Comparison of the impacts of different multimodalities on incidental L2 vocabulary learning

AHMET ÇEKİÇ[1] & MEHMET DEMİREZEN[2]

Sivas Cumhuriyet University[1], Ufuk University[2]

Abstract
Multimodality of input in incidental L2 vocabulary learning has recently been a topic of interest among language acquisition researchers, yet the results have been somewhat contradictory. This study seeks to compare the impacts of two different multimodalities on incidental L2 vocabulary learning, namely, reading-plus-watching (experimental group I) vs. reading-plus-listening (experimental group II), as compared to the reading only condition, which is included as a control measure. Experimental group I watched and read the transcriptions of four news texts with electronic glosses for the target words, while experimental group II read and listened to the same news texts again with electronic glosses for the same 20 target words. Next, the two experimental groups swapped roles with a new set of four news texts glossed for another group of 20 target words. The control group only read the same eight news texts without glosses. The results suggest that reading-plus-listening can be a more conducive multimodal presentation for incidental vocabulary learning as compared to reading-plus-watching. The results also challenge the validity of some principles of the *Cognitive Theory of Multimedia Learning* in incidental L2 vocabulary learning, while providing supporting evidence for some other principles.[1]
Keywords: Second language acquisition, incidental L2 vocabulary learning, electronic glosses, multimodality of input

1 Introduction

The first two decades of the 21st century witnessed a rapid digitalization process in many areas. Education, particularly L2 (second or foreign language) instruction (Bax 2011), is not an exception. In line with this, multimedia learning environments involving viewing videos, listening to songs, and playing video games have now become the most preferred learning media for L2 learners (Lindgren & Muñoz 2013, Peters 2018). Increasingly affordable digital devices and ubiquitous access to the Internet have inevitably transformed learning and speeded up a transition from unimodal to multimodal learning environments (i.e. from books to audio cassettes, CDs, projectors, PCs and tablet screens) especially for foreign language learning.

In spite of such transformation in language learning in practice, Second Language Acquisition (SLA) research has fallen behind this rapid change. For instance, except for a few empirical studies (Feng & Webb 2019, van Zeeland & Schmitt 2012, Vidal 2003, 2011, Webb & Chang 2015), incidental L2 vocabulary learning research has so far emphasized unimodal presentation (mostly reading) while somewhat overlooking the affordances multimedia learning environments can offer

[1] The article is a shortened and updated version of the PhD dissertation of the first author under the supervision of the second author.

for incidental L2 vocabulary learning. Previous research includes examinations of the multimodality of electronic glosses (Çakmak & Erçetin 2018, Jones & Plass 2002), multimodality (reading-plus-listening) of the texts vs. reading (Brown, Waring, & Donkaewbua 2008, Kelly 1992), and comparisons of different modes of input (Feng & Webb 2019) in incidental vocabulary learning research. However, there is a paucity of research into the effects of bimodality vs. multimodality of text on incidental vocabulary learning. To address this research gap, the current study compares the potential conduciveness of bimodal input (the reading-plus-listening condition) vs. multimodal input (the reading-plus-viewing condition) for incidental L2 vocabulary learning as compared to reading only, the latter being the most widely researched and traditional mode of input in incidental L2 vocabulary learning.

2 Literature review

The current study pertains both to L2 acquisition and multimedia instruction. Therefore, it will benefit from the review of relevant theories of multimedia learning and that of incidental L2 vocabulary learning research. Below, we provide a review of the principles of the cognitive theories of multimedia learning with specific reference to multimodality of input, as well as of empirical incidental L2 vocabulary acquisition research.

2.1 Theoretical foundations

There is now well-established literature on how to optimize multimedia learning (Chan, Lyons, Kon, Stine, Manley, & Crossley 2020), and several theories of how the human mind processes multimodal information (Baddeley 1986, Mayer 1997, Paivio, 1969, Sweller 1988) have been proposed for multimedia instruction in L1. In spite of decades of research into incidental L2 vocabulary acquisition, only a few studies have investigated the effects of different modalities (listening and viewing) in incidental L2 learning and even fewer studies have benefitted from cognitive theories of multimedia learning in their discussions.

However, several principles and theories of multimedia learning can provide a theoretical framework for interpreting and researching incidental L2 vocabulary learning in multimedia environments. Multimedia learning theories indicate that the reading-plus listening and the reading-plus-viewing conditions can yield different incidental L2 vocabulary gains. For example, the *dual-coding theory* (Paivio 1969) of memory and cognition proposes that information is better represented, grasped and thus recalled when it is conveyed both verbally and visually. Likewise, Baddeley's (1986) *theory of working memory* (WM) postulates that the WM is composed of two partially independent stores – a phonological loop and a visual-spatial sketch pad for storing and processing verbal and visual input. In addition, a more recent and comprehensive theory of multimedia learning proposed by Mayer (1997), the *Cognitive Theory of Multimedia Learning* (CTML) features three main assumptions and several principles. The *limited capacity assumption* assumes that there is a certain amount of processing capacity for each channel. Further, similar

to the *dual-coding theory*, the *dual-channel assumption* postulates that learning is facilitated when information is presented via both channels. Therefore, bimodal input can facilitate efficiency of the WM by extending its capacity beyond that of a single channel. The third assumption is the *Active Processing Assumption*, which posits that learning is an active process involving filtering, selection, organization and integration with preceding knowledge. As for the principles assumed in the CTML, the *congruency principle* postulates that learning is enhanced when the information presented via the two channels is congruent. In addition to this, the *spatial* and the *temporal contiguity principles* postulate that there should be spatial and temporal proximity in the presentation of the visual and verbal information. Finally, the *redundancy principle* posits that when visual text and auditory text are presented concurrently in presentation, information processing is hindered as either the visual text or the auditory text becomes redundant (Clark & Mayer 2016). Thus, refraining from narration in combination with an identical written text can enhance learning. The basic rationale is that people cannot focus when the same verbal message is presented via written text and narration (Hoffman 2006).

Research into multimedia instruction in L1 suggests that presenting verbal input as narration rather than as coexisting on-screen text can increase learning in an animated presentation (Mayer & Moreno 1998; Mayer Moreno, 2003). Their findings evidence the *dual-channel* and the *limited capacity assumptions* of the CTML (please note that despite its verbal character, on-screen is regarded as visual input since it initially impinges on the eyes). The fact that on-screen text uses the processing capacity of the visual channel can be deduced from the research suggesting that on-screen text and animated presentation overload the visual channel (Mayer & Moreno 1998). In sum, according to the *dual-channel assumption*, bimodal input (i.e. the reading-plus-listening condition) can facilitate learning in multimedia environments.

However, bimodal L2 input seems to violate the *redundancy principle*, which postulates that any redundant input can hinder learning. According to this principle, on-screen text is redundant with the auditory narration as both present the same information concurrently. The *redundancy principle* has been evidenced by many researchers (Mayer, Heiser, & Lonn 2001, Mayer & Johnson 2008) for multimedia instruction in L1 (mother tongue). Yet, SLA research into bimodality (reading-plus-listening) (Brown et al 2008, Horst, Cobb, & Meara 1998, Kelly 1992, Webb & Chang 2015) and multimodality, e.g. viewing videos with subtitles/captions (Montero Perez, Noortgate, & Desmet 2013, Peters, Heynen, & Puimege 2016) reports findings to the contrary. L2 research suggests enhancing effects of on-screen text with auditory narration. Likewise, Chan et al (2020) found that as it can enhance word decoding, on-screen full text facilitated retention more than on-screen summary text with foreign-accented narration in an animated presentation. The results of these studies suggest that presenting concurrent on-screen full text does not make it redundant with the auditory narration in the case of L2; on the contrary, it can have a scaffolding effect for lower proficiency EFL learners, who are challenged by the burden of decoding auditory input in a foreign language. In

sum, we thus conclude that given that auditory narration and on-screen texts make use of the processing capacities of different channels without violating the *redundancy principle* in the case of L2, verbal (auditory) and visual (on-screen text) presentation of information can enhance incidental L2 vocabulary learning.

The explanations that can be construed from the current theories of cognition for the case of the reading-plus-viewing condition are a bit more complicated. The *limited capacity principle* of the CTML points to the limited capacity of the visual- and the verbal channel. Thus, two sources of visual information (i.e. pictures in the video and the on-screen text) both impinging on the eyes can overload a single channel – the visual channel – and hinder information processing. According to this, too much information presented visually in the case of viewing videos with on-screen texts (i.e. the reading-plus-viewing condition) can overload the visual channel and impede incidental L2 vocabulary learning. However, as already mentioned, SLA research into the effects of the subtitles/captions on incidental L2 vocabulary learning (Montero Perez et al 2013, Montero Perez, Peters, Clarebout & Desmet 2014, Lin 2014, Peters et al 2016, Syodorenko 2010) has evidenced the positive effects of on-screen text (even if it is redundant with audio narration) especially only with the key words are presented or with the key words highlighted (Montero Perez et al 2014). In a similar study, Chan et al (2020) found that on-screen full text can enhance retention when the narration is foreign-accented, an effect which is not existent with native narration.

According to incidental L2 vocabulary learning research and the results of Chan et al (2020), with foreign-accented narration, even if on-screen text can overload the visual channel when combined with the visuals in the video, it can be facilitative rather than impeding in L2. However, whether presenting on-screen full text in bulk can contribute to L2 incidental learning still needs to be researched. To the best of our knowledge, no study has examined the effects of presenting on-screen text as an additional source of input to video and audio narration in EFL. Yet, some inferences for EFL can be made from the study by Chan et al (2020), who compared the effects of no text, summary text and full text with native vs. foreign-accented narration in learning from an animated PowerPoint presentation by using knowledge transfer and retention as measure. They reported that while presentation of full text enhanced word level decoding and retention of foreign-accented narration (in L1) compared to summary text, the participants performed better with no text compared to summary text in native narration.

Given the enhancing effects of captions in the case of EFL and the facilitative effect of on-screen full text with foreign-accented narration, we believe that the redundancy effect and the *limited capacity assumption* might not be as evident in ELF contexts as in other L1 multimedia learning environments. In sum, while the SLA literature validates the *dual-channel assumption* (in the reading-plus-listening condition), it provides evidence against the *redundancy principle*. As can be deduced from captioning studies and the facilitative effects of full text in foreign-accented narration, the results with regard to the *limited capacity assumption* are contradictory.

2.2 Incidental L2 vocabulary research

Apart from the theoretical foundations for the current study presented above, research into incidental L2 vocabulary learning is also relevant: decades of research into incidental L2 vocabulary learning has evidenced its veracity and indispensability for L2 learning. Incidental learning is defined as learning occurring when learners acquire new aspects of language without paying special attention to do so (van Zeeland & Schmitt 2012). However, it is also argued that incidental vocabulary learning is mostly be contingent on the extent of attention to both meaning and form (Resrtrepo Ramos 2015). The need for incidental L2 vocabulary learning is also emphasized by research into essential L2 vocabulary size, which indicates varying figures ranging from 2,000 to 4,000 word families (van Zeeland & Schmitt 2012) to 6,000 to 7,000 word families (Nation 2006) for understanding spoken language, and to as high as 9,000 word families for unassisted comprehension of written discourse (Nation 2006). Given the limited time normally allocated for L2 vocabulary learning in L2 educational contexts, it is now generally agreed that incidental vocabulary learning is indispensable in L2 learning, and thus it needs to be systematically incorporated into L2 instruction not only due to time constraints on explicit vocabulary learning, but also due to the incremental nature of L2 vocabulary acquisition and the multidimensional nature of lexical knowledge, even for words from the higher frequency bands.

Reading was regarded as the main source of incidental vocabulary learning both in L1 (e.g. Nagy, Anderson, & Herman 1987) and L2 (Horst et al 1998, Waring & Takaki 2003). As is well known, when learners meet new words in different contexts, they need to get familiar with words in their various dimensions, such as in the written form, as a certain part of speech, as part of a collocation and from various semantic perspectives. This multidimensional nature of lexical knowledge requires multiple encounters with words through extensive reading (Feng & Webb 2019).

Earlier studies on L2 (Day, Omura & Hiramatsu 1991, Pitts, White, & Krashen 1989, Saragi, Nation, & Meister 1978) examined the rate of incidental vocabulary gain from reading novels or short stories and found that reading enhances L2 vocabulary gain in interaction with the frequency of occurrence of words and the proficiency level of the participants (Horst et al 1998, Pellicer-Sanchez & Schmitt 2010, Waring & Takaki 2003). As graded readers include more than one occurrence of the same word for many times in the text (Nation & Wang 1999), they were used in many studies and studies reported various numbers ranging from 7 to 10 and even up to 20 as a necessary frequency of occurrence for acquisition (Chen & Truscott 2010, Webb 2007).

Listening was less studied in incidental vocabulary learning and was initially regarded as an additional input to reading (Brown et al 2008, Horst et al 1998, Kelly 1992, Webb & Chang 2015). Kelly (1992) studied the additional facilitative effects of adding listening to reading in learning vocabulary and found that dual modality can increase retention of vocabulary. In another study, Brown et al (2008) studied

incidental vocabulary learning from three different modes of input: reading, reading-plus-listening, and listening and found that both reading modes were more conducive to incidental vocabulary learning compared to listening only for the participants in their study. In a similar study, Webb and Chang (2015) studied incidental vocabulary learning by reading-plus-listening 10 graded readers. They reported relatively high rates of vocabulary learning (44%) similar to the result of Brown et al (2008) (45%) but not compatible with the 22% rate of learning reported by Horst et al (1998). In sum, studies into the effects of the reading-plus-listening condition indicate that this kind of bimodal presentation of information can have positive, or at least no negative, impacts on incidental vocabulary gain.

Although not directly related to our study, more recent studies focusing on listening only for incidental vocabulary learning indicate its potential for incidental vocabulary learning – though weaker as compared to reading only (Çakmak & Erçetin 2018, Jones 2006, Jones & Plass 2002, van Zeeland & Schmitt 2013, Vidal 2003, 2011). Vidal (2003) found that listening can be a source of vocabulary learning in interaction with the proficiency level and the frequency of occurrence of the target words. In another study, Vidal (2011) compared reading of, versus listening to, academic texts in terms of incidental vocabulary learning and found that readers outperformed listeners for all four proficiency levels with the gap narrowing with increased proficiency level. The results indicate that presentation of text can visually enhance incidental learning particularly for EFL with low proficiency level. However, Feng and Webb (2019) found that reading, listening and viewing led to similar incidental vocabulary gains. These contradictory results can be attributed to the participants with high L2 proficiency in Feng and Webb's study. As already mentioned in the review of multimedia learning theories, research into the effects of listening either in comparison with reading (Vidal 2011) or compared to reading-plus-listening (Brown et al 2008) indicate that EFL learners, particularly those with low proficiency levels, benefit from the presentation of full text even if it is redundant. This contradicts with the redundancy effect, which is well-document in multimedia instruction in L1.

Given the importance of incidental learning for L2, some researchers have also examined electronic glosses as a way of enhancing incidental vocabulary learning through reading online. The studies reported positive effects of electronic glosses (Abraham 2008). A number of studies examined the effects of multimodal presentation of glosses as compared to single channel glosses (Akbulut 2007, Al-Seghayer 2001, Ariew & Erçetin 2004, Chun & Plass 1996, Jones 2003, Jones & Plass 2002, Plass, Chun, Mayer & Leutner 1998) and multimodal presentation of words (Bisson, van Heuven, Conklin & Tunney 2013, 2014, 2015). These studies show contradictory results. Furthermore, some other studies examined the effects of adding glosses to listening texts. For example, Jones and Plass (2002) examined the effects of adding pictorial, written and pictorial-plus-written glosses as compared to listening without glosses and found that multimodality of glosses enhances vocabulary acquisition. In a similar and more recent study, Çakmak and Erçetin (2018) studied incidental vocabulary learning through mobile-assisted L2

listening with the same three different combinations of glosses and no glosses. Unlike Jones and Plass (2002), they found that glosses had a significant effect on vocabulary learning with type of gloss having no effect. In sum, previous research indicates the superiority of adding glosses as compared to no gloss condition regarding incidental vocabulary learning, but the findings into the effects of mono vs. multimodal glosses are inconclusive.

Research into the effects of viewing with L1 subtitles, with captions (L2 subtitles) or with reversed subtitles (i.e., L2 video with L1 subtitles) on incidental vocabulary learning reports positive effects of captions (Montero Perez et al 2013, Montero Perez et al 2014, Lin 2014, Peters et al 2016, Syodorenko 2010). Syodorenko (2010) compared video with audio and captions (VAC, multimodality), video with audio (VA, bimodality appealing to two different channels), and video with captions (VC, bimodality appealing to the visual channel only). The results revealed that groups with captions (VAC and VC) scored higher on written than aural recognition tests with the VA group performing reversely. It can be concluded form this study that on-screen text can enhance written recognition even if it is redundant with the audio. The results of written and aural recognition tests provide evidence for the *dual-channel assumption* and also for separate processing of information in the phonological loop and visual sketchpad.

Text-related and individual factors also interact with incidental vocabulary acquisition. Peters et al (2016) found that captions can enhance form learning, with learners' size of vocabulary and frequency of occurrence positively correlating with vocabulary gain. In contrast, the results of a meta-analysis by Montero Perez et al (2013) on the impacts of captioned video on vocabulary learning (10 studies), which indicated a large effect of captions on vocabulary acquisition, reported that proficiency is not a moderating factor for vocabulary learning. Neuman and Koskinen (1992), who examined incidental L2 learning of 90 target words by 129 middle school bilingual students in the U.S. over 12 weeks by studying science videos, via viewing, viewing with captions, reading-plus-listening or reading only, found that viewing with captions led to more vocabulary gain compared to reading-plus-listening, watching without captions and reading only in word. Just like other multimodality studies in L2 (Montero Perez et al 2014, Lin 2014, Syodorenko 2010, Wong & Samudra 2019), the results of Neuman and Koskinen's (1992) study strongly suggest that the redundancy effect of on-screen text is not existent in L2 instruction.

Also, their results are contradictory to some other principles of the CTML as well. For example, their results contradict the *dual-channel* and the *limited-capacity assumptions*. According to the *dual-channel assumption*, the participants in the reading-plus-listening condition should have outperformed the readers. In line with the *limited capacity assumption*, concurrent presentation of two visual information channels, i.e. viewing with captions, should have overloaded the visual channel. However, contrary to these two principles, viewing with captions led to more vocabulary gain than bimodal input and unimodal input in Neuman and Koskinen's study (1992). Not including the principles of the CTML in their interpretation of

the results, the researchers concluded that different modalities facilitate incidental learning rather than overwhelming attentional capacity by ignoring the lower performance in the reading-plus-listening condition, which is also multimodal (Neuman & Koskinen 1992). However, their result can also be attributed to the congruency and contiguity of the visuals and the on-screen text in the science videos they used.

In addition to multimodality of texts and glosses, some researchers have examined the effects of multimodal presentation of the words separately rather than in a text. For example, Wong and Samudra (2019) found that dual-coding of vocabulary on screen improved incidental vocabulary learning, particularly when learners have low proficiency, and the words and pictures are aligned. The results confirm the *dual-coding theory* and the *congruency principle*, but, just like a number of other studies, refutes the redundancy effect of on-screen texts in L2 instruction.

To sum up, research into modality in incidental L2 vocabulary learning reports contradictory results, and also challenges some of the principles of multimedia learning theories. More importantly, well-established theoretical foundations such as the *dual-coding theory* of Paivio (1986), Baddeley's *theory of Working Memory*, and the *Cognitive Theory of Multimedia learning* (Mayer, 2009), which have been extensively evidenced in multimedia instruction in L1, are not researched to a great extent for L2, and the CTML is largely ignored in the interpretations of L2 findings. Moreover, existing studies report conflicting results and are also contradictory to the well-evidenced principles of the CTML (Brown et al 2008, Feng & Webb 2019, Kelly 1992, Neuman & Koskinen 1992, Vidal 2011). Furthermore, no study so far has investigated the comparative effects of multimodality, which obviously warrant further attention due to increased use of multimedia in L2 learning environments. To address this gap, this quasi-experimental study seeks to compare the effects of reading-plus-watching videos (on-screen text with glosses plus audiovisual presentation) with verbal glosses vs. reading-plus-listening (on-screen text with glosses plus auditory presentation) news texts with verbal glosses on incidental L2 vocabulary learning. The reading only condition, which does not include glosses, is included as a control measure. To this end, the study addresses the following research questions:

- How do reading-plus-listening with glosses, reading-plus-watching with glosses and reading only compare in terms of their conduciveness to incidental vocabulary learning?
- How do reading-plus-listening with glosses, reading-plus-watching with glosses and reading only compare in terms of their conduciveness to incidental vocabulary retention?

3 Method

The study is a quasi-experimental study with two experimental groups and one control group and includes administration of a post-test and a retention test. Via

random cluster sampling, Turkish L1 EFL learners in an ELT department in Turkey were assigned to experimental group I, experimental group II and a control group.

3.1 Participants and setting

A total of 107 Turkish L1 upper-intermediate EFL learners took part in the study. All of the participants had passed a proficiency exam at B1 level, CEFR and were freshmen EFL students. Most of the subjects were female (n=75) and their ages ranged between 18 and 22. As some missed some sessions, a total of 95 students were left in the groups (see Table 1). They were informed about the aim of the study and verbally expressed their consent. Data were collected in a computer lab under the supervision of the first author, where each participant had access to one computer with headphones and the internet.

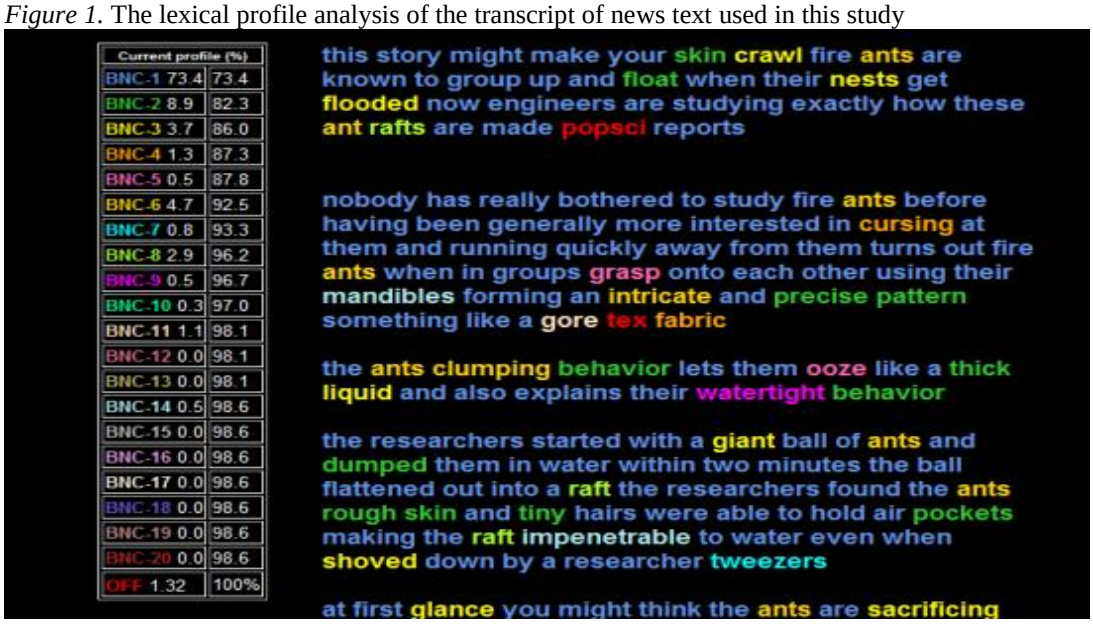
Table 1. Numbers of participants in each group

Group Name	**N**
Control group (Read-only)	29
Experimental group I (1. Watch-Read, 2. Listen-Read)	31
Experimental group II (1. Listen-Read, 2. Watch-Read)	35
Total	95

3.2 Materials

About 100 news texts (transcripts) from the science, health and technology sections of the newsy.com website ("Breaking News, World News, Online News – Newsy", n. d.) were analyzed using an online vocabulary profiler (Cobb, n. d.) to choose texts with possible target words (see Figure 1). Texts with an adequate amount of low frequency words were chosen.

Figure 1. The lexical profile analysis of the transcript of news text used in this study



Research suggests the lexical coverage necessary for adequate understanding of spoken text to be between 95 % (van Zeeland & Schmitt 2012) and 98% (Nation 2006). In this study, texts with a lexical coverage around 85% were chosen as the target words were glossed. The lower lexical coverage was deemed appropriate as the participants could see the meaning of the words thanks to electronic glosses. The students in both groups were given a list of possible target words and asked to check the words they knew. The groups received similar amounts of checks. The words (and the texts including them) known by more than 50% of any one of the groups were omitted.

The vocabulary levels of the participants were measured using the online version of the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham 2001), which revealed that the groups did not vary significantly ($t(64) = 0.73$, $p=0.94$). The results also suggest that the groups had mastered the 3000 words level, so we chose words from the lower frequency bands, between 3,000 and 9,000 words for the study.

As for the experimental groups I and II, the target texts were converted into web-pages with verbal glosses for the target words in the free and open-source learning management system: Modular Object-Oriented Dynamic Learning Environment (Moodle), developed in 2002 with the aim of providing a platform for educators and learners for educational purposes (Costa, Alvelos, & Teixeira 2012). The materials prepared for the experimental groups include:

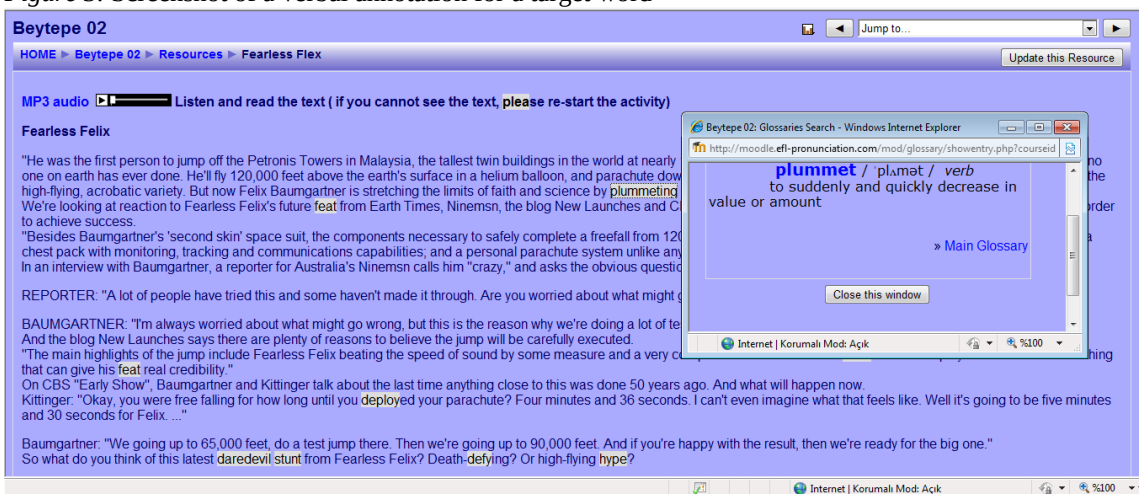
1. a total of eight news transcripts (texts) with verbal glosses for the target words,
2. 5 to 6 reading comprehension questions for each text,
3. verbal glosses for 40 target words in 8 texts, and
4. audio extracted from the videos.

An example of the reading-plus-watching (video with audio) condition is given in Figure 2. As for the reading-plus-listening condition, the audio files were extracted from the video using the software Audacity, version 1.3.13. On average, the videos, and thus the audio files, were 2.5 minutes long.

Figure 2. Screenshot of watching-plus-reading condition in Moodle

As can be seen in Figure 2, the target words are highlighted. The participants watched the news and saw not only the faces of the news reporters, but also moving pictures relevant to the content of the news texts. The verbal glosses for each word included a phonetic transcription, the relevant part of speech and a short L2 definition of the target word taken from *the Longman Dictionary of Contemporary English Online* (n. d.). To see verbal glosses, the participants could pause the video or audio and click on the word. We used the built-in hypertext technology in Moodle (see Figure 3). The ecological validity of making students watch and read news texts could be questioned, but for the reasons of comparability of the acquisition in two different modes, we had to keep other variables, such as the length, difficulty, genre and contextual features of the texts, in the experimental conditions constant.

Figure 3. Screenshot of a verbal annotation for a target word



3.3 Data collection tools

As in many other incidental vocabulary studies (Aldera & Mohsen 2013, Ariew & Erçetin 2004, Brown et al 2008, Çakmak & Erçetin 2017, Waring & Takaki 2003), no pre-test was given to avoid making implicit announcement of a test. As indication of a forthcoming test (Uchihara, Webb, & Yanagisawa 2018) can spoil the incidental learning condition – and instead trigger intentional learning, the participants were told that the glosses were given to support comprehension. In his meta-analysis of studies on glosses in incidental vocabulary learning, Abraham (2008: 223) noted that 17 of the studies involving the use of glosses did not use a pre-test. Sharing the concerns of the previous researchers that pre-testing can lead to intentional learning, we decided not to include a pre-test. The post-test required the participants to match 3 words with 6 short definitions. The test included 20 words from the first stage of the experiment and 20 words from the second stage of the experiment. Thus, the test was scored separately as post-test section I and post-test section II. In order to test the acquisition of 20 target words for each section, each correct answer was given five points for the ease of calculation and reporting, so the highest score to be obtained from one section of the post-test was 100. As Howell (2009: 62) states, linear transformation of results such as multiplying the score with a constant number does not affect the results of statistical tests.

Each set of 3 words was of the same part of speech, so that the participants could not find the correct answer just by guessing the part of speech of a word. The test was piloted with native speakers, ELT instructors and ELT trainees. As for face and content validity, a panel of experts, two native speakers of English and a faculty member teaching the participants, were asked to review the test items in the post-tests to confirm the content validity of the tests. As a result, some items (and /or distractors) were either replaced or omitted.

To validate the post-test, five educated native speakers of English, and five Turkish L1 English Language Instructors took the tests. Furthermore, 30 Turkish L1 upper intermediate EFL learners were asked to take the post-test for piloting the test. The native speakers also went through all items in the test and evaluated the correct answers and distractors, which helped us to further refine the test. Factor analysis indicated that the total variance explained in the pilot-test was 46.63%. Some items whose factor load was below .30 were not omitted in line with the views of experts. As a result, a post-test with 40 items and total variance level of 46.63% was obtained.

To determine the reliability of the post-test, the revised version of the test was again administrated to the same group of 40 test takers. The results of the reliability analysis indicated that Cronbach's Alpha value is .944. Furthermore, the discriminative power of the test was shown by the significant difference between the most successful test-takers (M=39.14, SD=5.49) and the least successful test-takers (M=11.28, SD=3.14), $t(1) = 11.64$, $p < .05$.

3.4 Experimental procedures

In the first stage of the treatment, the participants in experimental group I were exposed to 20 target words in 4 texts, which they watched (and listened to) and read the transcript on screen with verbal glosses, while the subjects in experimental group II listened to and read the same 4 texts with verbal glosses. In the second stage of the experiment, the experimental groups swapped roles. The subjects in the control group only read the 8 texts on paper with 40 target words without glosses and responded to the same reading comprehension questions as the experimental groups. We planned to add listening only as another group but lacked suitable subjects. The mean length of videos, and thus audios, was 2.5 minutes. Each text included 1800 to 2000 words. One of the researchers monitored the students in the lab and made sure that everyone referred to the glosses when they felt the need to do so. As the target words were chosen after a self-declared lack of knowledge in the pre-survey of vocabulary knowledge in all groups, most of the participants looked up the glossed words by pausing the video or audio. So, both the experimental sessions took 5 minutes longer than the control sessions. The participants in the experimental groups were given 20 minutes to process two texts in each session.

The instructional period in the study lasted for 5 weeks with a 40-minute session each week, with the first week allocated for the Moodle training of the experimental groups. In the current study, we used the transcriptions of news texts so that electronic glosses, which research suggests to be effective in vocabulary learning, could be added to the texts. An unannounced vocabulary test was given just after the experimental period and two weeks later. An overview of the study is provided in Table 2.

Table 2. Overview of the study

	Control Group	Experimental group I	Experimental group II
Week 1	No Instruction	Moodle Instruction	Moodle Instruction
Week 2	Read texts 1-2	Read & Watch texts 1-2	Read & Listen texts 1-2
Week 3	Read texts 3-4	Read & Watch texts 3-4	Read & Listen texts 3-4
Week 4	Read texts 5-6	Read & Listen texts 5-6	Read & Watch texts 5-6
Week 5	Read texts 7-8 and the Posttest	Read & Listen texts 7-8 and the Posttest	Read & Watch texts 7-8 and the Posttest
Week 7	Retention test	Retention test	Retention test

See Appendix A for the titles and target words in each text.

3.5 Data analysis

The study involved two experiments, each including 20 target words. Moreover, the experimental groups swapped roles as regards to the mode of input they received in each stage, and took the same test as the post-test and the retention test. Thus, there were two points to be analyzed: modality and testing point (time). Therefore, the data from the tests could have been analyzed more holistically using a two-way mixed ANOVA to reflect the role of modality and the two testing points, which

would also mean adding up the performances of the experimental groups in the same modality at two different stages. However, as different words are inherently more or less difficult to learn, the researchers chose to evaluate the performances of the experimental groups separately for each stage of the experiment and thus opted to analyze the data using ANOVA for the comparison of the three groups at each testing points and for each mode of presentation. This entailed the conduction of four ANOVA, and the groups' performances at each testing point and for each modality were compared using paired-samples t-tests.

4 Results

The results from the post-test were calculated separately in two sections as section 1 and section 2, respectively, because they reflect the performance in the first and second experiments. Table 3 presents a brief explanation of the experimental process.

Table 3. Content of section 1 and section 2 in post-test and retention test

Groups	Post-test Section 1 Retention Test Section 1	Post-test Section 2 Retention Test Section 2
Experimental Group I	Reading + watching	Reading + listening
Experimental Group II	Reading + listening	Reading + watching
Control Group	Reading	Reading

4.1 Analysis of the results of post-test section 1

A one-way ANOVA was conducted to compare the effects of the three different modes of input on incidental vocabulary learning. The analysis indicated that the different modes of input had significantly different effects on vocabulary gain [F (2.92) = 164. 985, p=0.001]. The Bonferroni test was used for post-hoc comparisons, which suggested that the differences between all groups are statistically significant. There was a significant difference between the experimental group I (*M*=49.03, SD=15.99) and the control group (*M*=23.79, SD=10.99) in favor of experimental group I. Moreover, experimental group II (*M*=83.42, SD=12.53) performed significantly better than experimental group I (*M*=49.03, SD= 15.99). Finally, the difference between experimental group II (*M*=83.42, SD= 12.53) and the control group (*M*=23.79, SD=10.99) was significant as expected.

4.2 Analysis of the results of the post-test section 2

The results of the second one-way ANOVA also showed that different modes of input had significantly different effects on vocabulary gain [F (2.92) = 98.612,

p=0.001] according to the test results obtained from section 2. The post-hoc comparisons using the Bonferroni test demonstrated that the performances of the groups differed significantly. In other words, experimental group I (*M*= 47.90, SD= 10.86) significantly outperformed the control group (*M*=18.44, *SD*=13.16). Besides, there was a significant difference between experimental group II (*M*=71.71, SD=19.24) and experimental group I (*M*= 47.90, SD=10.86) in favor of the former. Experimental group II (*M*=71.71, SD=19.24) performed significantly better than the control group (*M*=18.44, SD=13.16) on the post-test section 2 (see Table 4).

Table 4. Descriptive statistics for the post-test

	Post-test Section 1			Post-test Section 2	
	N	*M*	Sd.	*M*	Sd.
Experimental Group I	31	49.03	15.99	47.90	10.86
Experimental Group II	35	83.42	12.53	71.71	19.24
Control Group	29	23.79	10.99	18.44	13.16

4.3 Analysis of the results of the retention test section 1

The post-test was used as the retention test, as well. The results of one-way ANOVA also showed that the vocabulary retention rates differed significantly according to the modes of input [F (2.92) = 89.549, p=0.001]. The results of the Bonferroni test showed that the differences between all the groups were statistically significant. There was a significant difference between experimental group I (*M*= 37.25, *SD*= 13.46) and the control group (*M*=20.86, *SD*=11.73). In addition, experimental group II (*M*=66.00, *SD*= 15.37) significantly outperformed the experimental group I (*M*=37.25, *SD*= 13.46). Lastly, experimental group II (*M*=66.00, *SD*= 15.37) significantly outperformed the control group (*M*=20.86, *SD*=11.73).

4.4 Analysis of the results of the retention test section 2

The results of one-way ANOVA indicated that the groups differed significantly in terms of retention rates [F (2.92) = 58.383, p=0.001]. The Bonferroni test was conducted, which demonstrated that the differences between all groups were statistically significant. The experimental group I (*M*=35.48, *SD*=12.13) gained significantly more words than the control group (*M*=20.17, *SD*=11.45). Moreover, experimental group II (*M*=58.71, *SD*= 18.00) scored significantly higher compared to experimental group I (*M*= 35.48, *SD*=12.13). Finally, the difference between experimental group II (*M*=58.71, *SD*= 18.00) and the control group (*M*=20.17, *SD*=11.45) was also significant (see Table 5).

Table 5. Descriptive statistics for the retention test

	Retention Test Section 1			Retention Test Section 2	
	N	*M*	Sd.	*M*	Sd.
Experimental Group I	31	37.25	13.46	35.48	12.13
Experimental Group II	35	66.00	15.37	58.71	18.00

Control Group	29	20.86	11.73	20.17	11.45

4.5 Post-test and retention test comparisons

The results of the paired-samples t-test conducted to reveal differences in the performance of the control group on the post-test and retention test showed that there was not a significant difference between the control group's mean scores in the first section of the post-test (*M*=23.79, *SD*=10.40) and the retention test (*M*=20.86, *SD*= 11.73); t (28) = 0.99, p=0.329, and in the second section of the post-test (*M*=18.44, *SD*= 12.16) and the retention test (*M*=20.17, *SD*= 11.45); t (28) = .56, p=.577). However, the lack of significant difference should not be considered to mean that the control group retained words better. The subjects in the control group already scored very low in the post-test.

It is clear from Table 6 that there is a significant difference between the mean scores of experimental group I in the first and second sections of the post-test and the retention test. Based on these results, it can be concluded that experimental group I did not retain the same level of vocabulary gain both in the reading-plus-listening condition and in the reading-plus-watching condition.

Table 6. Post-test and retention test results of experimental group I

	N	***M***	**Sd.**	**T**	**p**
Experimental Group I					
Posttest 1	31	49.03	15.99	3.35	.002*
Retention test 1	31	37.25	13.46		
Posttest 2	31	47.90	10.86	6.03	.001*
Retention test 2	31	35.48	12.13		

***** $p < .05$

As seen in Table 7, the mean scores of experimental group II differ significantly in the first and second sections of the post-test and retention test. This shows that experimental group II did not retain the vocabulary gained in any of the modalities.

Table 7. Comparison of the Posttest and the Retention Test Results of the Experimental Group II

	N	*M*	Sd.	T	p
Experimental Group II					
Posttest 1	35	83.42	12.53	7.39	.001*
Retention test 1	35	66.00	15.37		
Posttest 2	35	71.71	19.24	5.43	.001*
Retention test 2	35	58.71	18.00		

***** $p < .05$

So, the rate of attrition between the two sections of the post-test and of the retention tests does not indicate any difference between the two experimental conditions. Although comparing the performance of the experimental groups in the two different sections of the post-test and the retention test comes with its drawbacks (as already discussed at the beginning of the results section the inherently different

nature of the words in the two sections can cause a problem), we decided that within subject comparisons could also shed more light on the comparison of the two different multimodalities of input. These comparisons indicate that only experimental group II performed significantly better in the reading-plus-listening-condition (*M*=83.42, *SD*=15.99) than in the reading-plus-watching condition (*M*=71.71, *SD*=10.86); t (34) = .593, p=.001, on the post-test. The performance of experimental group II reading-plus-listening condition (*M*=66.42, *SD*=15.37) and reading-plus-watching condition (*M*=58.71, *SD*=18.00); t (34) = 3.281, p=.002, differed significantly on the retention test as well. So, this consistent difference in the performance of the experimental groups in both testing points in favor of the bimodal (written-plus-audio) input can be considered to indicate the conduciveness of bimodal input over multimodal input (written-plus-audiovisual).

5 Discussion

Overall, the results of the study show that both of the experimental groups outperformed the control group both in the post-test and the retention test, but any difference between the control and the experimental groups would be difficult to interpret as we cannot tell whether the difference can be attributed to the mode of text or the lack of glosses. However, based on previous research indicating the positive effects of glossing (Abraham, 2008), significantly positive results in favor of the experimental groups as compared to the control group can be largely attributed to electronic glosses. The study aimed to compare two multimodalities, however, and not glosses, which have already been proven to be effective (Abraham, 2008). We included as reading only without glosses as control since it is the most common mode in EFL classes. Therefore, we will not be discussing the differences between the experimental groups and the control group any further as it is not our main aim. Besides, interpretation of the significant difference between the control and the experimental groups based on previous modality studies would not yield sound results since previous research does not include glosses and compares mostly single modalities; i.e. reading vs. listening (Brown et al. 2008, Kelly 1992, Vidal 2011) and viewing (Feng & Webb 2019).

Instead, our main discussion concerns the comparison of the different multimodalities of texts, which is somewhat controversial and challenging to interpret due to the paucity of previous research and results that contradict the assumptions made in the CTML. Therefore, the discussion of our results weakly indicating superiority of the reading-plus-listening condition as compared to reading-plus-viewing is limited to previous similar studies comparing multimodal and single mode texts (Brown et al 2008, Feng & Webb 2019, Horst et al 1998, Kelly 1992, Neuman & Koskinen 1992, Vidal 2011, Webb & Chang 2015), multimodal presentations in incidental word learning tasks (Bisson et al 2013, 2014, 2015), studies into using captions in videos (Montero Perez et al 2013, Montero Perez et al 2014, Lin 2014, Peters et al 2016, Syodorenko 2010) and relevant theories (Baddeley 1986, Mayer 2009, Paivio 1969).

In the first experiment, the mean score of experimental group II (reading-plus-listening) is significantly higher than the mean score of experimental group I (reading-plus-watching). The significant difference between experimental groups I and II in favor of experimental group II can be interpreted to mean that reading-plus-listening to news texts leads to better incidental vocabulary acquisition than reading-plus-watching news texts. However, our finding that reading-plus-listening is superior is not verified in the second experiment, where experimental group II, who was presented with the text in the reading-plus-watching modality, outperformed experimental group I in the reading-plus-listening condition.

These contradictory results in favor of experimental group II can be attributed to some uncontrolled variables, such as motivation, cognitive style differences, language proficiency etc. Although we gave some measures to ensure that the participants had the same level of proficiency and knowledge of the target words, we could not give a variety of pre-tests due to time constraints. We did not give language learning motivation scale, but we gave a cognitive style scale, which indicated that almost 90% of the participants were visualizers. For this reason, we did not include the data in the analysis and could not make assignments to conditions based on cognitive style. We made sure that the participants in both groups had similar language proficiency, and gave them a vocabulary level test, which revealed no significant difference between the groups. Their L2 proficiency levels were also expected to be similar because they had taken the same university entrance exam, the same prep exemption exam to enroll to the same department and were randomly assigned to the sections in the same grade.

The significantly higher performance of experimental group II in the reading-plus listening modality can also be attributed to different difficulty levels of the words in the different test sections. As can be seen in Table 4, all of the participants, including the control group, who were presented with the same single modality in both experiments scored higher in section 1 of the post-test and the retention test, which suggests that the words in section 2 are more difficult. The lower scores for all the participants in the second section of the post-test can also be attributed to a time effect. The participants who were more interested and motivated to learn words via annotated news texts might have lost their enthusiasm over time. This might in turn have caused their lower performance in the second modality they were exposed to. However, this time/order effect is not consistently evident in the performance of the groups. While experimental group I performed equally in both modalities, experimental group II performed better in the modality to which they were exposed first, i.e. the reading-plus-listening condition. Thus, in spite of the slight possibility that presentation time/order of modalities affected the results, the results can be attributed to different level of difficulty of the words sets in the test sections, to uncontrolled variables such motivation, aptitude etc. in addition to the modalities themselves.

To ensure that the words were evenly distributed in both sections of the test, the researchers randomly assigned the news videos and thus the target words to the sections in the post-test. They chose all the target words using frequency,

participants' lack of knowledge (50% of the participants did not know the words) and dissimilarity to the participants' L1 (Turkish). Yet, other features of the target words such as their orthography, morphology, phonology and semantics can also lead to differences in the words level of difficulty particularly for incidental acquisition.

On the one hand, the possibility that the words in Section 2 were more difficult might have contributed to the significantly higher post-test and retention test performance of experimental group II in the reading-plus-listening condition, because the words presented to them in this condition seem to be relatively easier. However, because of the higher performance of experimental group II in the reading-plus-watching condition compared to experimental group I in spite of the possibly more difficult words presented to experimental group II, the significantly higher performance of experimental group II compared to experimental group I are most likely due to uncontrolled variables (in favor of experimental group II) rather than to the different difficulty levels of the target words sets. However, the higher performance of experimental group II in the reading-plus-listening modality according to within-subject comparisons can be attributed partially to the modality and partially to the presentation of presumably easier words to them in this modality compared to reading-plus-watching modality.

On the other hand, given that the performance of experimental group I in both sections of the test was similar in spite of the more difficult words presented to them in the reading-plus-listening condition, the possibility that the words in section 2 were more difficult does not rule out the argument that the reading-plus-listening condition is more conducive to incidental vocabulary learning. This suggests that the reading-plus-listening condition can be more conducive to incidental vocabulary learning.

However, it should be noted that the deduction that the words in one section of the test are more difficult, the tentative assumption that the time/order of the presentation of modalities may have affected the results, and the interpretations that unmeasured variables might have contributed to the better performance of experimental group II are not based on solid statistical evidence. Therefore, we build our argument on more solid and statistically verified within-subject comparisons, which consistently and significantly indicate better performance of experimental group II in the reading-plus-listening condition on the post-test and the retention test. Although partially verified by between-subject comparisons, the answer for both research questions can thus be that the reading-plus-listening condition is more conducive to incidental vocabulary learning and retention compared to the reading-plus-watching condition. Yet, as indicated above, these suggestions should be read with caution as besides the different modalities, our results can be attributed to uncontrolled variables in favor of experimental group II and to the presumed difficulty level difference between the words presented in the different modalities.

In line with previous research indicating the facilitative effects of bimodality as in listening-plus-reading (Brown et al 2008, Horst et al 1998, Kelly 1992, Webb &

Chang 2015) in incidental L2 vocabulary learning, our results provide support for the *dual-channel assumption* of the CTML and the existence of the phonological loop and visual sketchpad (Baddeley 1986). Furthermore, the results consistently suggest that written text and auditory text are not redundant in L2 and thus that the *redundancy principle*, might not apply in the case of an L2. This can be attributed to the fact that on-screen or printed text enables lower proficiency L2 learners to decode words better and process the text at their own speed.

In a recent study, Chan et al (2020) reported that as it can enhance word decoding, on-screen full text facilitated retention more than on-screen summary text with foreign-accented narration in an animated presentation. Their result indicates that while overloading the visual channel by adding on-screen (or printed) text to pictures can hinder learning in L1 (as argued by the *redundancy* and the *limited-capacity principles*), it can facilitate learning in foreign-accented L1, which is difficult to decode just like L2. We also found that on-screen text can support the processing of foreign language narration, even if it is potentially redundant. This supports our finding that reading-plus-listening is conducive to vocabulary learning in spite of the assumed redundancy of the written text.

However, the facilitative effect of bimodality (reading-plus-listening) verified in L2 research is not verified for audiovisual bimodality. As the only study that compared multimodal (audiovisual) mode with single modes (written and spoken), Feng and Webb (2019) found that viewing television documentary (without captions) led to similar incidental vocabulary learning in all modes. Contrary to their expectations, multimodal input (audiovisual) did not yield better results for their participants (Chinese learners), which they attributed to the participants' unfamiliarity with the viewing mode as it is not a common part of EFL classes in China. In contrast to their study, adding video to onscreen and auditory text (reading-plus-watching) did not lead to similar vocabulary gains in our study; on the contrary, it yielded less incidental vocabulary gain as compared to bimodality (reading-plus-listening). We suggest that the limited gain in our study can be attributed to the overloading of the limited capacity of the visual channel by concurrent presentation of the on-screen text and moving pictures (video).

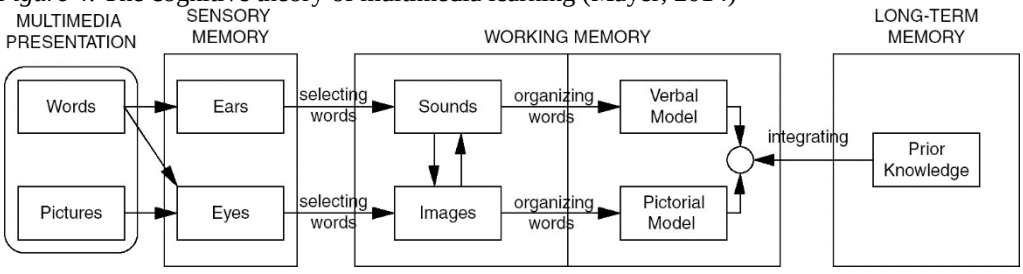
Another explanation for the lower performance in the reading-plus-watching condition is that moving pictures might have overloaded the visual channel of the learners, thus leading to the relatively lower vocabulary gain in our study or counterbalancing against the theoretically expected facilitative effects of dual-modality compared to single modes as in the study by Feng and Webb (2019). This indicates the existence of a very subtle balance between the facilitative effects of visuals and possibly overloading a channel. Evidence for this comes from the result of a gloss study (Chun & Plass 1996), which reported that video-annotations led to less vocabulary gain compared to standstill picture glosses. This suggests that cognitive overload is not just to do with simultaneous presentation of information to the same channel, but also with the amount of information presented at a time. It is better realized when we compare the example of the learning of lightening process, which Mayer (2002) use to explain the theory, with our experiment. In

Mayer's experiment, which shows facilitative effects of visuals added to narration, the pictures are fully congruent and presented as standstill pictures rather than moving pictures. The enhancing effects of visuals in his experiment seems to be contrary to our result, which indicates detrimental effects of visuals added to narration and on-screen text. However, when the results of Mayer's study and our result are compared regarding the *congruency principle* and the *limited capacity principle* of the CTML, the lower performance in the audiovisual-plus-written presentation (reading-plus-watching) as compared to reading-plus-listening condition can be attributed to the existence of the on-screen text, to the high amount of visuals (moving pictures) in the video and to lesser congruency and temporal-spatial contiguity in the news videos in our study as compared to standstill pictures in Mayer's experiment. Thus, we conclude that it is not merely the existence of visuals but their amount and congruency with verbal (auditory) information that determine their facilitative effects.

Depending on its content, audiovisual bimodality can violate the *congruence principle* of the CTML, which hypothesizes that higher correspondence between visual and verbal presentation enhances learning. Evidence for this principle is presented in a study by Wong and Samudra (2019), who found that young learners who watched *Sesame Street* picked up more words when they were dual-coded and presented congruently compared to non-congruent words in the same series. Given the relatively slower and more congruent nature of *Sesame Street* as compared to the news texts in our study and the documentary in Feng and Webb's study (2019), it is plausible that incongruence between audio and visuals may have hampered the learning process in our study (as compared to reading-plus-listening) or that it thus did not yield the expected superiority over single modes such as in Feng and Webb's study.

In line with the *limited capacity principle*, the reading-plus-watching condition involves overloading of the visual channel as both on-screen text (initially processed visually, though being verbal in essence; see Figure 4) and moving pictures (visual) are visually processed.

Figure 4. The cognitive theory of multimedia learning (Mayer, 2014)



According to the *limited capacity assumption*, concurrent presentation of on-screen text and video might have overloaded the visual channel in our study. However, this contradicts the results of Neuman and Koskinen (1992). Contrary to our findings, they found that audiovisual and written input was more conducive to

incidental vocabulary learning as compared to reading-plus listening. One potential explanation for the conflict between the results of our study and Neuman and Koskinen (1992) may be the contents of the videos and the different amount of text presented simultaneously in the two studies. In our study, we used news texts, which duly present the anchorperson's face at times, and thus achieve relatively less congruency between the on-screen text and the video compared to the science lesson videos they used. In their case, the researchers themselves pointed out the congruency of the visuals by stating that the science video provided an example of a concept first visually and then explained it in detail seconds later.

The contradictory results of the two studies can also be explained by the *temporal contiguity principle* of the CTML, which posits that humans learn better when corresponding words and pictures are presented simultaneously rather than successively and to the *spatial contiguity principle* of the CTML, which postulates that people learn better when corresponding words and pictures are presented in close proximity. The visuals and verbal presentations (captions) in the science lesson videos achieve more spatial and temporal contiguity compared to those in the news videos in our study. Moreover, we provided the whole transcriptions of news text in screen in relatively larger segments (see Figure 2), which position the words far from the visual and thus make it more difficult for participants to match the words with the visuals (as compared to captions). Therefore, the reading-plus-watching modality might have made it difficult for the participants to see both the video and the text simultaneously, instead having to switch between the two, which definitely would lower their attention and potentially their vocabulary gain as well.

Furthermore, on-screen full text added to the video in bulk might have overloaded the limited capacity of the visual channel more compared to the captions added to the video for each scene line by line. This interpretation is in line with the results of a meta-analysis study (Montero Perez et al 2013) suggesting facilitative effects of captions in incidental L2 learning. Facilitative impact of even shorter on-screen text is further evidenced in a captioning study (Montero Perez et al 2014), which reported that presenting only the keywords, or highlighting the key words in full captions, yielded higher vocabulary gain compared to full captions only. Thus, we suggest that on-screen text that is congruent with and (spatially and temporally) proximal to the visuals and presented in small manageable segments for easier processing (as in the case of captions) can be more conducive to incidental L2 vocabulary learning.

In sum, the discussion of our results in light of the existing body of research into multimodality in incidental vocabulary L2 vocabulary learning suggests that the design of L2 multimedia learning environments instruction can positively or negatively affect incidental L2 learning. In addition, L2 research challenges some of the principles of the CTML, which warrants further research.

6 Conclusion

Although the current study suffers from some limitations and from inadequate amount of previous research for comparison, our results suggest that:

- The redundancy effect might not be as evident in L2 learning as in multimedia learning in L1.
- The facilitative effects of auditory and written (text) dual-modality (reading-listening) can be seen in incidental L2 vocabulary learning.
- The facilitative effects of the audiovisual dual-modality (viewing-plus-listening) might not be as significant as for the auditory and written modality (reading-plus-listening), which we attribute to the affordances of the written text, which arguably facilitates the decoding and processing of the text particularly for lower proficiency L2 learners.
- Congruency, spatial and temporal contiguity between the visuals and on-screen text in written-plus-audiovisual modality can enhance incidental L2 vocabulary learning.
- The amount of the text (on-screen or printed) added to moving pictures (with audio) in written-plus-audiovisual modality should be limited to avoid overloading the limited processing capacity of the visual channel.

6.1 Pedagogical implications

The results of the current study lend support to multimodality of input (reading-plus-listening) and addition of electronic glosses for L2 vocabulary learning. However, the results also suggest that digital materials developers are to be cautious about adding moving pictures to written text input or vice versa. Given the lower vocabulary gains of experimental group II in the reading-plus-watching condition, cognitive load should be considered when presenting learners multimodal materials. Multimedia L2 materials should not present on-screen texts in bulk, but in smaller pieces suitable for processing. Also, multimedia materials that involve more congruency, and spatial and temporal contiguity between verbal and visual modes are preferable.

6.2 Limitations

The pedagogical implications of the study should be read in light of some limitations. The first limitation is that since this study aimed to reveal incidental vocabulary learning from listening versus watching authentic news texts, it was not quite appropriate and possible to use substitute or pseudo words, which would have provided more control over the possibility that the participants in different groups have varying levels of previous knowledge of the target words. As this is an experimental study, we asked the participants to read news transcripts, listen to them with transcripts and watch news with transcriptions, but this lacks ecological validity. Furthermore, the words could have been assigned to the experimental conditions (i.e. test sections) more evenly in terms of perceived difficulty by examining their semantic, orthographic and semantic features etc. Moreover, as an incidental vocabulary study, this research lacks a pre-test, which makes interpretations a bit challenging. Finally, the study is limited to findings of learners

at only one proficiency level, so it does not allow us to see how multimodality interacts with various levels of L2 proficiency.

6.3. Further studies

Having discussed the pedagogical implications and limitations of the study, we also have a number of suggestions for future studies. One interesting study would be to examine multimodal presentations in which captions are glossed as well as the difference of addition of full-captions vs. word specific glosses in interactive videos in incidental L2 vocabulary learning. In addition, it must be born in mind that modality is only one of the factors which can influence incidental vocabulary leaning. Thus, how multimodality of input interacts with variables such as frequency, contextual elaboration, proficiency level etc. needs to be further researched. Researchers can also study how individual differences like vocabulary size, listening proficiency, cognitive style, working memory capacity and word-related variables such as frequency of occurrence, contextual saliency etc. interact with multimodality of input in incidental L2 vocabulary gain.

On a methodological note, incidental L2 vocabulary studies should keep the period of experiment short and carry out the research under highly controlled conditions to prevent involvement of other uncontrollable variables like fluctuations in motivation and interest overtime, attrition and possible exposure to target words in other materials. Further, in spite of the existence of a number of incidental vocabulary studies without pre-tests (Abraham, 2008), we strongly recommend further studies to give participants vocabulary pre-tests that do not announce the post-test (which can be achieved by hiding the target words in a larger list of words and leaving a significant time gap between the pretest and the treatment). Due to the incremental development of lexical knowledge in incidental vocabulary learning conditions, further studies should include testing of multidimensional lexical knowledge.

In conclusion, a particularly interesting study as regards the generalizability of the results of this study would be to include learners at different proficiency levels and in different locations. Similarly, the congruency, and the contiguity of visual and verbal contents of the multimodal texts can be examined. This study is an initial attempt to discover the impacts of multimodality of input for incidental L2 vocabulary learning, yet as the amount of multimodality of input in L2 learning settings increases, multimodality of texts needs to be researched further to discover latent challenges and affordances it can have for incidental L2 vocabulary learning.

References

Abraham, Lee B. (2008), "Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis", *Computer Assisted Language Learning*, 21(3):199-226. https://doi.org/10.1080/09588220802090246

Akbulut, Yavuz (2007), "Effects of multimedia glosses on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language", *Journal of Instructional Science,* 35(6):499-517. https://doi.org/10.1007/s11251-007-9016-7

Aldera, Abdullah & Mohammed A. Mohsen (2013), "Annotations in captioned animation: Effects on vocabulary learning and listening skills", *Computers & Education,* 68: 60-75. https://doi.org/10.1016/j.compedu.2013.04.018

Al-Seghayer, Khalid S. (2001), "The effects of multimedia gloss modes on L2 vocabulary acquisition: A comparative study", *Language Learning and Technology*, 5(1): 202-232. http://dx.doi.org/10125/25117

Ariew, Robert & Gülcan Erçetin (2004), "Exploring the potential of hypermedia glosses for second language reading", *Computer Assisted Language Learning*, 17(2): 237-259. https://doi.org/10.1080/0958822042000334253

Audacity [Computer Software]. (2011), Retrieved from: https://www .audacityteam.org/

Baddeley, Alan (1986), *Working memory*. New York: Oxford University Press.

Bax, Stephen (2011), "Normalisation revisited: The effective use of technology in language education", *International Journal of Computer-Assisted Language Learning and Teaching*, 1(2):1-15. https://doi.org/10.4018/ijcallt. 2011040101

Bisson, Marie-Josée, Walter J. B. van Heuven, Kathy Conklin & Richard J. Tunney (2013), "Incidental acquisition of foreign language vocabulary through brief multi-modal exposure", *PLoS One*, 8(4):e60912. https://doi.org/10.1371/ journal.pone.0060912

Bisson, Marie Josée, Walter J. B. van Heuven, Kathy Conklin & Richard J. Tunney (2014), "The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary", *Language Learning*, 64(4): 855-877. https://doi.org/10.1111/lang.12085

Bisson, Marie Josée, Walter J. B. Van Heuven, Kathy Conklin & Richard J. Tunney (2015), "The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary", *The Quarterly Journal of Experimental Psychology*, 68(7):1306-1326. https://doi.org/10.1080/1 7470218.2014.979211

Breaking News, World News, Online News - Newsy. (n.d.) Retrieved from: https://www.newsy.com/

Brown, Ronan, Rob Waring & Sangrawee Donkaewbua (2008), "Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories", *Reading in a Foreign Language*, 20(2):136-163.

Chan, Kit Y., Claire Lyons, Lo L. Kon, Katelyn Stine, Melissa Manley & Anthony Crossley (2020), "Effect of on-screen text on multimedia learning with native and foreign-accented narration", *Learning and Instruction.* 67 (10):101305.

Chen, Chuntien & John Truscott (2010), "The effects of repetition and L1 lexicalization on incidental vocabulary acquisition", *Applied Linguistics* 31(5):693-713. http://doi.org/10.1093/applin/amq031

Chun, Dorothy M. & Jan L. Plass (1996), "Effects of multimedia glosses on vocabulary acquisition", *The Modern Language Journal*, 80(2):183-198. https://doi.org/10.2307/328635

Clark, Ruth C., & R. E. Mayer (2016), *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco, CA: John Wiley & Sons.

Cobb, Tom (n.d.), *Web VP / BNC-20 v 3.2 British National Corpus lists version (Beta)* Retrieved from: http://www.lextutor.ca/vp/bnc/

Costa, Carolina, Helena Alvelos & Leonor Teixeira (2012), "The use of Moodle e-learning platform: a study in a Portuguese University", *Procedia Technology*, 5:334-343.

Çakmak, Fidel & Gülcan Erçetin (2018), "Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening", *ReCALL*, 30(1):24-47. https://doi.org/10.1017/s0958344017000155

Day, Richard R., Carole Omura & Motoo Hiramatsu (1991), "Incidental EFL vocabulary learning and reading", *Reading in a Foreign Language*, 7(2):541-551.

Feng, Yanxue & Stuart Webb (2019), "Learning vocabulary through reading, listening, and viewing: which mode of input is most effective?" *Studies in Second Language Acquisition*, 42(3): 1-25. doi:10.1017/S0272263119000494

Hoffman, Bob (2006). *The Encyclopedia of Educational Technology*. San Diego, CA: Montezuma Press.

Horst, Marlise, Tom Cobb & Paul Meara (1998), "Beyond a clockwork orange: Acquiring second language vocabulary through reading", *Reading in a Foreign Language*, 11(2):207-223.

Howell, David C. (2009), *Statistical methods for psychology*. Belmont, CA: Cengage Learning.

Kelly, Peter (1992), "Does the ear assist the eye in the long-term retention of lexis?" *IRAL: International Review of Applied Linguistics in Language Teaching*, 30(2):137-145.

Jones, Linda C. (2003), "Supporting listening comprehension and vocabulary acquisition with multimedia glosses: The students' voice", *CALICO Journal*, 21(1):41-65. https://doi.org/10.1558/cj.v21i1.41-65

Jones, Linda C. (2006), "Effects of collaboration and multimedia glosses on vocabulary learning and listening comprehension", *CALICO Journal*, 24(1): 33-58. Retrieved from http://www.jstor.org/stable/24156293

Jones, Linda C., & John Plass (2002), "Supporting listening comprehension and vocabulary acquisition in French with multimedia glosses", *The Modern Language Journal*, 86(4):546-561. https://doi.org/10.1111/1540-4781.00160

Longman Dictionary of Contemporary English Online. (n.d.) Retrieved from: https://www.ldoceonline.com/

Lin, Phoebe M. S. (2014), "Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences", *System*, 42(1):164-176. https://doi.org/10.1016/j.system.2013.11.010

Lindgren, Eva & Carmen Muñoz (2013), "The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension", *International Journal of Multilingualism,* 10(1):105-129.

Mayer, Richard E. (1997), "Multimedia learning: Are we asking the right questions?" *Educational psychologist*, 32(1):1-19.

Mayer, Richard E. (2002), "Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction", *New directions for teaching and learning*, (89):55-71.

Mayer, Richard E. (2009), *Multimedia Learning* (2nd ed.) Cambridge: C.U.P.

Mayer, Richard E. (2014), Cognitive theory of multimedia learning. In R.E. Mayer (ed.), *The Cambridge handbook of multimedia learning* (2nd ed.):43-71, Cambridge: C.U.P. m

Mayer, Richard E., Julie Heiser & Steve Lonn (2001), "Cognitive constraints on multimedia learning: When presenting more material results in less understanding", *Journal of Educational Psychology,* 93(1):187-198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, Richard E. & Cherly I. Johnson (2008), "Revising the redundancy principle in multimedia Learning", *Journal of Educational Psychology,* 100(2): 380-386. https://doi.org/10.1037/0022-0663.100.2.380

Mayer, Richard E. & Roxana Moreno (1998), A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320. https://doi.org/10.1037/0022-0663.90.2.312

Mayer, Richard E. & Roxana Moreno (2003), "Nine ways to reduce cognitive load in multimedia learning", *Educational psychologist*, 38(1):43-52.

Montero Perez, Maribel, Wim Van Den Noortgate & Piet Desmet (2013), "Captioned video for L2 listening and vocabulary learning: A meta-analysis", *System*, 41(3): 720-739. https://doi.org/10.1016/j.system.2013.07.013

Montero Perez, Maribel., Elke Peters, Geraldine Clarebout, & Piet Desmet (2014). "Effects of captioning on video comprehension and incidental vocabulary learning", *Language Learning & Technology*, 18(1):118-141.

Nagy, William E., Richard C. Anderson & Patricia A. Herman (1987), "Learning word meanings from context during normal reading", *American Educational Research Journal*, 24(2): 237-270. http://doi.org/10.3102/00028312024002237

Nation, I. S. P. & Ming-Tzu K. Wang (1999), "Graded readers and vocabulary", *Reading in a Foreign Language*, 12(2):355-380.

Nation, I. S. Paul (2006), "How large a vocabulary is needed for reading and listening?" *Canadian Modern Language Review*, 63(1): 59-82. https://doi.org/10.3138/cmlr.63.1.59

Neuman, Susan B. & Patricia Koskinen (1992), "Captioned television as "comprehensible input": Effects of incidental word learning from context for language minority students", *Reading Research Quarterly,* 27(1):95-106. http://doi.org/10.2307.747835

Paivio, Allan (1969), "Mental imagery in associative learning and memory", *Psychological Review*, 76(3):241-263.

Paivio, Allan (1986), "*Mental representation; Dual-coding hypothesis*," Oxford: Oxford University Press.

Pellicer-Sánchez, Ana & Norbert Schmitt (2010), "Incidental vocabulary acquisition from an authentic novel: Do "things fall apart?" *Reading in a Foreign Language*, 22(1):31- 55.

Peters, Elke (2018), "The effect of out-of-class exposure to English language media on learners' vocabulary knowledge", *ITL - International Journal of Applied Linguistics*, 169(1):142-168.

Peters, Elke, Eva Heynen & Eva Puimege (2016), "Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions", *System*, 63: 134-148. https://doi.org/10.1016/j.system.2016.10.002

Pitts, Michael, Howard White & Stephen Krashen (1989), "Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers", *Reading in a Foreign Language*, 5(2):271-275.

Plass, Jan L., Dorothy M. Chun, Richard E. Mayer & Detlev Leutner (1998), "Supporting visual and verbal learning preferences in a second-language multimedia learning environment", *Journal of Educational Psychology*, 90(1): 25-36.https://doi.org/10.1037//0022-0663.90.1.25

Restrepo Ramos, Falcon D. (2015), "Incidental vocabulary learning in second language acquisition: A literature review", *Profile Issues in Teachers Professional Development, 17*(1), 157-166. https://doi.org/10.15446 /profile.v17n1.43957

Saragi, Thomas, I. S. Paul Nation & Gerry F. Meister (1978), "Vocabulary learning and reading", *System*, 6 (2):72-78. https://doi.org/10.1016/0346-251x(78)90027-1

Schmitt, Norbert, Diane Schmitt & Caroline Clapham (2001), "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test", *Language Testing*, 18 (1): 55-88. https://doi.org/10.1191/026553201668475857

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

Syodorenko, Tetyana (2010), "Modality of input and vocabulary acquisition", *Language Learning and Technology*, 14 (2):50-73. http://dx.doi.org/10125/44214

Uchihara, Takumi, Stuart Webb, & Akifumi,Yanagisawa (2019), "The Effects of Repetition on Incidental Vocabulary Learning: A meta-analysis of correlational studies", *Language Learning*, 69 (3):559-599.

van Zeeland, Hilde & Norbert Schmitt (2012), "Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?" *Applied Linguistics*, 34(4):457-479. https://doi.org/10.1093/ applin/ams074

van Zeeland, Hilde & Norbert Schmitt (2013), "Incidental vocabulary acquisition through L2 listening: A dimensions approach", *System*, 41(3):609-624. https://doi.org/10.1016/j.system.2013.07.012

Vidal, Karina (2003), "Academic listening: A source of vocabulary acquisition?" *Applied Linguistics*, 24(1): 56-86. https://doi.org/10.1093/applin/24.1.56

Vidal, Karina (2011), "A comparison of the effects of reading and listening on Incidental vocabulary acquisition", *Language Learning*, 61(10):219-258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Waring, Rob, & Misako Takaki (2003), "At what rate do learners learn and retain new vocabulary from reading a graded reader?" *Reading in a Foreign Language*, 15(2):130-163.

Webb, Stuart (2007), "The effects of repetition on vocabulary knowledge", *Applied Linguistics*, 28(1):46-65.

Webb, Stuart & Anna C-S. Chang (2015), "Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning?" *Language Teaching Research*, 19(6): 667-686. https://doi.org/10.1177/1362168814559800

Wong, Kevin M. & Preeti G. Samudra (2019), "L2 vocabulary learning from educational media: extending dual-coding theory to dual-language learners", *Computer Assisted Language Learning*, 1-23. 10.1080/09588221.2019.1666

Appendix A

Text No.	Title of the Text	Target words and their frequency in text
1	How to Sink a Fire Ant Raft	intricate, clump, ooze, impenetrable, porous, resilient, wreak havoc, swarm (2)
2	Fossilized Foot Bone Suggests Pre-humans Walked Upright	stiff (2), locomotion, residual, ambling,
3	One Cigarette is One Too Many	reproductive, implementation, complacency, clot, anchor, affiliate
4	The Brain and Sleep	slip, drawback, lapse, akin
5	Cops Use Data From TomTom to Set Speed Traps	doling, outrage, faze, outrage, savvy, lucrative, cynical, anonymity, debacles
6	INTEL 3D Chips	fin, unprecedented, stipulate, obsolete, breakthrough, debut
7	Airline Food Unsanitary	darn, impromptu, flatulence, stowaway, scrutiny
8	Fearless Flex to Attempt Free fall from Space	stun, plummet, feat (2), daredevil, stunt, hype