

REPLICATION RESEARCH

# Vocabulary learning at first exposure: Replication of Gullberg et al. (2012) and Shoemaker and Rast (2013)

Imma Miralpeix

Universitat de Barcelona, Barcelona, Spain  
Email: [imiralpeix@ub.edu](mailto:imiralpeix@ub.edu)

(Received 30 July 2022; accepted 31 July 2022)

## Abstract

This article puts forward several proposals for replicating two well-known FIRST EXPOSURE studies dealing with the earliest stages of adult second language acquisition. Both of them enquire into the word-level knowledge that complete beginners are able to extract from minimal input when exposed to a new language for the first time. They also focus on several input variables that may enhance learning from minimal input. However, the first, by Gullberg et al. (2012), uses audiovisual input in Dutch learners of Chinese to assess word recognition and word meaning after watching a short video; while the second, by Shoemaker and Rast (2013), uses oral input with French learners of Polish to measure word recognition before and after 6.5 hours of intensive classroom exposure. Close and approximate replications of these studies can help to re-evaluate and generalise the findings, as well as contributing additional relevant data to the field.

## 1. Introduction

Studies on how adults break into a foreign language system at first contact do not yet abound. So far, very little research has concentrated on the earliest stages of adult second language acquisition (SLA), even if it has been claimed that this work should be more central to second language (L2) research (e.g., Han & Rast, 2014; VanPatten, 2014): any SLA theory should account for all phases in the process so as not to be incomplete, including very initial stages in adult learners. At a practical level, research findings in this area can also inform communication and learning practices (e.g., to find the most adequate ways to boost the learning of new languages from the very beginning and help adult learners make the most of the input when initially exposed to a new language).

I refer to the studies in this line of research as FIRST EXPOSURE (FE) studies, where ‘data are collected from the very first moment of contact with the target language (TL) and within the first seconds, minutes and hours of subsequent exposure, and all TL input is controlled’ (Rast, 2008, p. 29). FE studies usually deal with adult L2 acquisition of *ab initio* learners (i.e., without any previous knowledge of the TL), and this is different from *en route* or *al fine* learners (i.e., learners who are already familiar with the TL). FE studies also often focus on incidental learning: ‘picking up’ different features from the input, such as unknown words, without deliberate attention (Hulstijn, 2013).

Although adult L2 acquisition is thought to be slow and laborious compared with first language (L1) acquisition, it has also been shown that in this population minimal L2 instruction can produce rapid change (e.g., McLaughlin et al., 2004). Discussions about age effects often focus on ultimate attainment, end states and nativelikeness instead of on the development process or rate (as mentioned in Ristin-Kaufmann & Gullberg, 2014). However, it has also been seen that ‘the adult brain is clearly capable of plasticity and of rapidly adjusting to new learning experiences’ (Gullberg et al., 2012, p. 258) and acquiring new languages may be one example. In vocabulary studies, research on implicit and

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

incidental learning has often been conducted with *en route* learners, who already have an interlanguage in place, instead of *ab initio* learners who are trying to break into a language when exposed to it for the very first time.

In the last 15 years, though, especially since the publication of *Foreign Language Input: Initial Processing* by Rast and the special issue of *Language Learning* on the topic edited by Gullberg and Indefrey (2010), very insightful FE studies have been conducted, and we propose several replications for two of them. The first study is entitled ‘What word-level knowledge can adult learners acquire after minimal exposure to a new language?’ by Gullberg, Roberts and Dimroth, which was published in *International Review of Applied Linguistics in Language Teaching (IRAL)* (2012). The second is ‘Extracting words from the speech stream at first exposure’ by Shoemaker and Rast, published in *Second Language Research (SLR)* the following year.

Both of them are significant contributions to the field and share several common traits: they are FE studies dealing with vocabulary learning from minimal input in a novel language. They both use either naturalistic or classroom input, although it is tightly controlled. Additionally, they focus on different factors that may help novice learners to process complex natural language input at the outset of learning (e.g., word frequency in the input or lexical transparency). However, the former examines vocabulary learning from audiovisual input and the second from minimal classroom exposure. The findings they present are promising, revealing significant learning from the very first moments of contact with completely unknown languages. Replications of these studies considering different input types, language combinations, intake levels or learner variables would be needed to improve the validity and reliability of the findings, as well as to help to generalise from them (e.g., see Porte & McManus, 2019, p. 13, on reasons to carry out replications). The authors themselves acknowledge some limitations and propose further research in their papers, suggesting different points to investigate in the future. Several of them can effectively be explored in approximate or close replication studies: these would not only assess the solidness of the results but also complement them with additional variables that can help us better to interpret the findings of the original studies.

## 2. Suggested study for replication: Gullberg et al. (2012)

### 2.1 Background to the study

Gullberg et al. (2012) has been widely cited and is one of the very few FE studies using continuous natural audiovisual speech as input. The study examines what word-related information learners can extract from very brief exposure to audiovisual input in a completely unknown language and what helps them to do so. More specifically, the paper aims to answer three research questions: (1) Can L2 learners extract information about word forms and word meanings after minimal exposure to input in the form of sustained speech in an unknown language? (2) What are the effects of item frequency and visual highlighting in the input? (3) How little exposure is enough to learn?

In order to answer these questions, Dutch L1 university students (mean age: 22) with no knowledge of Mandarin Chinese – or any typologically-related language – were asked to watch a short weather forecast in Mandarin, either once (‘one exposure’ condition) or twice (‘double exposure’ condition). Participants were tested individually by the researchers, who played the video to each of them on a TV screen after asking them to ‘watch the video’ without any further instructions. Therefore, participants were not aware they were going to be tested afterwards. Two experiments were conducted in the study: in Experiment One, participants took a Word Recognition Test (WRT) after watching the video, while in Experiment Two, a different cohort with the same characteristics took a Meaning Recognition Test (MRT). Participants also filled in a questionnaire with bio data after being tested.

The seven-minute video in Mandarin Chinese that participants were exposed to was in a language typologically and genetically unrelated to the L1 of the participants. Input was tightly controlled: 24 target words (TWs) were selected to construct the weather report text, which consisted in 292 word types organised into 120 clauses with one TW per clause. TWs were also distributed so that they appeared towards the beginning, middle and end of clauses to control for potentially irrelevant effects

of sentence position. TWs occurred in one of the four experimental conditions: frequently (eight times) or infrequently (two times), and were either accompanied by a gesture (highlighted) or not (not highlighted). Therefore, TWs could be frequent and highlighted (+F + H), frequent and not highlighted (+F – H), infrequent and highlighted (–F + H) or infrequent and not highlighted (–F – H). All the other words in the report ('padding words') were also controlled for frequency. The complete text is provided in the Appendix of the paper with detailed information. A total of six weather charts appeared in the video as visual support, showing the weather conditions reported in the text in a fictitious country. The report was presented by a female speaker of Mandarin Chinese, who read the text in Chinese characters off a prompter and had previously been trained to highlight the required TWs with a gesture.

In the first experiment, the WRT was used to examine whether participants were able to segment the Mandarin sound string after minimal exposure and the cues they used. Forty-one participants (21 having watched the video once and 20 twice) were tested on the recognition of the 24 TWs. In a sound-proofed experimental booth, the audio file with the test items was played to each participant. After each item was played, they answered whether they had heard the item in the weather report or not by pushing a button on the computer. The WRT took about 10 minutes to complete and is also provided in the Appendix. It has two versions in random order, each containing the TWs and 72 fillers. A score of 1 was given to correct responses and 0 to incorrect responses. Authors then ran a series of mixed-effect logistic regression models on the response data, with subjects and items as random effect factors and Accuracy of Response (correct/incorrect) as the outcome variable. As number of syllables and languages known by participants could influence perceptions, they were entered as control variables in the analysis. Results showed that L1 Dutch speakers were able to segment the Mandarin sound stream to correctly identify words from the video when presented in isolation. An effect was found for frequency and number of syllables, but not for gestural highlighting or amount of exposure: disyllabic items occurring eight times were more accurately recognised than infrequent monosyllabic ones.

In the second experiment, two new groups of Dutch native speakers (20 watching the video once and 20 twice) were assessed with the MRT, which took approximately 10 minutes to complete. It was a sound-to-picture matching task created by pairing each of the weather report icons for the eight TW nouns, once with its correct word form and once with an incorrect word form. Some filler items were also introduced, and the experimental list containing 41 experimental trials in total is provided in the Appendix of the paper. After watching the video, participants were presented with each icon and an auditory stimulus was played via headphones. They had to indicate with a computer button whether this stimulus was the correct word for the picture they had seen on the screen or not. Following the same procedure for data analysis as in the first experiment, it was observed that participants could also map meaning to the new words in Mandarin Chinese. An interaction between frequency and gestural highlighting was found and, again, number of syllables had an effect on the results: disyllabic items that were frequent (occurring eight times) and accompanied by gesture were matched with the correct meaning significantly above chance.

Altogether, results from the two experiments suggest that adults can extract considerable information from very brief exposure to audiovisual input in a completely unknown language. Amount of exposure (one vs. two viewings) did not affect the results in any of the experiments. However, it was revealed that frequency of occurrence and number of syllables (word length) had a robust effect, and that eight instances of a disyllabic word were enough to recognise it in speech and to match it with its meaning. The mapping of meaning to a word form seemed to require accumulative cues: in this case, frequency and word length needed to be combined with gestural highlights in order to see significant effects on learning.

## 2.2 Approach to replication

This study offers very interesting possibilities for close replications. Several ideas are already suggested by the authors in the 'further research' section of the paper: some of them can be explored by replicating the study in a number of ways. Replications would also provide useful data on how learning can be maximised

in minimal input conditions. Nowadays, a wide range of audiovisual materials (e.g., short videos, episodes of TV series, etc.) are freely available in multiple languages. Research so far has concentrated on the effects of audiovisual input in *en route* learners (i.e., learners who already have an interlanguage in place), but little is known about *ab initio* or novice learners, who have not been learning the TL before.

Most FE studies focus on incidental learning, as in the case in Gullberg et al. (2012). However, we do not know how much lexical information adults would intentionally be able to extract from audiovisual input. A good way to start analysing this issue would be by changing one variable in the study: the instruction that is given to participants before watching the video. Instead of simply 'watch the video', they could be told that they would watch a video in a language they do not know and that they should try to learn as much as possible from this new language. Another possibility would be telling them in advance that they would be tested on the new language after having watched the video (Hulstijn, 2013). Changing task orientations would surely change the participants' focus (VanPatten, 2014), although we do not know up to which extent it may change the results.

We may anticipate that participants would actually learn more if they were explicitly instructed to concentrate on learning as much as possible while watching the weather report. However, we may also find that they learned the same amount, as the input would be equally challenging and they would not be able to fall back on pre-existing knowledge, given the linguistic distance between the L1 and the TL. Furthermore, the limited time would prevent participants from developing and consistently applying effective learning strategies. It may also be the case that they learn less, as they might focus more on remembering a very limited number of items instead of on understanding the message. Moreover, maybe anxiety or fear of failure in the test could affect quantity and quality of intake while performing the task. A replication study would empirically show which of the three hypotheses is confirmed.

Predictions on the most plausible hypothesis are hard to make if we consider the very few studies available on FE and multimodal input. On the one hand, research by Bisson et al. (2014), in which L1 English speakers watched a 25-minute subtitled episode of *Sponge Bob* in Dutch, did not show significant vocabulary learning in an unannounced test after watching the episode, indicating that no incidental learning was taking place. On the other, a study by Miralpeix et al. (in press), where L1 Catalan/Spanish learners were asked to watch a short advert subtitled in Polish and explicitly asked to learn as much as possible from the new language, revealed significant vocabulary gains. However, in that study, the audio of the advert was in English, a language participants knew, and this facilitated intentional learning and making form-meaning connections. Therefore, a close replication of Gullberg et al. (2012) with intentional learning would be very valuable: it would help us to reassess the actual gains in the original study while also giving us more details on the advantages or disadvantages of incidental and intentional learning at FE.

In order to know more about optimal learning conditions, a second possibility for close replication would be adding subtitles to the weather report. In this case, as the TL is Mandarin Chinese, subtitles could be added in the Latin alphabet (transcription is found in the Appendices). Different theories such as the DUAL CODING THEORY (Paivio, 1986, 2007) and the COGNITIVE THEORY OF MULTIMEDIA LEARNING (Mayer, 2009, 2014) have pointed out that receiving information through different channels (verbal and non-verbal, as in audiovisual materials) facilitates learning and information recall because there is greater processing depth. However, the COGNITIVE LOAD THEORY (Chandler & Sweller, 1991; Sweller, 1994) has also suggested that multimodality can increase cognitive load. Subtitles, for example, can then be a 'double-edged sword': on the one hand, they can reduce cognitive load in language acquisition settings and may be a good aid to make sense of the input in a new language. On the other, they may make learning more difficult, as the cognitive load may then be too high. A replication using a subtitled video would prove whether subtitles enhance or hinder learning at these very first stages. Furthermore, it would be interesting to add to this replication a written word recognition test in Experiment 1 and a MRT with the written forms in Experiment 2.

It should also be noted that no FE studies have been conducted so far comparing learning from subtitled and unsubtitled videos. Therefore, to make any prediction on the results, we have to fall back on the few available studies for low-level adult learners, which are inconclusive: in d'Ydewalle and Pavakanun (1995), the inclusion of textual support was beneficial for meaning recognition in adult L1 Dutch learners

of English. It was also positive for learning word meanings in beginner university learners of Russian (Sydorenko, 2010). Nevertheless, adding subtitles did not have a significant effect in Raine (2012) with beginner Japanese university learners. Given these results with low-level learners, it is not yet clear what the outcome for *ab initio* learners might be: replications in this line of enquiry would be very welcome. Not only would we be able to compare the amount of uptake in the original study and the replication to re-evaluate learning, we would also have new experimental data on subtitled video watching complementing what is already available for *en route* learners.

Another possibility for close replication entails changing one of the languages involved: Gullberg et al. (2012, p. 259) conclude from their research that ‘the adult learning mechanism appears to be a great deal more powerful than typically assumed in the L2 acquisition literature’, but more consistent findings involving different languages would be very appreciated in the field. A straightforward study could be conducted with participants with a different L1. One example of why this kind of replication is relevant is speech segmentation (i.e., the process by which the brain determines where one meaningful unit ends and the next begins in continuous speech): we use different cues – prosodic, distributional, and so forth – to segment language depending on our L1. Quite often, our abilities to segment streams using L1 cues are extended to other similar languages we learn. Nevertheless, this is more difficult to transfer when learning a typologically unrelated language (e.g., Gómez et al., 2018). For instance, because Dutch lacks contrastive lexical tones, participants in the original study could not transfer tonal cues to recognise words in Mandarin, which is a tonal language. However, it would be necessary to confirm that participants with other L1s that also lack contrastive lexical tones (such as Dutch) obtain similar results when exposed to Mandarin, and that results are not due to language-specific syllabification preferences in Dutch or in other languages.

It would be equally interesting to change the TL of the study instead of participants’ L1. This would require more effort, as new materials would need to be produced following the guidelines provided. For example, the same study with the video in Vietnamese, also a tonal language, would show whether Dutch speakers without pre-existing knowledge of the TL behave in a consistent manner with tonal languages or whether other particular intrinsic characteristics of each TL language have an effect on word learning from minimal exposure. Obviously, the next step would be conducting approximate replications where both languages (L1 and TL) are different from the original study. For instance: would Mandarin native speakers be able to extract significantly more lexical information from the Vietnamese video given the fact that the two languages are typologically closer than those in the original study? How different would results be at FE when participants can bootstrap from the L1? We do not know yet up to what point the use of audiovisual materials can increase the learning rate of languages that are typologically similar to our mother tongue/s. Therefore, this kind of cross-linguistic replication would help us to assess the generalisability of the findings in the original study (i.e., to what extent they are the same or not when the languages involved are different). Furthermore, they would help us to understand both the potential and the limitations of multimodal input for novice learners depending on the typology of the languages they know and the language they are exposed to for the first time.

Finally, it is worth noting that all these replications would benefit from the addition of one variable, which is delayed testing. Administering the same post-test to the same participants after some time would reveal if (and how much) knowledge was retained after minimal exposure. As Gullberg et al. (2012) mention, the question remains whether the initial capacity adults seem to show for making sense of minimal input can help them to become better learners in the long run. One first step we can take in this direction is assessing for how long this knowledge is actually retained.

### 3. Suggested study for replication: Shoemaker and Rast (2013)

#### 3.1 Background to the study

Shoemaker and Rast (2013) analyse how novice learners break into the sound stream of a novel language after minimal exposure to classroom instruction. It is often cited in research papers on initial



processing, implicit learning, crosslinguistic influence and word segmentation. It explores the role of three variables in the development of word recognition strategies of L1 French participants after very few hours of classroom exposure to Polish. More specifically, the variables under study are word transparency (i.e., similarity between the words in the L1 and the TL), word position (i.e., beginning, middle or end of the utterance) and word frequency of occurrence in the input. These factors were not chosen at random: the selection was motivated by results in a previous study by Rast and Dommergues (2003), which assessed production rather than reception, and it found these variables to be worthy of further exploration at the receptive level. In these studies, French and Polish were chosen because there are many differences between these two languages at the segmental, suprasegmental and rhythmic levels. Regarding segmental inventories, their vocalic and consonantal systems present considerable differences. At the suprasegmental level, even if they share the same prosodic characteristic of fixed stress, they differ in where stress falls (in French usually in the last syllable, while in Polish usually in the penultimate syllable). Regarding rhythm, French is a syllable-timed language, whereas there is not an agreement on whether Polish is a stress-timed or a syllable-timed language. It should be acknowledged that the study is part of the larger VILLA Project (Varieties of Initial Learners in Language Acquisition) (Dimroth et al., 2013).

In the study, a group of 18 L1 French adults in their twenties was taught Polish for a total of 6.5 hours in an intensive course that lasted 5 days. All the participants had learned English and another Romance language as the L2, but none of them had any knowledge of any Slavic language. The native Polish teacher used a 'communicative-based method that excluded all use of metalanguage, as well as explicit explanations of grammar and pronunciation' (Shoemaker & Rast, 2013, p. 171). Therefore, the environment represented an authentic instructed language-learning situation, and learners were asked not to check dictionaries, grammar books or any other materials in Polish during the instruction and data-collection period. The course was fully recorded and input was transcribed in CHAT format (MacWhinney, 2000), so that the TWs could be carefully selected. Participants were tested twice: before starting the course (T1) and after 6.5 hours of exposure (T2).

For the perceptual word recognition task, a list of 16 TWs provided in the paper was compiled, taking into account their transparency and input frequency in the classroom. In order to rate transparency, 13 independent judges who were native speakers of French (and had no knowledge of Slavic languages) heard a list of 71 words and were asked to provide a French translation: words without any correct translation were classified as Low Transparency (LT) words and those with more than 50% correct translations were considered High Transparency (HT) words. Regarding frequency, words that were absent from classroom input were categorised as Low Frequency words (LF), while those appearing more than 20 times were considered High Frequency (HF) words. All test words were counterbalanced for the transparency and frequency categories, with four words appearing in each combination of categories (all of them were two or three syllables long and carried stress on the penultimate syllable). Finally, in order to assess word position in an utterance, 48 test sentences were created, including the 16 TWs in either initial, medial or final position. The test also included 33 distracter sentences, making a total of 81, and was recorded by a female native speaker of Polish and administered with E-prime (Schneider et al., 2002). After training with ten stimuli, participants were presented with the experimental sentences: they could hear each sentence in Polish, immediately followed by the word 'OK'; then they heard the word in isolation and they had to answer whether the word was present in the sentence or not by pressing a key on the computer keyboard. The test was not timed, it took participants about 15 minutes to complete and stimuli were presented in random orders.

Accuracy scores at each testing time were examined with ANOVAs according to Transparency and utterance Position. A significant effect for transparency was found at both testing times, and it was stronger in initial and medial positions (there was a probable ceiling effect in final positions, where accuracy rate was very high). Then, a repeated measures ANOVA was also conducted with Transparency, Frequency, Position and Testing Time as within-subjects factors. It was shown that there was a significant improvement between T1 and T2 and that sensitivity to low transparency words increased from T1 to T2 (the same was not found for high transparency words, possibly because

of the aforementioned ceiling effect). Post-hoc analyses also revealed that participants recognised words in initial and medial positions significantly better at T2 (the same was not found for final positions, probably because scores related to final position TWs were high at both testing times). Frequency did not prove to be a significant factor in the analyses.

Results suggest that novice learners rely on the edges of prosodic boundaries because those words in initial and final positions were more successfully recognised than others, and those in final position obtained the highest recognition scores. Furthermore, it was shown that the phonetic forms of transparent words in Polish were sufficient to activate the L1 forms in the mental lexicon and that transparent words were easier to extract from the very beginning, before any exposure to Polish at all. Learners most likely acquired progressive sensitivity to the general phonological forms and prosodic patterns of Polish (as less transparent items were often recognised at T2 but not at T1). However, frequent exposure did not play a role in learners' ability to recognise words, possibly indicating that it did not yet suffice at this point to help them better recognise individual words in continuous speech.

### 3.2 Approach to replication

Several replications of this study can be conducted to strengthen the validity of its results and better understand the language processing abilities of late *ab initio* learners. To start, we could perform close replications with adults of different ages, as we know from the cognitive literature that associative memory and processing speed are compromised by age, with a linear decline starting in early adulthood (see Birdsong, 2006 for a review). In the present research, the participants were on average 21.2 years old (range: 19–27); however, we do not know what the results might be with older adults in their thirties, forties or sixties: would their accuracy response rates be comparable to those of younger adults? Would transparency still be more influential than frequency for spoken word recognition in more mature learners? Would sensitivity to low-transparency words improve at the same rate between T1 and T2? These are empirical questions that replications could answer. Therefore, we would be examining the generalisability of the results in the original study while contributing new data to the field, which would allow us to know whether 'adult language acquisition' follows similar patterns or not according to the age at which adults are first exposed to a completely novel language. Studies on L2 learning and age have shown maturational constraints and a cognitive decline for certain aspects of acquisition, although once again this research has mostly been conducted with *en route* learners, and very little is known about what happens in FE situations.

A different possibility for a close replication would be conducting the study with French monolingual adults instead of multilingual participants. In the present study, young adults had learned English (as the L2) and knew another Romance language, so Polish was the fourth language they were exposed to. As has often been shown in the literature (e.g., Cenoz, 2013; Herdina & Jessner, 2002), bilinguals and multilinguals tend to have an advantage when learning other languages: counter to what happens with monolingual learners, there are several factors developed just by learners knowing more than one language (e.g., metalinguistic awareness, metacognitive strategies, etc.). These may give them an advantage in the first stages of acquiring an additional language, but this cannot be concluded from the present study because all participants already knew three languages. We might also find that linguistic processing could be more effortful for bilinguals/multilinguals, as they may need to resolve a competition between previously acquired languages that would slow down performance (bilingual advantages can depend on characteristics of the participants and task features, as shown by Bialystok et al., 2014). Therefore, by conducting the replication with monolingual speakers, we would know whether the 'number of languages known' variable affected the results in the original study and whether multilingualism is what facilitates word recognition at FE. These results could also inform language teaching practices in beginner language classes, where monolingual and multilingual learners are mixed in the same groups.

When carrying out these replications, it would be extremely useful to add data on individual variables. For example, in the present study dealing with learners' abilities to break into a novel acoustic

signal, data on learners' linguistic aptitude and working memory (WM) would be necessary to better account for the findings. Linguistic aptitude appears to be predictive of rate of progress at early language learning stages (Doughty, 2019), even if its influence tends to decrease as L2 proficiency and other cognitive skills and strategies related to language learning improve (Serafini & Sanz, 2016; Winke, 2013). Likewise, measuring WM may be relevant: in this experiment, participants need to retain a sentence they have never heard before so that some seconds later they tell whether a word they heard was present or not in the sentence they heard. WM is precisely what allows us to hold onto information for a brief period of time while doing something else, and it has also shown to be a good predictor in tasks where bilingualism was not (Ratiu & Azuma, 2015). Therefore, by administering an aptitude test assessing phonological memory and a WM span test, we could then relate participants' abilities to recognise oral stimuli in new strings of words to individual learner factors, giving us more insight into the factors intervening in the process of learning novel languages. It should also be acknowledged that Dimroth et al. (2013, p. 125) used a range of tests measuring individual differences in the VILLA Project.

Another suggestion that Shoemaker and Rast make at the end of their paper is including introspective data from participants on the word-recognition strategies they may have used. This would be an interesting variable to add to any of the replication studies proposed, as it would help to validate the research findings. Triangulation using different methods of data collection would be an excellent way of re-evaluating not only the results of the original study, but also those in the replication itself. Post-experimental verbal protocols could be conducted immediately after the word-recognition task has been completed, so that participants still remembered as much as possible about their thoughts during task performance and how they tried to process and segment the sentences in the new language. Open-ended interviews may be preferable to questionnaires with closed questions, as participants would not be constrained by the options given: for example, apart from transparency, frequency or word position in the utterance, they might have used other factors that helped them to decide. However, it would also be recommendable to ask explicitly about these three factors, as their answers could then be related to the accuracy scores in the 48 test sentences and the degree of significance of each factor in the ANOVAs. Because recalling strategies can sometimes be difficult (it requires awareness and good memory), another possibility instead of asking them 'in the abstract' would be choosing representative examples in the test and re-playing them, one at a time, to help them reflect on the strategies they employed to decipher continuous speech. These reports would be very valuable for the field because FE studies seldom make use of qualitative data.

Finally, two approximate replications would be needed with (1) speakers of only syllable-timed languages with a stress-timed language as a TL and with (2) speakers of only stress-timed languages with a stress-timed language as a TL. In syllable-timed languages, the syllables occur at regular intervals (as in French, Spanish, Mandarin or Turkish), whereas in stress-timed languages (like English, Norwegian, Russian or Arabic), the stresses are equal distances apart even though the number of syllables between each stress is not the same: therefore, some syllables have to be said very quickly if there are several between two stresses, while others are said slowly if there is one or none between two stresses. It is often the case that participants with L1 syllable-timed languages are 'deaf' to stress when they start learning stress-timed languages, due to the lack of lexical stress in their own L1.

In the study, L1 speakers of French (a syllable-timed language) learn Polish, and, as discussed in the paper, it is not clear whether Polish is a stress-timed or a syllable-timed language because there are arguments in favour and against the classification of Polish in either of these two categories. Therefore, even if authors hypothesise that participants have gained sensitivity to the overall rhythm of Polish, 'the mixed nature of Polish rhythmic structure means that any supposition should be approached with caution' (p. 177). Due to the nature of Polish, it is impossible to actually conclude whether results can be attributed to phonological knowledge acquired at the segmental level (Polish phonemic inventory), suprasegmental level (stress distribution) or both. This is why the study also mentions that the language pairing in the original study 'is problematic for a theory in which FE participants are using stress placement in the segmentation of running speech' (p. 178). Furthermore, we



should remember that all participants in the study knew English, a stress-timed language, as the L2, and we do not know if this may influence their segmental abilities in Polish. In order to disentangle the possible transfer of rhythmic patterns from the L1 to the new language, replications with a group of speakers of just stress-timed languages (as well as with a group of speakers of only syllable-timed languages) when learning a stress-timed language would provide evidence of facilitative or hindering L1 effects in recognising stress patterns in the TL at FE. This would also provide more accurate information regarding the influence of rhythm in the first stages of learning new syllable-timed or stress-timed languages. Even if the VILLA project tested participants from different source languages (namely Dutch, English, French, German and Italian), replications with other TLs would be worth performing.

#### 4. Conclusion

In sum, research on FE would benefit from replicating some of its most outstanding studies, such as the two we have proposed in this paper. Having precise, reliable information on the processes and mechanisms at work at the outset of learning an unknown language is of the utmost importance at not just a theoretical but also a practical level. Their original research questions are very relevant to the concerns and issues in the field and putting the original studies' conclusions to the test will always enrich and strengthen our knowledge about learning from minimal input.

Replications proposed are related to the areas that Rast (2008) identified as key in FE investigations: pre-existing linguistic knowledge brought by the individuals to the acquisition task (e.g., depending on them being monolinguals or multilinguals, on the typologies of the languages novice learners already know, etc.); learners' individual differences, especially those that have been shown to be very influential in SLA at the beginning of the learning process (e.g., age, aptitude, working memory, etc.); as well as TL input (e.g., unimodal, bimodal or multimodal). Replication studies along these lines would help to (dis)confirm previous findings and provide further data on the generalisability of the studies. Furthermore, changing variables such as intentional learning (as opposed to incidental), or adding new ones, such as delayed (vs. immediate) testing, would undoubtedly provide insights into the most effective ways of approaching the challenging task of breaking into a completely new language.

#### References

- Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychol Aging*, 29(3), 696–705. doi:10.1037/a0037254
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56(S1), 9–49. doi:10.1111/j.1467-9922.2006.00353.x
- Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(2), 399–418. doi:10.1017/S0142716412000434
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, 46(1), 71–86. doi:10.1017/S0261444811000218
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. doi:10.1207/s1532690xci0804\_2
- Dimroth, C., Rast, R., Starren, M., & Watorek, M. (2013). Methods for studying the acquisition of a new language under controlled input conditions. *Eurosla Yearbook*, 13, 109–138. doi:10.1075/eurosla.13.07.dim
- Doughty, C. (2019). Cognitive language aptitude. *Language Learning*, 69(1), 101–126. doi:10.1111/lang.12322
- d'Ydewalle, G., & Pavakanum, U. (1995). Acquisition of a second/foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in Europe: The state of the art – perspectives for the future* (pp. 51–64). Westdeutscher Verlag.
- Gómez, D. M., Mok, P., Ordín, M., Mehler, J., & Nespó, M. (2018). Statistical speech segmentation in tone languages: The role of lexical tones. *Language and Speech*, 61(1), 84–96. doi:10.1177/0023830917706529
- Gullberg, M., & Indefrey, P. (Eds.) (2010). The earliest stages of language learning. *Language Learning*, 60(S2), 1–283. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/lang.2010.60.issue-s2/issuetoc>

- Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *IRAL – International Review of Applied Linguistics in Language Teaching*, 50(4), 239–276. doi:10.1515/iral-2012-0010
- Han, Z. H., & Rast, R. (Eds.) (2014). *First exposure to a second language*. Cambridge University Press.
- Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism*. Multilingual Matters.
- Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 5, pp. 2632–2640). Wiley-Blackwell. doi:10.1002/9781405198431.wbeal0530
- MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk*. Lawrence Erlbaum Associates.
- Mayer, R. E. (2009). *Multimedia learning (second edition)*. Cambridge University Press.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, 7(7), 702–704. doi:10.1038/nn1264
- Miralpeix, I., Gesa, F., & Suárez, M. (in press). Vocabulary learning from subtitled input after minimal exposure. In B. L. Reynolds (Ed.), *Vocabulary learning in the wild*. Springer.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Lawrence Erlbaum Associates.
- Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.
- Raine, P. (2012). *Incidental learning of vocabulary through subtitled authentic videos* [Master's thesis, University of Birmingham]. <https://www.birmingham.ac.uk/index.aspx>
- Rast, R. (2008). *Foreign language input: Initial processing*. Multilingual Matters.
- Rast, R., & Dommergues, J. Y. (2003). Towards a characterisation of saliency on first exposure to a second language. *EUROSLA Yearbook*, 3, 131–156.
- Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, 27(1), 1–11. doi:10.1080/20445911.2014.976226
- Ristin-Kaufmann, N., & Gullberg, M. (2014). The effects of first exposure to an unknown language at different ages. *Bulletin Suisse de Linguistique Appliquée*, 99, 17–29.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Psychology Software Tools.
- Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38(4), 607–646. doi:10.1017/S0272263115000327
- Shoemaker, E., & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research*, 29(2), 165–183. doi:10.1177/0267658313479360
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. doi:10.1016/0959-4752(94)90003-5
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73. <http://dx.doi.org/10125/44214>
- VanPatten, B. (2014). Epilogue: Input processing by novices - issues in the nature of processing and in research methods. In Z. H. Han, & R. Rast (Eds.), *First exposure to a second language* (pp. 193–207). Cambridge University Press.
- Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *Modern Language Journal*, 97(1), 109–130. doi:10.1111/j.1540-4781.2013.01428.x

**Imma Miralpeix** is an Associate Professor at the University of Barcelona, where she obtained her Ph.D. in Applied Linguistics. Her main research interests include second language vocabulary acquisition, especially lexical development and assessment, and multilingualism. She has recently investigated the potential of multimodal input for L2 vocabulary learning in EFL settings at different proficiency levels.