Gorter, 2015, p. 4). The use of parallel AVI is an explicit practice of these multilingual principles.

The monolingual assumption has been 'increasingly challenged' in modern times (Hall and Cook, 2012). Multilingual education research has 'challenged previous ideas about [...] the isolation of languages' (Cenoz and Gorter, 2015, p3̇) and the use of two or more languages in education is on the rise (Lewis et al., 2012, p. 643). 'Pedagogical translanguaging' is a 'theoretical and instructional approach' based on embracing the linguistic resources of every student and has opposed school systems which have 'adopted monolingual ideologies and have isolated languages in the curriculum' (Cenoz and Gorter, 2021). More and more education systems around the world are implementing multilingual educational initiatives (see Calafato, 2020). However, one multilingual resource, used throughout much of the recorded history of language learning, remains understudied in SLA and underused in language education – the parallel text.

2.5 Parallel Texts

Parallel texts are instances of writing where the same thing is written in two or more languages side-by-side (see Figure 2.4). Parallel texts are everywhere—think food packets, clothes labels, and international airports.

2.5.1 Parallel Texts in History

Bilingual texts present versions of the same text in two languages, for example facing each other on either side of an open book. Bilingual texts have been around as long as writing itself: the earliest recorded lessons in literacy nearly 4,000 years ago were 'routinely bilingual' – they were reading and writing lessons in cuneiform that used both Sumerian (the language of prestige) and Akkadian (the vernacular) in parallel (Hardach, 2021).

Multilingual parallel texts too have existed for centuries. Multilingual inscriptions like the famous Rosetta Stone existed across the ancient world (Payne, 2006). Multilingual glossaries and grammars on clay tablets were used by civil servants and translators over 4,000 years ago in the Akkadian Empire, and later in the Hittite Empire and ancient Egypt (Franceschini, 2012). Word-lists with translations like this have aided translators 'from Ancient Mesopotamia to the late Middle Ages in Western Europe'; bilingual and multilingual glossaries today are still crucial in many industries (Bellos, 2012, pp. 94–95).

The term 'polyglot' can refer to a book which places several versions of the same

text in multiple languages side-by-side. This idea is around two thousand years old – Origen's *Hexapla* (Greek: *Sixfold*), created before AD 245, displayed six versions of the Old Testament in Hebrew and Greek (Britannica, 1998). The *Genoa Psalter of 1516* (see Figure 2.5) was 'the first true polyglot printing of any part of the Bible' (Giustiniani, 1516b). It 'display[ed] the words of the psalms laid out in eight columns – four per page – in Hebrew, a Latin paraphrase, the Vulgate Latin, the Septuagint Greek, Arabic, Aramaic ("Chaldean"), a Latin paraphrase, and the editor's notes, known as the scholia or gloss' (Giustiniani, 1516a).

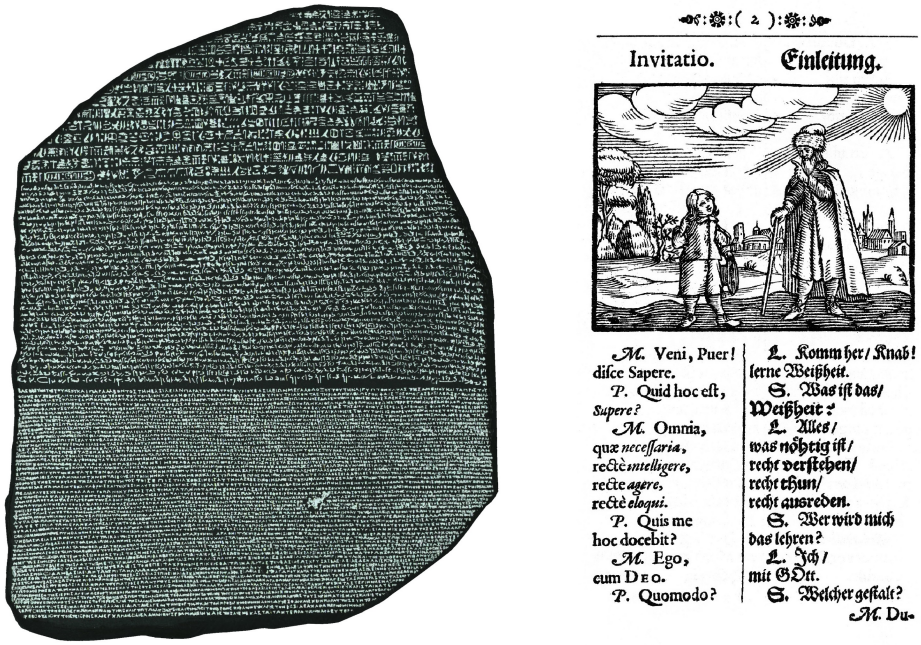

Figure 2.4: Two examples of parallel texts in history.
Left: The famous *Rosetta Stone*, an inscription from 196 BC, was key to deciphering the hieroglyphic and Demotic scripts of Ancient Egyptian.
Right: A page from a Latin–German edition of John Amos Comenius's *Orbis Pictus*, published in 1658. Considered the first children's picture book (Epstein, 1991), it was also published in many trilingual (Comenius, 1685) and quadrilingual editions (Comenius, 1679).

The parallel texts created automatically by the computer program of this thesis use subtitle timings (see Section 4.2.4) to align dialogue in multiple languages side-by-side, and bear striking resemblance to polyglot texts (Figure 2.6). The use of dialogues written in multiple languages in parallel have a long history—they are among the 'oldest of all language teaching aids'. The tradition of Latin–Greek bilingual dialogues in the Roman Empire was one 'revived by Renaissance scholars such as Erasmus' (Butzkamm and Caldwell, 2009, p. 142; Dickey and Ferri, 2012). The 'manières de langage' of late fourteenth and fifteenth century England taught mediaeval French using parallel dialogues and grammatical notes (Critten, 2023). For merchants in Europe, starting from the mid-15th century, parallel dialogues teaching conversational skills and compact polyglot dictionaries facilitated the self-study of up to eight languages in parallel (Franceschini, 2012; Hüllen, 2005).

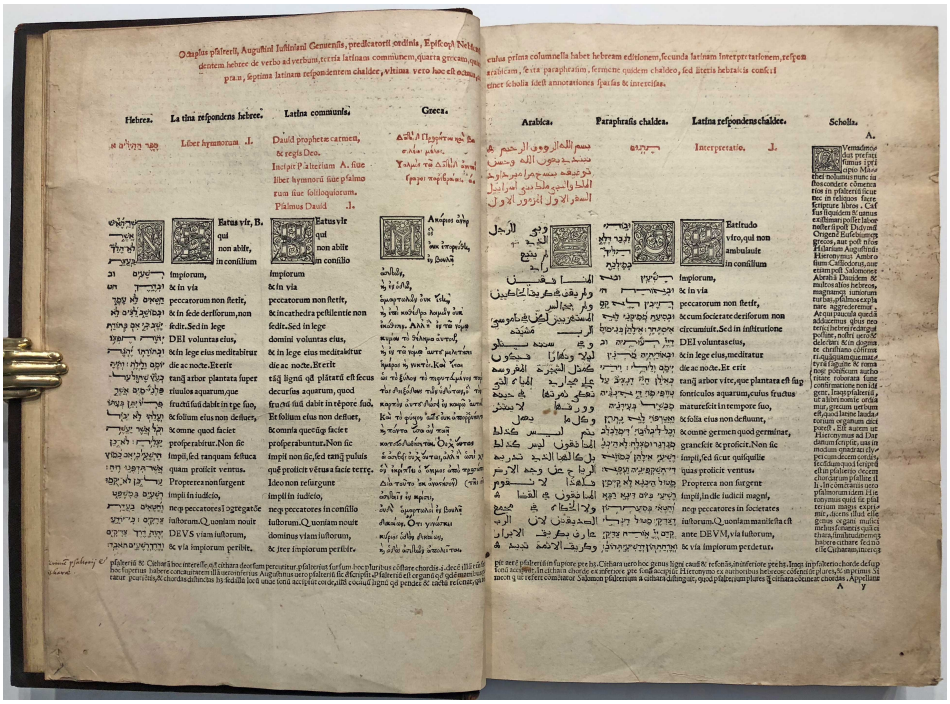

Figure 2.5: The *Genoa Psalter of 1516*, edited by Agostino Giustiniani, displays the psalms of the Bible in seven languages, with additional editor notes, in eight columns across two pages.
Source: X user @incunabula (2018). Retrieved from https://twitter.com/incunabula/status/1030345066605424640/photo/2

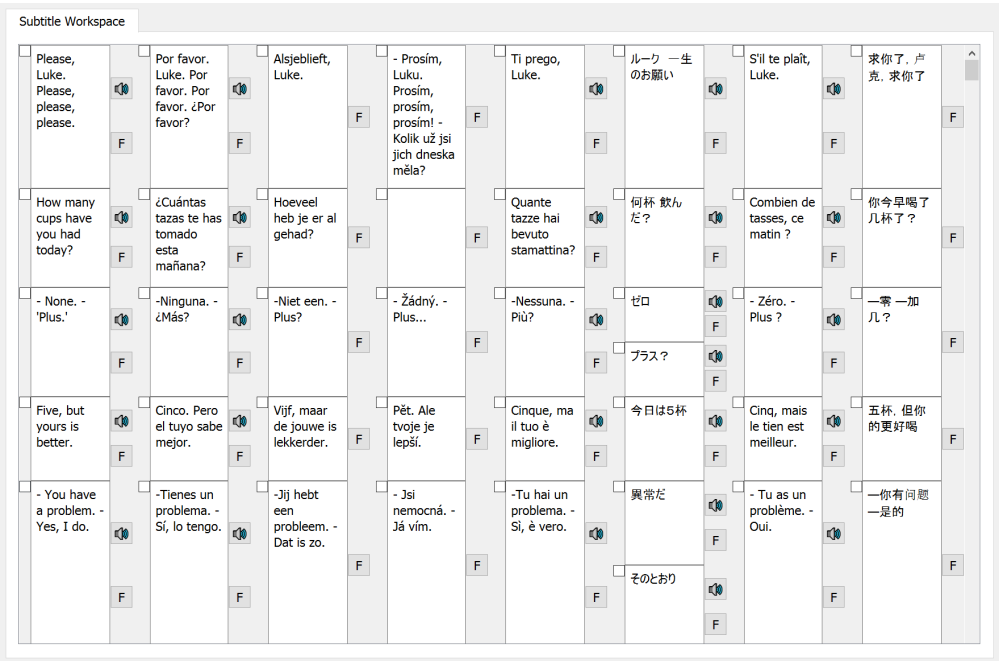

Figure 2.6: Subtitles from the pilot episode of *Gilmore Girls* in English, Spanish, Dutch, Czech, Italian, Japanese, French, and Chinese, aligned by the computer program in Section 4.2.4.
Note there is no corresponding audio track for Dutch, Czech, and Chinese.

2.5.2 Learning Languages with Parallel Texts

Using parallel texts to learn languages is fun, approachable, and effective. Access to the L1 version of a text provides comprehension of material, like previewing AVI in one's L1, and words and phrases can be disambiguated by comparison which saves time looking up words in a dictionary. Learners' L1 can serve as a 'scaffold' that facilitates understanding and acquisition of vocabulary from L2 materials—'the connections made between [...] languages [...] serve to facilitate new learning' (Ulanoff and Pucci, 1999).

Bilingual texts allow the understanding of L2 texts beyond a learner's current level (Zhang and Webb, 2019), offering learners 'richer, more authentic texts sooner [which] means more comprehensible input and faster acquisition' (Butzkamm, 2003). Bilingual books can promote bilingualism and biliteracy development, and make independent reading more enjoyable and more accessible to beginners (Ernst-Slavit and Mulhern, 2003). They can improve learners' overall linguistic and metalinguistic skills (Naqvi et al., 2013; Semmingson et al., 2015), help the development of vocabulary knowledge by facilitating the transfer of conceptual knowledge and skills across languages (Hu et al., 2012; Taylor et al., 2008), and have been shown to lead to more vocabulary gains than monolingual reading (Abdallah, 2021; Zhang and Webb, 2019).

Although bilingual books are widely available (Ernst-Slavit and Mulhern, 2003), common in countries like China and Spain (Zhang and Webb, 2019), and recommended in the US education system since 1951 (Bernard, 1951), there is very little L2 research on them (Zhang and Webb, 2019). This is reflective of the 'two solitudes' assumption of many language learning curricula – instruction in learners' L1 and L2 should be kept 'rigidly separate' (Cummins, 2008) – an insistence on strict separation of languages which stifles learning (Hopewell, 2011). The 'absence of effective bilingual techniques' in language classrooms is partly due to a 'general suspicion surrounding the use of [learners'] mother tongue' in language education (Butzkamm and Caldwell, 2009, p. 13). As explored in the previous section, teaching languages in isolation denies students the chance to 'use resources from their whole linguistic repertoire' (Cenoz and Gorter, 2020, p. 5).

The Bilingual Method (Dodson, 1967) is an exception, and its philosophy and methods are highly relevant to this thesis. It posits that the learner's first language is the most important linguistic resource at their disposal, 'the greatest asset people bring to the task of foreign language learning', and is 'the fastest, surest, most precise, and most complete means of accessing a foreign language' (Butzkamm, 2003). Teaching is done through dialogues with the explicit use of learners' L1 to support understanding, a direct analogue to the approach of this thesis.

2.5.3 Parallel Text Alignment

Text alignment techniques programmatically match corresponding segments of text in multiple languages. They are often used to create parallel corpora, large collections of parallel texts, which are crucial in many areas of linguistic research,

are a key resource for translators, lexicographers, and terminologists, and form the backbone of example-based machine translation (Giguet, 2005). Text alignment techniques can create parallel texts automatically, like the *Bitextual* tool for language learners that takes EPUBs (e-book files) in two languages and outputs a bilingual text (Bitextual).

Text alignment algorithms often work at the sentence level. Some methods rely on sentence length, measured by the number of tokens (Brown et al., 1991) or the number of characters (Gale and Church, 1991), while others focus on identifying word associations (Kay and Roscheisen, 1993) or use word features like frequency and position (Fung and McKeown, 1994). Modern methods using neural networks are now common (Forgac et al., 2023).

The computer program of this thesis creates parallel texts from the dialogue of captioned multilingual AVI. One major advantage captioned AVI provides over traditional text alignment techniques is the subtitle timings, embedded in subtitle files, which tell a video player when to display subtitles. They match the text shown on screen with the exact time in the video at which the text is spoken. When videos are dubbed in other languages, the spoken dialogue occurs at the same points in the video, and the corresponding subtitle timings in these languages will largely overlap with those of the original version. In a sense, this is a time-based sentence alignment method.

Many approaches for aligning subtitles in multiple languages have been designed for building parallel corpora used in training machine translation systems (see Tsiartas et al., 2013), such as the *OpenSubtitles* parallel corpora (Lison and Tiedemann, 2016). They include matching subtitles that have start times within 0.5 seconds of each other (Mangeot and Giguet, 2005), and match subtitles with the highest timing overlap (Tiedemann, 2007). Both the start time and highest overlap methods are implemented in the computer program of this thesis (see Section 4.2.4).

Subtitle alignment hasn't been extensively used for language learning purposes. One exception is the *Sublearning* website, which takes subtitles in the learner's target language from the *OpenSubtitles* corpora and asks for the equivalent subtitle in the learner's first language (Sublearning). Yet parallel multilingual dialogues were used for learning several languages simultaneously over 500 years ago (Franceschini, 2012, p. 2529). Today, there are two great advantages to the parallel multilingual dialogues created by aligning subtitle files, as implemented in the computer program discussed in Section 4.2.4: first, they are accompanied by matching recordings of native speech; and second, parallel texts can be generated automatically from any audiovisual input a learner desires. Aligning multilingual captions doesn't just create parallel texts, but parallel multimedia in any number of languages.

2.6 Spaced Repetition and Interleaving

This section explores concepts from psychology related to the two outputs from the computer program discussed in Section 4.4. Active recall and the spacing effect, which combine to give spaced repetition, are crucial to the effectiveness

of the multimedia flashcards of this system for learning vocabulary and phrases within a spaced repetition system (SRS), while interleaving is a crucial aspect of the multilingual 'condensed audio' output of the system. Each is an example of a 'desirable difficulty', a way to make learning more difficult in the short term that leads to superior long-term learning (Bjork, 1994).

2.6.1 Ebbinghaus, Active Recall, Spacing Effect

The experimental study of memory dates back to the work of Hermann Ebbinghaus, who used a system of nonsense three letter syllables to test different aspects of learning and forgetting (Ebbinghaus, 1885). His experiments, carried out on himself but replicated in the literature (Murre and Dros, 2015), unveiled the 'forgetting curve' – retention of new information quickly falls after learning, then decays more slowly (see Figure 2.7 (1)). He demonstrated that with subsequent relearning, a memory is strengthened and the forgetting curve becomes flatter. Figure 2.7 (2) shows what happens when we relearn some piece of information every day, as in Ebbinghaus's original experiments.

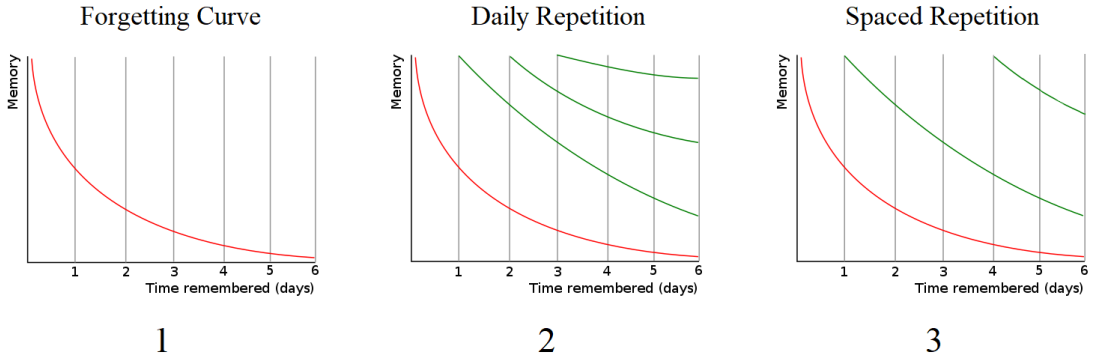

Figure 2.7: A visualisation of the forgetting curve (1), daily repetition (2), and spaced repetition (3). The red line is the initial memory; green lines are reinforced memories.

'Active recall', also known as 'retrieval practice' or the 'testing effect', is the act of deliberately trying to remember a piece of information. It is an excellent technique for increasing long term retention – it's more effective than 'passive' restudying for the same time and produces knowledge that can be flexibly recalled and transferred to different contexts (Roediger III and Butler, 2011). This relates to findings in intentional vocabulary learning – intentional study transfers to general language proficiency (Nation, 2013, p. 441). Active recall even aids learning when retrieval attempts are unsuccessful (Kornell and Vaughn, 2016).

While Ebbinghaus showed that memories are reinforced with repeated reviewing, his work also gave evidence that distributing instances of active recall over a space of time is more effective than 'massing' them at one time (Ebbinghaus, 1913, p. 89), now known as the 'spacing effect'. Retrieval practice and spacing have been widely shown to benefit both memory retention and transfer in a wide variety of domains across the lifespan, and are regarded as among the most robust findings in the psychology of learning (Carpenter et al., 2022).

Spacing and retrieval practice can be combined for further benefit – 'retrieving information repeatedly over spaced time intervals produces durable and long-lasting benefits on learning, compared to simply reviewing the information over the same time intervals.' (Carpenter et al., 2022). Daily repetition produces strong memories, but, for learning vocabulary, reviewing potentially thousands of words every day is unfeasible. The optimal strategy is to minimise the number of reviews in the long-term without forgetting.

This led to the idea of 'spaced repetition', or 'distributed practice'. As a memory is strengthened when it is actively recalled, we can instead try to increase the interval before our next review every time we successfully recall the information. This is shown in Figure 2.7 (3) – notice the retention rate never falls too low. This is due to the forgetting curve flattening out (i.e. creating a stronger memory over time) with each successful recall. We can exponentially increase the time we wait before next reviewing the information, meaning we can develop strong long-term memories with minimum repetitions. Even though learners typically underestimate the benefits of spaced repetition and tend to prefer massed practice (Logan et al., 2012), the superiority of distributed practice over massed practice has been verified by countless studies (Cepeda et al., 2006).

2.6.2 Flashcards and Spaced Repetition

A simple method to apply spaced repetition practice to a set of paper flashcards is Sebastian Leitner's system of 'Zettelkasten' or 'note boxes' (Leitner, 1972), shown in Figure 2.8. New cards are initially stored in Box 1 which learners go through and test themselves on perhaps once every day[1]. If they can successfully recall the answer, the back of the flashcard, they move it to Box 2, which is reviewed less frequently, say once every three days. Successful reviews of cards in Box 2 send them to Box 3, which is reviewed even less frequently, say once a week, and so on. In this way, newer cards and more difficult cards are reviewed more frequently, while easier cards and older and thus more familiar cards are reviewed less frequently, which saves time.

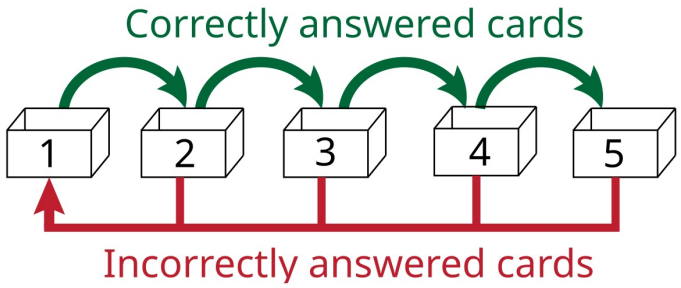

Figure 2.8: The 'Leitner System' for spaced repetition practice with paper flashcards

In Leitner's system, all incorrectly reviewed cards go straight back to Box 1; another option is moving them back a set number of boxes. In any case, each

[1]Leitner's original system used boxes with increasing physical length, e.g. 1 cm, 2cm, 5cm, etc. that were reviewed when they were filled with flashcards. In terms of subsequent boxes being reviewed in increasing intervals, the effect is the same.

card in a box is treated the same regardless of its past reviews. In addition, there is always a physical limit to the number of boxes in the system. After a successful review in the final box, one solution is to move the card to a 'retired box', of extremely well known cards, and perhaps review them once every 6 or 12 months.

A flexible and powerful modern extension to paper-based systems is digital flashcards, commonly encapsulated in a 'spaced repetition system', or 'SRS'. One key benefit they offer is sophisticated review scheduling – the time until the next review for a card scales dynamically based on the card's previous interval and whether the card was successfully reviewed or not. Crucially, the system can respond flexibly to cards of varying difficulty – the user can rate a successful card as more or less difficult and the interval will be adjusted accordingly. Learning parameters, i.e. the interval multiplier values, can be set by the user, to account for different aptitudes, and multiple decks, each with their own parameters, can be used in tandem for easier or more difficult subjects.

The first SRS, *SuperMemo*, was created by Piotr Woźniak in 1987, and was the first system to implement a digital spaced repetition algorithm, named *SM-2* (Woźniak, 1990). Woźniak pioneered mathematical formulas for computing optimal spaced repetition intervals based on user performance (Woźniak and Gorzelańczyk, 1994), while more sophisticated scheduling algorithms are an active area of research (Settles and Meeder, 2016; Su et al., 2023; Ye et al., 2022). Many spaced repetition systems have followed *SuperMemo*, such as *Mnemosyne* and *Quizlet*, and they have been implemented in many commercial language learning platforms like *Duolingo* (Settles and Meeder, 2016), as well as *FluentU* and *Memrise* (Seibert Hanson and Brown, 2020).

We focus our attention on *Anki*. It's free, open-source, has flexible and highly customisable flashcards, offers community-made shared flashcard 'decks' and application add-ons, and users' decks are synced between mobile and desktop applications. It's been shown to be an effective application to learn and retain vocabulary (Altiner, 2011; Seibert Hanson and Brown, 2020), while many other studies have demonstrated the all-around effectiveness of spaced repetition for long-term retention of L2 vocabulary (Bahrick et al., 1993; Bloom and Shuell, 1981; Ellis, 1995).

Anki leverages the natural long-term memory mechanisms of the brain. In language learning it is generally not intended as a primary learning source but as a supplementary daily activity. SRSs are designed for motivated individual learners – they work best when learners have the discipline to complete their flashcards reviews every day (Godwin-Jones, 2010).

2.6.3 Interleaving

Interleaving involves alternating between different concepts, skills, tasks, or categories while learning. Interleaving is another example of a 'desirable difficulty'. Like active recall and the spacing effect, it can boost both memory and transfer (Firth et al., 2021).

Interleaving has been shown to lead to better inductive learning. An example is

learning to differentiate the styles of different painters – learning is more effective when examples from each painter are shown interleaved, i.e. learners see examples from different painters mixed up, rather than consecutively, i.e. seeing all of one painter's paintings first followed by another (Kang and Pashler, 2012; Kornell and Bjork, 2008). This bears resemblance to the difference between videos shown to the Whole Group and Segments Group in the experiment in Section 3. In these painting styles studies, participants practising 'blocked practice' even had an 'illusion of competence' – they believed they performed better after seeing groups of examples consecutively, when in fact interleaving was more effective.

Similar benefits for interleaving have been found in experiments involving taxonomic classification in biology (Birnbaum et al., 2013; Tauber et al., 2013; Wahlheim et al., 2011), categorisation of chemical compounds using diagrams (Eglington and Kang, 2017), and identifying music composers' styles (Wong et al., 2020).

Interleaving has also been called the 'contextual interference effect' and has been shown to be more effective than blocked practice for students learning mathematics (Rohrer et al., 2014, 2020; Rohrer and Taylor, 2007), leads to enhanced memory and problem-solving in physics (Samani and Pan, 2021), and is an effective technique for training motor skills, such as in a wide variety of sports and in music performance (Carter and Grahn, 2016; Magill and Hall, 1990). Its benefit to language learning has only been demonstrated once, in learning the grammar of a foreign language (Pan et al., 2019).

The concept of interleaving is relevant in this study in relation to the parallelisation of AVI in multiple languages – instead of watching, reading, or listening to the same input the whole way through in each language, it's possible to view, read, or listen to shorter sections at a time, then reread, relisten, or review the same section in another language before moving on. An example is reading parallel sentences or paragraphs of a bilingual/multilingual book, instead of whole pages or full chapters, before changing languages.

This idea is relevant to the interleaved video shown to the Segments Group in the language learning experiment described in Section 3, and the interleaved condensed audio output of the language learning program described in Section 4.

2.7 Conclusion

The goal of this section was to answer **RQ1**, to develop the theoretical, historical, and practical precedent for this study, and to motivate the language learning experiment and computer program. This section also addressed **RQ2** by introducing sentence mining and highlighting its benefits for language learners.

Watching captioned audiovisual input is a highly effective and accessible way to learn a second language. Watching AVI can lead to incidental vocabulary acquisition, especially when viewed with L2 captions. Cognates and words that appear more often in a video are learned easier, and learners with more prior vocabulary knowledge tend to learn more vocabulary than other learners.

Prior knowledge of L2 input and the repeated viewing of interesting and familiar AVI content leads to higher comprehension, improved vocabulary gains, and overall improved language acquisition (Huckin and Coady, 1999; Tierney and Cunningham, 1984; Wang, 2012; Webb, 2015). For autonomous learners, we suggest watching a show the learner has already watched in their L1 – comprehension and enjoyment are likely to be much higher. A major gap in AVI research lies in the use of videos with multiple language versions. A wide variety of untested experimental techniques are possible—previewing input in learners' L1 for full comprehension, and reusing videos to learn multiple languages.

The opportunity for informal self-study of languages has never been better. One of the key advantages of the Internet for informal learners is extensive input – this includes opportunities for extensive reading, extensive watching, and extensive listening. Sentence mining combines the benefits of extensive input for language acquisition with highly effective intentional vocabulary learning through multimedia flashcards taken from context and spaced repetition algorithms. Its techniques are supported in language learning research. Its range of cutting-edge tools is exciting and thoroughly understudied, but yet to be expanded to a multilingual setting.

Despite the potential for multilingual language learning with modern technology and widespread multilingual language input, and given the extensive history of bilingual and multilingual language learning and traditions of parallel dialogues as language learning aids, it's surprising the field of SLA has overlooked the opportunity to use parallel AVI materials for learning multiple languages simultaneously. It is perhaps more evidence of the 'monolingual bias' that persists in language learning research.

Learners' L1 can serve as a 'scaffold' that makes 'richer, more authentic' L2 input more comprehensible and facilitates language acquisition (Butzkamm, 2003; Butzkamm and Caldwell, 2009, p. 25; Ulanoff and Pucci, 1999; Zhang and Webb, 2019). This is the principle behind showing AVI first in participants' L1 in the experiment, and the explicit use of learners' L1 in the multilingual parallel text and interleaved audio files of the computer program.

Multilinguals have an 'expanded and deepened language awareness' (Singleton and Aronin, 2007). Learning languages side-by-side, like when using parallel texts, is more effective as learners have the chance to use all their linguistic resources in parallel (Cenoz and Gorter, 2015, p. 4). This supports the idea of using multilingual AVI materials for multilingual learning. No AVI-based language learning platforms or sentence mining tools support learning multiple languages in parallel. A gap exists for a multilingual system that takes advantage of the potential to create parallel texts using subtitle timings and supports the creation of multimedia flashcards in multiple languages in the same interface.

Finally, to summarise the benefits from active recall, spacing, and interleaving: 'More complex and durable learning comes from self-testing, introducing certain difficulties in practice, waiting to re-study new material until a little forgetting has set in, and interleaving the practice of one skill or topic with another' (Brown et al., 2014). These principles underlie the potential outputs of the computer program - flashcards and interleaved audio.

3. Language Learning Experiment

This section details the first experiment measuring the impact of watching the same audiovisual input in multiple languages. In this proof-of-concept study, we examine what happens when learners are shown captioned audiovisual input in English, their native language, followed by viewings of the same material in two new languages, Dutch and Italian. Learning is assessed by vocabulary acquisition. The aim is to answer **RQ4**–**RQ6**: to determine if learners can incidentally pick up vocabulary in two languages at the same time and compare two viewing methods (watching in full videos or in short segments, visualised in Figure 3.1), to investigate the effect of words' frequency of occurrence and learners' prior vocabulary knowledge, and to observe participants' reactions to viewing audiovisual input repeatedly in several languages.

As no previous studies have investigated learning multiple languages with the same video, nor measured the effect of previewing AVI in one's L1 before watching in an L2, we first discuss potential experiments these ideas unveil, as an answer to **RQ3**. We then detail the study that was carried out, along with an analysis of the results.

3.1 Multilingual AVI Experiments

Introducing repeated viewings of the same audiovisual input in different languages leads to a multiplication of viewing options. None have been tested before, which means a number of fundamental studies are required in the future.

This section answers **RQ3** by detailing a number of experiments that were considered for this thesis, and will be described as variations on the experiment that was performed. The possibility of segmenting L2 videos as in the completed experiment is omitted, as is the possibility for different viewing orders and repeated viewings. Considering these options quickly leads to an intractable number of variables. A number of further experimental research directions with multilingual AVI are explored in Section 3.6 and Section 5.

3.1.1 L1 Previewing

First, the effect that previewing AVI in learners' native language has on a subsequent viewing in a new target language needs to be tested in isolation in a future experiment. One possible method would measure the incidental vocabulary acquisition in Dutch of two groups:

- No previewing: Watches a video in Dutch.
- Previewing: Watches a video first in English and then in Dutch.

As the viewing time doubles for the previewing group, showing the Dutch video twice to a third group would indicate what is a better use of an equal amount of

study time. However, a more effective and enjoyable use of study time might be to simply watch two different L2 videos.

3.1.2 Two L2s vs One

How does learning two L2s from a video compare to learning just one? An experiment could measure the incidental vocabulary acquisition in Dutch of three groups:

- One L2: Watches a video in English then Dutch.
- Two L2s: Watches a video in English, then Dutch, then Italian.

This experiment aims to assess how learning two languages simultaneously impacts the acquisition of one of the languages. It does not consider the learning gains in the second L2, Italian.

This could of course be expanded to learning three or four L2s from the same video, or learning without previewing.

3.1.3 Non-Beginner L2

Many of the participants in this experiment had prior experience studying Spanish, French, or German in school. Another experiment could compare vocabulary learning when one of the new languages, here an L3, is swapped with a previously studied L2, especially if it is related to the other new L3.

For example, participants who had previously studied Spanish could be split into two groups:

- Non-beginner L2 group: Watches a video in English, then Spanish, then Italian.
- Unrelated L3 group: Watches a video in English, then Dutch, then Italian.

Vocabulary acquisition is measured in this experiment for Italian. Watching in the L1 for full comprehension, then in a previously studied L2 with similar vocabulary to the new L3, might lead to more vocabulary gains than watching in another unrelated new L3.

This experiment is more difficult as it must account for individuals' L2 levels. It also doesn't measure possible learning gains in the non-beginner L2, or the unrelated L3.

A natural extension, even more complex, is watching the same video in multiple languages learners have varying proficiencies in, for example watching in English, Irish, Spanish, and Italian – some of the participants in this study had studied both Irish and Spanish in the past.

3.1.4 Same Content vs Different Content

Finally, is it more effective to reuse the same video for learning two languages, or watch a different video for each language? A possible experiment could measure vocabulary acquisition in both Dutch and Italian for two groups:

- Same content: Watches a video in English, Dutch, and Italian.
- Different content: Watches a video in English and Dutch, then a different video in English and Italian.

In this experiment, the same content group watches only three videos compared to the different content group's four.

3.2 Methodology

3.2.1 Experimental Design

This study measured incidental vocabulary acquisition from watching an episode of *Peppa Pig* shown in three languages – English, the participants' L1, and Dutch and Italian, two new languages. The study compared the learning gains from two separate viewing methods relative to a control group, discussed below. The viewing methods are illustrated in Figure 3.1: the Whole Group watched the episode the whole way through in English, then immediately after in Dutch, and finally Italian; the Segments Group watched in short segments that repeated in English, Dutch, and Italian, e.g. when Segment 1 in Italian ended, Segment 2 in English immediately began.

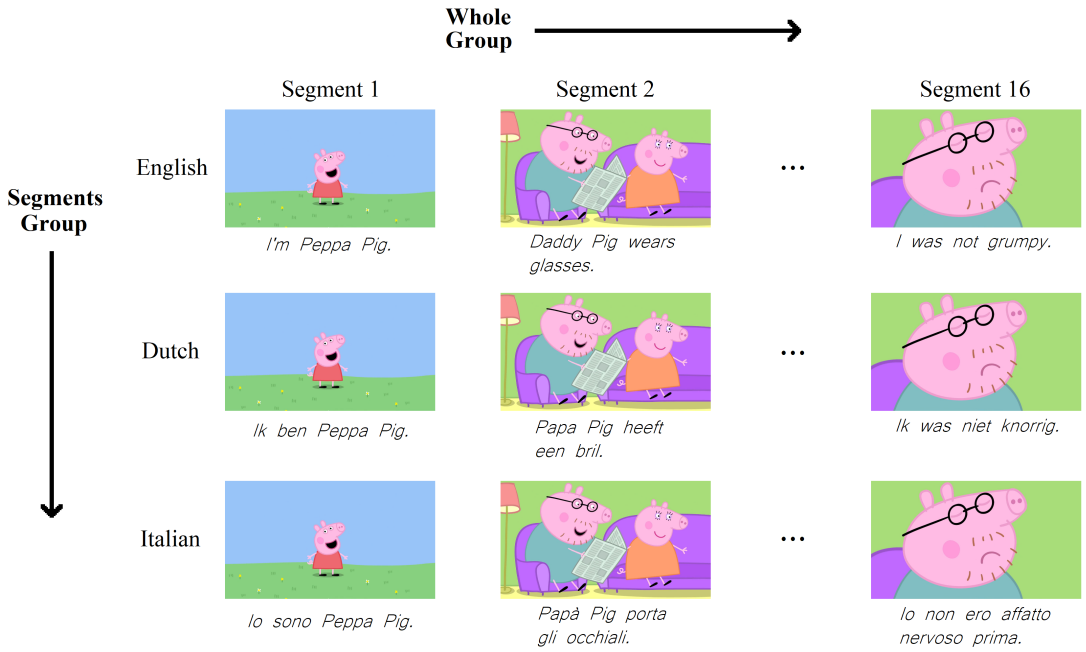

Figure 3.1: An illustration of the structure of the videos shown to the experimental groups.

3.2.2 Pre-Testing and Control Group

As prior vocabulary knowledge has a positive impact on L2 incidental vocabulary learning through watching AVI (see Feng and Webb, 2020; Montero Perez, 2022), a pre-test/post-test design was considered to more accurately measure vocabulary gains. However, given that this was a proof-of-concept study with volunteer

Table 3.1: Previous language experience by group.

Previous Language	Control Group	Whole Group	Segments Group
Dutch	0	0	0
Italian	1	2	1
Irish	5	8	12
Spanish	2	8	10
French	3	5	3
German	1	3	5
Arabic	1	0	0

participants, it was limited to one session. This meant there would be no vocabulary pre-test, only a post-test. A compromise was made – a control group of 20% of the participant group would complete the vocabulary tests without watching the episode – an approach used in other studies on incidental vocabulary learning from watching AVI. An example is Montero Perez (2020), with the caveat that this study tested the learning of pseudowords. With a big enough sample size, this approach shouldn't have posed a problem, however only 7 participants comprised the control group.

Showing the control group the episode in English only might have allowed for a fairer comparison with the experimental groups who had the context for the target items.

3.2.3 Participants

35 native English speakers (29 male, 6 female) from Ireland and the UK were personally recruited by the researcher and contacted via text message to participate in the experiment. Nearly all of the participants were 24 or 25 years old (M_{Age} = 24.57, SD = 1.77, Min = 19, Max = 31). All of the participants had graduated from secondary education; 20 had completed a Bachelor's degree and were in the workforce, while 9 had completed a Master's. The majority of participants were from Ireland, with 25 participants reporting some level of proficiency in Irish. 17 of the participants reported proficiency in Irish along with one of French, German, or Spanish which aligns with the usual path taken in Irish language education – compulsory Irish lessons throughout primary and secondary education, with 3–6 years of French, German, or Spanish in secondary education. Only 2 participants reported no previous second language experience. Previous language experience per group is shown in Table 3.1. The second most popular language was Spanish with 20 participants, followed by French with 11 and German with 9. Between the Whole Group and Segments Group there was a relatively even split for each of these non-Irish L2s.

The participants were randomly assigned in a 20/40/40 split to three groups: a Control Group (7 participants) who only took the vocabulary tests, a 'Whole Group' (14 participants) who watched the TV episode the whole way through in English, Dutch, and Italian, and a 'Segments Group' (14 participants) who watched the same episode cut into short segments repeated in English, Dutch, and Italian.

5 other participants took part but were omitted as they either did not finish the experiment, or they completed the experiment without watching the video the whole way through.

3.2.4 Materials

3.2.4.1 Audiovisual Input

Due to the limited scope of the study, the audiovisual material needed to be a short, self-contained TV episode, that in a single session might yield measurable vocabulary acquisition in two languages.

Using L2s previously studied by participants was avoided to prevent bias in the results, as having no pre-test to compare learning gains to would make it impossible to determine what vocabulary was learned from the input and what was participants' previous knowledge.

Dutch and Italian were chosen as the target languages, as most participants had limited or no experience in either language, yet they were similar to languages that many of the participants had already studied – German for Dutch, French and Spanish for Italian – which meant they were more likely to acquire some new vocabulary from watching a short video.

As intralingual subtitles (captions) and the viewing of animations seem to have the largest effect on vocabulary acquisition from watching captioned/subtitled video (Reynolds et al., 2022), the most appropriate choice for the experiment seemed a captioned animated children's TV show. Several series were looked at to find an episode at an appropriate level for complete beginners to a new language, ideally with fully matching captions.

The Peppa Pig series has been found to be appropriate input for pre-primary learners, especially for vocabulary development (Alexiou, 2015; Kokla, 2022, 2021; Scheffler et al., 2021). It was deemed an appropriate choice for complete beginners – the episodes are short (around 5 minutes) and follow simple stories, the narration and dialogue is slow and clear with lots of repetition, and it's a show participants were familiar with. The language used is predominantly high-frequency words, while the more uncommon words 'tend to be either names of the characters or words which are fully contextualised (they are discussed by the characters and they can often be seen)' (Scheffler et al., 2021). Imagery in AVI has been shown to help learners infer the meanings of unknown words (see Montero Perez, 2022), and Peppa Pig often contains visual reinforcement of each episode's themes (Scheffler et al., 2021).

Several episodes of *Peppa Pig* were available on YouTube in English, Dutch, and Italian. Two episodes were trialled with volunteers who didn't take part in the subsequent experiment. The episode 'Daddy loses his glasses' (Season 1 Episode 9, 2004) was chosen due its simple premise and the high amount of appropriate target words to form the vocabulary test.

Captions for the English version of the episode were found on YouTube with some minor corrections made, and two native speakers provided captions that matched

each line of dialogue in the Dutch and Italian dubs. The script for each language is provided in Appendix B.

The captions were then embedded in their corresponding videos. The Whole Group's video was a concatenation of the three videos achieved with the *ffmpeg* command-line multimedia tool. For the Segments Group, the episode was divided at natural breaks in the dialogue/narration to create segments of roughly the same length, ranging from 11 to 28 seconds with 2 to 5 lines of dialogue in each. Each segment's English, Dutch, and Italian versions were concatenated together, and finally all the three-language segments were combined to create the segmented video. Each full video was 14 minutes long. The whole and segmented videos can be viewed here and here respectively.

3.2.4.2 Target Items

Research has shown watching a single TV episode can lead to learners' incidental acquisition of single words (Montero Perez, 2022, p. 172). 13 Dutch and Italian target vocabulary were chosen to measure learning. They are shown in Table C.1 and Table C.2 respectively, in Appendix C, along with each word's:

- Translation into English.
- Frequency of occurrence in the episode.
- Frequency rank, e.g. '10' indicates the word is the 10th most common word in the language. Word frequencies were provided by monolingual frequency lists from the v2018 release of the *OpenSubtitles* corpora (Lison and Tiedemann, 2016).
- Cognateness with English, Spanish, French, German. Cognates are words in two languages with high overlap of form and meaning. Words from the same origin that share similar forms but differ in meaning are shown in brackets, like Italian "guardare" ("to watch") and Spanish "guardar" ("to save"). Cognates with Irish were not considered.
- Part of Speech, e.g. Noun, Verb, Adjective.

To choose the target items, the complete text of each version of the episode was extracted from the subtitle files and examined using the web-based corpus analysis tool *Voyant Tools*.

Both frequency of occurrence and cognateness are positively correlated with vocabulary learning (see Montero Perez, 2022, p. 175). To maximise the observed effect, only words with at least two occurrences were considered.

Content words rather than function words were chosen because of the potential benefit of imagery for vocabulary learning (e.g. Pujadas and Muñoz, 2023), and English translations for each target item were used in the same context in the English version of the episode.

No target items were selected from the first section of the episode, to give participants in the Segments Group a chance to see sections repeating in Dutch and Italian.

Some inference was required for two words: “slaap” (“sleep”) from “slaapkamer” (“bedroom”), and “letto” (“bed”) from “camera da letto” (“bedroom”).

Words whose frequency value is marked with an asterisk had multiple forms or were part of another word, e.g. “slaap”. The frequency shown is a sum of the frequencies of all the forms in the episode. All forms and their frequencies are included in Appendix C.2.

3.2.4.3 Distractor Items

Table C.3 and Table C.4 in Appendix C.3 show 14 distractor items in Dutch and Italian respectively. These are words which didn't appear in the episode but were included in the vocabulary tests for a number of reasons:

- Motivation – Typically, including easier non-target items helps motivate participants who are unable to get many of the target items.
- Challenge – Mixing distractor words into the vocabulary test makes it harder to guess target items' meaning by process of elimination. It could be argued this gives a truer sense if a participant has learned a target item.
- Prior Knowledge – Learners' scores on distractor items can act as an approximation of prior vocabulary knowledge, which is positively correlated with incidental vocabulary learning (Montero Perez, 2022, pp. 175–176).

The chosen words were common words (top 3,000 words by frequency as measured by the same *OpenSubtitles* corpora) and of a variety of parts of speech – 7 nouns, 4 verbs, and 3 adjectives in each language.

3.2.4.4 Vocabulary Tests

Different methodologies for testing incidental vocabulary learning were considered. The five-level description Vocabulary Knowledge Scale (VKS) (Paribakht and Wesche, 1993) is a self-report measure of vocabulary acquisition designed to reflect the stages in learning a word, from total unfamiliarity to recognition of the word form, knowledge of meaning, and finally the ability to use the word in context. However, it was deemed problematic for learners with no prior experience in a language – participants can only guess the meaning of a word provided they report having seen the word before, which doesn't allow for guessing unseen words through knowledge of cognates in participants' other languages.

Instead, individual form recognition and meaning recall questions for each word were used. For form recognition questions, 1 point was given if participants responded having seen the word before, and 0 points otherwise. For meaning recall, 1 point was given for a correct translation into English, with 0 points for an incorrect translation. An extract of the Dutch vocabulary test is shown in Figure 3.2.

These are frequently used vocabulary tests that tap into form-meaning mappings (Montero Perez, 2022). Using form recognition and meaning recall questions together was the approach taken in the first study on incidental vocabulary acquisition from watching a single full-length TV program (Peters and Webb, 2018).

The weakness of form recognition questions within a pre-test/post-test design noted by the authors – post-test answers are biased from exposure to words in the pre-test – was not relevant to this study as there was no pre-test.


For the following **Dutch** words, please indicate if you have seen the word before.
If you are able, please provide a translation into English.

	I have seen this word before		English translation
	No	Yes	I believe this word means
knorrig	○	○	
hebben	○	○	
slaap	○	○	
spelen	○	○	
muziek	○	○	

Figure 3.2: An extract of the Dutch vocabulary test shown to participants in the experiment. Note that target and distractor items are randomly shuffled together.

3.2.4.5 Post-Questionnaire Likert Scale Questions

The experimental groups were asked to answer a series of 11 Likert scale questions, shown in Figure 3.3.

3.2.5 Procedure

The experiment was conducted anonymously using a *Qualtrics* survey. The participants were given a link to complete the experiment in their free time, and were told the experiment was to investigate what happens when native English speakers watch a TV episode in English, then in Dutch, and finally Italian, and it would be followed by a short test that measures learning. They were asked to not rewind or rewatch the video.

The experiment began with a short questionnaire of basic demographics (age, gender, educational background). The participants were then asked to report the languages they had studied before, and self-assess their language proficiency in each, given a table of CEFR language proficiency descriptors. They were also explicitly asked if they had experience learning Dutch or Italian before, and, if so, to describe their experience – how long they studied it, in what context and how, and their motivation for learning. Finally, they were asked had they ever used AVI for language learning, and, if so, how often they use it to study.

Initially, each participant was randomly assigned to one of the three groups as discussed in Section 3.2.2. The Whole Group and Segments Group were told they would watch an episode of *Peppa Pig* to be shown in three languages –

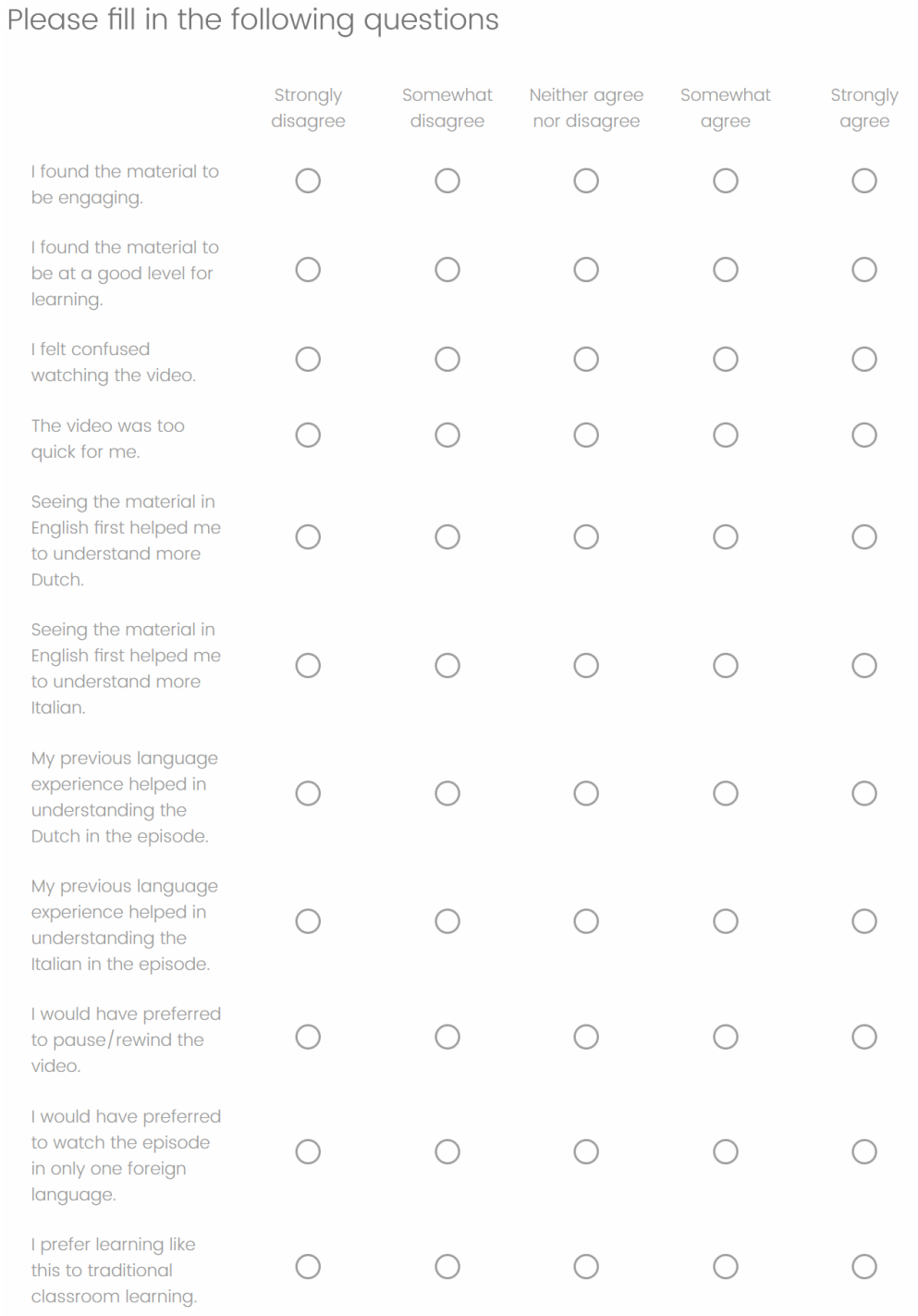

Figure 3.3: The Likert scale questions asked to the Whole Group and Segments Group.

English, Dutch, and Italian – and what they viewed would always be in that order. They were then described the structure of their respective video, and that the episodes had been combined into one video. They were asked to be in a quiet space where they could hear the audio, and make sure they could view the subtitles embedded in the video. Finally, they were asked to watch the video the whole way through once without stopping, and were presented with a link to their group's video.

After watching the video, each experimental group completed the vocabulary test in Dutch, followed by the one in Italian. The control group were shown the vocabulary tests immediately after the pre-questionnaire.

After the vocabulary tests came a post-questionnaire. The experimental groups completed the series of Likert scale questions, and were asked if they watched the video the whole way through, and if they had seen the episode before. Finally, all participants were given a debriefing about the purpose of the experiment.

For participants in the Whole Group and Segments Group, the experiment took approximately 45 minutes.

3.3 Hypotheses

Due to the high repetition of several target items in both languages in the episode, it was hypothesised that the experimental groups would learn at least some vocabulary in each language from watching the video, by comparing vocabulary scores with the control group. Nevertheless, a null hypothesis was posited that there would be no difference in target item vocabulary scores between the control and experimental groups.

It was difficult to predict whether repeatedly watching the whole episode or short segments would lead to more vocabulary gains as no previous work has explored previewing in learners' L1, rewatching the same AVI in multiple L2s, or watching in segments.

We first posit potential reasons why each approach could be more effective.

Benefits of watching the whole episode:

- It might be easier to follow the story when watching in the target languages, as viewers see the entire episode in English to begin with.
- It avoids the cognitive load of repeatedly switching languages, which could be taxing. For less able participants, working memory might be overloaded in the segmented approach, inhibiting learning – see Cognitive Load Theory (Sweller, 1988). We note working memory as a learner-related factor when learning from AVI has been understudied (Wei and Fan, 2022).

Benefits of watching in segments:

- Viewers always see the English version of each line of dialogue shortly before the target languages, which may make it easier to pick up vocabulary.
- Repeatedly rotating between languages can be regarded as a form of 'interleaving' (see Section 2.6.3), a 'desirable difficulty' which, although more difficult, leads to better learning.

A null hypothesis was posited that there would be no difference in vocabulary acquisition between the experimental groups.

We also suggest reasons why vocabulary acquisition might be higher in the first L2 (Dutch) or second L2 (Italian).

Dutch over Italian:

- As the Dutch episode/segments are shown directly after the English equivalent it might be easier to remember the meaning of words/phrases and so pick up more vocabulary.
- Dutch is typologically closer to English (both are Germanic languages) than Italian and shares more cognates (8 target items vs 5). The typological distance between previously known languages and a target language can influence acquisition (Cenoz and Gorter, 2015, p. 2).
- For the same reason, participants with experience in German have an advantage in Dutch over Italian (9 target item cognates vs 4).

Italian over Dutch:

- The Italian episode/segments are seen closer to the vocabulary test, especially for the Whole Group.
- Participants with experience in French or Spanish should have an advantage in Italian (all are Romance languages) over Dutch. French and Spanish share 11 and 7 target item cognates with Italian, and 1 each with Dutch.

There is a wide variety of factors that influence vocabulary learning from captioned AVI – see Montero Perez (2022) for a review. In addition to the predicted effect from cognates, higher vocabulary scores were expected for words with higher occurrences in the episode. A positive effect was also expected for learners with higher prior vocabulary knowledge, as approximated by their performance on the distractor words.

3.4 Data Processing

Some participants' responses were deleted as the time they spent completing the experiment was less than the duration of the video. 2 participants had seen the episode before but were included in the analysis nonetheless due to limited participant numbers and the fact that all experimental group participants watched the episode in English anyway. 4 participants had studied Italian before (1 in Control Group, 2 in Whole Group, 1 in Segments Group). Their scores on the Dutch test were kept, also due to limited participant numbers.

Several participants misinterpreted the form recognition task. Some participants only reported having seen a word if they had encountered it before the experiment, leaving out a number of target items they had correctly translated. Others only reported having seen a word if they guessed the meaning. Only the meaning recall answers from these participants were kept., which amounted to 5 participants in each of the experimental groups. Some participants reported recognising Italian words like "camera" and "fare" that resemble English words, but didn't know the correct meaning. The meaning recall scores were considered more reliable and thus are used exclusively for several analyses due to the higher certainty that participants interpreted the question correctly.

Meaning recall answers were accepted if they generally conveyed the correct meaning. Some specific rejections were “look” or “search” for “vinden” = “find”, “look” for “zoeken” = “look for”/“search for”, “bedroom” for “kamer” = “room”. Different verb forms were accepted, for example “seen” for “zien” = “see”, as were all correct translations for polysemous words, e.g. “kid” or “young” for “jong”.

Self-assessed proficiency was not always accurate or reliable. For example, some participants rated themselves as high intermediate in Spanish but missed easy Spanish cognates, e.g. ”C2 Spanish” but didn’t know basic words like ”mano”. These participants’ proficiency levels were omitted.

Overall, no participants had C2 proficiency in a language other than English. Only 2 participants reported a level of C1, both in Irish.

The results were analysed with the *Pandas* and *NumPy* packages for Python, with statistical models from *SciPy* and *Statsmodels*.

3.5 Results

3.5.1 Exploratory Data Analysis

3.5.1.1 Vocabulary Scores

Table 3.2 presents a breakdown of vocabulary scores. Having three groups, two languages, two measures of vocabulary knowledge, and both target and distractor items means an abundance of choices for graphing and tabulating data.

The table separates the scores on target and distractor items and presents the results of both the complete sample and the three groups, in 9 categories. The form recognition and meaning recall scores for both languages are combined to give the ‘Overall’ scores. The scores are then broken down by question type, e.g. ‘Recognition’ is form recognition scores for both languages combined, and by language, e.g. ‘Dutch’ is Dutch form recognition and meaning recall scores combined. The maximum possible scores for the ‘Overall’ category, for example, were 52 for target items (13 for Dutch meaning recall and form recognition, same for Italian) and 56 for distractor items.

Table 3.3 presents Cronbach’s alpha values for the various vocabulary measures, mirroring the structure of the previous table with the caveat that all groups’ scores were analysed together. Only the Italian target item meaning recall scores report less than high internal reliability.

The table of vocabulary scores shows that both experimental groups comprehensively outscored the control group on every target item measure, with the Control Group scoring an average of 1.5 points in the ‘Overall’ category compared to the 12.5 and 16.4 of the Whole Group and Segments Group respectively. However, the experimental groups also outscored the control group on every distractor item measure, albeit by far smaller margins. This suggests that the control group was weaker and had less prior knowledge of Dutch and Italian vocabulary.

The Whole Group and Segments Group scored similarly. The Segments Group’s average Dutch score was more than double the Whole Group’s average. The

Whole Group scored better on Italian words but by a smaller margin.

The closest average score on distractor items between the experimental groups and Control Group was for Dutch meaning recall – 3.29 for the Control Group, 4.07 for the Whole Group, 4.64 the for Segments Group. For target items, all 7 participants in the control group knew none of the meanings of the Dutch words in the episode while the experimental groups averaged over 3 correct answers (out of 13). This is also the measure with the most number of participants.

Average scores on Italian distractor words were almost always higher than for Dutch, suggesting higher prior knowledge. On meaning recall scores for target items however, Dutch scores were moderately better.

Scores on distractor items were higher, as were form recognition scores compared to meaning recall, which was expected.

Table 3.2: A summary of the vocabulary scores on target and distractor items.

			Target Items				**Distractor Items**			
		n	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**
Overall	**Total**	23	11.17	8.55	0	26	20.09	11.63	2	42
	Control	6	1.5	1.38	0	3	13.5	6.16	3	21
	Whole	8	12.5	7.19	3	24	21.62	12.0	7	39
	Segments	9	16.44	7.16	6	26	23.11	13.31	2	42
Recognition	**Total**	23	7.04	5.52	0	18	10.43	6.27	1	22
	Control	6	1	0.89	0	2	6.5	2.81	2	10
	Whole	8	8.62	5.42	2	18	12.12	6.92	3	22
	Segments	9	9.67	4.47	3	15	11.56	6.73	1	21
Meaning	**Total**	31	4.48	3.58	0	13	9.9	6.02	1	21
	Control	6	0.5	0.55	0	1	7	3.41	1	11
	Whole	12	4.75	3.19	1	13	10.42	6.67	3	21
	Segments	13	6.08	3.43	2	12	10.77	6.31	1	21
Dutch	**Total**	25	5.28	5.26	0	19	7.92	6.35	0	22
	Control	7	0.14	0.38	0	1	5	3.16	0	9
	Whole	9	4.67	3.2	0	8	7	6.08	0	17
	Segments	9	9.89	5.01	2	19	11.11	7.52	2	22

Continued on next page

Table 3.2: A summary of the vocabulary scores on target and distractor items. (Continued)

Italian	**Total**	23	5.7	4.5	0	16	11.78	7.19	0	25
	Control	6	1.33	1.51	0	3	8.67	5.57	3	18
	Whole	8	8	5.07	2	16	13.88	7.2	6	25
	Segments	9	6.56	3.36	2	11	12	8.09	0	21
Dutch Recognition	**Total**	25	3.32	3.11	0	10	4.08	3.4	0	11
	Control	7	0.14	0.38	0	1	1.71	1.89	0	4
	Whole	9	3.33	2.24	0	6	4.44	3.54	0	9
	Segments	9	5.78	2.86	1	10	5.56	3.47	1	11
Dutch Meaning	**Total**	35	2.51	3.12	0	13	4.14	3.94	0	13
	Control	7	0	0	0	0	3.29	1.98	0	6
	Whole	14	3	3.78	0	13	4.07	4.5	0	13
	Segments	14	3.29	2.61	0	9	4.64	4.24	0	13
Italian Recognition	**Total**	23	3.61	3.12	0	12	6.04	3.78	0	13
	Control	6	0.83	0.98	0	2	4.5	2.74	2	9
	Whole	8	5.38	3.81	1	12	7.25	4.03	3	13
	Segments	9	3.89	2.09	1	7	6	4.15	0	11
Italian Meaning	**Total**	31	2.16	1.44	0	5	5.9	3.4	0	13
	Control	6	0.5	0.55	0	1	4.17	2.86	1	9
	Whole	12	2.5	1.31	1	4	6.75	3.55	3	13
	Segments	13	2.62	1.33	1	5	5.92	3.43	0	10

Graphs showing the distribution of scores of all participants in the nine categories in Table 3.2 are presented in Appendix D. The graphs illustrate that the Control Group scored consistently weaker on distractor items. However, their exceptionally low scores on target items, in contrast to the experimental groups' more varied scores, strongly suggest many words were unfamiliar to the participants and the experimental groups learned vocabulary through watching the episode.

Figures E.1 and E.2 in Appendix E show the meaning recall scores in Dutch and Italian with strong participants in the experimental groups filtered out. These were participants who scored more on distractor items than the strongest member (in each language separately) of the control group. These subgroups had similar distractor scores, an approximation of prior knowledge, to the control group. Again, the experimental groups' scores are noticeably higher on target words and suggest vocabulary was learned from the video.

3.5.1.2 Word Analysis

This section analyses the performance of the experimental groups on the target items. Analysis of the effect of frequency of occurrence addresses part of **RQ5**.

Target items' average form recognition and meaning recall scores (scale of 0 to 1) for both experimental groups combined and separate were graphed – see Figure 3.4 for Dutch and Figure 3.5 for Italian. The words are arranged by their frequency in the episode, including multiple forms of the same word.

In Dutch, the meaning of each target item was known by at least one person in the experimental group, and higher frequency words were generally known more, as evidenced by the general fall in average score going left to right in the graph. For Italian, 6 of the 13 words weren't known. Even though "guardare" appeared 8 times, its meaning wasn't learned by any participant. All Italian target items however were recognised by at least one person.

Table 3.3: Cronbach's alpha values for the various vocabulary measures.

Vocabulary Category	**Target Items**	**Distractor Items**
Overall	0.90	0.93
Recognition	0.88	0.93
Meaning	0.82	0.90
Dutch	0.86	0.88
Italian	0.84	0.94
Dutch Recognition	0.80	0.82
Dutch Meaning	0.88	0.89
Italian Recognition	0.83	0.92
Italian Meaning	0.56	0.89

The most recalled Dutch word was “bril” (“glasses”), which was the most frequent word with 20 occurrences, and was the theme of the episode. It was known by 64% of participants in both experimental groups. “Occhiali” (Italian for “glasses”) appeared 24 times and was also known by 64% of participants. Both were likely learned from watching the video – the only cognate was the German “Brille”, and the words were entirely unknown to the Control Group.

The meanings of high frequency Dutch verbs were not learned so well. The next highest average recall score after “bril” was for “knorrig” (“grumpy”) which appeared 5 times, a word which participants were highly unlikely to know beforehand. In Italian however, the high scores for “nervoso” were likely due to its similarity to its English cognate “nervous”; it was also known by many Control Group participants.

Graphs with target words sorted by frequency when multiple word forms are not counted are included in Appendix F. For Dutch, they show an even steadier fall in average score, again suggesting a facilitative effect from high frequency of occurrence.

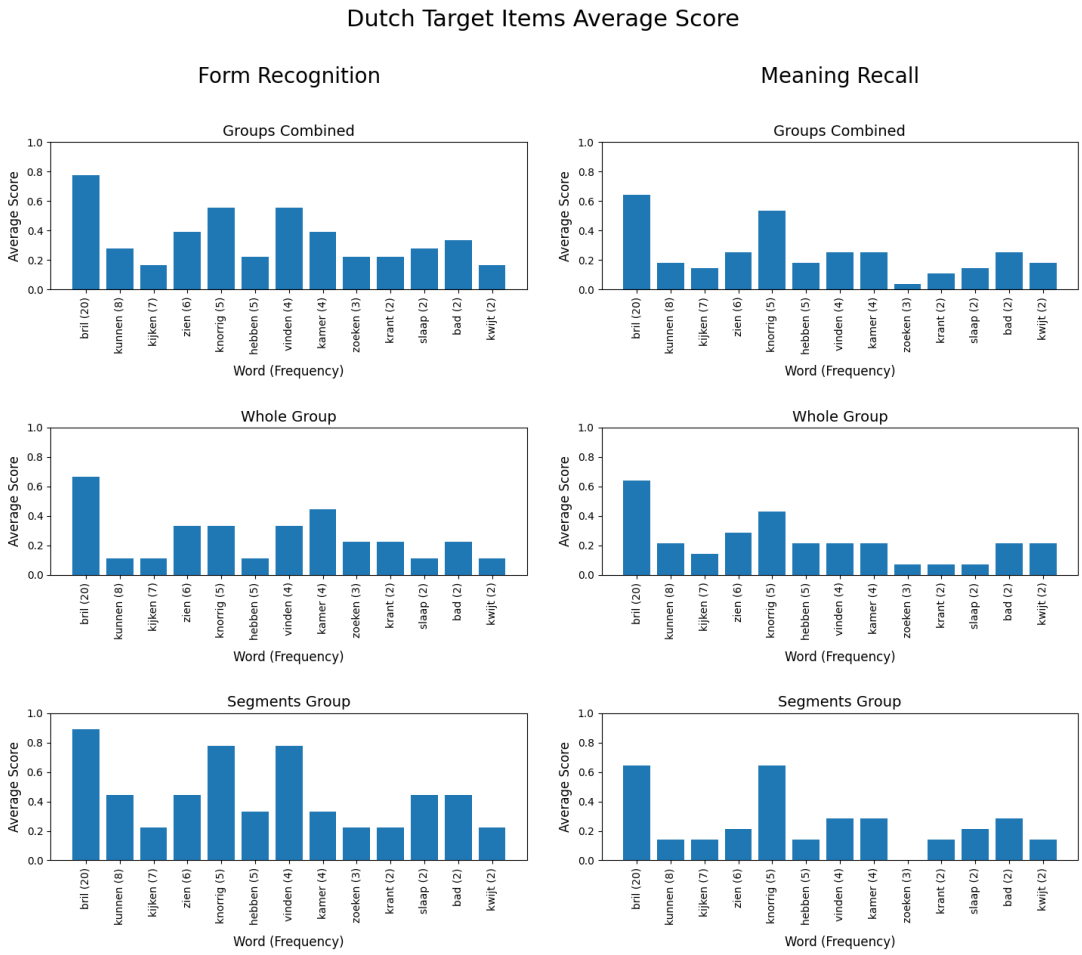

Figure 3.4: Dutch target items' average scores, arranged by frequency.

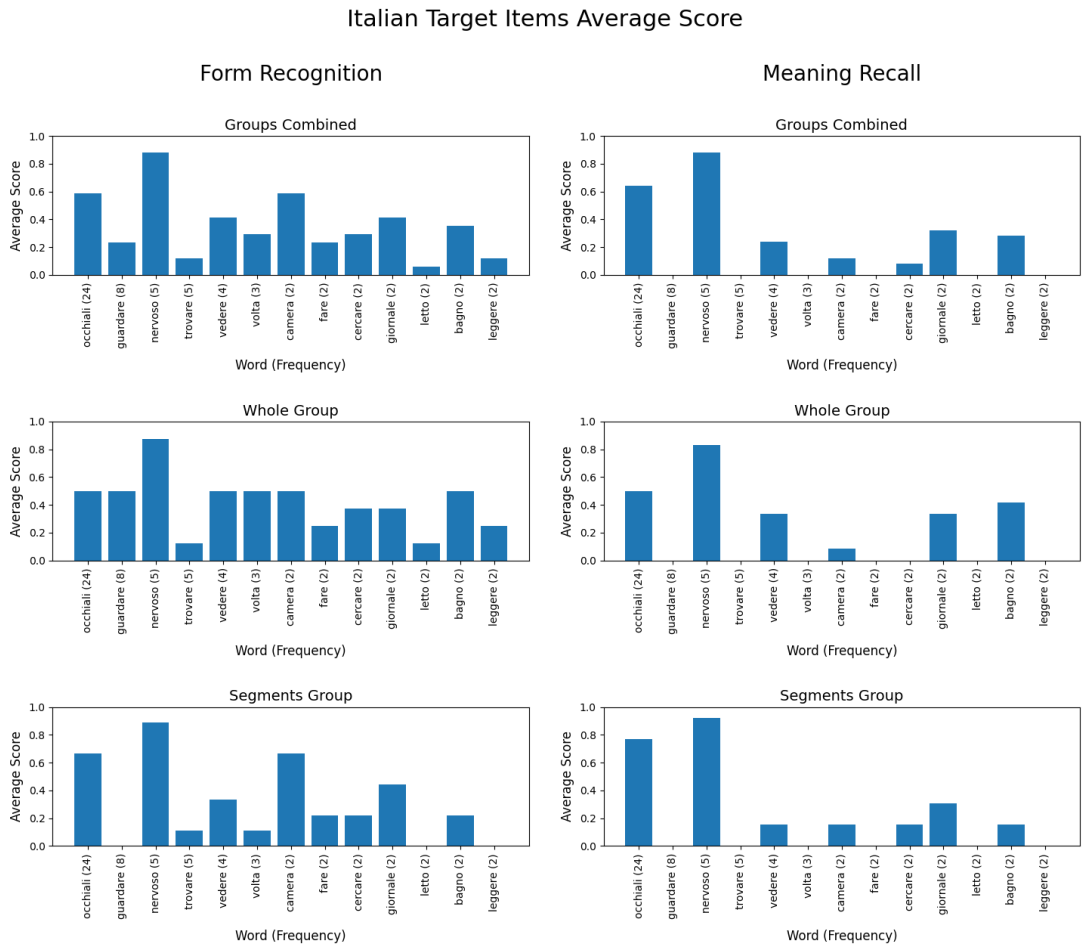

Figure 3.5: Italian target items' average scores, arranged by frequency.

To quantitatively analyse the effect of frequency of occurrence on learning, the correlation (both Pearson (r) and Spearman (ρ)) between the averages scores of the target items and their frequency of occurrence was calculated, shown in Table 3.4.

Table 3.4: Pearson and Spearman correlation coefficients for target items, measuring the correlation between items' frequency of occurrence and average form/meaning scores for both experimental groups.
Format: (r, ρ)

Language	**Recognition**		**Meaning**	
	Whole	**Segments**	**Whole**	**Segments**
Languages Combined	(0.34, 0.19)	(0.44, 0.32)	(0.50, 0.40)	(0.57, 0.23)
Dutch	(0.66, 0.23)	(0.58, 0.40)	(0.81, 0.57)	(0.58, 0.18)
Italian	(0.22, 0.43)	(0.35, 0.14)	(0.36, 0.27)	(0.57, 0.17)

The overall correlation between frequency of occurrence of all target items and their recognition scores across all experimental participants were (r=0.46, ρ=0.28) and for meaning recall were (r=0.56, ρ=0.34).

The moderate but wholly positive correlations suggest frequency of occurrence has a facilitative effect on vocabulary learning in multilingual AVI settings.

3.5.1.3 Previous Language Experience Subgroup Analysis

A short descriptive analysis using subgroups was carried out to explore the effect of previous language experience on learning two new L2s from captioned AVI. It is included as Appendix G.

3.5.2 Quantitative Analysis

3.5.2.1 Vocabulary Scores

To address **RQ4**, the target vocabulary scores of the three groups were analysed across different categories using ANOVAs and Kruskal-Wallis tests, depending on the data distribution. Form recognition and meaning recall were assessed separately as some form recognition results were removed from the dataset. The analyses included both scores for languages combined and separate.

Due to the limited group sizes, Q-Q plots were chosen as the most suitable method to assess the normality of vocabulary scores. The Q-Q plots for form recognition and meaning recall scores for the three groups are shown in Figure 3.6 (both languages combined), Figure 3.7 (Dutch only), and Figure 3.8 (Italian only).

The plots demonstrate the scores are approximately normally distributed with overall minor variance attributable to the small sample size. Although not perfect, they were deemed satisfactory for a pilot study. The one exception was Dutch meaning recall scores for the control group – as all participants scored 0, the use of ANOVA was inappropriate.

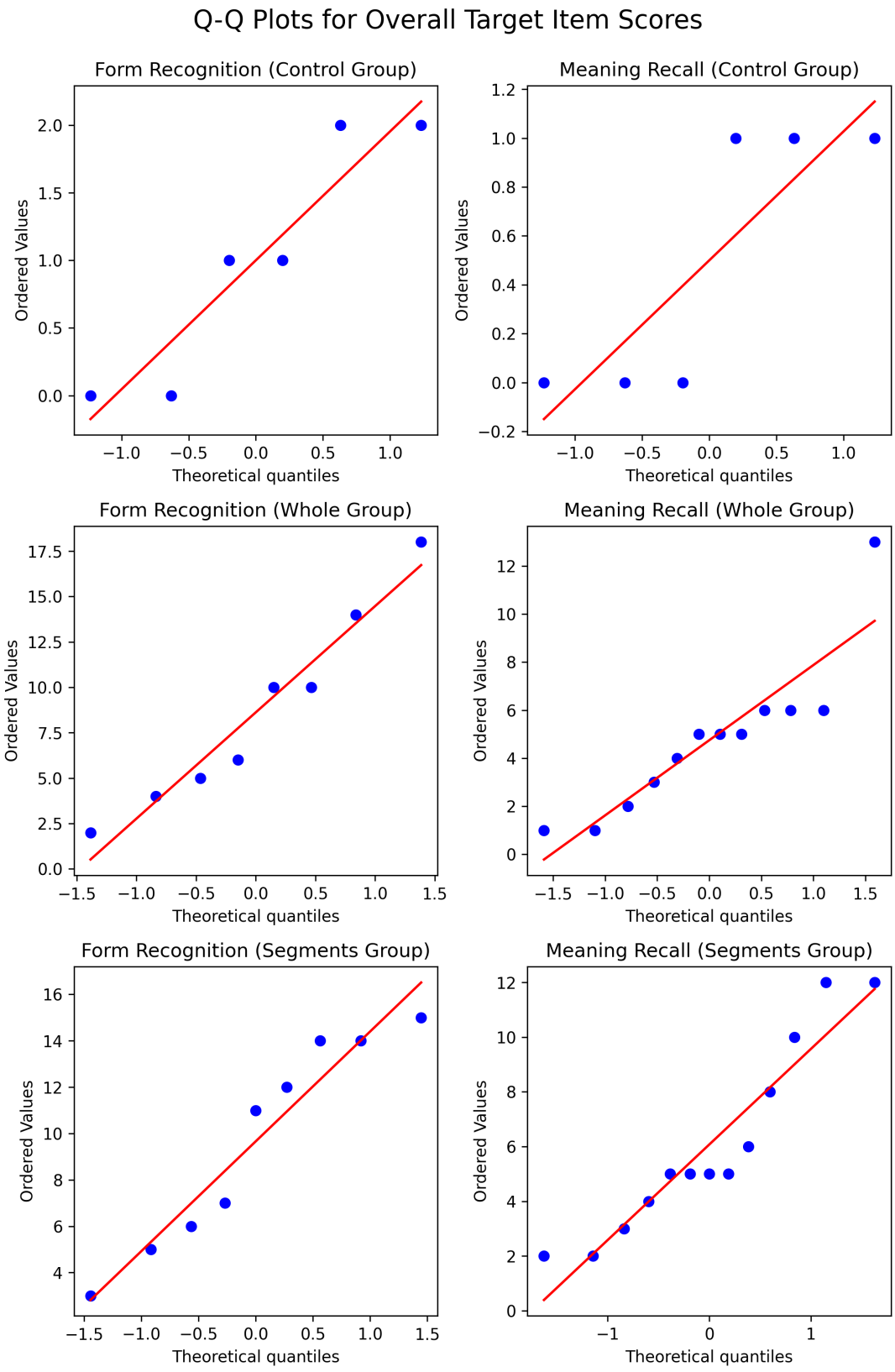

Figure 3.6: Q-Q plots for target item scores (both languages combined).

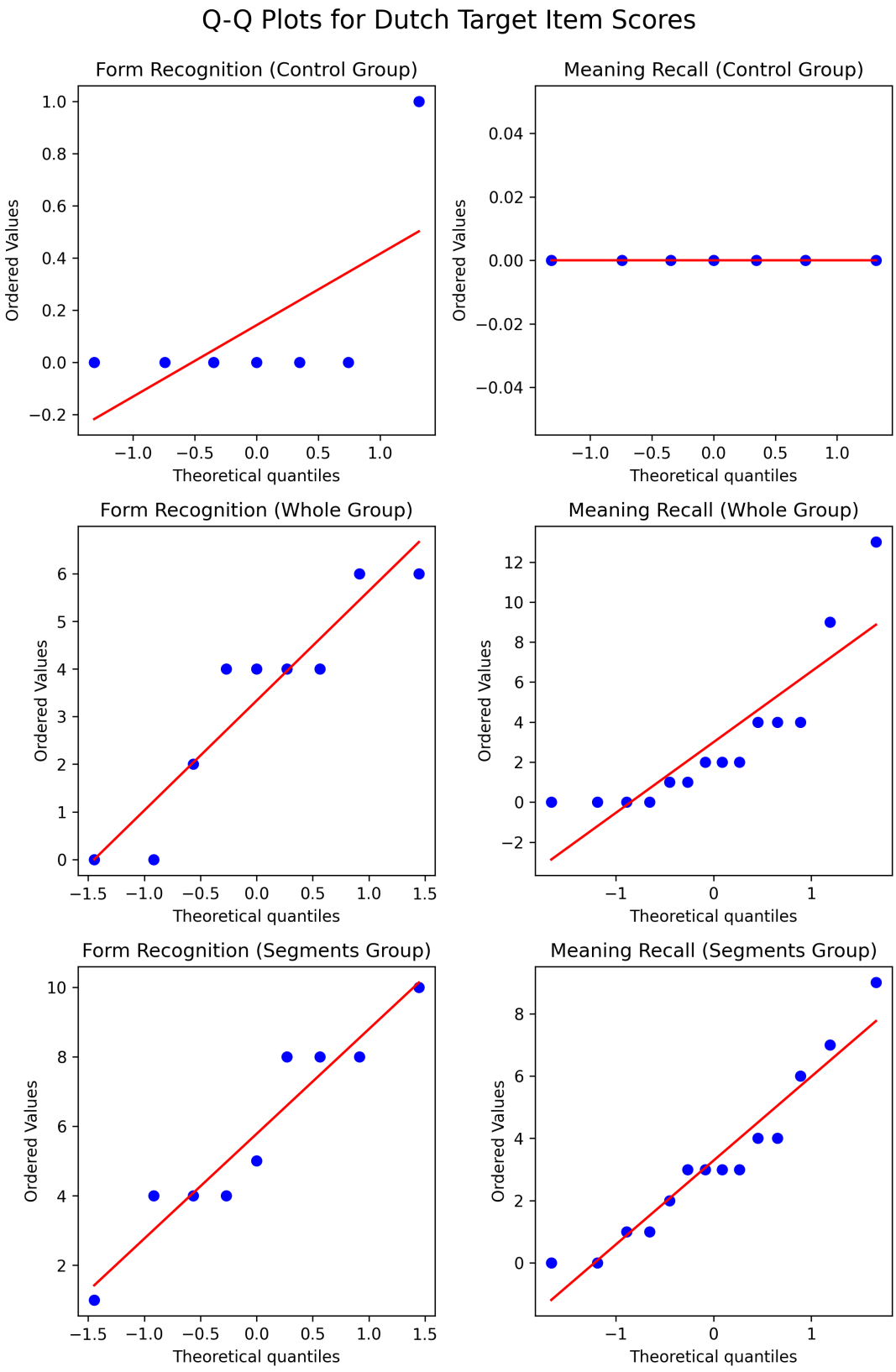

Figure 3.7: Q-Q plots for Dutch target item scores.

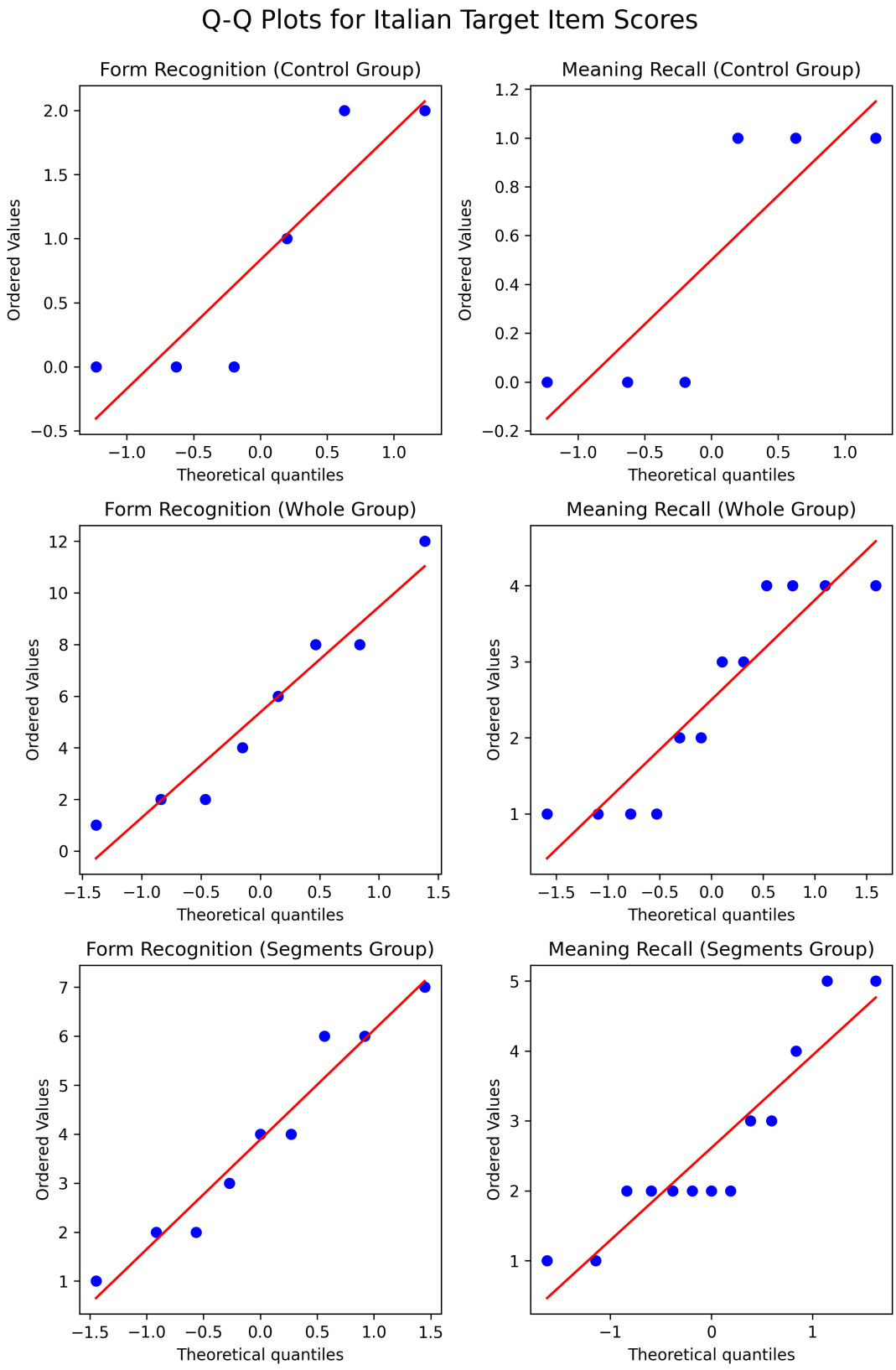

Figure 3.8: Q-Q plots for Italian target item scores.

The assumption of homogeneity of variances between the groups for the various vocabulary scores was tested using Levene's test. The results are shown in Table 3.5. Significant differences in variance between groups were found for all form recognition categories. No significant differences in variance were found for meaning recall for Italian and both languages combined. This meant the use of ANOVA to test for significant difference in vocabulary scores between the groups was appropriate only for meaning recall scores in Italian and with both languages combined. For form recognition scores, and Dutch meaning recall scores, the non-parametric Kruskal-Wallis test was used instead.

Additionally, Levene's test was carried out for the two experimental groups, excluding the Control Group's scores (see Table 3.6). No significant differences in variance were found, which meant the use of t-tests to test for significant difference between the vocabulary scores of these groups was appropriate.

Table 3.5: Levene's test results for target item scores (all 3 groups). Format: (Levene's test statistic, p-value).

Language	**Form Recognition**	**Meaning Recall**
Languages Combined	(5.02, 0.02)	(1.64, 0.21)
Dutch	(4.66, 0.02)	N/A
Italian	(5.32, 0.01)	(1.42, 0.26)

Table 3.6: Levene's test results for target item scores (2 experimental groups). Format: (Levene's test statistic, p-value).

Language	**Form Recognition**	**Meaning Recall**
Languages Combined	(0.22, 0.65)	(0.15, 0.70)
Dutch	(0.98, 0.34)	(0.37, 0.55)
Italian	(3.87, 0.07)	(0.48, 0.50)

The results of the ANOVAs and Kruskal-Wallis tests analysing the target item scores of the three groups are presented in Table 3.7. Significant differences were observed across every test, which includes form recognition and meaning recall scores for both languages independently and when combined.

T-tests comparing the vocabulary scores of the two experimental groups (shown in Table 3.8) found no significant differences, suggesting that the two viewing methods had similar effects on vocabulary acquisition. However, the results approached significance for Dutch form recognition.

Next, post-hoc tests were conducted on target item scores for all three groups. For the form recognition scores and the Dutch meaning recall scores analysed with the Kruskal-Wallis test, Dunn's test with Bonferroni correction was used. For the other meaning recall scores, analysed with ANOVAs, Tukey's HSD test was used. Only significant results are reported, shown in Table 3.9.

Table 3.7: ANOVAs & Kruskal-Wallis tests for target item scores (all 3 groups).
Format of Kruskal-Wallis results: (H-statistic, p-value).
Format of ANOVA results: (F-statistic, df (within groups), p-value).
*Kruskal-Wallis, not ANOVA.

Language	**Form Recognition (Kruskal-Wallis)**	**Meaning Recall (ANOVA)**
Languages Combined	(12.63, 0.002)	(7.10, 0.003, 28)
Dutch	(13.88, 0.001)	(11.31, 0.003)*
Italian	(9.15, 0.01)	(6.94, 0.004, 28)

Table 3.8: *t*-tests on target item scores (2 experimental groups).
Format: (*t*-value, p-value).

Language	**Form Recognition**	**Meaning Recall**
Languages Combined	(-0.43, 0.67)	(-0.10, 0.33)
Dutch	(-2.02, 0.06)	(-0.23, 0.82)
Italian	(1.01, 0.33)	(-0.22, 0.83)

Table 3.9: Post-hoc test results for target item scores.
*Dunn's test, not Tukey's HSD

Language	**Form Recognition (Dunn's test)**	**Meaning Recall (Tukey's HSD test)**
Languages Combined	Control vs Whole: 0.01 Control vs Segments: 0.002	Control vs Whole: 0.02 Control vs Segments: 0.002
Dutch	Control vs Segments: 0.0006	Control vs Whole: 0.02* Control vs Segments: 0.002*
Italian	Control vs Whole: 0.01 Control vs Segments: 0.04	Control vs Whole: 0.008 Control vs Segments: 0.004

3.5.2.2 Covariate Model with Prior Vocabulary Knowledge

Participants' scores on distractor items was used as an approximation of prior vocabulary knowledge in a covariate analysis included in Appendix H.

3.5.3 Qualitative Analysis

Finally, we analyse the results of the Likert scale post-questionnaire completed by the experimental group and summarise the feedback from participants to judge learners' reaction to watching AVI in multiple languages, which addresses **RQ6**.

The results from the post-questionnaire are summarised in Table 3.10. The five options – 'Strongly Disagree', ' to 'Strongly Agree' – were transformed to an integer scale (-2, -1, 0, 1, 2). The distributions of responses to each question are

shown in Appendix I.

Both groups found the content relatively engaging and at a good level for learning (Questions 1–4). The Whole Group rated the learning method of the experiment slightly more favourably when compared to traditional classroom learning than the Segments Group did (Question 11). The appropriate difficulty of the episode was noted by two participants.

Both groups responded that seeing the video first in English helped them understand the Dutch version of the episode, with nine participants in the Whole Group 'strongly agreeing'. Two members of the Segments Group noted finding the Dutch episode easier as it directly followed the English version.

The Whole Group also believed English previewing helped them understand the Italian video, while the Segments Group gave more mixed, but still broadly positive, responses. Additionally, two Segments Group participants noted the Italian video was more difficult as they had forgotten what the English had said – the Dutch segments were always shown before Italian. In the vocabulary tests, the Whole Group had slightly better knowledge of the Italian distractor items. They also recognised more target items but, interestingly, the Segments Group actually performed slightly better on Italian meaning recall scores on target items. The interleaved segments might have created a 'desirable difficulty', which gave the illusion that learning was less effective than it actually was (cf interleaving in Section 2.6.3).

The more polarised responses to Q7–8 suggest previous language experience had a relatively significant effect in understanding of the Dutch and Italian in the video. A number of participants noted their experience in German helped them in understanding Dutch, and Spanish in Italian. Spanish learners were especially likely to say their previous language experience helped in Italian and had high distractor item scores in Italian, while German learners 'strongly disagreed' their previous language experience helped in Italian (see Appendix G).

Many participants would have preferred to watch the video to learn only one language, with half of participants in the Segments Group 'strongly agreeing'. The Segments Group also would have preferred to pause/rewind the video more than the Whole Group.

Table 3.10: A summary of the results of the Likert scale post-questionnaire.

#	**Question**	**Whole Group** μ (σ)	**Segments Group** μ (σ)
1	I found the material to be engaging.	0.86 (1.03)	0.57 (1.02)
2	I found the material to be at a good level for learning.	1.14 (0.77)	1.29 (0.83)
3	I felt confused watching the video.	-0.86 (1.17)	-0.71 (1.2)
4	The video was too quick for me.	-0.5 (1.29)	-0.57 (1.4)
5	Seeing the material in English first helped me to understand more Dutch.	1.36 (1.08)	1.36 (0.5)
6	Seeing the material in English first helped me to understand more Italian.	1.57 (0.85)	0.86 (1.03)
7	My previous language experience helped in understanding the Dutch in the episode.	0.07 (1.33)	0.07 (1.9)
8	My previous language experience helped in understanding the Italian in the episode.	0.86 (1.29)	0.29 (1.59)
9	I would have preferred to pause/rewind the video.	0.36 (1.15)	0.86 (1.23)
10	I would have preferred to watch the episode in only one foreign language.	0.71 (0.99)	0.93 (1.44)
11	I prefer learning like this to traditional classroom learning.	0.79 (1.12)	0.29 (1.44)

3.6 Discussion

This was the first experiment to investigate simultaneous language learning from audiovisual input in multiple languages. It used three novel techniques in AVI research:

- Previewing a target video in learners' L1.
- Repeating a video in another language to learn two languages at once.
- Dividing a video into short segments that repeat in multiple languages.

The findings indicate that repeated exposure to the same video can lead to simultaneous vocabulary acquisition in two languages. Previewing the material in English (the participants' L1) facilitated comprehension of the video and likely led to improved vocabulary gains in both languages. Dividing the video into segments that repeat led to similar vocabulary gains as watching the full video through in each of the languages.

RQ4 – The significant difference in vocabulary scores between the Control Group and both experimental groups, even when controlling for prior knowledge, suggests captioned AVI repeated in multiple languages can lead to incidental vocab-

ulary acquisition in two languages at once.

No significant difference was found between the two viewing methods – viewing whole episodes or short segments – indicating both are valid methods for learning from parallel AVI. When controlling for prior knowledge, the Segment Group's combined meaning recall score was higher but by an insignificant amount.

RQ5 – Word- and learner-related factors that influence learning vocabulary from captioned AVI appear to also apply when learning two languages at once, with analysis focused on the influence of frequency of occurrence and prior vocabulary knowledge.

Words with a higher frequency of occurrence were generally learned better in both languages. In particular, the word for "glasses" in Dutch ("bril") and Italian ("occhiali"), which both appeared at least 20 times, were learned by the majority of participants. The unusual Dutch word "knorrig" ("grumpy"), which appeared 5 times, was also learned by a large number of participants.

Prior vocabulary knowledge, approximated by participants' combined distractor item meaning recall scores on both languages, was a significant predictor of combined target item meaning recall scores. It seems likely that prior knowledge facilitated vocabulary acquisition as in normal AVI watching conditions, but without a pre-test it can't be determined what target items were learned and what were known before the experiment.

RQ6 – Participants reacted positively to the video, generally finding it engaging and at a good level for learning. The Segments Group especially would have preferred to watch the video in one language only, and to have the option of pausing or rewinding the video.

The participants found watching the material first in English helped them to understand both the Dutch and Italian versions. Some Segments Group participants noted that watching in Dutch was easier as it came right after the English version, and that watching in Italian was harder as they forgot what was said in English.

Participants reported finding Dutch and Italian easier or harder to understand depending on their previous language experience. Spanish learners in particular agreed their previous language experience helped in understanding the Italian episode, while German learners strongly disagreed. Several participants noted their experience in German helped understanding Dutch, and others that Spanish helped in understanding Italian. The effect of cognates was also noted by several participants.

3.6.1 Experimental Design

Incorporating a pre-test and having a larger sample size would greatly benefit a follow-up study of this kind. They would allow for more reliable results and measurable learning gains. A delayed post-test would allow measuring retention of learning.

The study used a control group to answer vocabulary tests without seeing the

episode, aiming to estimate the baseline knowledge of target items across the sample, similar to Montero Perez (2020). The use of new languages for the participants aimed to minimise the influence of prior knowledge. Unfortunately, the control group was small (only seven participants) and they performed significantly worse on the distractor items, indicating lower proficiency. This made it difficult to accurately estimate the learning gains of the experimental groups. This limitation was somewhat mitigated by the covariate analysis which controlled for prior knowledge.

There were higher meaning recall scores for Dutch words from the episode despite evidence of higher prior knowledge in Italian. This was perhaps due to watching in Dutch first, which several participants noted helped them understand the Dutch, with others noting it hindered their understanding in Italian. Reversing the order of the Dutch and Italian episodes/segments should also be considered in a follow-up study. During the study's design phase, it became evident that introducing multiple languages, viewing methods, and different viewing orders can quickly result in an unmanageable number of groups and variables, especially for a pilot study.

A within-subjects design was also considered using two episodes, where each participant would watch one shown in full and the other in segments. Four groups would start with different episode/method combinations, serving as controls for each other to somewhat balance out bias from prior knowledge. If one method works significantly better, it should result in greater vocabulary gains across both episodes given a large enough sample size. This approach would also likely provide better feedback, as each participant would experience both viewing methods.

Finally, future work could explore different viewing orders (e.g. L3->L2->L1) and repeated viewings (L1->L2->L1->L3 or L2->L1->L2). The repeated viewing of videos can enhance learning from AVI (Webb, 2015).

3.6.2 L1 Previewing

Participants reported that the English previewing aided understanding, particularly for the Dutch video/segments that followed immediately after the English. The fact that correct meaning recall answers often relied on English equivalents (e.g., "grumpy" for "knorrig") suggested the English previewing helped participants learn target vocabulary.

Studies in bilingual reading suggest that the use of the L1 to 'scaffold' learning when reading bilingual texts facilitates language acquisition (Butzkamm, 2003; Ulanoff and Pucci, 1999; Zhang and Webb, 2019). The extent to which L1 previewing 'scaffolds' learning from L2 videos needs to be tested in isolation in a follow-up study, such as the one suggested in Section 3.1.1.

The technique of 'sandwiching' (Dodson, 1967) could be applied to segmented videos by playing a segment in an L2, then in the L1, and again in the L2.

Studies in bilingual reading have shown that both L1->L2 and L2->L1 reading orders can lead to similar learning (Zhang and Webb, 2019). This approach could

be applied to AVI.

Finally, the Control Group should have perhaps been shown the English version of the episode. Familiarity with the context of the target items likely made the vocabulary tests somewhat easier for the experimental groups.

3.6.3 Survey

There were a number of weaknesses to the anonymous, online survey. Performing a confidential survey instead would have allowed for contacting participants to address incorrectly filled out answers, e.g. 'D2' level of language proficiency. An in-person experiment would have ensured consistent viewing conditions for every participant, avoided participants misinterpreting the form recognition question, and could have yielded more detailed feedback.

3.6.4 Participants

An advantage of the study was its uniform sample, comprising predominantly 24/25-year-old monolingual English speakers from Ireland and the UK, who had similar educational backgrounds.

The use of brand new languages was intended to minimise prior knowledge bias. Dutch proved to be a suitable choice, whereas participants demonstrated a higher prior knowledge of Italian. Using completely unfamiliar languages in a future experiment was suggested by one of the participants.

Most participants were not regular language learners, and hadn't learned a language since school. This study gives evidence that even complete beginners to a language can develop vocabulary in two languages simultaneously with multilingual AVI, in less than 15 minutes.

The omission of Dutch scores of participants who studied Italian before might have been more appropriate, as viewing the episode in Italian may have affected their Dutch learning differently to other participants.

Proficiency levels of participants were not used due to the small sample size. Overall proficiencies were generally low. Future analysis could explore the impact of varying levels of previous language experience.

3.6.5 Materials

Peppa Pig proved to be an appropriate choice for the experiment. The episode led to vocabulary learning gains for complete beginners and participants noted it was material at an appropriate level. This correlates with research suggesting *Peppa Pig* is excellent input for vocabulary development in early learners (Alexiou, 2015; Kokla, 2022, 2021; Scheffler et al., 2021).

Many participants learned the unusual word "knorrig". This ties in with research suggesting uncommon words used in *Peppa Pig* tend to be 'fully contextualised' and thus easier to learn (Scheffler et al., 2021).

3.6.6 Vocabulary Tests

Given the limited scope of the study, the hybrid use of distractor words for participant motivation and an approximation of prior knowledge was practical. Ideally, prior knowledge would have been measured by a proficiency test separately before the experiment. However, performance on distractor words as an approximation of prior vocabulary knowledge functioned as a significant predictor of target items scores.

There were a number of issues with selected target items:

- Including the English cognate "nervoso", although one of the most frequent target items in the episode, was problematic as it skewed the word frequency correlation analysis.
- Participants reported recognising Italian words like "camera" and "fare" that resemble English words, but didn't know the correct meaning.

The form recognition question should have been more clearly defined to avoid misinterpretation. Alternatively, simply using meaning recall by itself could have led to a more simplified data analysis.

The Vocabulary Knowledge Scale (VKS) (Paribakht and Wesche, 1993) would have been less open to misinterpretation by participants, but it does not allow guessing the meaning of unrecognised words, which could be understood because of cognates in a participants' other language(s). This was an important factor to consider as the study measured learning in new languages for participants. There were 30 instances where participants identified the meaning of words they said they had not seen before (25 for Dutch and 5 for Italian). Nearly all were control words.

Should vocabulary measures combine languages or assess each separately? Combining scores on each language could provide an overall picture of learning gains; leaving them separate could reveal variations in learning outcomes per language.

Short vocabulary tests (only 13 target items in each language) were used to minimise the experiment's duration. However, participants could have learned other content words that weren't tested. A solution could have been to include a text box for participants to list any other words they recalled, and their meanings. This approach could also introduce an additional form recognition point for remembered words and a meaning recall point for correctly identified meanings.

3.6.7 Data Analysis

The high number of variables made it more difficult to process, describe, and graph the results.

Choosing to combine the frequencies of multiple forms of a word was problematic. For example, "kunnen" ("to be able to") had a frequency of 8 when 5 of its occurrences were of the form "kan" ("can"), which beginners may perceive quite differently as it closely resembles its English cognate. Combining frequencies impacted the frequency of occurrence analysis in Section 3.5.1.2. For example,

"knorrig", the second most learned Dutch word by meaning, was fifth in number of occurrences with this method, but second if multiple forms weren't counted together. It's difficult to know how to account for words with multiple forms all used in the same video.

Also, around half of the Italian target items occurred twice in the episode, which somewhat skewed the frequency of occurrence correlation analysis results.

3.6.8 Segmenting

The results suggest that both viewing methods lead to vocabulary development in two languages at once.

Using greater differences in segment size, e.g. 10 seconds, 3 minutes, and 60 minutes (for a TV show), would likely reveal more pronounced differences between groups.

The point in the episode where target items occurred may have affected learning differently in the two groups. For example, an Italian target word that occurred once in the first minute of the episode would be viewed around 4 minutes before the end of the entire video in the Whole Group; for the Segments Group this word would be viewed more than 11 minutes before the end.

Many interesting experiments are possible using segmenting. One could use an L1 previewing for comprehension, followed by a segmented episode where segments play in alternating languages, then the segmented episode again with the languages of the segments swapped.

Many segments of the video did not have target items in them. An interesting experiment might be to highlight specific words in each segment to learn, potentially the same word across many languages. This technique, 'textual enhancement', is supported in research (Montero Perez, 2022, p. 173).

3.6.9 Word-Related and Learner-Related Factors

The facilitative effect of various word- and learner-related factors on vocabulary acquisition from AVI is well-documented (see Montero Perez, 2022), such as cognateness, frequency of occurrence, and prior knowledge. These factors, especially frequency of occurrence and prior knowledge, were explored in a multilingual setting in this study. Although a cognate analysis was not conducted, several participants noted the impact of cognates.

Additionally, research on multilingualism supports the idea that experience with previous languages, especially those similar to the L2, aids learning (Cenoz and Gorter, 2022; Festman, 2021). The facilitative effect of previous languages similar to Dutch or Italian was also documented in this study, however a comprehensive analysis of the interaction between previous language experience and the target languages was not feasible for this small-scale study.

In line with modern practices in measuring the effect of word-related and learner-related factors (Montero Perez, 2022), a larger multilingual AVI study could use logistic regression or a mixed effects model to include all these variables in one

comprehensive model. To account for prior study in other languages that may facilitate learning through language transfer, proficiency test results could be incorporated into such a model.

3.6.10 Future Experiment Ideas

Finally, we conceptualise some future multilingual AVI experiment ideas. Future study could investigate:

- Allowing pausing, rewinding, or rewatching videos.
- Pre-reading of the captions of an episode in parallel.
- Trilingual subtitles overlaid on all three viewings.
- The learning of multi-word units.
- Testing understanding of full L2 sentences in the input, which learners may be able to match to equivalent lines in the L1 version of a video.
- Whether learners exhibit an 'illusion of competence' when learning through whole episodes (resembling blocked practice) versus segmented episodes (resembling interleaved practice), as found in interleaving studies (see Section 2.6.3).

4. Language Learning Computer Program

The central achievement of this thesis is the creation of a multilingual study tool, which breaks down audiovisual input with multiple language audio tracks and matching subtitle files into an interactive multilingual text, making the vast collection of multilingual videos available online accessible for simultaneous multilingual learning.

As described in Appendix A, sophisticated sentence mining tools exist that support language study from audiovisual content through pop-up dictionaries for subtitles and one-click creation of media-rich vocabulary flashcards. However, no system allows for the simultaneous study of multiple languages.

With the computer program introduced in this section, learners can create interactive parallel texts showing dialogue in multiple languages from any target video, with access to translation and dictionaries in multiple languages to support learning. Users can listen to segments of dialogue and create multimedia flashcards for learning vocabulary and phrases that include images and audio from the video being studied. Learners can also create condensed audio files from AVI dialogue that can interleave segments in multiple languages.

The description of the computer program and its use throughout this section answers **RQ8**. The explanations of how subtitle files and multilingual videos files are broken down, aligned across languages, and manipulated to create both parallel texts of dialogue and interleaved condensed audio files form the answer to **RQ7**. The versatile flashcard format and gallery of possible flashcards that can be created with the system detailed in Section 4.4.1 answer **RQ9**.

First, we present a high-level overview of the program. We introduce the three 'modes' available to learners, the software architecture, the technologies used in building the system, and the languages currently supported.

The subsequent sections detail chronologically, from input to output, how the system works, and how it is used. We look at the input to the system and how it is processed (Section 4.2), the components of the user interface (UI) (Section 4.3), and present the potential outputs of the system (Section 4.4).

Future development opportunities of the system are acknowledged in Appendix J.

4.1 System Overview

The system has three modes to choose from:

- The **Text Mode** is the basic study mode of the application, used to aid reading target language texts (websites, e-books, etc.). Users can save sentences found while reading and create basic text flashcards, facilitating 'sentence mining' from foreign texts. It is described in Section 4.3.1.

- The **AVI Mode** is the main study mode. It consists of an interactive multilingual text created from any audiovisual input.
- The **Media Export Mode** allows learners to create practice audio files from any audiovisual input.

Upon launching the application, users are prompted to select their preferred settings via the 'Startup Dialog' window. Figure 4.1 shows the steps of the Startup Dialog when choosing the Text Mode – users choose their L1 (only English for now), L2, and dictionaries for their L2.

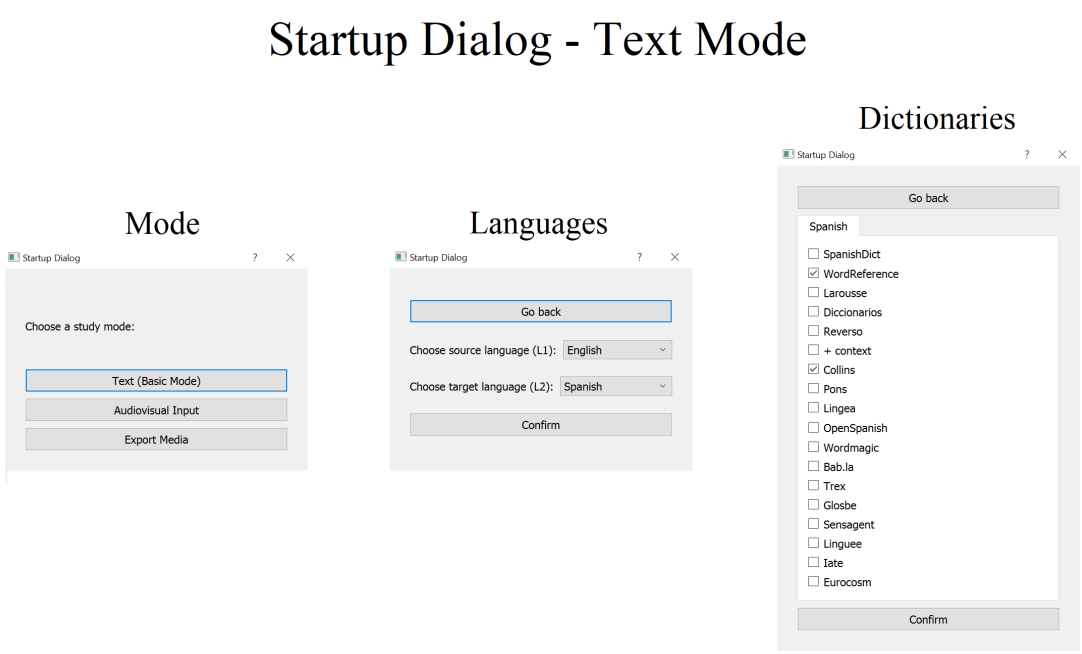

Figure 4.1: The steps of the Startup Dialog when setting up the Text Mode.

4.1.1 Software Architecture

The architecture of the system broadly follows the Model-View-Controller design pattern.

The Model manages the internal representation of information. When studying from AVI, the Model reads the given subtitle files (discussed in Section 4.2.3) and aligns the subtitles across the given languages (discussed in Section 4.2.4). It can also group subtitles together to break the input's dialogue into segments (discussed in Section 4.2.5).

The View is the user interface (UI), what the user sees and interacts with. It consists of all the elements of the UI as discussed in Section 4.3.

The Controller encapsulates the core functionality of the system. It manages the interaction between the View and the Model, passing the subtitle alignment created by the Model to the View to display the subtitles of the input. It also sets up and interacts with various auxiliary components to translate queries, open dictionary searches, extract audio segments and screenshots, and export flashcards.

4.1.2 Technologies Used

The system is built in the Python programming language, with the help of several external libraries:

- The user interface is built with **PyQt5**, a set of Python bindings for the graphical user interface (GUI) toolkit *Qt*. It is a powerful, modern, and cross-platform GUI library.
- **FFmpeg** is a free, open-source suite of tools for processing audio and video files. It is used for extracting the various audio tracks from the chosen multilingual video file, and for extracting the audio that matches a specific subtitle.
- The **Pydub** library is used to create the condensed audio files from any given AVI input file. It's used to extract the segments of audio where dialogue is taking place, i.e. corresponding to the subtitles of the episode, and compile a single output audio file for listening practice.
- **OpenCV** (Open Source Computer Vision Library) is used to extract screenshots to be used in flashcards.
- **BeautifulSoup4**, a library for parsing HTML, was used for processing the data scraped from the *Lexilogos* website to extract the various dictionary links.

4.1.3 Supported Languages

The system works with the 29 languages currently supported by DeepL: Arabic, Bulgarian, Chinese, Czech, Danish, Dutch, Estonian, Finnish, French, German, Greek, Hungarian, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, Turkish, and Ukrainian.

Dictionary data from the website *Lexilogos* was extracted for these languages and more, described in Section 4.3.4. In addition, users can also add their own dictionary links. English is not currently supported as a target language.

4.2 Input

We now discuss the steps required in processing videos with multiple audio tracks and corresponding subtitles files, which pertains to the AVI Mode only. The system currently supports MP4 video files and SRT subtitle files.

4.2.1 File Selection

First, Figure 4.2 shows the AVI Mode's file selection page in the Startup Dialog for two different study sessions. The user chooses:

- The folder containing the video file and subtitle files.
- The video file to study from, inside this folder.
- The 'reference subtitle file', which serves as the 'ground truth' subtitle file. Typically, it corresponds to the original language of the video, and is trusted for its well-timed subtitles and comprehensive coverage of spoken content. Other subtitle files are aligned against this reference.
- The source language, which is the user's L1. Currently, only English is supported, as it is the only language with bilingual dictionaries in each of the target languages.
- The first target language, the first language of study.
- The second target language, which is an optional second language of study. As will be shown later, the system can be trivially extended to more than two target languages.
- The audio tracks of the different languages in the video, which are retrieved from metadata in the video file. This parameter is optional.
- The subtitle files for each language. This is required for each target language. If no reference subtitle is chosen, the first subtitle file selected is used as the reference.

The next step is selecting the dictionaries for each target language, shown in Figure 4.3.

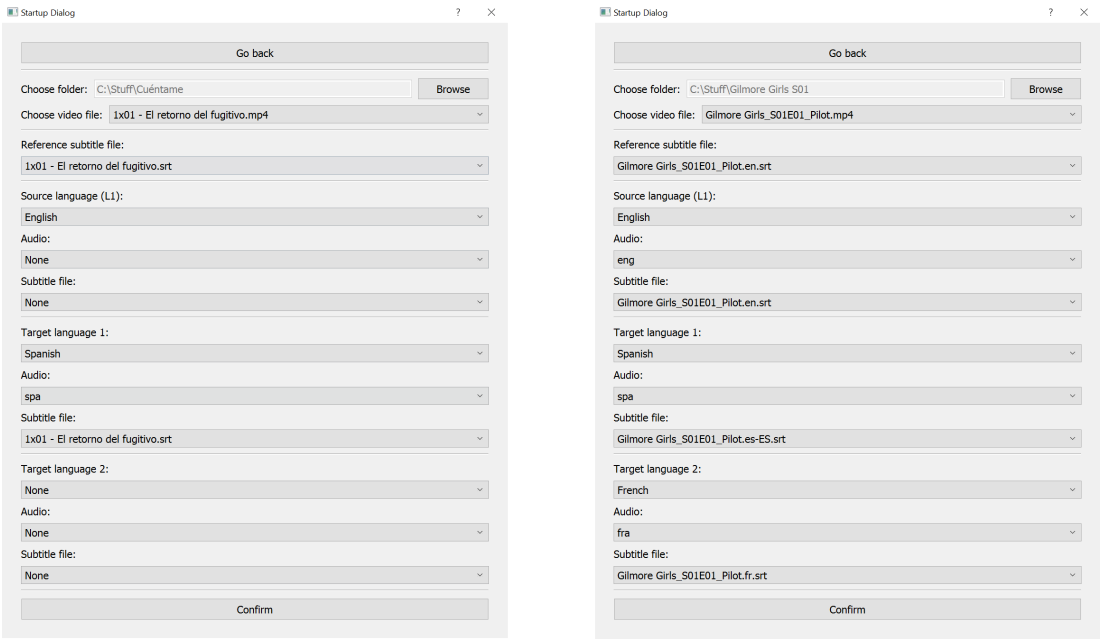

Figure 4.2: Two versions of the file selection screen in the Startup Dialog.
Left: Selecting the first episode of the Spanish series *Cuéntame*, where only Spanish audio and subtitles are available.
Right: Selecting the first episode of *Gilmore Girls*, where subtitles and audio are available for English, Spanish, and French.

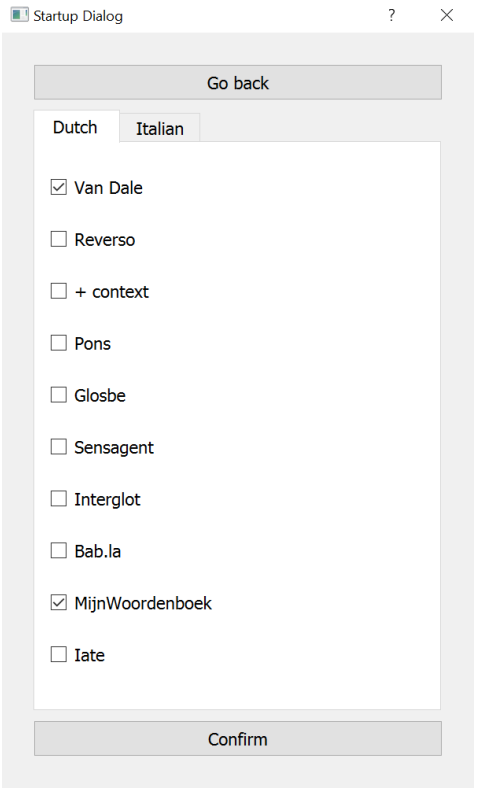

Figure 4.3: Selecting dictionaries to study the episode of *Peppa Pig* from the experiment.

4.2.2 Extracting Audio Tracks

The Controller passes the video file and selected audio tracks to an Audio Extractor object which extracts the audio tracks from the video into standalone audio files for easier access/manipulation later. This step is skipped if a user chooses to study from subtitle files without matching audio tracks.

4.2.3 Subtitle Parsing

The process of parsing subtitle files is described in Figure 4.4. The Model parses each subtitle file by processing its lines one-by-one as detailed by the algorithm in Figure 4.5. This process creates a 'Subtitle Model' for each file. Each Subtitle Model is a list of 'Subtitle' objects, with each object containing the start time, end time, and text of a subtitle.

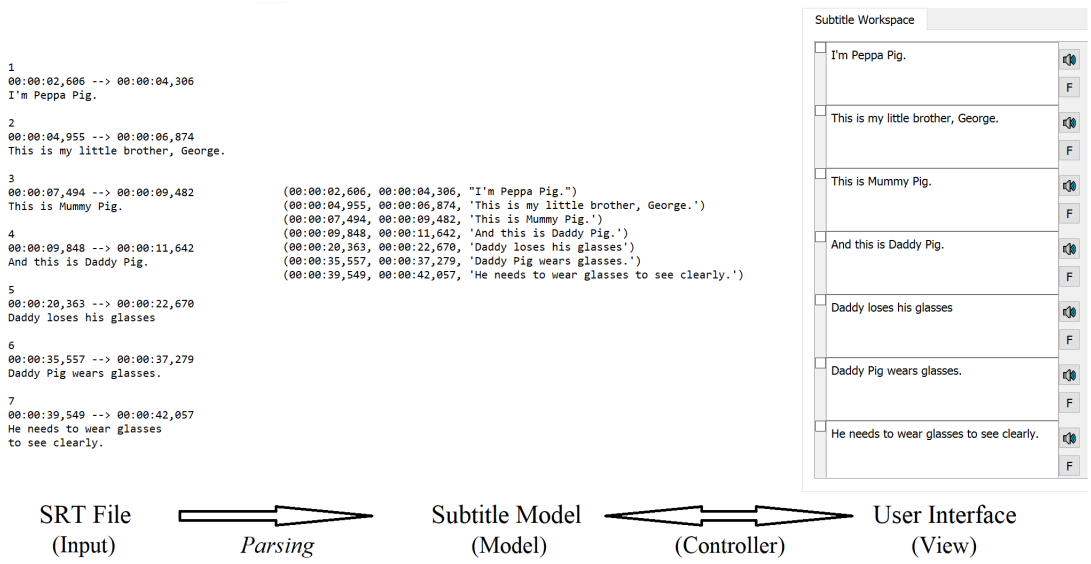

Figure 4.4: Subtitle Models are created by parsing each SRT subtitle file. The Controller manages the interaction between the Subtitle Models and the user interface.

Converting from other subtitle formats to SRT (SubRip) can be done using online tools. The system assumes there are no temporal differences or irregularities in the various subtitle files for a single video. If discrepancies exist, users can manually realign subtitles with tools like *SubtitleEdit*. As subtitles and dubbed lines are often created independently, they can be contradictory. Future solutions could be to use speech-to-text models to create new subtitles, or even to use AI text-to-speech models to record new dubbed audio from the subtitle text.

Algorithm 1 parse_subtitle_file(subtitle_lines, non_speaking_symbols)

Require: *subtitle_lines*: list of str, *non_speaking_symbols*: list of str
Ensure: *subtitles*: list of Subtitle
  subtitles ← empty list
  start_time ← minimum datetime value ▷ Start time of current subtitle
  end_time ← minimum datetime value ▷ End time of current subtitle
  text ← empty string ▷ Text of current subtitle
  for $i \leftarrow 0$ **to** len(*subtitle_lines*) − 1 **do**
    line ← *subtitle_lines*[*i*].strip()
    if line matches "[0-9]+" **then** ▷ Matching a whole number
      if $i = 0$ **or** *subtitle_lines*[$i - 1$].strip() = "" **then**
        continue ▷ Skip subtitle numbers (always at first line or after an empty line)
      else
        text += " " + *line* ▷ Otherwise it's a subtitle of just a number
      end if
    else if " → " is in *line* **then**
      start_time, *end_time* ← parse_subtitle_timing(*line*) ▷ Read subtitle timings
    else if *line* = "" **then** ▷ Seeing an empty line means we have read a full subtitle
      if *text* ≠ "" **then**
        continue ▷ Skipping if subtitle has no text
      end if
      if subtitles ≠ empty **and** (*start_time* = subtitles[−1].*start_time* **or** *start_time* < subtitles[−1].*end_time*) **then** ▷ Merge with previous subtitle if overlapping
        subtitles[−1].*text* += " " + text.strip()
        if *end_time* > subtitles[−1].*end_time* **then**
          subtitles[−1].*end_time* ← *end_time* ▷ Lengthening previous subtitle if necessary
        end if
        text ← empty string ▷ Flushing the text
        continue
      end if
      subtitles.append(Subtitle(*start_time*, *end_time*, text.strip()) ▷ Adding the current subtitle
      text ← empty string ▷ Flushing the text
    else if any symbol in *line* is in *non_speaking_symbols* **then**
      continue ▷ Skip non-speaking subtitles, e.g. that contain "♪" for music
    else
      text += " " + *line* ▷ Otherwise, add the line to our current subtitle's text
    end if
  end for
  return *subtitles*

Figure 4.5: Subtitle parsing algorithm. Subtitle files are parsed line-by-line by creating and saving 'Subtitle' objects. The strict structure of SRT files means a simple rule-based algorithm is all that's required. Users can add 'non-speaking symbols' to omit captions for music or sound effects, which can be written with different symbols

4.2.4 Aligning Subtitles Across Languages

Once each subtitle file has been parsed, the central computational challenge is aligning the subtitles across each of the languages. The Model builds a 'Multilingual Alignment' object, consisting of multilingual subtitles, shown in Figure 4.6. These aligned subtitles are used to create the parallel text.

The Multilingual Alignment begins as a copy of the Subtitle Model of the reference subtitle file. New languages are added to the Multilingual Alignment in turn, one subtitle at a time. The Model collects all the multilingual subtitles with an overlap in subtitle timings with the current subtitle being added, and chooses from among these candidates by one of two methods – selecting the one with the highest overlap in subtitle timings, or the one with the closest start time. The latter has been found to be better in practice so is the default method. Multiple subtitles from a given language may be assigned to the same multilingual subtitle (e.g. the last Japanese subtitle in Figure 4.6). Finally, if no candidate multilingual subtitles are found, a new multilingual subtitle is created, demonstrated in Figure 4.7.

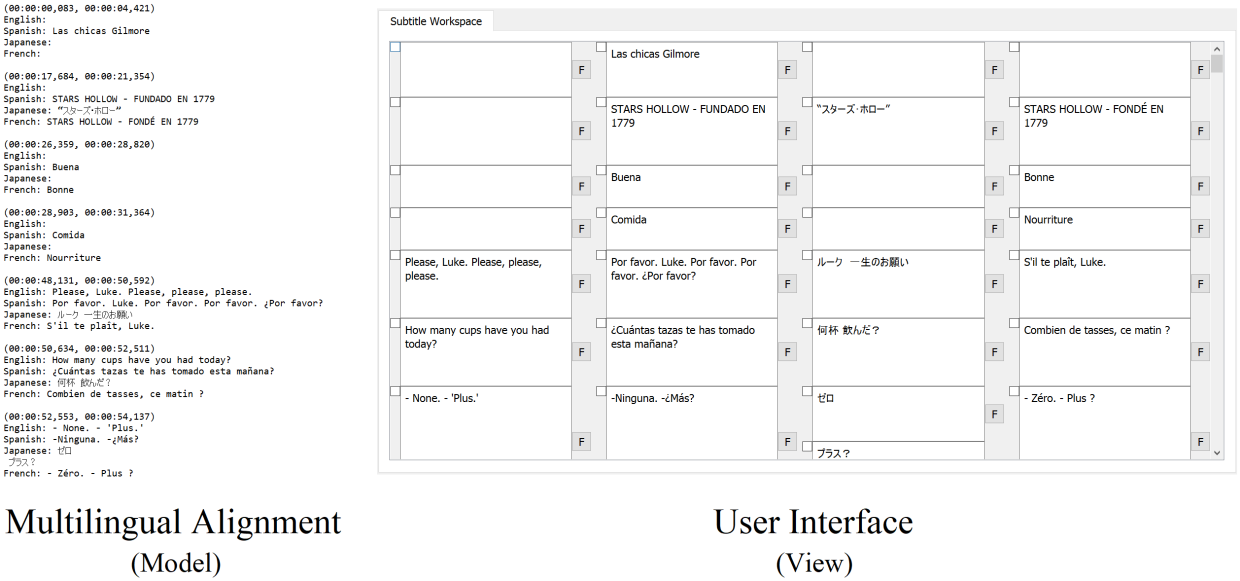

Figure 4.6: An example of the Multilingual Alignment data structure and the corresponding View in the UI.

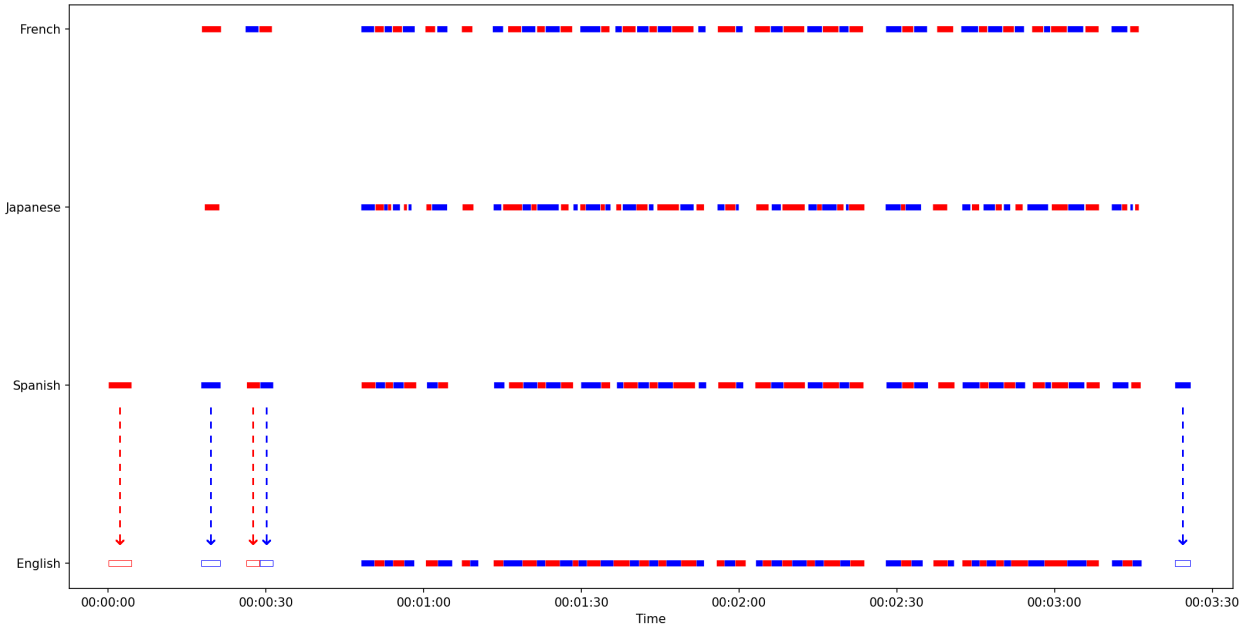

Figure 4.7: A visualisation of the subtitles of the opening scene of the series *Gilmore Girls* in English, Spanish, French, and Japanese. With English (the show's original language) as the reference, new multilingual subtitles are created when aligning Spanish subtitles with no matching English subtitles (shown by the dotted lines). This is reflected in the Multilingual Alignment in Figure 4.6 above.

4.2.5 Dialogue Segmenting

An optional final step is to split the dialogue into segments using a 'seconds between segments' parameter, which the user can freely choose. Each segment is built by combining consecutive subtitles whenever the difference between the end time of a subtitle and the start time of the following subtitle is less than this parameter. When the difference is exceeded, a new segment is created. This result is demonstrated in Figure 4.8.

An additional 'maximum segment length' parameter can be used to prevent segments from becoming overly long. Any segment whose length exceeds this parameter is divided evenly, such that each sub-segment is shorter than the 'maximum segment length'.

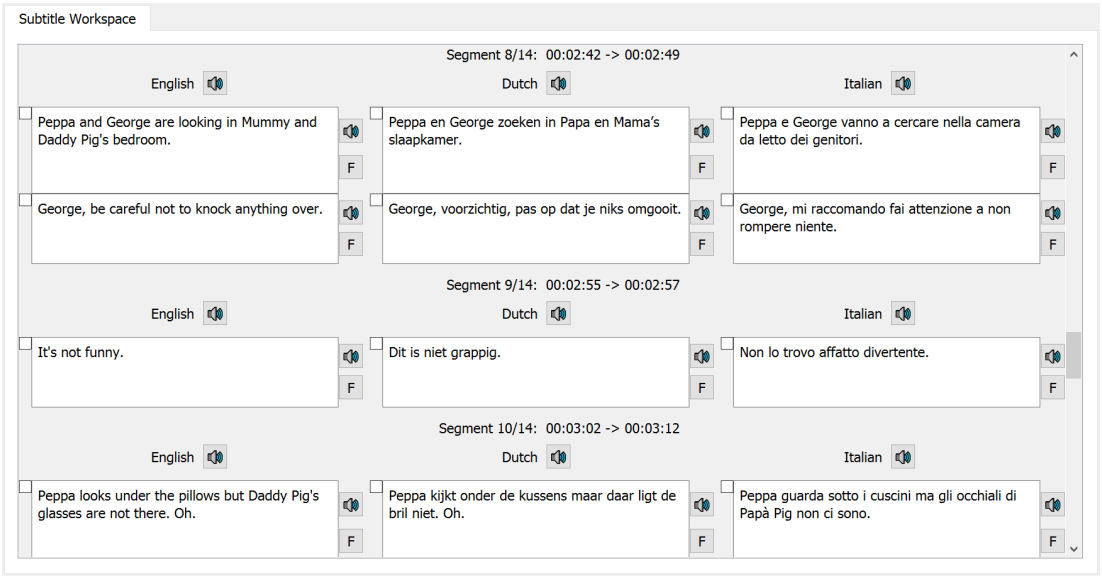

Figure 4.8: Segments created for the *Peppa Pig* episode from the experiment using 3 seconds between segments.

4.3 User Interface

When the various audio tracks have been extracted, and the Model has parsed the subtitle files and built an alignment of the subtitles across the languages, the Controller presents the user interface (UI) to the learner. Here the learner interacts with their chosen learning material. They can read the multilingual text, listen to individual subtitles or segments, translate between languages and look up words, and make a wide variety of flashcards.

The UI is broken down into four windows (shown in Figure 4.9):

- The **Study Materials** window shows the multilingual parallel text when using the AVI Mode, or the user's 'Saved Sentences' when using the basic Text Mode. These are discussed in Section 4.3.2 and 4.3.1 respectively.
- The **Translation Workspace**, discussed in Section 4.3.3, allows users to translate words and phrases.
- The **Dictionary Lookup**, discussed in Section 4.3.4, allows users to look up words.
- The **Flashcard Workspace**, discussed in Section 4.3.5, is where the user creates and customises flashcards for later study.

We first discuss the Text Mode, the basic form of the application. We then discuss the 4 windows of the UI in depth while using the AVI Mode.

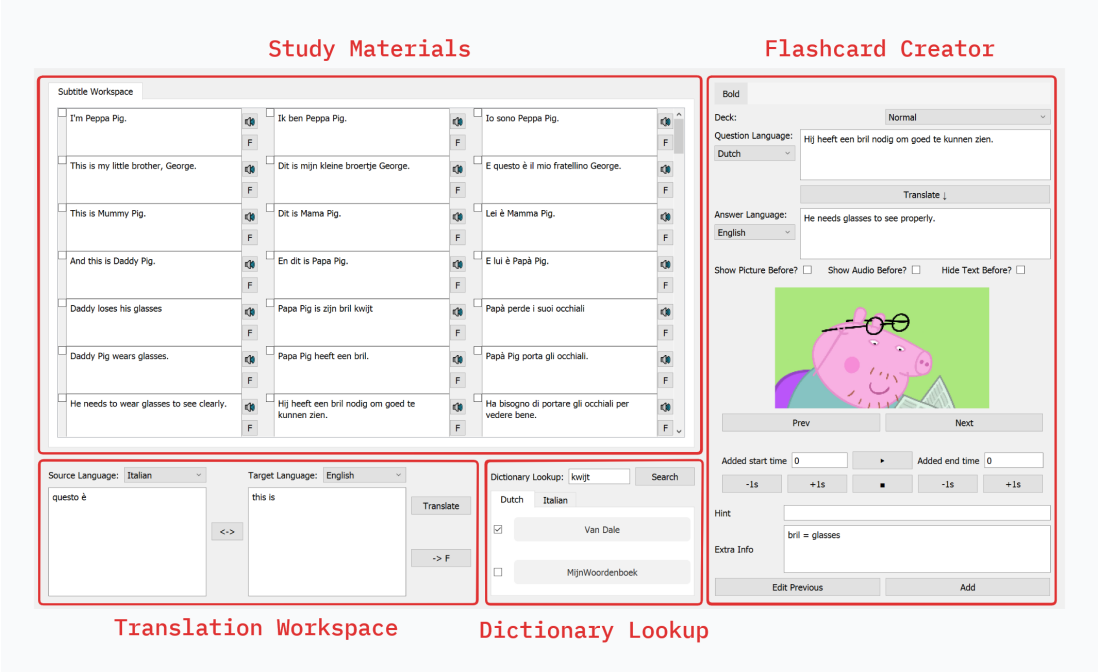

Figure 4.9: The four windows of the user interface (UI).

4.3.1 Text Mode

Using the Text Mode is described in Figure 4.10. Learners can save sentences while reading foreign texts using a free-floating 'Sentence Bin', or save them directly from the clipboard. These sentences are saved in the Study Materials window (titled 'Saved Sentences') and can be translated or moved to the Flashcard Workspace, where basic text flashcards can be created (discussed in Section 4.3.5).

4.3.2 Study Materials (AVI Mode)

The Study Materials window while using the AVI Mode consists of the multilingual parallel text of aligned subtitles (titled 'Subtitle Workspace'). The parsed and aligned subtitles are shown in individual boxes. Three versions of the UI in use are shown in Figure 4.11. Learners can copy words or phrases to translate or look up in dictionaries. If audio tracks were selected for a given language, the user can click to listen to individual subtitles or whole segments. Finally, clicking the 'F' button opens the selected subtitle in the 'Flashcard Workspace' where users can create a flashcard for later study, with matching audio and an image from that point in the video.

Text Mode

Saving Sentences

User Interface

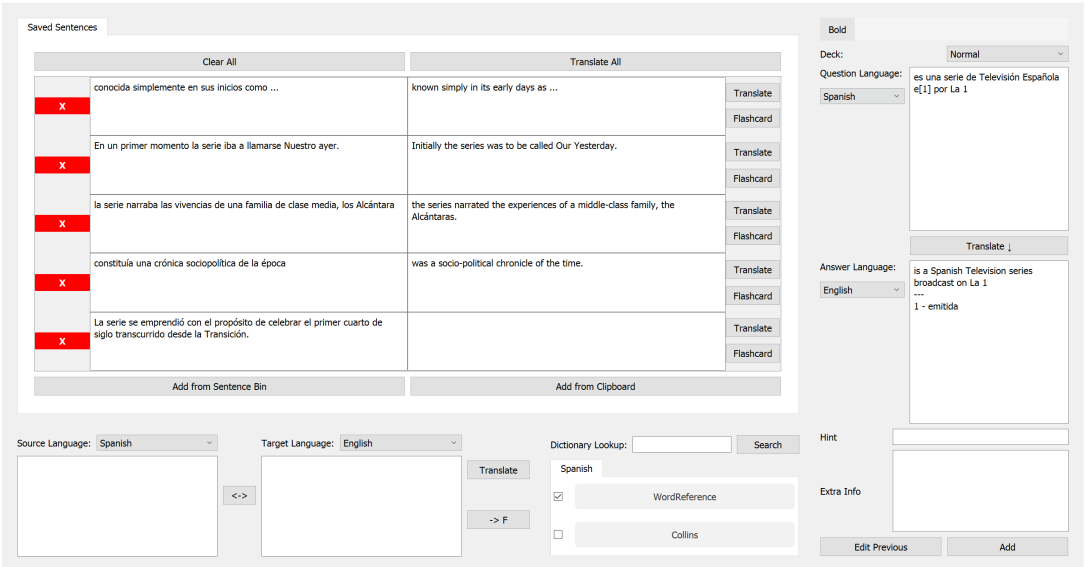

Figure 4.10: The Text Mode being used for sentence mining from Spanish Wikipedia. Sentences pasted into the 'Sentence Bin' are saved in the UI, where a flashcard is already being created.

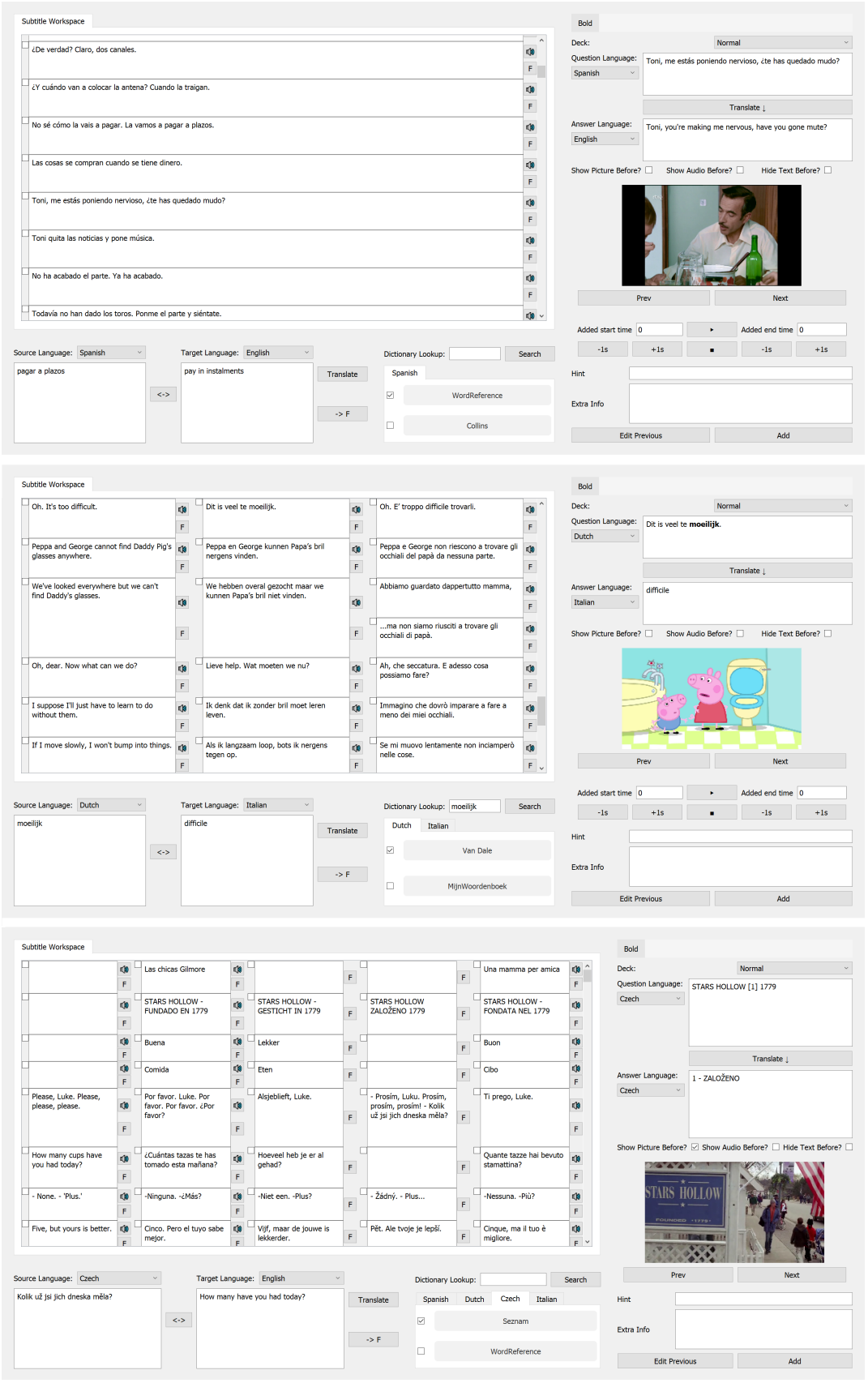

Figure 4.11: Three sessions of the AVI Mode in use – monolingual Spanish study with *Cuéntame*, studying Dutch and Italian with the *Peppa Pig* episode from the experiment, and *Gilmore Girls* in four target languages. Notice the three different types of cards being created.

4.3.3 Translation Workspace

The Translation Workspace (Figure 4.12) supports translation between any of the languages of the current study session. When the user clicks the 'Translate' button, the system translates the source text (the left-hand box) using the machine translation service *DeepL* and returns the translated text (target text) in the right-hand box. Users can also create a simple text flashcard by moving the translation pair to the Flashcard Workspace (discussed in Section 4.3.5) using the '-> F' button.

Translations are requested through *DeepL*'s API (application programming interface). The free version of the API has a 500,000 monthly character limit – users need to replace the API key with their own, which they can sign up for for free.

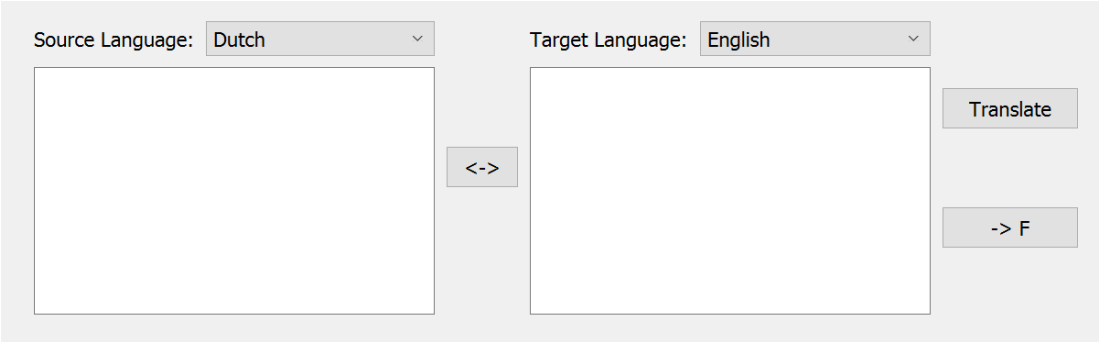

Figure 4.12: The Translation Workspace component of the UI.

4.3.4 Dictionary Lookup

The Dictionary Lookup window (Figure 4.13) allows users to find definitions for any word in any of the languages being studied. The dictionaries displayed are those the user chose in the 'Startup Dialog'. When the user clicks the 'Search' button, the system opens the currently selected dictionary websites – those with a filled checkbox beside them – in the user's web browser, with separate tabs for each website. This allows the user to easily access definitions from a variety of sources. The user can also click on individual dictionaries to open a single browser tab.

The dictionary links were taken from the *Lexilogos* website which contains useful resources for a wide number of the world's languages. This website was accessed using Python's *requests* library, and individual dictionary links for each language were extracted using the *BeautifulSoup4* library, a Python library useful for processing HTML files. Users can also add their own custom dictionary links for any language (not yet in the UI).

Currently, the system only supports L2 to English look-ups, although English to L2 and monolingual (L2 only) dictionary links from *Lexilogos* will be supported in the future. Monolingual dictionaries are only recommended for learners once their vocabulary size in an L2 is 2000-3000 words (Webb and Nation, 2017).

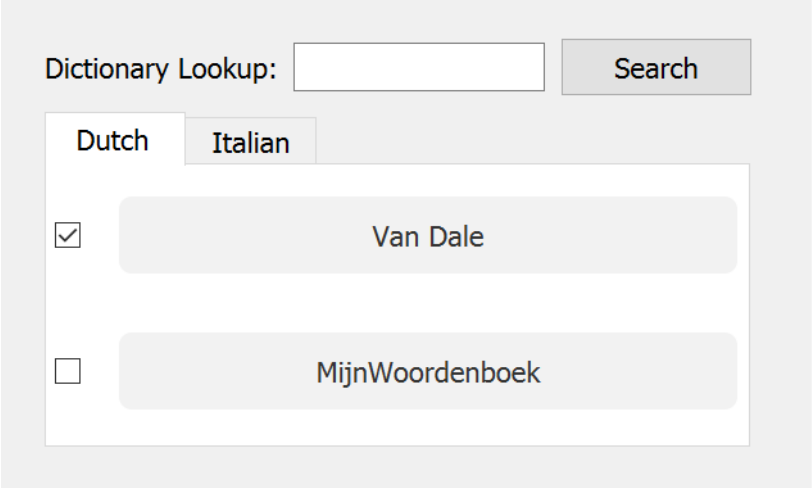

Figure 4.13: The Dictionary Lookup window of the UI.

4.3.5 Flashcard Workspace

The Flashcard Workspace is where learners create flashcards for words, phrases, and sentences they wish to study later using an SRS like *Anki*. Two examples of the Flashcard Workspace and their corresponding flashcards in *Anki* are shown in Figure 4.14.

For text flashcards, any text can be used, such as text taken from the Saved Sentences window while using the basic Text Mode or from the Translation Workspace. Learners can translate the given text to create the answer or insert their own answer, and can add extra information to the back of the card if desired.

For flashcards taken from subtitles (by pressing the 'F' button on any subtitle), the system extracts five screenshots and the audio from the specific point in the video. Users can choose the best screenshot, review the extracted audio, and lengthen the start/end of the audio if the subtitle timing isn't extensive enough to cover all the spoken dialogue.

Users can save easier/harder cards to different decks. The cards of each deck are saved into CSV files where they can be imported into *Anki*. The screenshots and audio are saved in the user's file system and must be moved to *Anki*'s media folder.

The Flashcard Workspace constitutes a versatile flashcard creator. The generic Question/Answer Language fields means any combination of languages can be used (e.g., L2->L1, L2->L2, L2->L3, L1->L2, L3->L2 cards). The many varieties of flashcards that are possible are detailed in Section 4.4.1.

The option to edit previous flashcards has not yet been implemented and will be included in a future update.

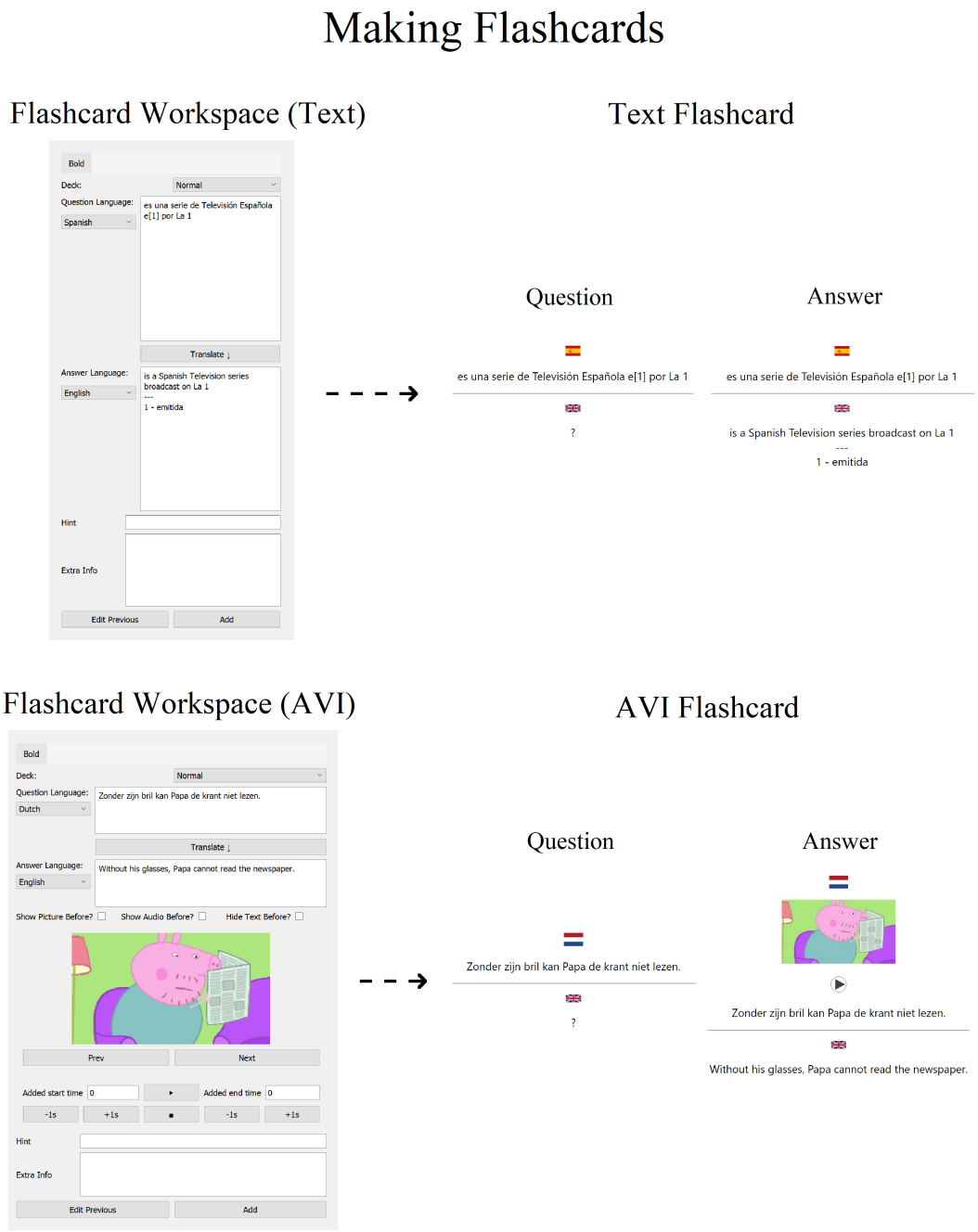

Figure 4.14: The Flashcard Workspace while using both Text and AVI Modes, and the corresponding cards in Anki.

4.4 Output

The system has two outputs – multilingual multimedia flashcards and audio files of condensed dialogue for listening practice.

4.4.1 Flashcards

Flashcards are created using the Flashcard Workspace described in Section 4.3.5. Each card generated by the system contains the following fields:

- **Question Text**: The text shown on the front of the flashcard.
- **Question Language**: The language of the question text, which is represented on the card by a language flag.
- **Answer Text**: The text shown when the back of the flashcard is revealed.
- **Answer Language**: The language of the answer text, also represented by a language flag.
- **Hint**: An optional hint shown on the front of the card.
- **Extra Info**: A field where extra information, like example sentences, pronunciation, or other notes, can be revealed on the backside of the card.
- **Picture**: A picture taken from the input.
- **Audio**: The audio from the input that matches the given subtitle's time and language audio track.
- **Show Picture Before**: If selected, the picture is shown on the front of the flashcard.
- **Show Audio Before**: If selected, the audio is shown on the front.
- **Hide Text Before**: If selected, no text is shown on the front of the flashcard, facilitating an audio-only or even picture-only front side.

The versatility provided by these fields allows for a wide variety of flashcards to be created. This section answers **RQ9**. A gallery of possible cards is shown below in Table 4.1.

Table 4.1: A non-exhaustive list of possible flashcard outputs created for the sentence 'Zonder zijn bril kan Papa de krant niet lezen', taken from the video in the experiment.

Card Type	Front	Back
Word Understanding	krant ? 'Zonder zijn bril kan Papa de krant niet lezen.'	 krant 'Zonder zijn bril kan Papa de krant niet lezen.' newspaper

Continued on next page

Table 4.1: A non-exhaustive list of possible flashcard outputs created for the sentence 'Zonder zijn bril kan Papa de krant niet lezen', taken from the video in the experiment. (Continued)

Card Type	Front	Back
Word Understanding (Monolingual)	krant ? 'Zonder zijn bril kan Papa de krant niet lezen.'	 krant 'Zonder zijn bril kan Papa de krant niet lezen.' Een papier met nieuws.
Word Generation	newspaper ? 'Zonder zijn bril kan Papa de krant niet lezen.'	 newspaper 'Zonder zijn bril kan Papa de krant niet lezen.' krant

Continued on next page

Table 4.1: A non-exhaustive list of possible flashcard outputs created for the sentence 'Zonder zijn bril kan Papa de krant niet lezen', taken from the video in the experiment. (Continued)

Card Type	Front	Back
Sentence	Zonder zijn bril kan Papa de krant niet lezen. ? Zonder zijn bril kan Papa de krant niet lezen. Without his glasses, Papa cannot read the newspaper.	
Word in Sentence	Zonder zijn bril kan Papa de **krant** niet lezen. ? Zonder zijn bril kan Papa de **krant** niet lezen. newspaper	

Continued on next page

Table 4.1: A non-exhaustive list of possible flashcard outputs created for the sentence 'Zonder zijn bril kan Papa de krant niet lezen', taken from the video in the experiment. (Continued)

Card Type	Front	Back
Cloze	 Zonder zijn bril kan Papa de [1] niet lezen. ? ?	 Zonder zijn bril kan Papa de [1] niet lezen. bril = glasses 1 - krant
Word in Sentence (L3)	Zonder zijn bril kan Papa **de krant** niet lezen. ?	 Zonder zijn bril kan Papa **de krant** niet lezen. il giornale

Continued on next page

Table 4.1: A non-exhaustive list of possible flashcard outputs created for the sentence 'Zonder zijn bril kan Papa de krant niet lezen', taken from the video in the experiment. (Continued)

Card Type	Front	Back
Audio	? Zonder zijn bril kan Papa de krant niet lezen. Without his glasses, Papa cannot read the newspaper.	
Audio Cloze	Zonder zijn bril kan Papa de [1] niet lezen. ?	Zonder zijn bril kan Papa de [1] niet lezen. 1 - krant

The 'Word in Sentence' and 'Cloze' card types can also be used to learn multiple words in the same card, for example to also learn 'bril' or 'lezen' for the given sentence. More advanced students can opt to use L2 explanations, definitions, or synonyms in place of L1 explanations.

In the future, more fields will be required for different languages, e.g. character readings (pronunciation) and stroke order for Chinese vocabulary (Ngc et al., 2021).

The language flags used in the flashcard format are sourced from Wiktionary's language flags list.

4.4.2 Condensed Audio

The second output from the system is audio files of condensed dialogue which can be created from any audiovisual input using the Media Export Mode. Using subtitle timings, the system extracts all dialogue audio from the given input video, omitting all non-speaking sections.

The benefits of extensive listening were explored in Section 2.3.3. Creating audio files from target language AVI containing only the sections of spoken dialogue allows for efficient extensive listening practice, serving as a complement to reading AVI dialogue in parallel and intentional vocabulary study using flashcards.

A key innovation of this thesis is a multilingual interleaving algorithm, used to create audio files where segments of dialogue repeat in multiple languages, as in the *Peppa Pig* episode watched by the Segments Group in the experiment. Exporting condensed/interleaved videos will be supported in the future, for language learners and researchers.

Figure 4.15 shows four sessions of the Media Export Mode in use. As in the Startup Dialog, the user selects a folder, then a video file and reference subtitle file. The various settings available offer a wide variety of export options, which will be clarified as the process of creating the audio files is explained.

The steps to create practice audio files are visualised in Figure 4.16 for the pilot episode of Gilmore Girls, and are described as follows:

1. The system parses the reference subtitle file. The subtitle timings for the first few minutes of the episode are shown with alternating red/blue rectangles.
2. 'Padding' is added to the start/end of each subtitle to account for subtitle timings that may miss parts of the dialogue. The user may adjust this setting.
3. The resulting sections of audio are extracted and concatenated to create simple condensed audio files, for example for each of the Spanish, French, and Japanese audio tracks of the video. Choosing the 'Separate files' option would create three separate files, while 'Combine everything' would create one file of all the dialogue in Spanish, then French, then Japanese.
4. The padded subtitles of the reference file can be broken down into segments. Segments are created by concatenating consecutive subtitles until the specified 'segment length' is reached (15 seconds in the example), after which a new segment begins. Each segment is ended by a short fade-out of audio after its last line of dialogue.
5. By exporting segments as separate files using the 'Separate files' option, learners can navigate between segments, and can listen to random segments of dialogue in shuffle mode. Learners can replay segments for 'intensive listening practice' (Flowerdew and Miller, 2005). If segmented files for multiple audio tracks are placed in the same folder, learners can shuffle between segments of dialogue in multiple languages.

6. Interleaving takes segments of dialogue and repeats them in multiple languages, as in the Segments Group's episode in the experiment. By default, the segments of each language are kept as separate files – users can listen through the entire interleaved audio, or again shuffle between random segments of audio in all the languages. Choosing to 'Combine interleaved segments' merges the various language versions of each individual segment together as shown in the diagram. Listening in shuffle mode would then rotate between these multi-language segment groups.

7. Finally, users can adjust the speed of the dialogue for each language. This can save time, for example by having one's L1 played sped up first in an interleaved segment for full comprehension. It also allows users to adjust for proficiency levels in different languages. The same language could be used twice when interleaving, for example to play the learner's weakest language at the start and end, or repeated twice at different speeds. The example shown matches the settings in the bottom right 'Media Exporter' in Figure 46.

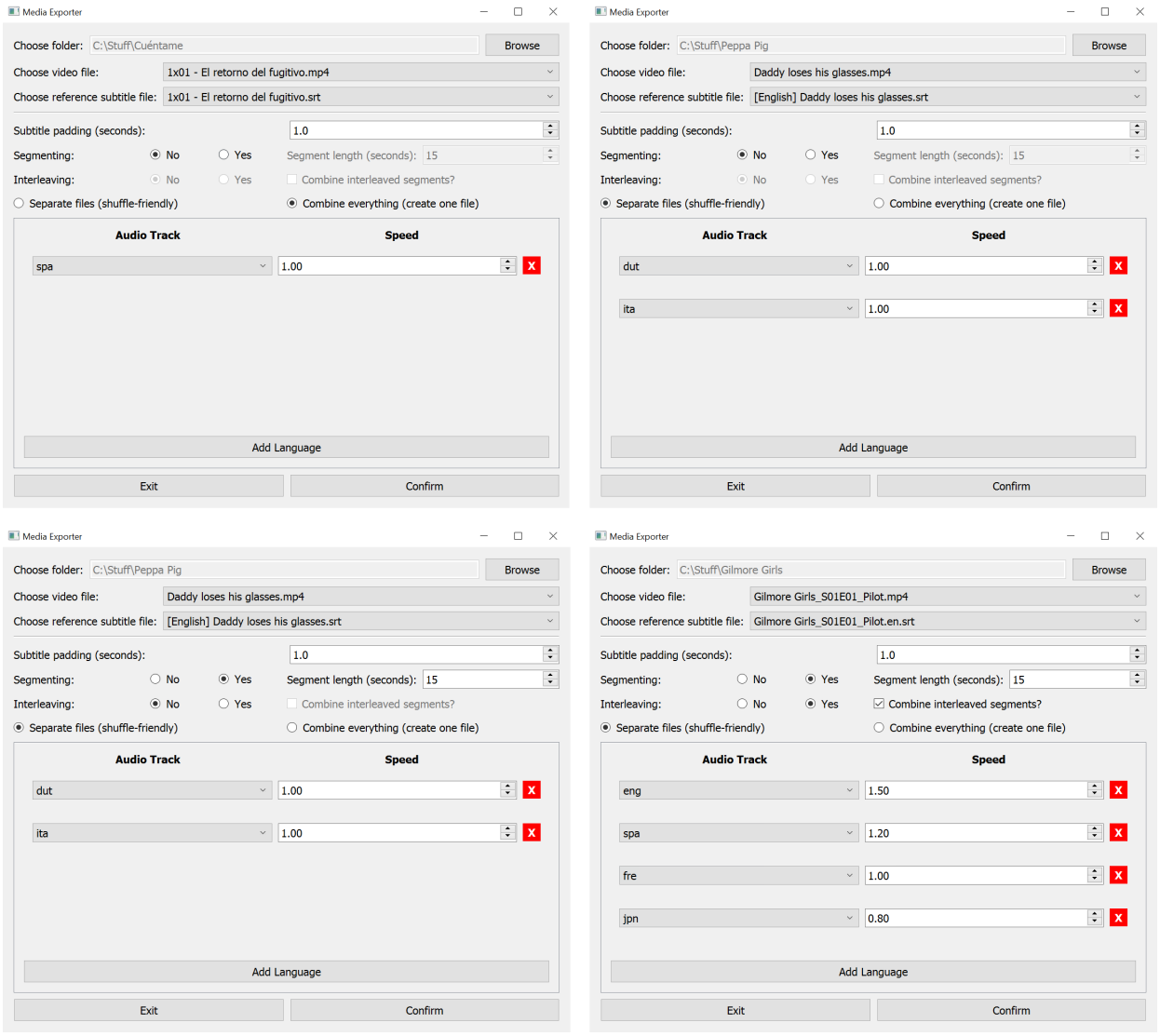

Figure 4.15: Four versions of the Media Export Mode in use.

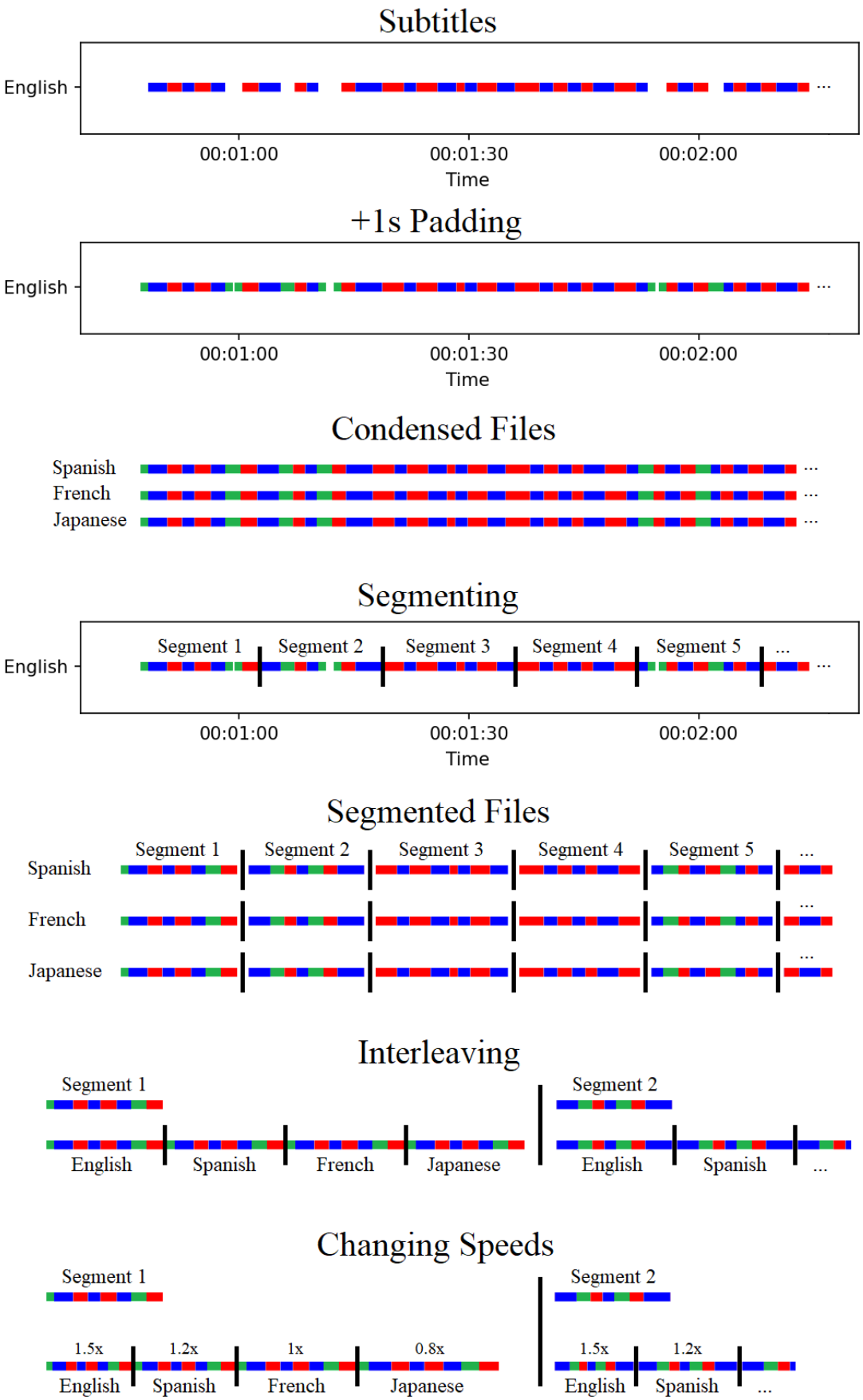

Figure 4.16: A visualisation of the steps involved in creating condensed audio files.

Automatically creating optimal segments of dialogue is a non-trivial problem. Ideally, segments would coincide with full scenes in the video, with longer scenes divided evenly at natural breaks in the dialogue. For example, Segments 1 and 2 above would be improved by earlier cuts in the spaces between exchanges of dialogue. An improvement on the current naive solution could be to look for the longest gap in dialogue in a specific range (say 10-30 seconds). Top-down approaches to segmenting could divide a video's dialogue starting with the longest gaps in dialogue, somewhat similar to the segmenting approach described in Section 4.2.5. Bottom-up approaches could merge the closest lines of dialogue followed by the division of segments exceeding the maximum allowed segment length.

Automatically generating coherent segments of dialogue using subtitle timings is an exciting challenge. Ultimately, the correct strategy and balance of parameters (segment lengths, appropriate gaps between segments) depends on the pacing and amount of dialogue in a given video.

Interleaved multilingual audio presents a wealth of research opportunities. Factors to investigate, which all likely interact in complex ways, include optimal segment length, number of languages being studied, listening order of languages, types of languages to pair, learners' proficiency across various languages, the effect of repetition and speed of playback, randomising the order of languages in segments, having segments overlap so the end of a segment is replayed at the start of the next, and even the effect of adding a level of noise or distortion to the input.

5. Conclusion

Using audiovisual input to learn multiple languages at the same time presents exciting opportunities for language learning research and the development of language learning technology. This thesis explored the idea of multilingual, or parallel AVI as a resource for learning multiple languages simultaneously from three perspectives: theoretical, experimental, and technological.

5.1 Theoretical

TV shows and movies are an incredibly rich source of input for language learners—they're vivid, limitless, memorable, and accessible anywhere. Sentence mining tools complement immersion in L2 input by offering context-rich multimedia flashcards created at the touch of a button. This approach stands at the intersection of extensive input and efficient vocabulary learning through flashcards. However, few tools that support language learning explicitly support the learning of multiple languages simultaneously.

The vast video content on the Internet is highly multilingual, with popular series and films recorded and subtitled in many languages. Captions that provide a foothold into the language being spoken on screen provide a unique multilingual learning opportunity – the subtitles files in multiple languages for the same video can be parallelised, creating a multilingual parallel text linking written transcriptions with native dialogue.

Printed multilingual dialogues for language learning existed in the past, but the benefits and potential of using multiple languages in parallel are generally underappreciated in language research which suffers from a 'monolingual bias' (Buendgens-Kosten, 2020; Cenoz and Gorter, 2011). Opportunities to apply multilingual practices to learning from AVI have been overlooked until now, in research and in technology.

5.2 Experimental

The language learning experiment in Section 3 explored three new techniques in AVI research: previewing a video in learners' first language for full comprehension, repeating a video to learn several languages simultaneously, and watching a video cut into segments that repeat in multiple languages. The experiment demonstrated the feasibility of simultaneous vocabulary acquisition through parallel audiovisual input, with both full videos and segmented videos leading to vocabulary gains.

A pre-test and larger sample size are essentials in a study like this in the future.

It's surprising that previewing has never been attempted in AVI research before, given the well-documented benefits of L1 'scaffolding' when using bilingual texts, the long history of using one's L1 in learning through written dialogues,

and the frequent assertion in AVI research that comprehension facilitates acquisition.

Interesting and familiar content improves L2 vocabulary gains from AVI (Wang, 2012). It seems a natural next step to reuse input to learn multiple languages. Segmenting can be likened to reading a parallel text, where one reads an L2 sentence or paragraph and compares it to the L1 version – the amount of text is analogous to the length of the segment.

Many new language experiments are unlocked with these techniques. Crucial next steps are to determine the effect of previewing, and of reusing input to learn more than one language.

5.3 Technological

The thesis presented a multilingual learning tool in Section 4 that creates parallel study materials for any multilingual video, and any number of languages.

The tool parses and aligns subtitles across languages, programmatically building a parallel text of dialogue in multiple languages in seconds. Using languages side-by-side means learners can use all their linguistic resources in parallel (Cenoz and Gorter, 2015, p. 4). The matching audio for each subtitle makes it a media-rich multilingual reader.

The program supports the sentence mining practice of creating multimedia flashcards taken from immersion materials. The condensed audio files combine the benefits of extensive listening, 'passive immersion', and interleaving.

The concept of segmenting is again applied to both the segmenting of dialogue in the parallel text, and the segmenting of condensed audio that can interleave languages, like the segmented video in the experiment. The concept of interleaving is fascinating, and was applied to a novel interleaving algorithm for condensed audio in multiple languages.

The expanded and deepened language awareness that comes with multilingual practice is cultivated and the interdependence of proficiency across languages is respected, with the ability to create a flexible variety of flashcards set up for multilingual learning, and custom condensed audio files to cater for different language levels.

5.4 Future Research

The techniques introduced in the study mean a wide variety of new variables can be investigated in the future – the effect of different segment lengths, order of languages, number of languages being studied.

Crucial follow-up studies would isolate the effects of previewing content in learners' L1, assess the impact of learning multiple languages simultaneously on learning outcomes, and evaluate the effectiveness of reusing content to learn multiple languages.

Other studies could measure the effect of watching a video again in a new language, not immediately after, but some time later. Learners could enhance the effects of positive language transfer by studying the first language version with flashcards and condensed audio files between viewings. An interesting prospect is to include other languages in interleaved listening practice after watching a video in a specific language, and then return to that video in one of those other languages at a later stage.

While there have been studies on the long-term effects of AVI on language learning, e.g. on grammar (Pattemore, 2022), greatly missing from research is longitudinal studies on informal language learning through long-term immersion in AVI. In addition, longitudinal studies could study the effects of:

- The use of flashcards to complement long-term immersion in an L2. This could provide insights into the effectiveness of different flashcard types.
- Using familiar material to learn languages. How much more effective is it compared to unfamiliar material? How much more effective is it to use the same material to learn a second, third language?
- Long-term use of condensed audio.

Finally, while listening to previously watched AVI content is recommended for developing vocabulary and listening comprehension (Butzkamm and Caldwell, 2009, p. 191), there is a lack of studies investigating its effects. Examining the benefits of repeated listening practice using interleaved condensed audio is also recommended.

5.5 A New View

Modern technology provides opportunities for all forms of extensive input – reading, watching, and listening. This thesis introduces a novel parallel multilingual learning method based on watching, reading, and listening in multiple languages simultaneously using audiovisual input. Watching an L2 television program, reading a transcript of its dialogue, or listening to the program can all lead to incidental vocabulary acquisition (Feng and Webb, 2020). All these methods are explored in this thesis in a multilingual setting through the language experiment, the parallel multilingual reader, and the use of interleaved condensed audio files.

The use of familiar materials is overlooked in language research and education. The opportunity to use interesting audiovisual content for teaching languages, and to reuse content to teach multiple languages could be transformative. This includes creating flashcards and condensed audio for both intentional vocabulary study and passive immersion.

Images, captions, and repetition all enhance vocabulary retention from L2 content (Wang, 2012). The language tool of this thesis incorporates images in flashcards, captions while reading dialogue in parallel, and repeated listening to dialogue from AVI. It integrates engaging content into a comprehensive learning package for simultaneous multilingual study through audiovisual input. It needs to be shared with language learners.

Appendices

A AVI & Sentence Mining Tools

A number of tools developed by language learning researchers in the past supported language learning from sources of extensive input. Pujolă (2002) developed a multimedia language learning program that incorporated bilingual and monolingual dictionaries for Spanish learners of English, to self-learn from native materials (radio, newspapers, television). Hadzilacos et al. (2004) developed a language learning tool where learners subtitle AVI themselves, simulating a professional translation environment, with the help of online dictionaries in six languages. The integration of L2 dictionaries with an interface for self-learning through L2 content is a key aspect of the computer program of this thesis.

One of the earliest tools developed to exploit the language learning potential of L2 videos with matching captions was the 'L1' system (Bird, 2005). Bird's 'L1' system took the subtitles bundled with DVDs and created a video player where language learners could click on words and phrases to see dictionary definitions. His use of subtitles as a 'bridge' between 'highly engaging core materials'—movies—and an interface of language learning tools was ahead of its time. The system aimed to maximise engagement with L2 input while providing language learning tools as a support, like the system developed for this thesis. The subtitle-by-subtitle navigation and list of previous and upcoming subtitles inside a video player are today key features of the popular[1] *Language Reactor* Chrome extension (previously *Language Learning with Netflix*). The ability to save dictionary look-ups including their context in the given movie was a precursor to sentence mining technologies.

Today, several key innovations and technologies created by members of the sentence mining community, which greatly accelerate the process of creating flashcards while immersing, have influenced and inspired the design of the computer program of this thesis:

- *Yomichan*, and its successor *Yomitan*, are browser extensions that provide pop-up dictionary definitions for Japanese words[2]. With one click, a learner can save a given sentence to their *Anki* deck for later review. The front of the flashcard created shows the 'mined' Japanese sentence with the target word in bold, while the back reveals the meaning of the word, its 'reading' (i.e. the Japanese pronunciation), and (optionally) text-to-speech audio of the sentence.
- *mokuro*, *Textractor*, and *Game2Text* are tools that work in conjunction with Yomitan to extract sentences from manga, visual novels, and video games.
- *subs2srs* takes a target language video file, with matching target and source language subtitle files, and automatically creates an *Anki* deck of flashcards. Each line of dialogue gets its own card, with L2 subtitle text on the front and the L1 text on the back, together with matching screenshots and audio.
- The *Voracious* video player, the *Anacreon* and *mpvacious* extensions for the *mpv* video player, the *Migaku* language learning website, and many

[1] *Language Reactor* has over 2 million users as of July 2024.

[2] *Dictionariez* is a similar browser extension that works with a wider range of languages.

other tools, allow learners to quickly create *Anki* cards while watching L2 videos with matching L2 subtitles. They automatically retrieve dictionary definitions and extract dialogue audio and a matching screenshot for any sentence the learner wishes to 'mine'. *jidoujisho* is an alternative that works on Android.

With these technological advancements, learners can immerse in L2 input and make high-quality multimedia flashcards taken from context at the touch of a button. The tools mentioned were developed especially for the study of Japanese, but their principles apply to any language. However, no sentence mining tool until now has realised the potential of multilingual AVI for both extensive immersion and the creation of flashcards, in multiple languages in parallel.

B Script for 'Daddy Loses His Glasses' in English, Dutch, and Italian

Section 1

English

I'm Peppa Pig.
This is my little brother, George.
This is Mummy Pig.
And this is Daddy Pig.

Dutch

Ik ben Peppa Pig.
Dit is mijn kleine broertje George.
Dit is Mama Pig.
En dit is Papa Pig.

Italian

Io sono Peppa Pig.
E questo è il mio fratellino George.
Lei è Mamma Pig.
E lui è Papà Pig.

Section 2

English

Daddy loses his glasses
Daddy Pig wears glasses.
He needs to wear glasses to see clearly.

Dutch

Papa Pig is zijn bril kwijt
Papa Pig heeft een bril.
Hij heeft een bril nodig om goed te kunnen zien.

Italian

Papà perde i suoi occhiali
Papà Pig porta gli occhiali.
Ha bisogno di portare gli occhiali per vedere bene.

Section 3

English

When Daddy pig wears his glasses everything looks fine.

But when Daddy Pig takes his glasses off, he can't see things clearly.
Everything looks a bit soft and fuzzy.
So, it is very important that Daddy Pig knows where his glasses are.
Hmm.

Dutch

Als Papa Pig zijn bril opheeft, ziet hij alles duidelijk.
Maar als hij zijn bril afzet, kan hij niet goed zien.
Dan ziet alles er een beetje wazig uit.
Het is dus heel belangrijk dat Papa Big weet waar zijn bril is.
Hmm.

Italian

Quando Papà Pig porta gli occhiali è tutto a posto.
Ma quando se li toglie, non riesce a vedere bene le cose.
Tutto appare sfumato e confuso.
Quindi è molto importante che Papà Pig sappia sempre dove sono i suoi occhiali.
Hmm.

Section 4

English

Sometimes, Daddy Pig loses his glasses.
Peppa, George, have you seen Daddy Pig's glasses?
He can't find them anywhere.
No, Mummy.

Dutch

Soms is Papa Pig zijn bril kwijt.
Peppa, George, hebben jullie Papa's bril misschien gezien?
Hij kan hem nergens vinden.
Nee Mama.

Italian

Qualche volta capita che Papà Pig perda gli occhiali.
Peppa, George, avete visto per caso gli occhiali di Papà Pig?
Non riesce a trovarli da nessuna parte.
No, io no mamma.

Section 5

English

Peppa and George do not know where Daddy Pig's glasses are.
Oh, dear!

Daddy Pig cannot see a thing without them...
...and it makes him very grumpy.

Dutch

Peppa en George weten niet waar Papa's bril is.
Lieve help!
Papa kan niets zien zonder zijn bril...
...en dat maakt hem erg knorrig.

Italian

Peppa e George non sanno dove siano gli occhiali di Papà Pig.
Oh, non ci voleva!
Papà Pig non riesce a vedere nulla senza gli occhiali...
...e questo lo rende molto nervoso.

Section 6

English

Without his glasses on, Daddy Pig cannot read his newspaper.
This is ridiculous!
I can't see anything.

Dutch

Zonder zijn bril kan Papa de krant niet lezen.
Dit is toch belachelijk.
Ik kan niets zien.

Italian

Senza gli occhiali Papà Pig non riesce a leggere il giornale.
Questo è assolutamente ridicolo!
Non riesco a vedere proprio nulla.

Section 7

English

Somebody must have put my glasses somewhere.
Do you remember where you last put them, Daddy Pig?
When I don't wear them I always put them in my pocket.
But they aren't there now.

Dutch

Iemand moet hem ergens hebben neergelegd.
Weet je nog waarje hem het laatst op had, Papa Big?

Als ik hem niet op heb, heb ik hem altijd in mijn zak.
Maar daar is hij nu niet.

Italian

Qualcuno deve aver messo i miei occhiali da qualche parte.
Non ti ricordi dove li hai messi l'ultima volta, Papà Pig?
Quando non li indosso per leggere li metto sempre in una tasca.
Ma sta volta lì non ci sono.

Section 8

English

Daddy, can we help find your glasses?
Good idea, Peppa.
If you find them, Daddy will stop being so grumpy.
I'm not grumpy.

Dutch

Papa, zullen we je helpen je bril te zoeken?
Goed idee Peppa.
Als jullie hem vinden, is Papa tenminste niet meer zo knorrig.
Ik ben niet knorrig.

Italian

Papà senti, possiamo aiutarti noi a cercare gli occhiali?
E' un'ottima idea, Peppa.
Se riuscite a trovarli, Papà Pig non sarà più così nervoso.
Io non sono affatto nervoso!

Section 9

English

Peppa and George are looking for Daddy's glasses.
Peppa looks under the newspaper, but Daddy Pig's glasses are not there.
George looks on top of the television...
...but Daddy Pig's glasses are not there.

Dutch

Peppa en George zoeken de bril van Papa.
Peppa kijkt onder de krant, maar daar ligt hij niet.
George kijkt op de televisie...
...maar daar ligt hij ook niet.

Italian

Peppa e George vanno alla ricerca degli occhiali di Papà Pig.
Peppa guarda sotto il giornale, ma gli occhiali di Papà Pig non ci sono.
George guarda sopra il televisore...
...ma gli occhiali di Papà Pig non sono neanche lì.

Section 10

English

Let's look upstairs in Mommy and Daddy's bedroom.
Peppa and George are looking in Mummy and Daddy Pig's bedroom.
George, be careful not to knock anything over.
It's not funny.

Dutch

Laten we in de slaapkamer van Papa en Mama kijken.
Peppa en George zoeken in Papa en Mama's slaapkamer.
George, voorzichtig, pas op dat je niks omgooit.
Dit is niet grappig.

Italian

Guardiamo di sopra nella camera da letto di mamma e papà.
Peppa e George vanno a cercare nella camera da letto dei genitori.
George, mi raccomando fai attenzione a non rompere niente.
Non lo trovo affatto divertente.

Section 11

English

Peppa looks under the pillows but Daddy Pig's glasses are not there. Oh.
George looks in Daddy's slippers but the glasses are not there, either.

Dutch

Peppa kijkt onder de kussens maar daar ligt de bril niet. Oh.
George kijkt in Papa's sloffen, maar daar ligt die ook niet.

Italian

Peppa guarda sotto i cuscini ma gli occhiali di Papà Pig non ci sono.
George guarda nelle pantofole del papà, ma gli occhiali non sono neanche lì.

Section 12

English

Let's look in the bathroom.

Peppa and George are looking in the bathroom.
The glasses are not in the bath... Oh!
The glasses are not in the toilet. Oh.
Oh. It's too difficult.

Dutch

Laten we in de badkamer kijken.
Peppa en George kijken in de badkamer.
De bril ligt niet in het bad... Oh!
De bril ligt ook niet in de wc. Oh.
Dit is veel te moeilijk.

Italian

Andiamo a guardare se sono nel bagno.
Peppa e George vanno a guardare nel bagno.
Gli occhiali non sono nella vasca... Oh!
Gli occhiali non sono nel water. Oh!
Oh. E' troppo difficile trovarli.

Section 13

English

Peppa and George cannot find Daddy Pig's glasses anywhere.
We've looked everywhere but we can't find Daddy's glasses.
Oh, dear. Now what can we do?

Dutch

Peppa en George kunnen Papa's bril nergens vinden.
We hebben overal gezocht maar we kunnen Papa's bril niet vinden.
Lieve help. Wat moeten we nu?

Italian

Peppa e George non riescono a trovare gli occhiali del papà da nessuna parte.
Abbiamo guardato dappertutto mamma, ma non siamo riusciti a trovare gli occhiali di papà.
Ah, che seccatura. E adesso cosa possiamo fare?

Section 14

English

I suppose I'll just have to learn to do without them.
If I move slowly, I won't bump into things.
There they are!
Daddy's glasses!
Daddy Pig! You were sitting on them all the time!

Dutch

Ik denk dat ik zonder bril moet leren leven.
Als ik langzaam loop, bots ik nergens tegen op.
Daar is die!
Papa's bril.
Papa Big! Je zat er al die tijd op!

Italian

Immagino che dovrò imparare a fare a meno dei miei occhiali.
Se mi muovo lentamente non inciamperò nelle cose.
Li ho trovati!
Ho trovato gli occhiali.
Papà Pig! Sei stato seduto sopra gli occhiali tutto il tempo.

Section 15

English

Silly Daddy!
I don't know how they got there.
I wonder how!

Dutch

Malle Papa!
Ik weet niet hoe die daar is gekomen.
Dat vraag ik me af.

Italian

Papà sciocchino.
Non so proprio come siano finiti lì.
Me lo sto chiedendo anche io.

Section 16

English

Well, you may have been a bit silly Daddy Pig,
...but at least you're not grumpy any more.
I was not grumpy.

Dutch

Nou, je bent misschien een beetje raar Papa Big,
...maar je bent tenminste niet knorrig meer.
Ik was niet knorrig.

Italian

Bè, forse sei un po' smemorato Papà Pig,
...ma almeno ora hai smesso di essere così nervoso.
Io non ero affatto nervoso prima.

C Experiment Vocabulary Items

C.1 Target Items

Table C.1: Dutch target items.

Word	English Translation	Frequency	Frequency Rank	Cognate				Part of Speech
				English	Spanish	French	German	
bril	glasses	20	2575				Brille	Noun
kunnen	(to be able to) can	8*	92	can			können	Verb
kijken	to look	7*	253					Verb
zien	to see	6*	103	see			sehen	Verb
knorrig	grumpy	5	26302					Adjective
hebben	to have	5*	57	have			haben	Verb
vinden	to find	4	192	find			finden	Verb
kamer	room	4*	404	chamber	cámara	chambre	Kammer	Noun
zoeken	to search	3	368	seek			suchen	Verb
krant	newspaper	2	1563					Noun
slaap	sleep	2*	814	sleep			Schlaf	Noun
bad	bath	2*	1754	bath			Bad	Noun
kwijt	lost	2	635					Adjective

Table C.2: Italian target items.

Word	English Translation	Frequency	Frequency Rank	Cognate				Part of Speech
				English	Spanish	French	German	
occhiali	glasses	24	2706					Noun
guardare	to watch	8*	754		(guardar)	regarder		Verb
nervoso	grumpy	5	2466	nervous	nervioso	nerveux	nervös	Adjective
trovare	to find	5*	343			trouver		Verb
vedere	to see	4	196		ver	voir		Verb
volta	time	3	162		(vuelto)			Noun
giornale	newspaper	2	2358	journal		journal		Noun
bagno	bathroom (/bath)	2	788	bath	baño	bain	Bad	Noun
camera	room	2	833	chamber	cámara	chambre	Kammer	Noun
letto	bed	2	355			lit		Noun
leggere	to read	2	1355		leer	lire	lesen	Verb
fare	to make/do	2	60		hacer	faire		Verb
cercare	to look for	2	613	search		chercher		Verb

C.2 Notes on Target Items

Below are words with multiple forms that featured in the episode (marked with * in Tables C.1 and C.2). The frequency values shown in the table took into account the occurrences of all forms of these words. The exceptions are ”slaap”, ”bad”, and ”kamer”, which only appeared inside ”slaapkamer” and ”badkamer”.

Dutch Target Items:
zien 4 (+2 for ziet)
kijken 3 (+4 for kijkt)
hebben 3 (+2 for heb)
kunnen 3 (+5 for kan)
kamer 0 (+2 for slaapkamer, +2 for badkamer)
slaap 0 (+2 for slaapkamer)
bad 0 (+2 for badkamer)

Italian Target Items:
guardare 2 (+4 for guarda, +1 for guardiamo, +1 for guardato)
trovare 2 (+3 for trovarli)

Other low-frequency cognates were found but not included in the table as participants were unlikely to know them:
kijken – “kieken” colloquial in NE Germany
kwijt – “quitt” is a regional word in Germany

C.3 Distractor Items

Table C.3: Dutch distractor items.

Word	English Translation	Frequency Rank	Cognate				Part of Speech
			English	Spanish	French	German	
stad	city	371				Stadt	Noun
muziek	music	851	music	música	musique	Musik	Noun
toren	tower	2950	tower	torre	tour	Turm	Noun
vriend	friend	261	friend			Freund	Noun
melk	milk	1992	milk			Milch	Noun
hond	dog	656	hound			Hund	Noun
huis	house	163	house			Haus	Noun
spelen	play	446				spielen	Verb
openen	open	1580	open			offen	Verb
winnen	win	720	win			gewinnen	Verb
denken	think	300	think			denken	Verb
jong	young	955	young	joven	jeune	jung	Adjective
nieuw	new	622	new	nuevo	nouveau	neu	Adjective
klein	small	507				klein	Adjective

Table C.4: Italian distractor items.

Word	English Translation	Frequency Rank	Cognate				Part of Speech
			English	Spanish	French	German	
casa	house	118		casa	chez		Noun
gatto	cat	1953	cat	gato	chat	Katze	Noun
famiglia	family	276	family	familia	famille	Familie	Noun
acqua	water	526	aqua	agua	eau		Noun
strada	street	447	street	estrada		Straße	Noun
mano	hand	370		mano	main		Noun
città	city	637	city	ciudad	cité		Noun
cantare	sing	2292	chant	cantar	chanter		Verb
volare	fly	2608		volar	voler		Verb
saltare	jump	1473		saltar	sauter	Salto	Verb
perdere	lost	573		perder	perdre		Verb
grande	big	247	grand	grande	grand	groß	Adjective
lento	slow	5235		lento	lent		Adjective
nuovo	new	206	new	nuevo	nouveau	neu	Adjective

D Vocabulary Score Graphs by Group

Figure D.1: Distribution of overall, form recognition (both languages combined), meaning recall (both languages combined), Dutch (form and meaning combined), and Italian (form and meaning combined) scores.

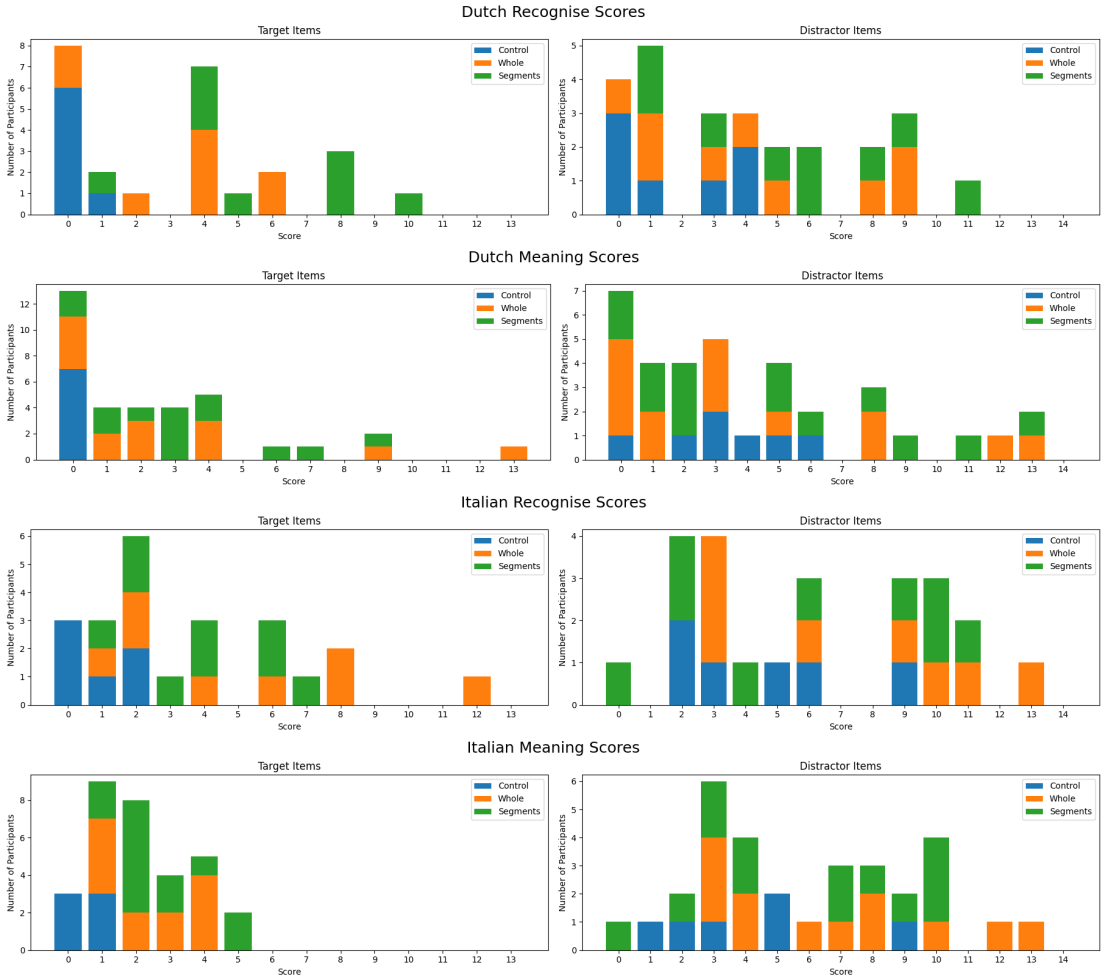

Figure D.2: Distribution of form recognition and meaning recall scores for Dutch and Italian separately.

E Dutch and Italian Meaning Recall Scores – Strong Participants Filtered Out

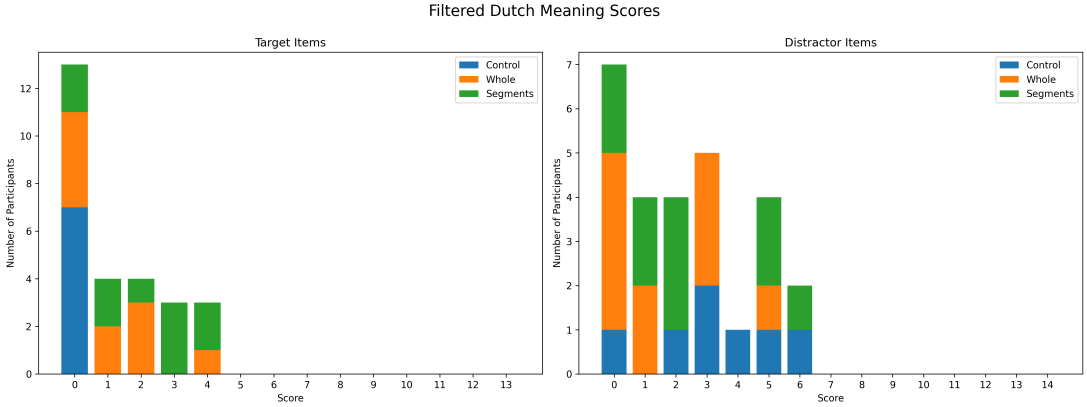

Figure E.1: Distribution of Dutch meaning recall scores, with candidates in the experimental groups showing higher prior knowledge filtered out.

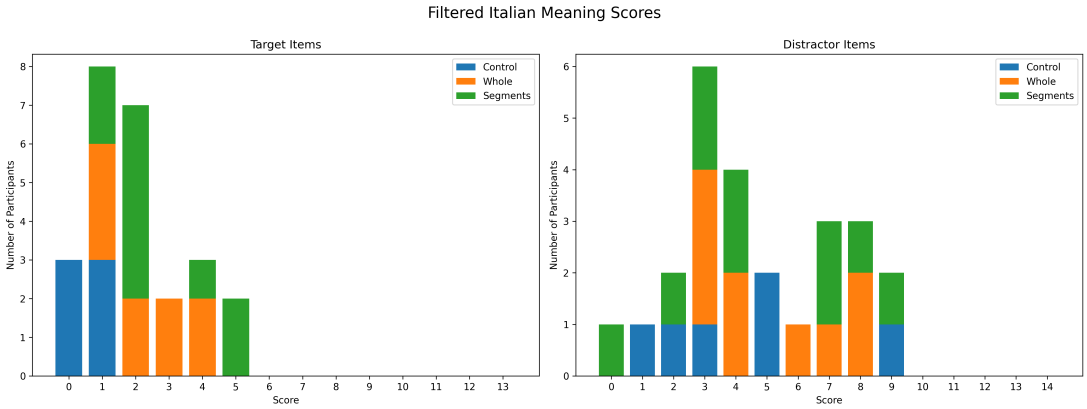

Figure E.2: Distribution of Italian meaning recall scores, with candidates in the experimental groups showing higher prior knowledge filtered out.

F Target Item Average Scores Sorted by Frequency (Multiple Word Forms not Counted)

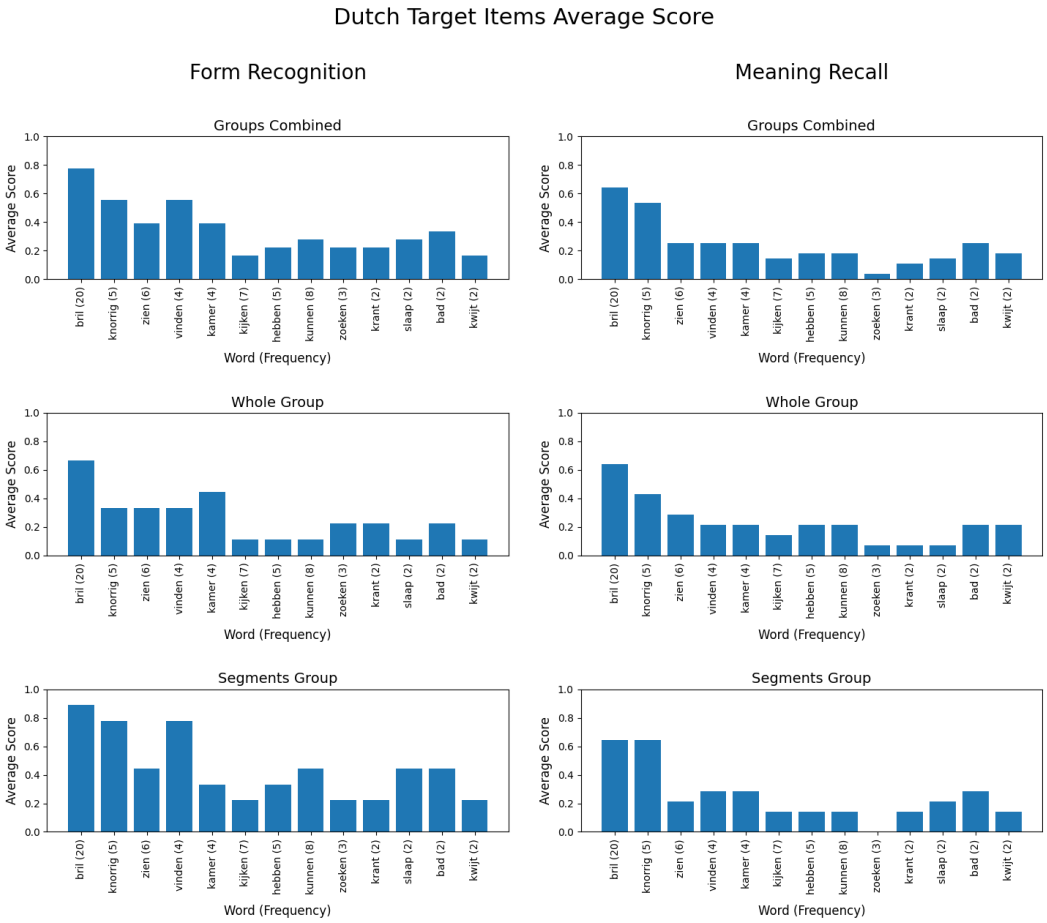

Figure F.1: Dutch target items' average scores, arranged by frequency (multiple word forms not counted).

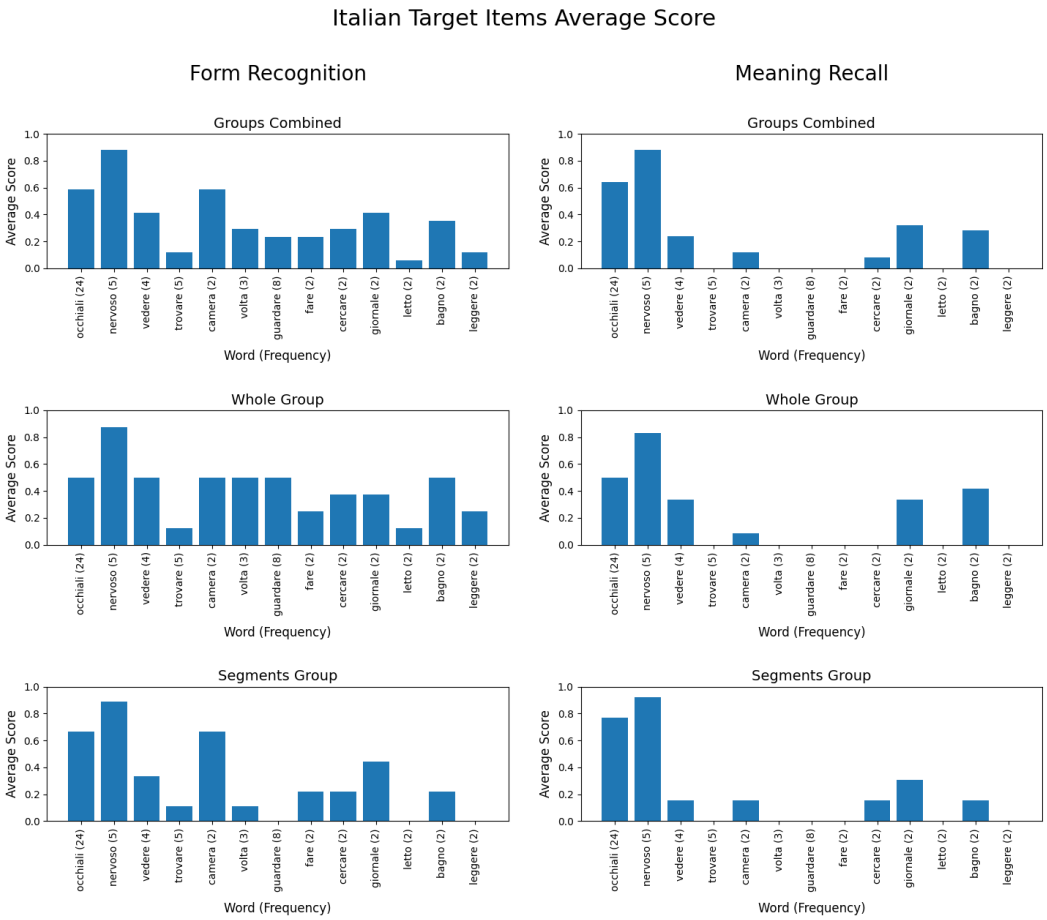

Figure F.2: Italian target items' average scores, arranged by frequency (multiple word forms not counted).

G Previous Language Experience Subgroup Analysis

A short descriptive analysis was carried out to explore the effect previous language experience has on learning two new L2s from captioned AVI. Cognates have a facilitative effect on incidental vocabulary learning from captioned AVI (Montero Perez, 2022), and knowing similar languages to a target language facilitates learning (Cenoz and Gorter, 2022; Festman, 2021).

Table G.1 below shows the meaning recall scores and self-reported benefit of previous language experience of several subgroups in the experiment. The two experimental groups were combined, and the control group and participants with previous study in Italian were omitted. The 'Spanish', 'French' and 'German' subgroups consisted of participants who reported previous language experience in that language but not the other two. A final subgroup consisted of participants who reported previous experience with three or more languages.

Table G.1: Meaning recall scores and answers to post-questionnaire Q7–8 for various subgroups of different prior language experience. The answers to the post-questionnaire were mapped to integer values (-2 to 2) as discussed in Section 3.5.3.
Each cell consists of mean and standard deviation – μ, σ.

Subgroup	**n**	**Meaning Recall Scores (Target Items)**		**Meaning Recall Scores (Distractor Items)**		**Questionnaire – Previous Language Experience Helped:**	
		Dutch	**Italian**	**Dutch**	**Italian**	**Dutch**	**Italian**
Average	25	2.88, 2.67	2.56, 1.29	4.28, 4.08	6.32, 3.44	0.08, 1.61	0.44, 1.47
Spanish	10	2.40, 1.96	2.80, 1.32	3.10, 3.18	7.50, 3.34	-0.40, 1.71	1.30, 0.67
French	3	0.67, 1.15	2.67, 1.15	2.67, 2.52	4.67, 2.08	-1.00, 1.00	-0.67, 1.53
German	4	5.75, 2.75	3.00, 1.83	8.75, 5.00	6.00, 2.16	0.75, 1.89	-1.75, 0.50
3+ Previous Languages	4	4.50, 3.11	2.50, 1.29	6.50, 4.43	8.00, 5.42	1.50, 0.58	0.50, 1.73

Most participants hadn't studied languages since school (+6 years ago). Very few reported language proficiencies at B1 (lower intermediate) level or higher: only 9 for Spanish/French/German combined, and 9 participants for Irish. No notable differences were found for those who had watched AVI before for language learning (13 participants).

Unfortunately, splitting the dataset like this resulted in limited participant numbers. With this caveat in mind, the table shows:

Participants who had studied Spanish scored slightly higher than average on Italian, which was predicted. They strongly perceived that their prior language experience helped in understanding the Italian in the video

Participants who had studied German scored significantly higher on Dutch questions and reported that their previous language experience was relatively helpful in understanding the Dutch in the video.

Participants with experience in three or more languages scored higher on average, albeit with more variation. They perceived helpful effects especially for Dutch.

A more serious analysis could combine a subgroup analysis with an analysis of performance on cognates, potentially factoring in different levels of proficiency in previous languages.

H Covariate Model with Prior Vocabulary Knowledge

The consistently lower vocabulary scores on the distractor words in the control group suggested a disparity in prior vocabulary knowledge between the groups. To address this, participants' scores on the distractor items—comprising high-frequency nouns, verbs, and adjectives—were used as an approximation of prior vocabulary knowledge. This approximation was then included as a covariate in an ANCOVA model to test for significant differences in target item scores while accounting for variations in learners' prior knowledge. Covariate models have been used in similar AVI studies before, e.g. (Peters and Webb, 2018). The ANCOVA model used here was OLS Regression.

Only meaning recall scores were used. Due to a moderate correlation between Dutch and Italian scores (Pearson/Spearman correlation of 0.41/0.38, respectively), and a better fit with ANCOVA assumptions, the Dutch and Italian distractor item scores were combined into a single previous vocabulary knowledge predictor for combined target items scores. In the remainder of this section 'test scores' will refer to combined meaning recall scores on target items and 'prior vocabulary knowledge' to scores on distractor items.

This analysis addresses **RQ4**: if a significant difference in test scores exists between groups when accounting for prior knowledge, it could indicate that vocabulary can be acquired by watching AVI in multiple languages, and if watching whole episodes or in segments is more effective. It also addresses **RQ5** by measuring if prior vocabulary knowledge is a significant predictor of vocabulary learning in a multilingual setting, as measured by test scores.

To meet the assumptions of ANCOVA, several steps were required. First, the scatterplot in Figure H.1 was used to investigate if a linear relationship exists between prior knowledge (the covariate) and test scores for the three groups.

The scatterplot indicates a clear linear relationship between prior knowledge and test scores, particularly noticeable in the experimental groups, whose lines of best fit exhibit similar slopes. The weaker linear relationship of the Control Group can be explained by the fact that these participants didn't watch the episode and had no opportunity to learn the target items.

To assess the homogeneity of regression slopes, a regression analysis was performed with test scores as the dependent variable, prior vocabulary knowledge as covariate, and group membership as the independent variable. The interaction term between prior vocabulary knowledge and group membership was not statistically significant ($p > 0.05$).

Homogeneity of variances of residuals was checked with scatterplots of residuals against both predicted values and covariate (Figure H.2), as well as Levene's test. Although the Control Group's points exhibit a somewhat linear trend in the second graph, with no other notable patterns observed in the scatterplots and a non-significant result in Levene's test ($p > 0.05$), the assumption of equal variances across groups was deemed sufficiently supported.

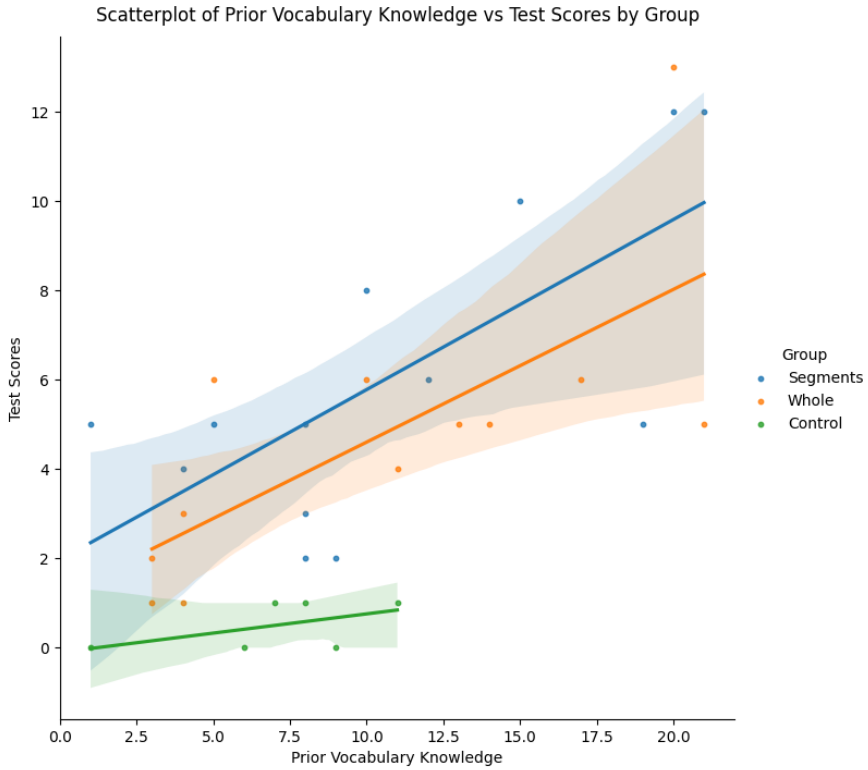

Figure H.1: Scatterplot showing the relationship between meaning recall scores on distractor items (prior vocabulary knowledge) and meaning recall scores on target items.

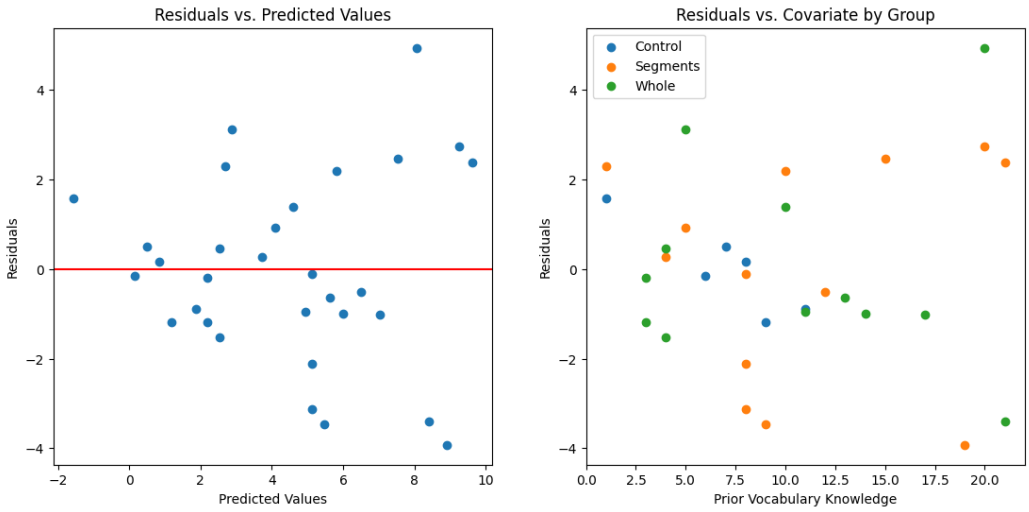

Figure H.2: Scatterplots of residuals against both the predicted values and covariate.

Finally, the histogram and Q-Q plot of residuals in Figure H.3 indicate the residuals are approximately normally distributed—although not perfect, they are satisfactory to continue with the analysis. This is supported by a non-significant Shapiro-Wilk test value (0.73), testing for the normality of residuals.

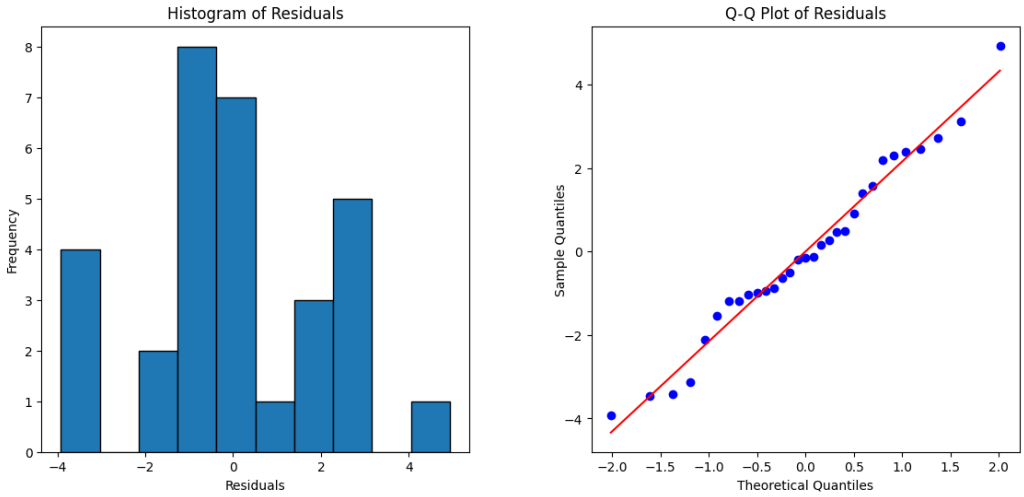

Figure H.3: Histogram and Q-Q plot of residuals.

With all assumptions for ANCOVA satisfied, we present the results in Table H.1.

Table H.1: Results of the ANCOVA with test scores as the dependent variable, prior knowledge as covariate, and group as an independent variable.

	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
Intercept	-1.9203	1.024	-1.875	0.072	-4.022	0.182
Whole Group	3.0687	1.131	2.713	0.011	0.748	5.389
Segments Group	4.2737	1.122	3.808	0.001	1.971	6.577
Covariate	0.3458	0.069	5.005	0.000	0.204	0.488

The ANCOVA model was statistically significant ($F(3, 27) = 17.14$, $p < 0.001$) with 65.6% of the variance in test scores explained by group membership and prior knowledge (R-squared value of 0.656; adjusted R-squared value of 0.617).

The results suggest that both group membership and prior vocabulary knowledge are significant predictors of test scores, after controlling for each other.

Participants in the Segments Group have an estimated mean score that is 4.27 points higher than those in the Control Group when controlling for prior knowledge, with a significant p-value ($p < 0.001$). Similarly, participants in the Whole Group have an estimated mean score that is 3.07 points higher ($p = 0.011$).

For each additional point in the combined meaning recall score on distractor items (prior vocabulary knowledge), the estimated test score increased by 0.35 points, with a highly significant p-value ($p < 0.001$), suggesting higher prior knowledge may have led to higher vocabulary gains.

The results of ANCOVA with the Control Group removed are presented in Table H.2.

The ANCOVA model was again statistically significant ($F(2, 22) = 11.9$, $p < 0.001$) with 52% of the variance in test scores explained by group member-

Table H.2: The results of an ANCOVA when excluding the Control Group.

	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
Intercept	2.1857	1.066	2.051	0.052	-0.024	4.395
Whole Group	-1.1995	0.962	-1.247	0.225	-3.194	0.795
Covariate	0.3613	0.077	4.679	0.000	0.201	0.521

ship and prior knowledge (R-squared value of 0.520; adjusted R-squared value of 0.476).

Participants in the Whole Group have an estimated mean score that is 1.2 points lower than those in the Segments Group, but with an insignificant p-value ($p = 0.225$).

Prior knowledge was again a significant predictor of vocabulary gains, with estimated test scores increasing by 0.36 points for each additional point in the combined meaning recall score on distractor items ($p < 0.001$).

I Likert Scale Post-Questionnaire Responses

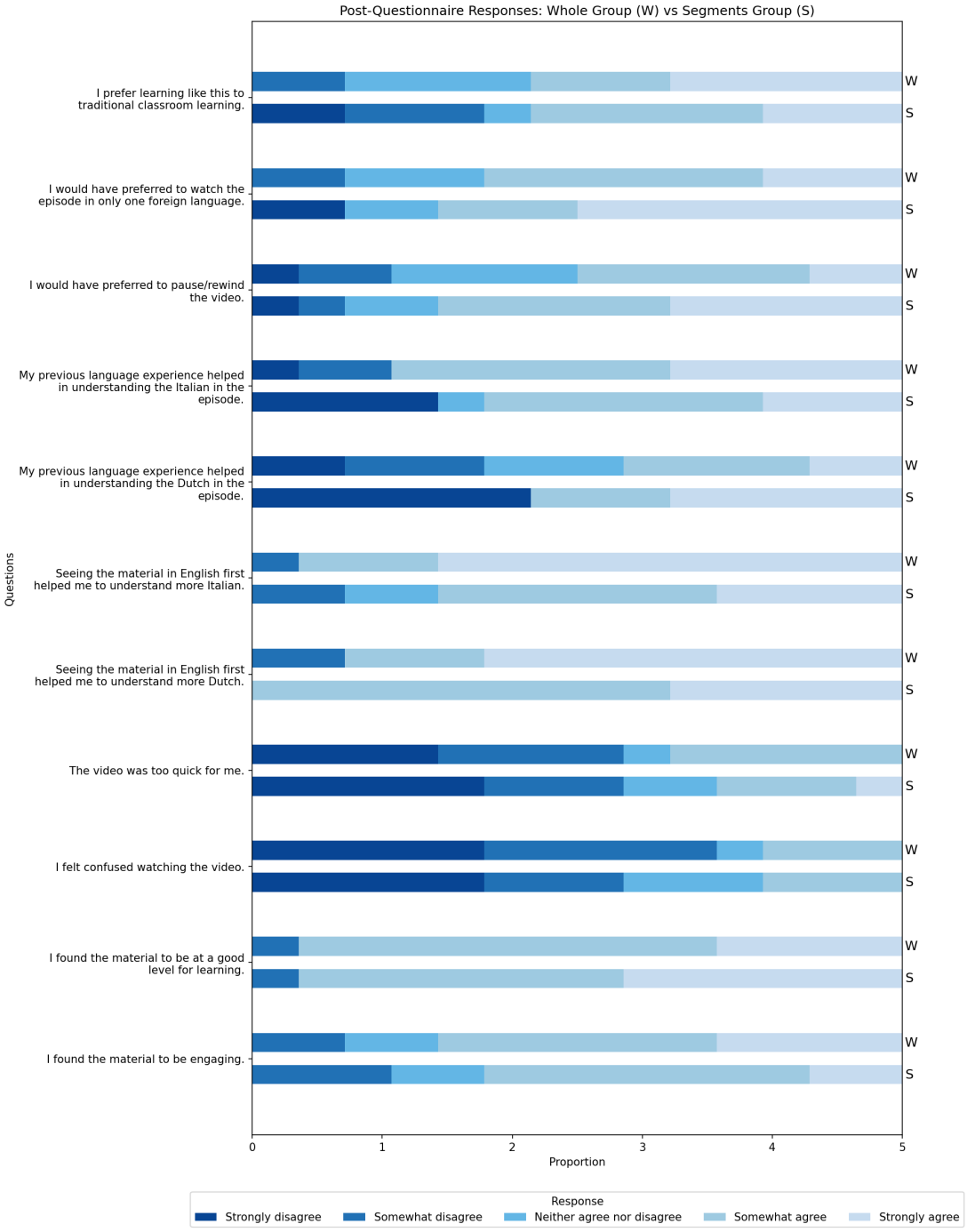

Figure I.1: Responses to the questions of the Likert scale post-questionnaire.

J Future Development of the Computer Program

The next stage of development of the program will transform the system into a comprehensive language learning platform for learning from multilingual AVI. Users will be able to save their progress through an episode/movie or even an entire series, preserving extracted audio tracks and their history of dictionary searches and translations. Linking to a user's *Anki* deck will allow the system to track learned vocabulary, while integrating an AVI player with access to a user's library of study material will allow learners to dip in and out of different content in different languages, and view occurrences of words in other videos.

J.1 Quality of Life Improvements

Several quality of life (QoL) improvements will also be implemented in the next version of the system:

- A search feature for quicker navigation.
- Moving between modes and study materials without closing the application.
- The ability to edit previous flashcards.
- The ability to click multiple subtitles in a row to create flashcards for a longer section.
- Currently, each dictionary search opens a new browser tab, which takes the user out of the application. Ideally, a pop-up dictionary would reveal the definitions for a given word, with separate tabs for the different dictionaries.
- The option to 'favourite' dictionaries to automatically use for a given language, and to add or remove dictionaries in the current session.
- Saving cards directly to *Anki* using the *AnkiConnect* add-on, instead of manually importing cards and moving audio files and images.

J.2 Future Input

For now, the system only supports learning from AVI with matching subtitle files, and a basic mode for learning from sentences found while reading texts like websites and ebooks.

The next step is to support learning from generic text files, such as text files, PDFs, and EPUBs, as well as from images using OCR. This would involve extracting text and tokenising sentences. The user could then alternate between a reader mode with pop-up dictionaries, and a study mode where sentences/phrases saved while reading are shown as in the basic Text Mode of the current system.

Aligning texts in different languages, such as different translations of the same book, is not as easy as with AVI where we can rely on subtitle timings. An extra layer of difficulty is added when matching sentences from audiobooks to written texts which could be aligned across languages just like in the program's AVI Mode.

The system could also support learning from listening materials using automatic speech recognition. The goal is to support multilingual extensive input and sentence mining from any L2 media – movies, websites/articles, YouTube videos, podcasts.

J.3 Artificial Intelligence

Artificial intelligence (AI) could enhance the program in several ways.

An AI assistant could provide definitions, grammar explanations, example sentences, explanations of idiomatic phrases/vocabulary, and could be used to automatically generate flashcards to save users time. The ability of AI to explain sentences that aren't understandable, even after using dictionaries and machine translation, offers great benefit to self-learners.

Glosses, 'short definition[s] or note[s] in order to facilitate reading and comprehension processes for L2 learners' (Lomicka, 1998), can help learners understand new words quickly and bridge their prior knowledge with current material. Glosses help minimise interruption for learners (Ko, 2005). They minimise the need to consult dictionaries or translate specific sentences/phrases and could be automatically created by AI.

Multimedia glosses—glosses using images, audio, or video—have been shown to help L2 vocabulary acquisition and reading comprehension (Mohsen and Balakumar, 2011). They are now entirely possible to create with generative AI, which can create images, audio, or video to suit any situation. Large language models (LLMs) offer the possibility of creating different flashcards for specific target vocabulary every time they are presented.

AI could be used for intelligent segmenting. Machine learning models could be trained to notice cuts between scenes, or identify long sequences of dialogue by one character. The possibility to clone a person's voice in other languages (Staniszewski, 2023) even means AI could create entirely new audio tracks to dub AVI into the user's desired target language.

Bibliography

Andira Abdallah. Impact of Using Parallel Text Strategy on Teaching Reading to Intermediate II Level Students. *International Journal on Social and Education Sciences*, 3(1):95–108, January 2021. ISSN 2688-7061. doi: 10.46328/ijonses.48. URL https://www.ijonses.net/index.php/ijonses/article/view/48. Number: 1.

Thomaï Alexiou. VOCABULARY UPTAKE FROM PEPPA PIG: A CASE STUDY OF PRESCHOOL EFL LEARNERS IN GREECE THOMAÏ ALEXIOU. pages 285–301. January 2015.

Cennet Altiner. Integrating a computer-based flashcard program into academic vocabulary learning. January 2011. URL https://dr.lib.iastate.edu/handle/20.500.12876/24388.

Antimoon. Sentence items (cards) in SRS, June 2002. URL https://www.antimoon.com/how/usingsm-makeitems-sentence.htm.

Yuliya Ardasheva, Tao Hao, and Xue Zhang. Pedagogical Implications of Current SLA Research for Vocabulary Skills. In *Research-Driven Pedagogy*. Routledge, 2019. ISBN 978-1-351-04328-1. Num Pages: 20.

Alexander Arguelles. Parallel Methods for Multiple Languages, June 2022. URL https://www.alexanderarguelles.com/question-answer/parallel-methods-for-multiple-languages/.

Harry P. Bahrick, Lorraine E. Bahrick, Audrey S. Bahrick, and Phyllis E. Bahrick. Maintenance of Foreign Language Vocabulary and the Spacing Effect. *Psychological Science*, 4(5):316–321, September 1993. ISSN 0956-7976. doi: 10.1111/j.1467-9280.1993.tb00571.x. URL https://doi.org/10.1111/j.1467-9280.1993.tb00571.x. Publisher: SAGE Publications Inc.

Colin Baker and Gwyn Lewis. A Synthesis of Research on Bilingual and Multilingual Education. In *The Handbook of Bilingual and Multilingual Education*, pages 109–126. John Wiley & Sons, Ltd, 2015. ISBN 978-1-118-53340-6. doi: 10.1002/9781118533406.ch7. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118533406.ch7. Section: 7 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118533406.ch7.

Ken Beatty. *Teaching & Researching: Computer-Assisted Language Learning*. Routledge, London, 2 edition, 2010. ISBN 978-1-315-83377-4. doi: 10.4324/9781315833774.

David Bellos. *Is That a Fish in Your Ear? The Amazing Adventure of Translation*. Penguin Books, 2012.

Walter Bernard. Psychological Principles of Language Learning and the Bilingual Reading Method. *The Modern Language Journal*, 35(2):87–96, 1951. ISSN 0026-7902. doi: 10.2307/318669. URL https://www.jstor.org/stable/

318669. Publisher: [National Federation of Modern Language Teachers Associations, Wiley].

Stephen A. Bird. Language Learning Edutainment: Mixing Motives in Digital Resources. *RELC Journal*, 36(3):311–339, December 2005. ISSN 0033-6882. doi: 10.1177/0033688205060053. URL https://doi.org/10.1177/0033688205060053. Publisher: SAGE Publications Ltd.

Monica S. Birnbaum, Nate Kornell, Elizabeth Ligon Bjork, and Robert A. Bjork. Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, 41(3):392–402, April 2013. ISSN 1532-5946. doi: 10.3758/s13421-012-0272-7. URL https://doi.org/10.3758/s13421-012-0272-7.

Bitextual. Bitextual: bilingual book generator. URL https://bitextual.net/.

Robert A Bjork. Institutional impediments to effective training. In *Learning, remembering, believing: Enhancing human performance.*, pages 295–306. National Academy Press, Washington, D.C., 1994.

Kristine C. Bloom and Thomas J. Shuell. Effects of Massed and Distributed Practice on the Learning and Retention of Second-Language Vocabulary. *The Journal of Educational Research*, 74(4):245–248, March 1981. ISSN 0022-0671. doi: 10.1080/00220671.1981.10885317. URL https://doi.org/10.1080/00220671.1981.10885317. Publisher: Routledge _eprint: https://doi.org/10.1080/00220671.1981.10885317.

The Editors of Encyclopaedia Britannica. Hexapla, 1998. URL https://www.britannica.com/topic/Hexapla.

Peter C. Brown, Henry L. Roediger III, and Mark A. McDaniel. *Make it stick: The science of successful learning*. Make it stick: The science of successful learning. The Belknap Press of Harvard University Press, Cambridge, MA, US, 2014. ISBN 978-0-674-72901-8. Pages: xi, 313.

Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer. Aligning sentences in parallel corpora. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, ACL '91, pages 169–176, USA, June 1991. Association for Computational Linguistics. doi: 10.3115/981344.981366. URL https://dl.acm.org/doi/10.3115/981344.981366.

Judith Buendgens-Kosten. The monolingual problem of computer-assisted language learning. *ReCALL*, 32(3):307–322, September 2020. ISSN 0958-3440, 1474-0109. doi: 10.1017/S095834402000004X. URL https://www.cambridge.org/core/journals/recall/article/abs/monolingual-problem-of-computerassisted-language-learning/3DEDB208ED096528253FFA0E0EC8896C.

Wolfgang Butzkamm. We only learn language once. The role of the mother tongue in FL classrooms: death of a dogma. *The Language Learning Journal*, 28(1): 29–39, December 2003. ISSN 0957-1736. doi: 10.1080/09571730385200181.

URL https://doi.org/10.1080/09571730385200181. Publisher: Routledge _eprint: https://doi.org/10.1080/09571730385200181.

Wolfgang Butzkamm and John Caldwell. *The Bilingual Reform. A Paradigm shift in Foreign Language Teaching*. January 2009. ISBN 978-3-8233-6492-4.

Raees Calafato. Evaluating teacher multilingualism across contexts and multiple languages: validation and insights. *Heliyon*, 6(8):e04471, August 2020. ISSN 2405-8440. doi: 10.1016/j.heliyon.2020.e04471. URL https://www.sciencedirect.com/science/article/pii/S2405844020313153.

Shana K. Carpenter, Steven C. Pan, and Andrew C. Butler. The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, 1(9):496–511, September 2022. ISSN 2731-0574. doi: 10.1038/s44159-022-00089-1. URL https://www.nature.com/articles/s44159-022-00089-1. Number: 9 Publisher: Nature Publishing Group.

Christine E. Carter and Jessica A. Grahn. Optimizing Music Learning: Exploring How Blocked and Interleaved Practice Schedules Affect Advanced Performance. *Frontiers in Psychology*, 7:1251, August 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.01251. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4989027/.

Sandro Caruana. An Overview of Audiovisual Input as a Means for Foreign Language Acquisition in Different Contexts. *Language and Speech*, 64(4):1018–1036, December 2021. ISSN 0023-8309. doi: 10.1177/0023830920985897. URL https://doi.org/10.1177/0023830920985897. Publisher: SAGE Publications Ltd.

Jasone Cenoz and Durk Gorter. A Holistic Approach to Multilingual Education: Introduction. *The Modern Language Journal*, 95(3):339–343, October 2011. doi: 10.1111/j.1540-4781.2011.01204.x.

Jasone Cenoz and Durk Gorter. Towards a holistic approach in the study of multilingual education. pages 1–15, June 2015.

Jasone Cenoz and Durk Gorter. Teaching English through pedagogical translanguaging. *World Englishes*, 39(2):300–311, 2020. ISSN 1467-971X. doi: 10.1111/weng.12462. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/weng.12462. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/weng.12462.

Jasone Cenoz and Durk Gorter. Pedagogical Translanguaging. *Elements in Language Teaching*, December 2021. doi: 10.1017/9781009029384. URL https://www.cambridge.org/core/elements/pedagogical-translanguaging/67802C1E5AE4A418AE3B8E2DEFBAD30A. ISBN: 9781009029384 9781009014403 Publisher: Cambridge University Press.

Jasone Cenoz and Durk Gorter. Enhancing bilingual resources in third language acquisition: towards pedagogical translanguaging. *Educational Linguistics*, 1(2):338–357, November 2022. ISSN 2748-9329. doi: 10.1515/

eduling-2022-0009. URL https://www.degruyter.com/document/doi/10.1515/eduling-2022-0009/html?lang=en. Publisher: De Gruyter Mouton.

Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3):354–380, May 2006. ISSN 0033-2909. doi: 10.1037/0033-2909.132.3.354.

Evgeny Chukharev-Hudilainen and Tatiana Klepikova. The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *CALICO Journal*, 33:334–354, August 2016. doi: 10.1558/cj.v33i3.26055.

J. Cole. *Foreign language learning in the age of the internet: A comparison of informal acquirers and traditional classroom learners in central Brazil*. http://purl.org/dc/dcmitype/Text, University of Oxford, 2015. URL https://ora.ox.ac.uk/objects/uuid:db80473a-2075-4e91-bb07-a706bb6a433f.

Jason Cole and Robert Vanderplank. Comparing autonomous and class-based learners in Brazil: Evidence for the present-day advantages of informal, out-of-class learning. *System*, 61:31–42, October 2016. doi: 10.1016/j.system.2016.07.007.

John Amos Comenius. *Orbis sensualium pictus quatrolinguis*. Michaelis & Joh. Friderici Endterorum, Noribergae [Nuremberg], 1679.

John Amos Comenius. *Orbis sensualium pictus trilinguis*. S. Brewer, Levoca, 1685.

Rory G. Critten, editor. *French Lessons in Late-Medieval England: The "Liber Donati" and "Commune Parlance"*. Arc Humanities Press, 2023. doi: 10.17302/fou-9781802701029. URL https://library.oapen.org/handle/20.500.12657/62011. Accepted: 2023-03-29T12:17:27Z.

David Crystal. *How Language Works*. Penguin Books, 2005. ISBN 978-0-14-051538-1.

Jim Cummins. Teaching for Transfer: Challenging the Two Solitudes Assumption in Bilingual Education. In Nancy H. Hornberger, editor, *Encyclopedia of Language and Education*, pages 1528–1538. Springer US, Boston, MA, 2008. ISBN 978-0-387-30424-3. doi: 10.1007/978-0-387-30424-3_116. URL https://doi.org/10.1007/978-0-387-30424-3_116.

Eleanor Dickey and Rolando Ferri. A New Edition of the "Colloquium Harleianum" Fragment in P.Prag. 2.118. *Zeitschrift für Papyrologie und Epigraphik*, 180:127–132, 2012. ISSN 0084-5388. URL https://www.jstor.org/stable/41616924. Publisher: Dr. Rudolf Habelt GmbH.

C. J. Dodson. *Language Teaching and the Bilingual Method*. Pitman Publishing Co, 1967. ERIC Number: ED113975.

H. Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Teachers College Press, 1885. doi: 10.1037/10011-000.

Hermann Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Memory: A contribution to experimental psychology. Teachers College Press, New York, NY, US, 1913. doi: 10.1037/10011-000. Pages: viii, 128.

Luke G. Eglington and Sean H. K. Kang. Interleaved Presentation Benefits Science Category Learning. *Journal of Applied Research in Memory and Cognition*, 6(4):475–485, December 2017. ISSN 2211-3681. doi: 10.1016/j.jarmac.2017.07.005. URL https://www.sciencedirect.com/science/article/pii/S2211368117300578.

Nick C. Ellis. The Psychology of Foreign Language Vocabulary Acquisition: Implications for Call. *Computer Assisted Language Learning*, 8(2-3):103–128, June 1995. ISSN 0958-8221. doi: 10.1080/0958822940080202. URL https://doi.org/10.1080/0958822940080202. Publisher: Routledge _eprint: https://doi.org/10.1080/0958822940080202.

Rod Ellis. *Learning a Second Language through Interaction*. John Benjamins, December 1999. ISBN 978-90-272-9887-4. doi: 10.1075/sibil.17. URL https://www.jbe-platform.com/content/books/9789027298874.

Rod Ellis. Instructed Second Language Acquisition: A Literature Review. 2005. URL https://www.semanticscholar.org/paper/Instructed-Second-Language-Acquisition-A-Literature-Ellis/a146a783014f021df9eaf592207a62f500788363.

Connie C. Epstein. *The Art of Writing for Children: Skills and Techniques of the Craft*. Archon Books, 1991. ISBN 978-0-208-02297-4 978-0-208-02296-7. URL http://archive.org/details/artofwritingforc0000epst.

Gisela Ernst-Slavit and Margaret Mulhern. Bilingual books: Promoting literacy and biliteracy in the second-language and mainstream classroom. *Reading Online*, 7, January 2003.

Yanxue Feng and Stuart Webb. LEARNING VOCABULARY THROUGH READING, LISTENING, AND VIEWING: WHICH MODE OF INPUT IS MOST EFFECTIVE? *Studies in Second Language Acquisition*, 42(3):499–523, July 2020. ISSN 0272-2631, 1470-1545. doi: 10.1017/S0272263119000494. URL https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/abs/learning-vocabulary-through-reading-listening-and-viewing/415CF118C5A114BB2D9C3B79B9656AFF.

Julia Festman. Learning and Processing Multiple Languages: The More the Easier? *Language Learning*, 71(S1):121–162, 2021. ISSN 1467-9922. doi: 10.1111/lang.12437. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/lang.12437. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lang.12437.

John Field. *Listening in the Language Classroom*. Cambridge Language Teaching Library. Cambridge University Press, Cambridge, 2009. ISBN 978-0-521-68570-2. doi: 10.1017/CBO9780511575945. URL https://www.

cambridge.org/core/books/listening-in-the-language-classroom/2119E4D62387E5A3B43B0A95E0BBAABE.

Jonathan Firth, Ian Rivers, and James Boyle. A systematic review of interleaving as a concept learning strategy. *Review of Education*, 9 (2):642–684, 2021. ISSN 2049-6613. doi: 10.1002/rev3.3266. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rev3.3266. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rev3.3266.

John Flowerdew and Lindsay Miller. *Second Language Listening: Theory and Practice*. Cambridge University Press, 2005.

Frantisek Forgac, Dasa Munkova, Michal Munk, and Livia Kelebercova. Evaluating automatic sentence alignment approaches on English-Slovak sentences. *Scientific Reports*, 13(1):20123, November 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-47479-w. URL https://www.nature.com/articles/s41598-023-47479-w. Publisher: Nature Publishing Group.

Rita Franceschini. History of Multilingualism. November 2012. ISBN 978-1-4051-9473-0. doi: 10.1002/9781405198431.wbeal0511.

Pascale Fung and Kathleen McKeown. Aligning Noisy Parallel Corpora Across Language Groups: Word Pair Feature Matching by Dynamic Time Warping. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, Columbia, Maryland, USA, October 1994. URL https://aclanthology.org/1994.amta-1.11.

William A. Gale and Kenneth W. Church. A program for aligning sentences in bilingual corpora. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, ACL '91, pages 177–184, USA, June 1991. Association for Computational Linguistics. doi: 10.3115/981344.981367. URL https://dl.acm.org/doi/10.3115/981344.981367.

Ofelia García and Hugo Baetens Beardsmore. *Bilingual education in the 21st century: a global perspective*. Wiley-Blackwell Pub., Malden, MA, 2009. ISBN 978-1-4443-5978-7. URL http://catalog.hathitrust.org/api/volumes/oclc/225852417.html. OCLC: 679479618.

Nina Garrett. Technology in the Service of Language Learning: Trends and Issues. *The Modern Language Journal*, 75(1):74–101, 1991. ISSN 0026-7902. doi: 10.2307/329836. URL https://www.jstor.org/stable/329836. Publisher: [National Federation of Modern Language Teachers Associations, Wiley].

Emmanuel Giguet. Multi-grained alignment of parallel texts with endogenous resources. January 2005.

Betsy Gilliland. Listening Logs for Extensive Listening Practice. In *Language Learning Beyond the Classroom*. Routledge, 2014. ISBN 978-1-315-88347-2. Num Pages: 10.

Agostino Giustiniani, editor. *Genoa Quadruplex Psalter*. Petrus Paulus Porrus, 1516a. URL https://library.chethams.com/collections/101-treasures-of-chethams/genoa-quadruplex-psalter/.

Agostino Giustiniani, editor. *Polyglot Psalter*. Petrus Paulus Porrus, 1516b. URL https://www.joh.cam.ac.uk/library/special_collections/early_books/psalter.htm.

Robert Godwin-Jones. EMERGING TECHNOLOGIES FROM MEMORY PALACES TO SPACING ALGORITHMS: APPROACHES TO SECOND-LANGUAGE VOCABULARY LEARNING. 2010. URL https://www.semanticscholar.org/paper/EMERGING-TECHNOLOGIES-FROM-MEMORY-PALACES-TO-TO-Godwin-Jones/0369accfd3f3a0766fa819dac0cb8ed64439005c.

Robert Godwin-Jones. Emerging technologies: Mobile apps for language learning. *Language Learning and Technology*, 15:2–11, June 2011.

Christine C. M. Goh and Larry Vandergrift. *Teaching and Learning Second Language Listening: Metacognition in Action*. Routledge, New York, October 2011. ISBN 978-0-203-84337-6. doi: 10.4324/9780203843376.

Maja Grgurović, Carol A. Chapelle, and Mack C. Shelley. A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2):165–198, May 2013. ISSN 0958-3440, 1474-0109. doi: 10.1017/S0958344013000013. URL https://tinyurl.com/cambridge-core.

Hugh Grigg. Sentence mining, 2020. URL https://www.chineseboost.com/blog/sentence-mining/.

Thanasis Hadzilacos, Spyros Papadakis, and Stavroula Sokoli. Learner's Version of a Professional Environment: Film Subtitling as an ICTE Tool for Foreign Language Learning. pages 680–685. Association for the Advancement of Computing in Education (AACE), 2004. ISBN 978-1-880094-54-9. URL https://www.learntechlib.org/primary/p/11399/.

Graham Hall and Guy Cook. Own-language use in language teaching and learning. *Language Teaching*, 45(3):271–308, July 2012. ISSN 1475-3049, 0261-4448. doi: 10.1017/S0261444812000067. URL https://www.cambridge.org/core/journals/language-teaching/article/abs/ownlanguage-use-in-language-teaching-and-learning/6F479F6C8DD0CF628A53131D5CCE7B5D. Publisher: Cambridge University Press.

Tao Hao, Zhe Wang, and Yuliya Ardasheva. Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis. *Journal of Research on Educational Effectiveness*, 14(3):645–667, July 2021. ISSN 1934-5747. doi: 10.1080/19345747.2021.1917028. URL https://doi.org/10.1080/19345747.2021.1917028. Publisher: Routledge _eprint: https://doi.org/10.1080/19345747.2021.1917028.

Sophie Hardach. *Languages Are Good For Us*. Head of Zeus, 2021. ISBN 978-1-78954-394-0. URL https://books.google.nl/books?id=3YCpDwAAQBAJ.

Md Masudul Hasan and Tan Bee Hoon. Podcast Applications in Language Learning: A Review of Recent Studies. *English Language Teaching*, 6(2):128–135, 2013. ISSN 1916-4742. URL https://eric.ed.gov/?id=EJ1076962. Publisher: Canadian Center of Science and Education ERIC Number: EJ1076962.

Philip Herdina and Ulrike Jessner. A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics. In *A Dynamic Model of Multilingualism*. Multilingual Matters, January 2002. ISBN 978-1-85359-554-7. doi: 10.21832/9781853595547. URL https://www.degruyter.com/document/doi/10.21832/9781853595547/html.

Jan Hokenson and Marcella Munson. *The Bilingual Text: History and Theory of Literary Self-translation*. St. Jerome Pub., 2007. ISBN 978-1-900650-93-9. Google-Books-ID: TOI_AQAAIAAJ.

Susan Hopewell. Leveraging bilingualism to accelerate English reading comprehension. *International Journal of Bilingual Education and Bilingualism*, 14 (5):603–620, September 2011. ISSN 1367-0050. doi: 10.1080/13670050.2011.564274. URL https://doi.org/10.1080/13670050.2011.564274. Publisher: Routledge _eprint: https://doi.org/10.1080/13670050.2011.564274.

Ran Hu, Xiaoning Chen, and Xiuping Li. Exploring bilingual books with five Chinese first graders: Children's responses and biliteracy development. *Reading Horizons: A Journal of Literacy and Language Arts*, 52(1):4, 2012.

Chester S. J. Huang, Stephen J. H. Yang, Tosti H. C. Chiang, and Addison Y. S. Su. Effects of Situated Mobile Learning Approach on Learning Motivation and Performance of EFL Students. *Journal of Educational Technology & Society*, 19(1):263–276, 2016. ISSN 1176-3647. URL https://www.jstor.org/stable/jeductechsoci.19.1.263. Publisher: International Forum of Educational Technology & Society.

Thomas Huckin and James Coady. INCIDENTAL VOCABULARY ACQUISITION IN A SECOND LANGUAGE: A Review. *Studies in Second Language Acquisition*, 21(2):181–193, June 1999. ISSN 1470-1545, 0272-2631. doi: 10.1017/S0272263199002028. URL https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/abs/incidental-vocabulary-acquisition-in-a-second-language/8843680A0DD18CDA6B264C6A2D421F25.

Werner Hüllen. *Kleine Geschichte des Fremdsprachenlernens*. Schmidt, Erich, Berlin, 1 edition, May 2005. ISBN 978-3-503-07946-9.

Kyeong-Ouk Jeong. Facilitating Sustainable Self-Directed Learning Experience with the Use of Mobile-Assisted Language Learning. *Sustainability*, 14(5):2894, January 2022. ISSN 2071-1050. doi: 10.3390/su14052894. URL https://www.mdpi.com/2071-1050/14/5/2894. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Sean H. K. Kang and Harold Pashler. Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, 26(1):97–103, 2012. ISSN 1099-0720. doi: 10.1002/acp.1801. URL

https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.1801. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1801.

Martin Kay and Martin Roscheisen. Text-Translation Alignment. *Computational Linguistics*, 19(1):121–142, 1993. URL https://aclanthology.org/J93-1006. Place: Cambridge, MA Publisher: MIT Press.

Khatzumoto. All Japanese All The Time: 10,000 Sentences, April 2007. URL https://web.archive.org/web/20070427062210/http://www.alljapaneseallthetime.com/blog/10000-sentences-how.

Lid King. The Impact of Multilingualism on Global Education and Language Learning, 2018. URL https://www.cambridgeenglish.org/Images/539682-perspectives-impact-on-multilingualism.pdf.

Allison Kimberly Kingman Rosero. *Influence of the sentence mining technique on the advancement of speaking competence of the english language*. Bachelor's Thesis, Universidad de Guayaquil Facultad de Filosofía, Letras y Ciencias de la Educación, April 2022. URL https://repositorio.ug.edu.ec/items/24933e34-07a4-4f4a-9e64-ed94d3d9e67c.

Myong Hee Ko. Glosses, Comprehension, and Strategy Use. *Reading in a Foreign Language*, 17(2):125–143, October 2005. ISSN 1539-0578. URL https://eric.ed.gov/?id=EJ759831. ERIC Number: EJ759831.

Anastasia Kokla. *The Phenomenon Peppa Pig: A preschool TV programme that promotes EFL learning, prosocial behaviour, multiculturalism and gender equality*. Διδακτρικη Διατριβη, Αριστττελει Πανεπιστημι Θεσσαλνικης (ΑΠΘ). Σχλη Φιλσφικη. Τμημα Αγγλικης Γλωσσας και Φιλλγιας, 2022. URL http://hdl.handle.net/10442/hedi/53222. Accepted: 2023-02-21T10:16:59Z Journal Abbreviation: Τ φαινμεν Πεππα τ γυρυνακι: ενα τηλεπτικ πργραμμα πρσχλικης ηλικιας πυ πρωθει την εκμαθηση των Αγγλικων, την πρκινωνικη συμπεριφρα, την πλυπλιτισμικτητα και την ιστητα των φυλων.

Natassa Kokla. Peppa Pig: An innovative way to promote formulaic language in pre-primary EFL classrooms (pp. 76-92), Natassa Kokla. *Research Papers in Language Teaching and Learning*, 11:76–92, 2021. URL https://tinyurl.com/koklapaper.

Nate Kornell and Robert A. Bjork. Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science*, 19(6):585–592, June 2008. ISSN 0956-7976. doi: 10.1111/j.1467-9280.2008.02127.x. URL https://doi.org/10.1111/j.1467-9280.2008.02127.x. Publisher: SAGE Publications Inc.

Nate Kornell and Kalif E. Vaughn. How Retrieval Attempts Affect Learning: A Review and Synthesis. In Brian H. Ross, editor, *Psychology of Learning and Motivation*, volume 65, pages 183–215. Academic Press, January 2016. doi: 10.1016/bs.plm.2016.03.003. URL https://www.sciencedirect.com/science/article/pii/S0079742116000165.

Stephen D. Krashen. *Principles and practice in second language acquisition*. Language teaching methodology series. Pergamon Press, Oxford, 1 edition, 1982. ISBN 978-0-08-028628-0. URL https://bac-lac.on.worldcat.org/oclc/300403590. OCLC: 7734300.

W. E. Lambert, I. Boehler, and N. Sidoti. Choosing the languages of subtitles and spoken dialogues for media presentations: Implications for second language education. *Applied Psycholinguistics*, 2(2):133–148, 1981. ISSN 1469-1817. doi: 10.1017/S0142716400000904. Place: United Kingdom Publisher: Cambridge University Press.

Sebastian Leitner. *So lernt man lernen*. Herder, 1972. ISBN 978-3-451-16265-7. Google-Books-ID: sD3AAQAACAAJ.

Michael Levy. *Computer-Assisted Language Learning: Context and Conceptualization*. Clarendon Press, March 1997. ISBN 978-0-19-823631-3. Google-Books-ID: RRGgrjteVjUC.

Gwyn Lewis, Bryn Jones, and Colin Baker. Translanguaging: origins and development from school to street and beyond. *Educational Research and Evaluation*, 18(7):641–654, October 2012. ISSN 1380-3611. doi: 10.1080/13803611.2012.718488. URL https://doi.org/10.1080/13803611.2012.718488. Publisher: Routledge _eprint: https://doi.org/10.1080/13803611.2012.718488.

Li Li. *New Technologies and Language Learning*. Bloomsbury Publishing, April 2017. ISBN 978-1-137-51768-5. Google-Books-ID: khtHEAAAQBAJ.

Eva Lindgren and Carmen Muñoz. The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10(1):105–129, February 2013. ISSN 1479-0718. doi: 10.1080/14790718.2012.679275. URL https://doi.org/10.1080/14790718.2012.679275. Publisher: Routledge _eprint: https://doi.org/10.1080/14790718.2012.679275.

Pierre Lison and Jörg Tiedemann. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, 2016.

Jessica M. Logan, Alan D. Castel, Sara Haber, and Emily J. Viehman. Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, 7(3):175–195, December 2012. ISSN 1556-1631. doi: 10.1007/s11409-012-9090-3. URL https://doi.org/10.1007/s11409-012-9090-3.

Lara L. Lomicka. "To gloss or not to gloss": An investigation of reading comprehension online. January 1998. ISSN 1094-3501. URL http://hdl.handle.net/10125/25020. Publisher: University of Hawaii National Foreign Language Resource Center.

Richard A. Magill and Kellie G. Hall. A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, 9(3):241–289, September

1990. ISSN 0167-9457. doi: 10.1016/0167-9457(90)90005-X. URL https://www.sciencedirect.com/science/article/pii/016794579090005X.

Mathieu Mangeot and Emmanuel Giguet. Multilingual Aligned Corpora From Movie Subtitles. Research Report, LISTIC, December 2005. URL https://hal.science/hal-00968632.

Stefano Maranzana. Intermediate Learner Opinions on Captioned Video Cartoons for Language Acquisition. In *Emerging Concepts in Technology-Enhanced Language Teaching and Learning*, pages 232–252. IGI Global, 2022. ISBN 978-1-79988-981-6. doi: 10.4018/978-1-7998-8981-6.ch012. URL https://tinyurl.com/2d7et5rz.

Richard E. Mayer. Introduction to Multimedia Learning. In Richard E. Mayer, editor, *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, pages 1–24. Cambridge University Press, Cambridge, 2 edition, 2014. ISBN 978-1-139-54736-9. doi: 10.1017/CBO9781139547369.002. URL https://tinyurl.com/multimediamayer.

Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, 3 edition, 2020. URL https://www.cambridge.org/highereducation/books/multimedia-learning/FB7E79A165D24D47CEACEB4D2C426ECD#overview.

Tayebeh Mosavi Miangah. Mobile-Assisted Language Learning. *International Journal of Distributed and Parallel systems*, 3(1):309–319, January 2012. ISSN 22293957. doi: 10.5121/ijdps.2012.3126. URL http://www.airccse.org/journal/ijdps/papers/0112ijdps26.pdf.

Mohammed Ali Mohsen and Mohammed Balakumar. A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*, 23(2):135–159, 2011. Publisher: Cambridge University Press.

Maribel Montero Perez. INCIDENTAL VOCABULARY LEARNING THROUGH VIEWING VIDEO: THE ROLE OF VOCABULARY KNOWLEDGE AND WORKING MEMORY. *Studies in Second Language Acquisition*, 42(4):749–773, September 2020. ISSN 0272-2631, 1470-1545. doi: 10.1017/S0272263119000706. URL https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/abs/incidental-vocabulary-learning-through-viewing-video/CA580AAB6C71179403EF440AA70083B8.

Maribel Montero Perez. Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, 55(2):163–192, April 2022. ISSN 0261-4448, 1475-3049. doi: 10.1017/S0261444821000501. URL https://tinyurl.com/perezavi. Publisher: Cambridge University Press.

Maribel Montero Perez and Michael P. H. Rodgers. Video and language learning. *The Language Learning Journal*, 47(4):403–406, August 2019. ISSN 0957-1736. doi: 10.1080/09571736.2019.1629099. URL https://doi.org/10.1080/09571736.2019.1629099. Publisher: Routledge _eprint: https://doi.org/10.1080/09571736.2019.1629099.

Jaap M. J. Murre and Joeri Dros. Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLoS ONE*, 10(7), July 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0120644. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4492928/.

Carmen Muñoz. Audiovisual input in L2 learning. *Language, Interaction and Acquisition*, 13(1):125–143, August 2022. ISSN 1879-7865, 1879-7873. doi: 10.1075/lia.22001.mun. URL https://www.jbe-platform.com/content/journals/10.1075/lia.22001.mun. Publisher: John Benjamins.

Rahat Naqvi, Keoma J Thorne, Christina M Pfitscher, David W Nordstokke, and Anne McKeough. Reading dual language books: Improving early literacy skills in linguistically diverse classrooms. *Journal of Early Childhood Research*, 11(1):3–15, 2013. URL https://journals.sagepub.com/doi/abs/10.1177/1476718x12449453. Publisher: Sage Publications Sage UK: London, England.

Paul Nation. Helping learners take control of their vocabulary learning. *GRETA*, 6(1):9–18, 1998. URL https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/1998-Autonomy-GRETA.pdf.

Paul Nation. *Learning Vocabulary in Another Language*. Cambridge Applied Linguistics. Cambridge University Press, Cambridge, 2 edition, 2013. doi: 10.1017/CBO9781139858656. URL https://www.cambridge.org/core/books/learning-vocabulary-in-another-language/25C481C771758930E8EBBE811231FEB7.

Hiu Lê Ngc, Xuân Trn Khai, Ngo Hue Phuong, Ung Phu Long, Trieu Trach Nam, and Nguyen Thien Ngan. Flashcard Design for Helping Students in Learning Chinese Vocabulary. *Journal of World Englishes and Educational Practices*, 3(10):10–23, October 2021. ISSN 2707-7586. doi: 10.32996/jweep.2021.3.10.2. URL https://al-kindipublisher.com/index.php/jweep/article/view/2238. Number: 10.

Eugene A Nida. Learning by listening. *Innovative approaches to language teaching*, pages 42–53, 1982.

Hiromi Nishioka. My Korean language teachers are YouTubers: learning Korean via self-instruction. *Computer Assisted Language Learning*, 36(3):346–374, March 2023. ISSN 0958-8221. doi: 10.1080/09588221.2021.1928227. URL https://doi.org/10.1080/09588221.2021.1928227. Publisher: Routledge _eprint: https://doi.org/10.1080/09588221.2021.1928227.

Lourdes Ortega. New CALL-SLA Research Interfaces for the 21st Century: Towards Equitable Multilingualism. *CALICO Journal*, 34(3):285–316, August 2017. ISSN 2056-9017. doi: 10.1558/cj.33855. URL http://journal.equinoxpub.com/Calico/article/view/7636. Number: 3.

Sue E. K. Otto. From Past to Present. In *The Handbook of Technology and Second Language Teaching and Learning*, pages 10–25. John Wiley & Sons, Ltd, 2017. ISBN 978-1-118-91406-9. doi:

10.1002/9781118914069.ch2. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118914069.ch2. Section: 2 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118914069.ch2.

Allan Paivio. *Mental Representations: A Dual Coding Approach*. Oxford University Press, 1986. ISBN 978-0-19-503936-8.

Steven C. Pan, Jahan Tajran, Jarrett Lovelett, Jessica Osuna, and Timothy C. Rickard. Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology*, 111(7):1172–1188, 2019. ISSN 1939-2176. doi: 10.1037/edu0000336. Place: US Publisher: American Psychological Association.

T. S. Paribakht and M. B. Wesche. Reading Comprehension and Second Language Development in a Comprehension-Based ESL Program. *TESL Canada Journal*, 11(1):09–29, October 1993. ISSN 1925-8917. doi: 10.18806/tesl.v11i1.623. URL https://teslcanadajournal.ca/index.php/tesl/article/view/623.

Anastasia Pattemore. *Learning grammatical constructions from audio-visual input*. PhD thesis, May 2022.

Annick Payne. Multilingual inscriptions and their audiences: Cilicia and Lycia. *Margins of Writing, Origins of Cultures*, pages 121–136, 2006. Publisher: The Oriental Institute.

Elke Peters and Stuart Webb. INCIDENTAL VOCABULARY ACQUISITION THROUGH VIEWING L2 TELEVISION AND FACTORS THAT AFFECT LEARNING. *Studies in Second Language Acquisition*, 40(3):551–577, September 2018. ISSN 0272-2631, 1470-1545. doi: 10.1017/S0272263117000407. URL https://tinyurl.com/petersandwebb. Publisher: Cambridge University Press.

Gill Philip. Classroom Concordancing in the 21st Century: The New Generation. *Language Forum*, 36(2), 2010. URL https://www.academia.edu/1090423/Classroom_Concordancing_in_the_21st_Century_The_New_Generation.

Karen Price. Closed-captioned TV: An untapped resource. *MATSOL*, pages 1,8, 1983. URL https://www.matsol.org/assets/documents/Currentsv12no2Fall1983.pdf.

Geòrgia Pujadas and Carmen Muñoz. Measuring the visual in audio-visual input: The effects of imagery in vocabulary learning through TV viewing:. *ITL - International Journal of Applied Linguistics*, March 2023. ISSN 0019-0829. doi: 10.1075/itl.22019.puj. URL https://benjamins.com/catalog/itl.22019.puj. Publisher: John Benjamins Publishing Company.

Joan-Tomás Pujolă. CALLing for help: researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, 14 (2):235–262, November 2002. ISSN 1474-0109, 0958-3440. doi: 10.1017/S0958344002000423. URL https://tinyurl.com/callingforhelppujola.

Tatsumoto Ren. Passive immersion, March 2023. URL https://tatsumoto-ren.github.io/blog/passive-immersion.html.

Barry Lee Reynolds, Yanwen Cui, Chian-Wen Kao, and Nathan Thomas. Vocabulary Acquisition through Viewing Captioned and Subtitled Video: A Scoping Review and Meta-Analysis. *Systems*, 10(5):133, October 2022. ISSN 2079-8954. doi: 10.3390/systems10050133. URL https://www.mdpi.com/2079-8954/10/5/133. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Jack C. Richards and Theodore S. Rodgers. *Approaches and Methods in Language Teaching*. Cambridge University Press, April 2014. ISBN 978-1-107-67596-4. Google-Books-ID: HrhkAwAAQBAJ.

Olly Richards. How To Learn Languages With Harry Potter: A Step-By-Step Guide, June 2017. URL https://storylearning.com/blog/learn-languages-with-harry-potter. Section: Reading Vs Writing.

Henry L. Roediger III and Andrew C. Butler. The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1):20–27, January 2011. ISSN 1364-6613. doi: 10.1016/j.tics.2010.09.003. URL https://www.sciencedirect.com/science/article/pii/S1364661310002081.

Doug Rohrer and Kelli Taylor. The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6):481–498, November 2007. ISSN 1573-1952. doi: 10.1007/s11251-007-9015-8. URL https://doi.org/10.1007/s11251-007-9015-8.

Doug Rohrer, Robert F. Dedrick, and Kaleena Burgess. The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, 21(5):1323–1330, October 2014. ISSN 1531-5320. doi: 10.3758/s13423-014-0588-3.

Doug Rohrer, Robert F. Dedrick, Marissa K. Hartwig, and Chi-Ngai Cheung. A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*, 112(1):40–52, 2020. ISSN 1939-2176. doi: 10.1037/edu0000367. Place: US Publisher: American Psychological Association.

Joshua Samani and Steven C. Pan. Interleaved practice enhances memory and problem-solving ability in undergraduate physics. *npj Science of Learning*, 6 (1):1–11, November 2021. ISSN 2056-7936. doi: 10.1038/s41539-021-00110-x. URL https://www.nature.com/articles/s41539-021-00110-x. Number: 1 Publisher: Nature Publishing Group.

Paweł Scheffler, Christian Jones, and Anna Domińska. The Peppa Pig television series as input in pre-primary EFL instruction: A corpus-based study. *International Journal of Applied Linguistics*, 31 (1):3–17, 2021. ISSN 1473-4192. doi: 10.1111/ijal.12298. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ijal.12298. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijal.12298.

Norbert Schmitt. Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3):329–363, July 2008. ISSN 1362-1688. doi: 10.1177/1362168808089921. URL https://doi.org/10.1177/1362168808089921. Publisher: SAGE Publications.

Norbert Schmitt. Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, 64(4):913–951, 2014. ISSN 1467-9922. doi: 10.1111/lang.12077. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/lang.12077. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lang.12077.

Aroline E. Seibert Hanson and Christina M. Brown. Enhancing L2 learning through a mobile assisted spaced-repetition tool: an effective but bitter pill? *Computer Assisted Language Learning*, 33(1-2):133–155, January 2020. ISSN 0958-8221. doi: 10.1080/09588221.2018.1552975. URL https://doi.org/10.1080/09588221.2018.1552975. Publisher: Routledge _eprint: https://doi.org/10.1080/09588221.2018.1552975.

Peggy Semmingson, Kathryn Pole, and Jodi Tommerdahl. Using bilingual books to enhance literacy around the world. *European Scientific Journal*, January 2015.

Burr Settles and Brendan Meeder. A Trainable Spaced Repetition Model for Language Learning. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1848–1858, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1174. URL https://aclanthology.org/P16-1174.

Jason Sharp. Sentence Mining: An Essential Tool for Language Learning, November 2010. URL https://web.archive.org/web/20110106193332/http://beyondbounds.org/2010/11/sentence-mining-an-essential-tool-for-language-learning/.

David Singleton and Larissa Aronin. Multiple Language Learning in the Light of the Theory of Affordances. *Innovation in Language Learning and Teaching*, 1(1):83–96, April 2007. ISSN 1750-1229. doi: 10.2167/illt44.0. URL https://doi.org/10.2167/illt44.0. Publisher: Routledge _eprint: https://doi.org/10.2167/illt44.0.

Skritter. Sentence Mining: Increase Vocabulary with Context < Skritter Blog, April 2012. URL https://blog.skritter.com/2012/04/sentence-mining/.

Mati Staniszewski. Multilingual Text Reader Tools, July 2023. URL https://elevenlabs.io/blog/multilingual-text-reader-tools/.

Jingyong Su, Junyao Ye, Liqiang Nie, Yilong Cao, and Yongyong Chen. Optimizing Spaced Repetition Schedule by Capturing the Dynamics of Memory. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10085–10097, October 2023. ISSN 1558-2191. doi: 10.1109/TKDE.2023.3251721. URL https://ieeexplore.ieee.org/document/10059206. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

Sublearning. Sublearning. URL https://sublearning.com.

John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988. Publisher: Elsevier.

Sarah K. Tauber, John Dunlosky, Katherine A. Rawson, Christopher N. Wahlheim, and Larry L. Jacoby. Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, 20(2):356–363, April 2013. ISSN 1531-5320. doi: 10.3758/s13423-012-0319-6. URL https://doi.org/10.3758/s13423-012-0319-6.

Taurus. Advice about sentence mining, November 2008. URL http://www.japanesepod101.com/forum/viewtopic.php?t=3641.

Lisa K. Taylor, Judith K. Bernhard, Suchi Garg, and Jim Cummins. Affirming plural belonging: Building on students' family-based cultural and linguistic capital through multiliteracies pedagogy. *Journal of Early Childhood Literacy*, 8 (3):269–294, December 2008. ISSN 1468-7984. doi: 10.1177/1468798408096481. URL https://doi.org/10.1177/1468798408096481. Publisher: SAGE Publications.

Jörg Tiedemann. Building a multilingual parallel subtitle corpus. *Proc. CLIN*, 14, 2007. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0d0eb34ab56f7b48b6d611b5d0767bc59ba8b9fc. Publisher: Citeseer.

R. J. Tierney and J. W. Cunningham. Research on Teaching Reading Comprehension. In D. Pearson, editor, *Handbook of Reading Research*. Longman, New York, 1984.

Leon Trotsky. *My Life*. Charles Scribner's Sons, New York, 1930.

Andreas Tsiartas, Prasanta Ghosh, Panayiotis Georgiou, and Shrikanth Narayanan. High-quality bilingual subtitle document alignments with application to spontaneous speech translation. *Computer Speech & Language*, 27(2):572–591, February 2013. ISSN 0885-2308. doi: 10.1016/j.csl.2011.10.002. URL https://www.sciencedirect.com/science/article/pii/S088523081100060X.

Sharon H. Ulanoff and Sandra L. Pucci. Learning Words From Books: The Effects of Read-Aloud on Second Language Vocabulary Acquisition. *Bilingual Research Journal*, 23(4):409–422, October 1999. ISSN 1523-5882. doi: 10.1080/15235882.1999.10162743. URL https://doi.org/10.1080/15235882.1999.10162743. Publisher: Routledge _eprint: https://doi.org/10.1080/15235882.1999.10162743.

Robert Vanderplank. Video and Informal Language Learning. In *The Handbook of Informal Language Learning*, pages 181–201. John Wiley & Sons, Ltd, 2019. ISBN 978-1-119-47238-4. doi: 10.1002/9781119472384.ch12. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119472384.ch12. Section: 12 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119472384.ch12.

Bill VanPatten and Megan Smith. *Key Questions in Second Language Acquisition: An Introduction*. Cambridge University Press, 2019.

Christopher N. Wahlheim, John Dunlosky, and Larry L. Jacoby. Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, 39(5):750–763, July 2011. ISSN 1532-5946. doi: 10.3758/s13421-010-0063-y. URL https://doi.org/10.3758/s13421-010-0063-y.

Yu-Chia Wang. Learning L2 Vocabulary with American TV Drama "From the Learner's Perspective". *English Language Teaching*, 5(8):217–225, May 2012. ISSN 1916-4742. URL https://eric.ed.gov/?id=EJ1079740. Publisher: Canadian Center of Science and Education ERIC Number: EJ1079740.

Stuart Webb. Extensive Viewing: Language Learning through Watching Television. *Language Learning Beyond the Classroom*, pages 159–168, January 2015. URL https://ir.lib.uwo.ca/edupub/107.

Stuart Webb and Paul Nation. *How Vocabulary is Learned*. Oxford University Press, 2017.

Stuart Webb, Takumi Uchihara, and Akifumi Yanagisawa. How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, 56(2):161–180, April 2023. ISSN 0261-4448, 1475-3049. doi: 10.1017/S0261444822000507. URL https://tinyurl.com/webbmeta.

Rong Wei and Lin Fan. On-Screen Texts in Audiovisual Input for L2 Vocabulary Learning: A Review. *Frontiers in Psychology*, 13:904523, May 2022. doi: 10.3389/fpsyg.2022.904523.

Sarah Shi Hui Wong, Amanda Chern Min Low, Sean H. K. Kang, and Stephen Wee Hun Lim. Learning Music Composers' Styles: To Block or to Interleave? *Journal of Research in Music Education*, 68(2):156–174, July 2020. ISSN 0022-4294. doi: 10.1177/0022429420908312. URL https://doi.org/10.1177/0022429420908312. Publisher: SAGE Publications Inc.

Piotr A Woźniak. Optimization of learning. *Unpublished master's thesis, Poznan University of Technology. Poznan, Poland*, 1990.

Piotr A. Woźniak and Edward J. Gorzelańczyk. Optimization of repetition spacing in the practice of learning. *Acta Neurobiologiae Experimentalis*, 54(1):59–62, 1994. ISSN 1689-0035. Place: Poland Publisher: Nencki Inst of Experimental Biology.

Xamuel. Sentence Mining, September 2009. URL https://web.archive.org/web/20110427030357/http://www.xamuel.com/sentence-mining/.

Junyao Ye, Jingyong Su, and Yilong Cao. A Stochastic Shortest Path Algorithm for Optimizing Spaced Repetition Scheduling. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pages 4381–4390, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 978-1-4503-9385-0. doi: 10.1145/3534678.3539081. URL https://doi.org/10.1145/3534678.3539081.

Zhiying Zhang and Stuart Webb. The Effects of Reading Bilingual Books on Vocabulary Learning. *Reading in a Foreign Language*, 31(1):109–139, April 2019. URL https://eric.ed.gov/?id=EJ1212816. ERIC Number: EJ1212816.

Andrew Zuo. i + 1 Sentence Mining Explained, September 2021. URL https://medium.com/litany-language-learning/i-1-sentence-mining-explained-1d0aee1e4af.