CrossMark

Studies in Second Language Acquisition **42** (2020), 653–663
doi:10.1017/S0272263120000145

Commentary

MULTIMODAL INPUT IN SLA RESEARCH

Maribel Montero Perez *

KU Leuven

INTRODUCTION

The studies in this special issue, *Language Learning from Multimodal Input*, investigate the potential of multimodal input for language learning from different angles. They focus on different types of linguistic outcomes, such as grammar, vocabulary (both single words and formulaic sequences), pronunciation, and comprehension but also on L2 learners' processing of multimodal input. In doing so, these studies examine research questions that have, so far, received little attention in the research literature and, thus, provide an important contribution to the field.

My commentary on this special issue consists of three parts. First, I discuss aspects related to the multimodality of input and the potential role of the multimedia learning theory (Mayer, 2001) in studies on multimodal L2 input. Next, I put forward three main reasons that may explain the importance of multimodal input for SLA researchers. I conclude this commentary by discussing the contribution of this special issue to theory, methodology, and pedagogy.

DEFINING MULTIMODAL INPUT

As language users, we are constantly surrounded by multimodal input (see Kress & Van Leeuwen, 2001 on multimodality). For instance, when we watch TV, we deal with incoming sound and images. In everyday conversations, we process sound, gestures, body language, and so forth. Being able to extract meaning from multimodal input is thus a crucial component of everyday life. Drawing on the multimodality of everyday life, the Douglas Fir Group (2016) has defined language learning as semiotic learning, which means that we can rely on a wide range of semiotic resources in our L2 meaning-making

I would like to thank the guest editors of this special issue, Carmen Muñoz and Elke Peters, for giving me the opportunity to write this commentary.

* Correspondence concerning this article should be addressed to Maribel Montero Perez, KU Leuven, Department of Linguistics and ITEC, imec Research Group at KU Leuven, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium. E-mail: maribel.monteroperez@kuleuven.be

© *The Author(s), 2020. Published by Cambridge University Press*

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

and learning processes, such as nonverbal information (e.g., facial expressions, body language), but also textual, graphic, or auditory resources. Language learning is thus necessarily multimodal learning.

During the last decade, we have witnessed a remarkable number of studies on the role of multimodal input for aspects of L2 acquisition. Even though studies already explored the potential of multimodal input in the nineties and the early 2000s (see Vanderplank, 2010 for a comprehensive review), we can clearly notice a renewed and reinforced interest into the role of multimodal input around 2009–2010. First of all, there is the publication of a number of corpus-based studies (e.g., Webb & Rodgers, 2009; Webb, 2010a, 2010b) that highlight the potential of TV as an important source of L2 input by investigating the vocabulary demands of television programs such as TV series and films to determine target vocabulary sizes needed for adequate comprehension of those programs. Second, a number of empirical studies such as the ones carried out by Winke et al. (2010) and Sydorenko (2010) on captioned video or Brown et al. (2008) and Chang (2009) on reading-while-listening can be considered important triggers in the study of multimodal input for L2 learning. After 10 years of intensive research, multimodal input has now earned its place in both SLA (Second Language Acquisition) and CALL (Computer Assisted Language Learning) research and takes up a central position on the international research agenda. This is not only evident from the current special issue, but also from previous initiatives such as the special issue on video and language learning for the *Language Learning Journal* (Montero Perez & Rodgers, 2019) and several symposia on multimodal input and language learning at international conferences (e.g., EuroSLA, 2018 and SLRF, 2018).

An analysis of the studies that have been published during the last decade shows that there is a wide range of possibilities to operationalize multimodal input for L2 learners. First of all, we notice that an increasing number of studies on multimodal input is based on audio-visual input (e.g., Rodgers & Webb, 2017) such as documentaries, TV series, films, or news clips that presents a combination of visuals (i.e., dynamic images) and sound. The bulk of studies has, however, looked into the role of audio-visual input enhanced with on-screen text. Previous research has shown that on-screen text can come in many different forms such as subtitles in the L1 (e.g., Peters et al., 2016), subtitles in the L2 or captions (see Montero Perez et al., 2013 for a meta-analysis), or other types of captions such as captions with enhanced target words or grammatical constructions (e.g., Cintrón-Valentín et al., 2019), captions with glosses (e.g., Montero Perez et al., 2018), or keyword captions (e.g., Montero Perez et al., 2014). Written text can also be made multimodal by adding, for instance, images and/or the accompanying audio, as is the case for reading-while-listening activities (e.g., Serrano & Pellicer-Sánchez, 2019; Tragant & Pellicer-Sánchez, 2019; Webb & Chang, 2015). Another option consists in enhancing written text with hyperlinks that link words in the text to multimodal annotations or glosses that contain word-meaning information in the form of text, sound, and/or pictures (e.g., Boers et al., 2017; Warren et al., 2018). These types of multimodal input have been studied from different perspectives and may influence not only L2 comprehension but also other aspects of L2 development such as L2 word learning or speech perception (see Montero Perez, 2019b, for a comprehensive review).

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

From a theoretical point of view, SLA research on multimodal input has often been linked with Mayer's (2001) Cognitive Theory of Multimedia Learning that builds on Paivio's dual-coding theory (1986). The central idea behind this theory is that human cognition presents separate channels for the processing of visual and verbal information. Even though these channels work separately, multimedia learning theory posits that learning outcomes can be enhanced when both channels are activated simultaneously. This is at the basis of the multimedia learning principle, which holds that "people learn better from words and pictures than from words alone" (Mayer, 2001, p. 63). Numerous studies have empirically confirmed this principle in the context of first language science learning (see Mayer, 2005 for an overview). Yet, it can also be seen that an increasing number of studies in the field of L2 learning from multimodal input draw on principles of this theory to frame research questions or explain research findings (e.g., Rodgers, 2018). It should, however, be noted that the cognitive theory of multimedia learning was originally not conceived as a theory of L2 learning from multimodal input. This implies that multimedia learning principles might be different for L1 learning contexts than for L2 learning contexts. As we see in the outcomes of previous research as well as in the results of studies in this special issue, this is clearly the case for the redundancy principle, which holds that "students learn better from animation and narration than from animation, narration, and text" (Mayer, 2001, p. 147). Yet, numerous studies have shown that combining audio or audio-visual input with accompanying (redundant) text results in better comprehension of the L2 input as well as in more linguistic benefits such as vocabulary learning. This is due to the fact that the presence of written input helps learners to decode and segment the speech stream and supports the development of learners' bottom-up listening processes (Charles & Trenkic, 2015). The more a learner's decoding skills are developed, the more time learners have to interpret the message or to notice unknown words. As argued by Plass and Jones (2005, p. 480), "[R]eading and listening are two competencies that both need to be developed, and in many cases one is used as input enhancement for the other."

Other multimedia learning principles, however, seem to operate similarly in L1 and L2 learning contexts. Recent research into the role of imagery for L2 learning from audio-visual input has, for instance, shown that the spatial and temporal contiguity principle, which holds that "students learn better when corresponding words and pictures are presented near rather than far from each other" (Mayer, 2001, p. 81), that is when they are "presented simultaneously rather than successively" (ibid., p. 96), can play an important role when selecting video materials for L2 learners. In this respect, a corpus-based study (Rodgers, 2018) revealed that 65% of potential target words (i.e., words that might be learned) in a documentary appeared simultaneously with the image, which might stimulate the learning of those words. This hypothesis was empirically investigated by Peters (2019, p. 1023) who reported that "words with on-screen imagery are almost three times more likely to be picked up incidentally than words without imagery" when the image appeared within a 5-second time frame before or after the presentation of the word in the auditory trace. This study thus provides empirical support for the spatial and temporal contiguity principle as well as for the multimedia learning principle. More importantly, this study also reveals that learners are able to use different sources of input (images, on-screen text, audio) offered by multimodal input in their meaning-making process, which may in turn lead to L2 learning gains.

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

EXPLAINING THE IMPORTANCE OF MULTIMODAL INPUT

There are at least three reasons that may explain the growing importance of multimodal input for SLA researchers and practitioners.

A first reason undoubtedly lies in the overall availability and easy accessibility of multimodal L2 input. Webb (2015) argued that the "greatest value of television for language learning might be its potential to provide large amounts of L2 spoken input" (p. 159). In addition, the quickly growing market of streaming services such as Netflix, HBO, or Amazon Prime Video; the availability of other resources such as YouTube; and an increasing number of TV channel websites offer enormous selections of TV programs, series, and films in different languages. We also see that the aforementioned services often come with numerous options to modify the input (see Chapelle, 2003 on input modification), which consist in the addition of information such as on-screen text to stimulate comprehension and language learning. One example is the Google Chrome extension "language learning with Netflix," which offers language support in the form of, for instance, subtitles in the target language. Words in those subtitles are clickable in the sense that they are linked to a dictionary that provides learners with access to word-meaning information. It also enables learners to create their own dictionary based on the series and programs they watched. Together, these tools and services highlight the role of technological advances. Indeed, new technologies have taken the realization of multimodal input to the next level by offering numerous options that were not available through more traditional (e.g., paper-based) resources.

The overall availability of L2 multimodal input is inextricably linked to findings on the role of out-of-class exposure for L2 development. In this respect, a number of studies on the role of English language exposure outside the classroom reported that watching TV was one of the most important out-of-class activities (Kuppens, 2010; Lindgren & Muñoz, 2013; Peters, 2018). In addition, it was found that watching foreign language movies or TV programs outside of the classroom was positively related to aspects of L2 development. Studies showed, for instance, that learners who often watched English language TV programs or movies had a larger receptive (Peters, 2018; Puimège & Peters, 2019) or productive (Bollansée et al., in press) vocabulary than learners who were less frequently exposed to this type of L2 input. Together, these findings indicate that multimodal input is not only easily accessible but frequent exposure to this type of input may also benefit learners' L2 development. In addition, it provides an interesting approach to stimulate contact with the L2 in contexts where exposure to the L2 is limited to the classroom.

A second reason why multimodal input has become an important research topic is theoretical in nature. Research into the effectiveness of different types of multimodal input is part of a larger research track that studies the impact of input and its characteristics on L2 acquisition from a cognitively oriented perspective. As can be seen in this special issue, multimodal input provides numerous options to study SLA constructs such as salience, noticing, and input enhancement. Studies have, for instance, used eye-tracking to investigate whether input enhancement in the form of visually enhanced target items or salient grammatical structures draws learners' attention to target items (Lee & Révész, 2018; Montero Perez et al., 2015). Multimodal input also offers multiple options to study learners' attention allocation to different types of materials such as reading-while-listening (Serrano & Pellicer-Sánchez, 2019) or captioned video (Montero Perez,

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

2019a) to investigate how learners divide their attention over different input modes (e.g., text and images) and how learners' attention allocation relates to linguistic outcomes (see Pellicer-Sánchez et al., 2020).

Finally, the third reason is a pedagogical one and concerns the fact that multimodal input may provide language teachers with effective options for classroom practice, which go beyond typical comprehension activities. In instructed SLA contexts (i.e., contexts where teaching time is typically limited to 2 or 3 hours of instruction per week), however, extensive exposure to multimodal input is not always possible. Yet, research has also highlighted the value of short clips for stimulating various aspects of L2 learning. Classroom-based use of multimodal input could, for instance, focus on the training of listening fluency. To become a fluent listener, researchers have argued that we need a successful interaction of learners' bottom-up (i.e., decoding) and top-down (i.e., learners' conceptual knowledge) processes (Rost, 2014). While top-down processes might be stimulated by the presence of, for instance, imagery and related storylines (Rodgers, 2013), multimodal input also presents numerous ways to train bottom-up listening skills. As shown by Wisnieska and Mora (2020) as well as by previous research (Charles & Trenkic, 2015), frequent exposure to captioned audiovisual input can be considered an effective technique to stimulate learners' speech segmentation and decoding skills in that it does not only aid their processing of the captioned input but also of newly encountered input (Charles & Trenkic, 2015). By providing a real-time written account of the auditory trace, we cannot only reduce learners' cognitive processing load but also give them more time to process and interpret the meaning. This may in turn also lead to better noticing of new words in the input (Winke et al., 2010) or stimulate other aspects of L2 learning such as grammar (Cintrón-Valentín et al., 2019). In addition, by processing multiple input modes (cf. multimedia learning principle), we may not only boost L2 learning but also learners' retention of the information.

CONTRIBUTION OF THIS SPECIAL ISSUE

This special issue illustrates that multimodal input can be operationalized in many different ways. Studies have investigated the value of audio-visual input (Feng & Webb, 2020; Puimège & Peters, 2020), audio-visual input with subtitles in the L1 (Pujadas & Muñoz, 2020), subtitles in the L2 (Lee & Révész, 2020; Pujadas & Muñoz, 2020), and L2 subtitles with enhanced (i.e., highlighted) target constructions (Lee & Révész, 2020). It also offers a study on audio-assisted reading activities, that is, reading while listening (Pellicer-Sánchez et al., 2020). In what follows, I discuss some of the main implications[1] of the aforementioned studies for theory, methodology, and pedagogy.

THEORETICAL CONTRIBUTION

A first theoretical contribution of this special issue is that it investigates aspects of L2 learning that have, so far, received little attention. More particularly, it expands upon existing studies that focused mainly on the learning of single words through multimodal input by looking into the potential of multimodal input for other aspects of L2 learning such as grammar (Lee & Révész, 2020), pronunciation (Wisnieska & Mora, 2020), and formulaic sequences (Puimège & Peters, 2020). In addition, this special issue provides an

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

answer to the long-standing call for more longitudinal research (Vanderplank, 2016) on the effects of multimodal input. In Pujadas and Muñoz (2020) the effects of watching a TV series on L2 learners' comprehension were investigated in an 8-month extensive viewing intervention program (or 515 minutes of total exposure). An 8-week intervention program (or 320 minutes of exposure to audiovisual input) was used in Wisnieska and Mora (2020) to investigate its effects on L2 speech segmentation, processing, and pronunciation.

Second, this special issue corroborates the beneficial impact of captioned video on L2 learning (Lee & Révész, 2020; Wisnieska & Mora, 2020) and provides additional evidence to reject the redundancy principle of the multimedia learning theory in the context of L2 learning. In addition, it sheds new light on the impact of L2 text on learners' auditory processing of the input. While Diao et al. (2007) reported negative effects of captioned audio on L2 listeners' auditory processing, Wisnieska and Mora (2020) provide further evidence that both uncaptioned and captioned video can facilitate L2 learners' speech segmentation and processing.[2]

A third theoretical merit of this special issue is that it contributes to our understanding of incidental learning processes that may occur while processing multimodal input. Feng and Webb (2020) do so by comparing the potential of written, auditory, and audiovisual input to determine the relative value of these three input modes for incidental vocabulary learning (of single words). Puimège and Peters (2020) investigated whether and to what extent incidental learning from audio-visual input is mediated by a combination of learner-related factors (i.e., learners' prior vocabulary knowledge) and factors related to the target formulaic sequences (i.e., collocate-node relationship, congruency, transparency, and statistical measures). They also included participants' prior knowledge of target items at the level of meaning recall, which was found to predict learners' uptake of word form. By including these factors, this study does not only provide information on the mediating role of learner- and word-related factors for incidental learning from multimodal input but they also provide evidence for the incremental nature of (incidental) vocabulary learning. In addition, these findings highlight the value of exposure to audiovisual input as a way to consolidate and refine learners' L2 vocabulary knowledge.

A fourth theoretical contribution concerns the fact that this special issue adds to the much discussed role of attention for L2 learning from input and the relationship between attention and aspects of L2 learning. The study by Pellicer-Sánchez et al. (2020) contributes to this discussion by investigating how learners process multimodal input. More particularly, they examine whether and to what extent the presence of audio influences young English as a second language (ESL) learners' reading (as measured in terms of eye movements) of an illustrated graded reader by comparing a reading-only and a reading-while-listening group. In addition, by combining data from eye-movement analyses and a comprehension test, they provide information on whether and how attention on text and images is related to learners' scores on the comprehension test. Lee and Révész (2020) expand on previous findings on the role of input enhancement (i.e., perceptually salient structures) by investigating its role in captioned audiovisual input. In their study, input enhancement was operationalized by highlighting target grammatical constructions (i.e., verbs in the present perfect and past simple) in the captions to direct learners' attention to these constructions. They found that textual enhancement does not only have the potential to draw L2 learners' attention, as measured

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

in terms of learners' eye fixations, but they also reported a positive relationship between learners' eye movements and their learning gains.

METHODOLOGICAL CONTRIBUTION

At the methodological level, I believe this special issue highlights, at least, two main features that future studies into the role of multimodal input for L2 learning might need to consider. First, it illustrates how different data-collection techniques ranging from paper-and-pencil tests (e.g., comprehension or vocabulary tests), interviews, or questionnaires (e.g., on language exposure) to more recent techniques such as eye-tracking can be used to investigate different components of L2 learning from multimodal input. More particularly, the special issue provides interesting examples of fine-grained (Pujadas & Muñoz, 2020) and multimodal comprehension tests (Pellicer-Sánchez et al., 2020). It also highlights the role of tests that target different aspects of L2 knowledge development (e.g., Wisnieska & Mora, 2020) and that reveal partial learning gains (e.g., Feng & Webb, 2020; Puimège & Peters, 2020). We also find two examples of how eye-tracking can be used to study multimodal input and which types of eye-tracking metrics we might need to consider (Lee & Révész, 2020; Pellicer-Sánchez et al., 2020). While eye-tracking is a particularly interesting technique to study multimodal input because it provides an online unobtrusive account of learners' attention allocation, the multimodal character of the input also has a number of implications for data collection procedures and analyses. For a detailed description on eye-tracking methodology in studies on multimodal input, we refer to Godfroid (2019) and Conklin et al. (2018).

Second, this special issue as well as other recently published studies demonstrate that L2 research into multimodal input should take into account a variety of factors in the study design. With regard to learner-related factors, it has been found that proficiency level (e.g., Suárez & Gesa, 2019), prior vocabulary knowledge (e.g., Feng & Webb, 2020; Peters & Webb, 2018; Puimège & Peters, 2020), age (Muñoz, 2017), or working memory (e.g., Gass et al., 2019; Montero Perez, 2020) may impact learning gains and the processing of multimodal input. Studies also indicate that including factors related to the target items (e.g., frequency of occurrence, part of speech, visual representation) or to the input (e.g., lexical coverage of TV episodes) is not only important to control for these variables but it also indicates whether these factors influence the learning process from multimodal input (Peters & Webb, 2018; Puimège & Peters, 2020). It is therefore not only crucial to include a detailed description of the multimodal input, individual differences, and the target items in the study methodology but also to use appropriate statistical techniques that allow for the inclusion of learner- and item-related factors such as logistic regression or mixed-effects models (e.g., Puimège & Peters, 2020; Pujadas & Muñoz, 2020).

PEDAGOGICAL CONTRIBUTION

This special issue presents research findings on the effectiveness of different types of multimodal input (e.g., audiovisual input, reading while listening) for L2 teaching and learning, how L2 learners process multimodal information, and whether this relates to

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

learning. It therefore offers a number of pedagogical recommendations that language learners and teachers might want to consider.

With regard to the use of audio-visual input (without on-screen text) for L2 vocabulary learning, this special issue shows that it does not only stimulate the learning of single words (Feng & Webb, 2020) but also of formulaic sequences (Puimège & Peters, 2020). In addition, it shows that the uptake of new vocabulary from audio-visual input may be comparable to other types of input such as written text and audio (Feng & Webb, 2020). These findings are encouraging: they show that more extensive exposure to video can yield L2 vocabulary development and confirm the role of audiovisual materials as a valuable means of increasing L2 learners' contact with the L2 in limited-input instructional settings. In addition, watching video can be considered a motivating alternative for L2 learners who are less inclined to use L2 written or listening input in out-of-class contexts (Peters, 2018).

Because captioned video does not only seem to aid speech decoding and pronunciation but also facilitates other aspects of L2 development such as vocabulary and grammar (Lee & Révész, 2020), language teachers should be encouraged to further explore the value of this type of multimodal input in their teaching and stimulate students to engage with those materials as out-of-class activity. In addition, they might also consider the use of other types of captions such as captions with enhanced target structures if the aim is to draw learners' attention to specific linguistic features (e.g., grammatical structures) and stimulate their learning of those structures (Lee & Révész, 2020).

Multimodal input is not only beneficial for intermediate to advanced students (e.g., Lee & Révész, 2020; Puimège & Peters, 2020; Wisnieska & Mora, 2020) but can also be used with young and adolescent beginning L2 learners (Pellicer-Sánchez et al., 2020; Pujadas & Muñoz, 2020). Both reading while listening (with supportive images) as well as videos with subtitles in the L1 have the potential to motivate beginning L2 learners to engage with this type of L2 input and facilitates their comprehension of those materials. Pedagogical approaches that stimulate the use of these materials early on in the learning process should therefore be promoted, both inside and outside the classroom (see Tragant et al., 2016, for an example).

Finally, it is worth noting that the study by Lee and Révész (2020) differs from many other studies in the field by implementing videos in an actual pedagogical approach, namely task-based language teaching. One of the merits of this study is that it clearly illustrates how short videos can be used in meaningful input-based tasks that are effective in stimulating L2 grammar learning.

Drawing on the previously mentioned implications and contributions of this special issue, I hope that researchers are inspired to conduct further research on the role of multimodal input for L2 learning. This will undoubtedly continue to refine our understanding of the numerous interesting findings and research questions put forward by the studies in this special issue.

NOTES

[1]This overview is not exhaustive.

[2]Positive effects of captioned video on speech decoding were previously reported by Mitterer and McQueen (2009) and Charles and Trenkic (2015).

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

REFERENCES

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, *30*, 709–725. https://doi.org/10.1080/09588221.2017.1356335

Bollansée, L., Puimège, E., & Peters, E. (in press). Watch out! Behind you is the enemy! An exploratory study into the relationship between extramural English and productive vocabulary knowledge. In V. Werner & F. Tegge (Eds.), *Learning languages through pop culture*. Routledge.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*, 136–163.

Chang, A. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, *37*, 652–663. http://doi.org/10.1016/j.system.2009.09.009

Chapelle, C. A. (2003). *English language learning and technology*. John Benjamins Publishing Company.

Charles, T. J., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bi-modal input. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning: Principles, strategies, and practical experiences* (pp. 173–198). Peter Lang.

Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, *47*, 439–459. http://doi.org/10.1080/09571736.2019.1615978

Conklin, K., Pellicer-Sanchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

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

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*, 19–47. https://doi.org/10.1111/modl.12301

Feng, Y., & Webb, S. (2020). *Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective*? *Studies in Second Language Acquisition*, *42*, 499–523. https://doi.org/10.1017/S0272263119000494

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning and Technology*, *23*, 84–104.

Godfroid, A. (2019). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Kuppens, A. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*, 65–85. https://doi.org/10.1080/17439880903561876

Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal*, *102*, 557–577. https://doi.org/10.1111/modl.12503

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42*, 625–651. https://doi.org/10.1017/S0272263120000108

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129. http://doi.org/10.1080/14790718.2012.679275

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One*, *4*, e7785. https://doi.org/10.1371/journal.pone.0007785

Montero Perez, M. (2019a). Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *The Language Learning Journal*, *47*, 460–478. https://doi.org/10.1080/09571736.2019.1638623

Montero Perez, M. (2019b). Technology-enhanced listening: How does it look and what can we expect? In N. Arnold & L. Ducate (Eds.), *Engaging language learners through CALL: From theory and research to informed practice* (pp. 141–178). Equinox Publishing.

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*. Advance online publication. https://doi.org/10.1017/S0272263119000706

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

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

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

Montero Perez, M., & Rodgers, M. P. H. (Eds.) (2019). Video and language learning [special issue]. *The Language Learning Journal*, *47*, 403–518.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720–739.

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, *67*, 77–86.

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

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*, 577–598. https://doi.org/10.1017/S0272263120000091

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

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*, 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., Heynen, E., & Puimège, E. (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

Peters, E., & Muñoz, C. (2020). Introduction to the special issue: Language learning from multimodal input. *Studies in Second Language Acquisition*, *42*, 489–497. https://doi.org/10.1017/S0272263120000212

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/S0272263117000407

Plass, J. L., & Jones, L. C. (2005). Second language acquisition with multimedia. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 476–488). Cambridge University Press.

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, *69*, 943–977. https://doi.org/10.1111/lang.12364

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*, 525–549. https://doi.org/10.1017/S027226311900055X

Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, *42*, 551–575. https://doi.org/10.1017/S0272263120000042

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes and captions* (Unpublished doctoral dissertation). Victoria University, Wellington, New Zealand.

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL—International Journal of Applied Linguistics*, *169*, 191–211.

Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-Language television programs. *CALICO Journal*, *34*, 20–38. https://doi.org/10.1558/cj.29522

Rost, M. (2014). Developing listening fluency in Asian EFL settings. In T. Muller, J. Adamson, P. Shigeo Brown, & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 281–296). Palgrave Macmillan.

Serrano, R., & Pellicer-Sánchez, A. (2019). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*. Advance online publication. https://doi.org/10.1515/applirev-2018-0102

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press

Suárez, M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *The Language Learning Journal*, *47*, 497–517, https://doi.org/10.1080/09571736.2019.1617768

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*, 50–73.

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing young learners' input: An intervention program. *Canadian Modern Language Review*, *72*, 234–257. https://doi.org/10.3138/cmlr.2942

Tragant, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye-movements. *System*, *80*, 212–223. https://doi.org/10.1016/j.system.2018.12.002

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*, 1–37. https://doi.org/10.1017/S0261444809990267

Vanderplank, R. (2016). "Effects of" and "effects with" captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*, 235–250. https://doi.org/10.1017/S0261444813000207

Warren, P., Boers, F., Grimshaw, G., & Siyanova-Chanturia, A. (2018). The effect of gloss type on learners' intake of new words during reading. *Studies in Second Language Acquisition*, *40*, 883–906. https://doi.org/10.1017/S0272263118000177

Webb, S. (2010a). Using glossaries to increase the lexical coverage of television programs. *Reading in a Foreign Language*, *22*, 201–221.

Webb, S. (2010b). Pre-learning low-frequency vocabulary in second language television programmes. *Language Teaching Research*, *14*, 501–515. http://doi.org/10.1177/1362168810375371

Webb, S. (2015). Extensive viewing. Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.

Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*, 651–675.

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, *59*, 335–366.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*, 65–86.

Wisniewska, N., & Mora, J. C. (2020). *Can captioned video benefit second language pronunciation?* Studies in Second Language Acquisition, *42*, 599–624. https://doi.org/10.1017/S0272263120000029

https://doi.org/10.1017/S0272263120000145 Published online by Cambridge University Press