Prominence in speech and gesture favour second language novel word learning

Olga Kushch, Alfonso Igualada & Pilar Prieto

To cite this article: Olga Kushch, Alfonso Igualada & Pilar Prieto (2018) Prominence in speech and gesture favour second language novel word learning, Language, Cognition and Neuroscience, 33:8, 992-1004, DOI: 10.1080/23273798.2018.1435894

To link to this article: https://doi.org/10.1080/23273798.2018.1435894

View supplementary material

Published online: 11 Feb 2018.

Submit your article to this journal

Article views: 118

View Crossmark data

Citing articles: 1 View citing articles
Prominence in speech and gesture favour second language novel word learning

Olga Kushcha, Alfonso Igualada and Pilar Prieto

Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain; Psychology and Education Science Studies, Universitat Oberta de Catalunya, Barcelona, Spain; Department of Translation and Language Sciences, Universitat Pompeu Fabra, Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

ABSTRACT
While recent research has shown that iconic gestures and the so-called pitch gestures (or gestures that mimic melody in speech) favour word learning in a second language, little is known about (a) the potential benefits of beat gestures (or hand gestures that accompany prosodic prominence) for second language novel word learning, and (b) the contribution of prosodic prominence (independently or in combination with gestural prominence) to this effect. This study investigates the effects of prosodic prominence (e.g. focal pitch accent) and visual prominence (e.g. beat gesture) on L2 novel vocabulary learning. In a within-subjects design, 96 Catalan-dominant native speakers were asked to learn 16 Russian words in four conditions, namely the presence or absence of prosodic prominence in speech (L+H* pitch accent) combined with the presence or absence of visual prominence (beat gesture). The results of recall and recognition tasks conducted after a training session showed that the strongest effect corresponded to target words presented with visual prominence together with prosodic prominence; by contrast, the condition involving visual prominence with no prosodic prominence triggered smaller effects than the condition involving prosodic prominence alone. Thus, beat gestures produced naturally (that is, accompanied by focal pitch accent in speech) favour second language vocabulary learning. The results have implications for second language instruction practices and multisensory integration and working memory models.

1. Introduction

There is broad consensus in the literature that humans use multimodal cues in their communication. We express information and convey our cognitive processes through both speech and gesture modalities (e.g. manual and facial gestures, body posture). Research over the past decades has shown that speech and manual co-speech gestures are tightly integrated at both the phonological (i.e. temporal) and semantic-pragmatic levels (e.g. Kendon, 1980; Levinson & Holler, 2014; McNeill, 1992, 2005; Poggi, 2007) and form part of a single communicative system (e.g. Bernardis & Gentilucci, 2006; Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992). Co-speech gestures are produced as part of an intentional communicative act, are constructed within speech and are not a functional act on an object or person (McNeill, 1992).

A good number of studies have demonstrated the positive role of representational gestures (which have also been referred to as iconic gestures, as the form of the gesture bears a close relation to the semantic content of speech; McNeill, 1992) in enhancing word memory recall and thus facilitating lexical access in first languages (e.g. So, Sim Chen-Hui, & Low Wei-Shan, 2012). However, much less is known about the cognitive effects of beat gestures. Beat gestures are a type of rhythmic hand and arm movement that are typically associated with prominent prosodic positions in speech; their function is non-referential and they are generally used in language to signal informational focus (e.g. McNeill, 1992; Shattuck-Hufnagel, Ren, Mathew, Yen, & Demuth, 2016).

In human communication, the focussing of information is commonly achieved through prosodic prominence. Recent research in audiovisual prosody has shown that prosodic prominence in speech is typically produced simultaneously with prominence expressed with gestural features (e.g. such as head nods, eyebrow movements, hand beat gestures, or exaggerated articulation, see, for example, Prieto, Puglesi, Borràs-Comes, Arroyo, & Blat, 2015; Dohen, 2009; Ekman, 1979; Shattuck-Hufnagel et al., 2016; Swerts & Krahmer, 2008). There is a strong temporal connection between the presence of prosodic prominence (or pitch accentuation) and beat gestures. Typically, the most prominent
part of co-speech gestures (the gesture stroke or apex) is temporally aligned with prominent parts of speech (i.e. accented syllables) (e.g. McNeill, 1992; see Esteve-Gibert & Prieto, 2013 for a review). Yasinnik, Renwick, and Shattuck-Hufnagel (2004) showed that during a narration more than 90% of instances of the gesture apexes in English occurred together with a pitch-accented syllable (see also Jannedy & Mendoza-Denton, 2005 for a review).

1.1. Effects of beat gestures on memory in L1

With respect to the mnemonic effect of beat gestures, there are contradictory results in the literature. First, Feyereisen (2006) argued that beat gestures might not enhance memory recall. In his experiment, he examined the mnemonic effect of producing meaningful gestures (i.e. representational gestures) vs. nonmeaningful gestures (such as beats), and detected no effect of nonmeaningful gestures on memory. However, indefinite gestures (i.e. gestures in which the referent was hard to identify) and beat gestures were both grouped together as "nonmeaningful" gestures and it was therefore impossible to analyze their effects separately. More recent results obtained by So et al. (2012) revealed that while iconic gestures enhanced memory recall in adults and children, beat gestures played a positive role only for adults. However, we must note that in So et al.'s study words accompanied by beat gestures were presented as a sequence of isolated words and not within a discourse context, whereas one of the crucial functions of beat gestures is precisely to highlight the most prominent part of a discourse (e.g. McNeill, 1992; Shattuck-Hufnagel et al., 2016). By contrast, a more recent study by Igualada, Esteve-Gibert, and Prieto (2017) carried out with 3- to 5-year-old children showed that children were better at recalling words when they were presented with a beat gesture than when they were not. The results support the idea that beat gestures do help children to recall prominent words presented in a relevant discourse. Similarly, Austin and Sweller (2014) showed that beat gestures can be of help in the recall of spatial directions in 3- to 4-year-old children. In their experiment, children recalled information about spatial directions better when the spatial directions were accompanied by beat gestures. A recent study by Llanes-Coromina, Vilà-Giménez, Kushch, and Prieto (under revision) has also shown that preschool children remember and comprehend information better in a set of target discourses when it is accompanied by prominence in both speech and gesture.

However, all the empirical evidence suggesting that beat gestures can bootstrap mnemonic processes has so far come from studies where participants were asked to recall information in their native language, and rather less is known about the effects of beat gestures on learning new words in a second language. In the following subsection we summarise the research that has explored the beneficial effects of gestures on second language word memorisation.

1.2. Effects of gestures on vocabulary learning in L2

Research on the role of gestures in L2 word learning has thus far centred largely on the role of representational gestures rather than beat gestures. A number of studies have shown that items accompanied by meaning-related representational gestures can be learned faster in an L2, as they facilitate word-meaning associations (e.g. Kelly, McDevitt, & Esch, 2009; Macedonia, Müller, & Friederici, 2011; Quinn-Allen, 1995; Tellier, 2008). The study by Macedonia et al. (2011) compared the effect of iconic vs. meaningless gestures produced by instructors on noun recall in students of a foreign language. In this study, 33 German-speaking participants were trained to remember 92 nouns from a corpus of artificial words that was created according to Italian phonotactic rules. In the training session, words were accompanied by either iconic or meaningless gestures. Participants were trained over four days and then performed a recall test on the fifth and sixth days. The results showed significantly better recall of words accompanied by iconic gestures in both short- and long-term memory tests. Tellier (2008) conducted a study to test the impact of iconic gestures on L2 word learning by children. Twenty French-speaking children were presented with eight common English words, half of them accompanied by iconic gestures and the other half by pictures. When asked to recall the words, the children performed significantly better in the gesture condition. Further support for the beneficial effect of iconic gestures on L2 word learning was presented by Kelly et al. (2009). In their study, 12 Japanese verbs with common everyday meanings were presented to adults with no previous knowledge of Japanese. The words were presented in blocks of three in the following four conditions: (a) speech, (b) speech + congruent gesture (for example, showing the gesture of drinking while saying the verb "drink"), (c) speech + incongruent gesture (showing the gesture of washing one’s face while saying the verb "drink"), and (d) repeated speech. The results showed that the most positive effect on word learning was achieved when the items were presented in the speech + congruent gesture condition.

Another line of research has explored the beneficial use of pitch gestures, or gestures that mimic or represent
the melody of speech, on L2 lexical tone discrimination and word learning in a target tonal language. For example, in the study by Morett and Chang (2014), 57 English monolingual participants were asked to learn a total of 20 Mandarin words that were accompanied by either hand gestures illustrating the shape of the tone pitch (pitch gesture), semantic (representational) gestures conveying the meaning of the word, or no gesture. The results showed that pitch gestures helped subjects distinguish between the meanings of Mandarin words that varied only in tone. These findings provide evidence that the visuospatial features of such pitch gestures might be facilitating the discrimination between Mandarin words differing in the lexical tones and thus indirectly enhance L2 word learning. However, in Morett and Chang’s study the availability of pitch gestures failed to improve the participants’ performance in a pitch discrimination task. By contrast, a recent study by Baills, Suárez-González, González-Fuente, and Prieto (under revision) has confirmed that observing and producing pitch gestures favours both tone discrimination and lexical word identification and recognition by non-tonal learners of Chinese.

1.3. The current study and theoretical implications

To our knowledge, none of the research carried out thus far has specifically focussed on the issue of whether beat gestures (gestures that convey prosodic rather than semantic information) can also enhance L2 word learning (but see Hirata, Kelly, Huang, & Manansala, 2014 and Gluhareva & Prieto, 2017 for the use of beat gestures to teach L2 pronunciation). Moreover, given that beat gestures are typically associated with prominent prosodic positions, previous experimental studies have not teased apart whether the effects of beat gestures are mainly due to the presence of concomitant prosodic prominence or not. In sum, while iconic gestures encode semantic information that can help learners memorise novel words, it is not clear whether beat gestures, and also their concomitant speech prominence features, can also be of help in the vocabulary learning process.

The present study has the goal of assessing the effects of the presence or absence of visual prominence (i.e. beat gestures) combined with prosodic prominence (i.e. focal pitch accents) on L2 novel vocabulary learning. To this end, we presented participants with novel L2 vocabulary in which the target words were shown in one of the following four multimodal conditions: no prominence either in speech or in gesture; prominence in both speech and gesture; prominence in speech but not in gesture; and prominence in gesture but not in speech. Our hypothesis is that vocabulary presented with redundant visual prominence together with prosodic prominence would be the most beneficial condition, followed by prosodic prominence without visual prominence. We expect the prosodic prominence without visual prominence condition to show positive results as it is comparable to a natural production of speech prominence (e.g. Bock & Mazzella, 1983; Fraundorf, Watson, & Benjamin, 2010). By contrast, visual prominence without the support of prosodic prominence represents quite an unnatural cross-modal combination, and we therefore hypothesised that it would entail no benefit relative to the control condition, where speech lacked either prosodic or gestural prominence marking.

That said, it is worth pointing out that there exist different theories about the effects of multimodal encoding of information on working memory. A number of studies have explained the memory enhancement triggered by gestures in terms of the depth of encoding. For example, Quinn-Allen (1995) states that gestures provide a context for verbal language that results in deep processing and internalisation of the verbal information. Tellier (2008) explains the depth of encoding in terms of multimodality and refers to the support of dual coding theory (Baddeley, 1990; Paivio, 1971). According to the dual coding theory, learning processes can be improved when both auditory and visual modalities work together. Baddeley’s (1990) model claims that information is better coded using the two modalities because the combination of different modalities leaves more traces in the memory system. The model posite three independent components: (1) the articulatory loop, which is a speech-sound-based storage system of a limited quantity of phonological items; (2) the visuospatial sketchpad, which encodes non-verbal visual and spatial information; and (3) the central executive device, which coordinates the two other components and directs attention to incoming stimuli. Thus, according to Baddeley (1990), working memory consists of separate auditory and visual working memories, and consequently the representation of auditory and visual information occurs in independent systems. By this logic, the addition of redundant visual information (gestural prominence) to speech information (regardless of whether the visual information is semantically related with the associated speech or not) should create multimodal memory traces and learning can be improved when the information is presented visually and auditorily. Also, incorporating an additional redundant visual modality (beat gestures) to auditory modality (prosodic prominence, in our case) should also lead to improvement. On the other hand, the independence between the two systems in this model should make it less likely
that cross-modal interactions exist in memory. In our case, it is possible that the multimodal benefit of beat gestures might critically depend on whether redundant/matched information is being encoded or not, e.g. shared prosodic prominence information. In sum, a dual coding theory like the one proposed by Baddeley (1990), which crucially maintains a strict separation between modality-specific subsystems, would predict that adding a redundant modality to an existing one (either audio or visual) in a novel words presentation task should lead to their better memorisation. By contrast, it would not predict any cross-modal interactions. In our experiment, Baddeley’s (1990) model would expect a benefit of adding visual prominence in both conditions namely the one which includes redundant/matched information regarding prominence in gesture and prominence in speech, and the one including non-matched information regarding prominence in gesture and non-prominence in speech.

By contrast, the grounded or embodied cognition theory (Barsalou, 2008) would make a different set of predictions in our experiment. A main principle of grounded or embodied cognition theory (Barsalou, 2008) is that cognition is grounded in multiple ways and that it shares mechanisms with not only perception and introspection but also with action. A set of studies within this paradigm highlight the importance of the body in cognition, as bodily states can cause cognitive states and be the result of them (e.g. Barsalou, Simmons, Barbey, & Wilson, 2003; Barsalou, 2008). And gesture is considered an important form of embodiment in language, and it is closely linked to memory (Barsalou, 2008). Recent work on embodied cognition states that language and body movements are supported by the same neural substrates (e.g. Glenberg & Kaschak, 2002; Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005). The cognitive system utilises the environment and the body as external informational structures that support internal representations (Barsalou et al., 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). There is neurophysiological evidence that self-performing a gesture when learning verbal information forms sensorimotor networks that represent and store the words in both native (Masumoto, 2006) and foreign languages (Macedonia et al., 2011). In addition to this, there is also evidence that not only gesture production, but also gesture observation leads to the formation of motor memories in the primary motor cortex (Stefan et al., 2005), which is considered a likely physiological step in motor learning. Thus, in contrast with dual coding theories, such as Baddeley’s (1990) model, embodied cognition theory would suggest that it is the integration of these components into one system that leads to memory improvement (see also Quak, London, & Talsma, 2015). Thus, embodied cognition theories would predict that the integration of gesture and speech (that is when beat gestures add naturally congruent/matching visuospatial information to prosodic prominence) can favour information coding. Following this theory we could expect that words presented with beat gestures combined with no prosodic prominence (e.g. a crossmodally incongruent interaction) would not have the same beneficial effect as a congruent audiovisual combination (e.g. beat gestures associated with prosodic prominence). In summary, our hypothesis is that vocabulary presented with redundant visual prominence together with prosodic prominence would be the most beneficial condition, followed by prosodic prominence without visual prominence. We expect prosodic prominence without visual prominence condition to show positive results as it is comparable to natural productions of speech prominence (e.g. Bock & Mazzella, 1983; Fraundorf et al., 2010). Finally, the condition which presents visual prominence without the support of prosodic prominence represents an unnatural crossmodal combination, because in natural life situations beat gestures are produced together with prosodic prominences (Shattuck-Hufnagel et al., 2016), and function in congruent prosodic contexts (Dimitrova, Chu, Wang, OzuyeK, & Hagoort, 2016). We thus hypothesise that the multisensory matching (and not mismatching) integration between visual and prosodic prominence will be central for optimal working memory processes.

2. Method

The study has a within-participant design with two factors, Prosodic Prominence (No prominence; Prominence) and Gesture Prominence (No prominence; Prominence), combined in a 2 × 2 Latin square design.

2.1. Participants

An initial group of 157 students at the Universitat Pompeu Fabra in Barcelona, Catalonia, volunteered to participate in the experiment. Following the administration of a screening questionnaire to determine whether the subjects were Catalan-dominant or not, 61 had to be eliminated, as they were Spanish-language-dominant (e.g. they reported using Spanish for more than 70% of their daily verbal communication needs). Due to the fact that in the experiment both the instructions and the stimuli themselves were in Catalan, we decided to eliminate Spanish dominant participants (who sometimes have little contact with Catalan) in order to be sure that our participants had a
homogeneous linguistic profile. The remaining 96 subjects (mean age = 18.77; SD = 1.33; range 18–22; 64 females, 32 males) proceeded to take part in the study. They reported using Catalan for on average 75.2% (SD = 8.9) of their daily communication needs and having no previous formal experience with the Russian language. All subjects provided written informed consent giving permission to process their data.

2.2. Materials

2.2.1. Audio-visual recordings: Discourse Completion Task

In order to find out which types of beat gestures and intonation are used in contexts of second language teaching, 11 Catalan-speakers recruited from the University Pompeu Fabra participated in the Discourse Completion Task including eight discourse situations2 (Blum-Kulka, House, & Kasper, 1989; Billmyer & Varghese, 2000; Félix-Brasdefer, 2010). They were asked to imagine that they were English teachers and that they were teaching new words to their students. They were also instructed to strongly emphasise the target English word to help the imagined student remember the word in English. No specific instructions were given as to how they were to produce emphasis on the English words or whether they were to produce gestures. An example of a target sentence is as follows: Finestra es diu “window” en anglès (“Window is called [target word] in English”). They were recorded with a PMD660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone as they carried out this production task.

This resulted in a total of 88 sentences produced by our participants (11 participants × 8 discourse situations). In each of the 88 recorded tokens, the target word in English was isolated and then analyzed for gestural and prosodic information. From a prosodic point of view, the most common pitch accent was found to be L+H* on the target word (98.20%). From a gestural point of view, only 52.27% of the utterances were accompanied by gestures. Different kinds of gestures were detected in these sentences, for a total of 46 gestures. Of these, 32.60% were iconic gestures, 8.64% were deictic gestures (i.e. pointing gestures), and 27% were beat gestures. As the aim of the study was to elicit beat gestures, we focussed on the 12 beat gestures that were detected during the completion of the task. A type of rhythmic hand and arm movement that is typically associated with prominent prosodic positions was considered as a beat gesture (McNeill, 1992). Beat gestures varied in their form and size, for example, some were produced as energetic flicks of a slightly opened palm, and others were produced by raising the hands up from the elbows, etc. The selection of the beat gesture used in the stimuli was based on the most frequent type of beat gesture, the palm-up bimanual gesture, that appeared in 8 out of 12 beat gestures (see Figure 2).

2.2.2. Preparation of the stimuli

The stimulus materials consisted of 64 test items prepared on the basis of 16 Russian words (all nouns3) conveying common everyday meanings (16 target words × 4 multimodal conditions; see Appendix 1). All words had a disyllabic CVCCV structure with stress on the first syllable and complied with the phonotactic constraints of the Catalan language.

For each Russian word we created four 2 s videos corresponding to the four experimental conditions that result from combining the two factors (1) prominence in speech and (2) prominence in gesture: no prominence in either speech or gesture: (Condition 1 – baseline); prominence in both speech and gesture (Condition 2); prominence in speech but not in gesture (Condition 3); and prominence in gesture but not in speech (Condition 4). The stimulus videos consisted of an instructor producing the target Russian word in a standard context sentence, e.g. Bossa es diu “sumka” en rus “Bag is called ‘sumka’ in Russian”.

A Catalan-Russian bilingual speaker was videotaped producing the 16 stimulus sentences in two conditions: (1) no prominence in either speech or gesture, and (2) prominence in both speech and gesture. To create the first of these conditions, the instructor produced the target word with a non-focal L* pitch accent and kept her hands still (see Figure 1). To create the second condition, the instructor produced the target word with a focal L+H* pitch accent and a beat gesture realised with her two hands held with the palms open (see Figure 2). Videos were recorded with a PMD660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone. During the stimuli preparation we made sure that in all the conditions the speaker kept the same facial expression and we controlled that no additional markers of prosody (for example head nods or eye-brow movements) were present. The Catalan-Russian bilingual instructor was specifically trained for this.

To generate conditions 3 and 4 (prominence in speech but not gesture, prominence in gesture but not speech), the audio recording of the target word in Condition 1 (no prominence) was replaced by the one in Condition 2 (prominence in speech) and vice versa by using Premiere Pro CS6 software. To check that the resulting audio-visual combinations did not differ from the remaining videos in perceived naturalness (synchrony), we asked 10 native Catalan-speakers to rate the naturalness of the videos from 1 (no synchronisation)
to 5 (totally synchronised). All videos were rated as highly synchronised ($M = 4.69$, $SD = 0.46$). Also, there was no significant difference noticed between the conditions in naturalness rating. The ratings were the following: Condition 1 ($M = 4.71$, $SD = 0.45$); Condition 2 ($M = 4.65$, $SD = 0.48$); Condition 3 ($M = 4.75$, $SD = 0.46$); Condition 4 ($M = 4.68$, $SD = 0.46$) ($p > .5$).

2.3. Procedure

The experiment consisted of two parts: training and testing sessions. Participants were trained and tested in a single day in four groups of 24.

2.3.1. Training

In the training session, all 96 participants were exposed to words in a within-subjects design (i.e. each participant was exposed to all four conditions). Each subject was exposed to a total of 16 words repeated 4 times ($16 \times 4 = 64$). Each word was assigned to one condition. Importantly, for a given participant this condition remained the same for each word during the four repetitions. Across subjects, words were assigned to different conditions to prevent a potential effect of the lexical items, thus creating a total of 4 presentations. For example in Presentation 1 (i.e. participant 1) the word “bag” was assigned to Condition 1, in Presentation 2 (i.e. participant 2) the same word “bag” was assigned to Condition 2, in Presentation 3 the word “bag” was seen under Condition 3, in Presentation 4 the word “bag” was presented in Condition 4, etc. A total of 24 participants were assigned to one of the presentations (e.g. $24 \times 4 = 96$ participants).

In order to avoid order effects, the following steps were taken. First, the 16 target words were organised in four blocks within one repetition. Each block contained one word in each condition (i.e. four words). The order of the words within the block was counterbalanced, and the position of the blocks was also counterbalanced across repetitions. For example, for a given...
subject the word “bag” (assigned to Condition 1) appeared in position 1 in a given block, in position 2 in the second block, etc., and moreover, these blocks of four words appeared in position 1 in the first repetition, in position 2 in the second repetition, etc.

At the beginning of the training session, participants were asked to attentively follow the instructions presented in the first five slides of a PowerPoint presentation, which contained a detailed explanation of the experiment structure. After that, the 16 Russian target words were presented, as follows.

For each target word, participants first saw the written Catalan equivalent on the screen for 3 s (e.g. *bossa* “bag”). They were then shown a 2 s video in which the instructor uttered the carrier sentence with the Russian translation embedded in it (e.g.* Bossa es diu ‘sumka’ en rus “Bag is called ‘sumka’ in Russian”). After all 16 words were presented in this fashion, there was a 2 min break during which participants were asked to complete a distraction task.4 Three further word presentations followed, using the same procedure but with the order of blocks and the order of words within each block changed and the same kind of distraction task filling the 2 min intervals between presentations.

Altogether, the training session lasted approximately 10 min. After the final training phase, there was a 5 min break, which was followed by the testing session.

### 2.3.2. Testing

The testing session consisted of two memory tasks. In the first task (a free recall test) participants were presented with the audio recording (with no video input) of the Russian words heard during the training sessions. The words were presented twice and in an order, which was different from the order in the training session. The same audio file was used in training and testing sessions, so the condition of the target word heard in the training session (prominent vs. non-prominent accent) was the same as participants heard during the testing sessions. After hearing each word twice the participants wrote the Catalan translation.

In the second task (a recognition test) the participants were auditorily presented with the Russian words heard during the training session (words were presented in a different order from the recall test). Participants heard the audio twice. For each word, participants had to choose between four possible Catalan translations of the Russian word, which included the correct translation and the translations of three other words heard during the training session, and circle the right answer from the four available options. Words used as distractors were from different blocks, but they were presented under the same experimental condition as in the training session.

The total time for training and testing was approximately 20 min.

### 3. Results

Responses by subjects for the two tests were first marked “0” or “1”, with 1 indicating in the recall task that they had recalled the correct Catalan translation of the Russian word they heard and “0” that they had not, and “1” indicating in the recognition task that they had selected the correct translation from the four alternatives while “0” indicated that they had chosen the wrong translation. All such responses obtained (3072 trials in total) in both tests were submitted to a Generalized Linear Mixed Model (GLMM), using IBM SPSS Statistics 21. The dependent variable was a combination of Response 1 (in a recall test) and Response 2 (in a recognition test), numerical measures (1-right reply; 0-wrong reply). The fixed factors were Prosodic Prominence (prominence in speech vs. no prominence in speech), Gestural Prominence (prominence in gesture vs. no prominence in gesture), Task (Two levels: free recall test and recognition test), as well as two-way and three-way interactions of those factors (e.g. Prosodic Prominence × Gestural Prominence; Prosodic Prominence × Task; Gestural Prominence × Task; Prosodic Prominence × Gestural Prominence × Task). Participant, Item, and Item Position were set as random factors. Figure 3 shows the results obtained from the GLMM analysis.

The GLMM results revealed a significant main effect of Task ($F(1,3064) = 332,697, p < .001$), confirming that participants performed better in the recognition test.
than in the free recall test. This is not surprising, since presumably it is easier to select a correct answer from a list of four than independently recall a translation from memory. Importantly, a main effect of Prosodic Prominence was found \( (F(1,3064) = 30.487, \ p < .001) \), showing that items accompanied by prominence in speech were remembered better. There was no main effect for Gestural Prominence \( (F(1,3064) = 0.358), \ p = .55) \), but a significant interaction was found between Prosodic and Gestural Prominence \( (F(1,3064) = 4.885), \ p < .05) \), indicating that only beat gestures produced with prosodic prominence had a significant positive effect on L2 word recall. Post-hoc analyses showed that when the gesture was prominent, scores for the condition with prosodic prominence were higher in comparison to the condition without prosodic prominence \( (p < .001) \), and when the gesture was not prominent, the condition with prosodic prominence scored higher than the condition without prosodic prominence too \( (p < .05) \). By contrast, when prosody was prominent the condition with gesture prominence scored higher than the condition without gesture prominence \( (p < .05) \), but crucially when prosody was not prominent the condition with gesture prominence scored similarly to the condition without gesture prominence \( (p = .27) \). The fact that there were no significant interactions with task (all \( ps > .40 \)) indicates that these patterns of results regarding the effects of prosodic prominence or gestural prominence are the same in both the free recall and recognition tasks. Figure 3 shows the mean proportion of memorised words across the four training conditions in the two tasks (free recall and recognition), separated by the prosodic and gestural conditions (prominence vs. no prominence).

4. Discussion and conclusions

The aim of our study was to investigate the potential positive effects of prosodic and gestural prominence on second language novel word learning. While previous research on lexical learning in an L2 has fully acknowledged the positive role of representational gestures (e.g. Kelly et al., 2009; Macedonia et al., 2011; Quinn-Allen, 1995; Tellier, 2008), little is known about the potential effects of beat gestures. Our study is the first to test whether accompanying the target words with beat gestures (whether accompanied by prominent prosody or not) is beneficial in the context of novel word learning in a second language. Moreover, the experimental design in our study allowed us to independently assess the potential effects of prosodic and gesture prominence on novel word learning.

In this respect, one of the important results that comes out of our experiment is the asymmetry between the effects of the combination of prosodic and gestural prominence. First, the presence of prosodic prominence alone had a clear beneficial effect on novel word learning in our data in comparison with the no prominence condition. Participants in the study remembered more words in the focal pitch accent condition (prominence in speech) than in the non-focal condition (no prosodic prominence), both in free recall and recognition tests, demonstrating that prominence in speech helps learners to acquire novel words in a second language. This effect is consistent with previous studies that have reported a positive role of prosodic prominence on information comprehension and memorisation in an L1 (e.g. Bock & Mazzella, 1983; Fraundorf et al., 2010; Kushch & Prieto, 2016). By contrast, presenting words with beat gestures but without prosodic prominence did not have beneficial effects on word learning. In the free recall test the mean proportion of successfully memorised words in the no prominence in speech but prominence in gesture condition was even slightly lower than in the baseline no-prominence condition. Importantly, the presence of beat gestures (i.e. gestural prominence) had an (optimal) effect only when it was accompanied by prosodic prominence. Also, when target words were produced with both gestural and prosodic prominence the beneficial effects were strongest in comparison to other conditions.

On the one hand, the optimal effects of the joint association of visual (gestural) and prosodic prominence (as compared with prosodic prominence alone) found in the present study confirm previous research in the L1 field that shows that beat gestures aid word recall in both adults (So et al., 2012) and children (Igualada et al., 2017; Llanes-Coromina et al., under revision). Though the abovementioned studies featured beat productions naturalistically accompanied by prominent prosody, thus not controlling for the possible interaction between the two factors, their results point to the fact that naturally produced beat gestures do have a positive impact on lexical recall. In the last of these studies, (Llanes-Coromina et al., under revision) compared the effects of beat gestures and prosodic prominence on information memorisation in contrastive discourse in a first language by adults. Following the design used by Fraundorf et al. (2010), 20 participants were asked to listen to 48 video-recorded stories where the target items were presented in a contrastive focus discourse. Each story contained two pairs of items in the introductory sentence. In the next sentence only one of the items was mentioned. This item was considered the target word and was presented under one of the following
two experimental conditions: prominence in both speech and beat gesture, and prominence in speech alone. The results showed that participants performed significantly better in the recall task when the target item was associated with both prosodic and visual gestural (beat gesture) prominence than when it was associated with prosodic prominence alone. All in all, this research provides evidence that not only gestures that contain semantic information (such as representational gestures) but also gestures that help focus the attention on the important part of the discourse (beat gestures) have beneficial effects for word memorisation.

An explanation for the enhancing effects of beat gestures is related to the attention processes and also language-related processes triggered by these manual gestures. A neurophysiological study conducted by Hubbard, Wilson, Callan, and Dapretto (2009) investigated whether the presence of beat gestures impacted speech perception at the neural level, controlling for the presence of prosodic prominence. Thirteen adult subjects underwent an fMRI while being exposed to videos with spontaneously-produced speech accompanied by either beat gestures, nonsense hand movements, or a still body. Their findings suggested that adding gestural prominence in the form of beat gestures to prosodic prominence (a) causes greater activity in bilateral nonprimary auditory cortex, suggesting a common neural substrate for processing speech and gesture; and (b) causes an increase of activity in the areas responsible for speech intelligibility, namely the left anterior areas of left superior temporal gyrus and sulcus. Thus, Hubbard et al.’s (2009) results suggest that beat gestures, as mentioned above, may help focus the viewer’s attention on speech (see also Biau & Soto-Faraco, 2013; Dimitrova et al., 2016). Several studies assessing neurological activations during observations of beat gestures support the hypothesis that beat gestures might increase attention processes and activations of language-related brain areas (e.g. Biau & Soto-Faraco, 2013; Holle et al., 2012; see also Hubbard et al., 2009). The functional neuroimaging study by Biau & Soto-Faraco, 2015 showed that beat gestures activated different brain areas in comparison to other non-related movements. Depending on whether speech was synchronised with beat gestures or with other non-gestural stimuli (discs/dots moving on a screen) different brain areas were activated. Beat gestures activated language-related areas of the brain, while non-gesture stimuli activated visual perception areas. Hubbard et al. (2009) found that beat gestures, and not nonsense movements or still images, enhanced auditory processing of speech. These studies support the idea that beat gestures can be distinguished from other potential visual highlighters because of their direct integration in the language system. However, if the beneficial effect of beat gestures was exclusively due to attention, we would expect the target items presented with beats be learnt at the expense of others. Interestingly, there is some evidence that points to the contrary. In a recent experiment with preschoolers, Llanes-Coromina et al. (under revision) assessed the memorisation of a list of nouns within a child-directed discourse context. While a beneficial effect was seen for the items associated with beat gestures and prosodic prominence, the results also showed that there was no negative effect for the items presented with no beats within the same lists. These results show that the positive effects of beats are probably not only due to attentional saliency effects.

Our results have clear implications for the models of multimodal cognition and learning reviewed above, namely a classical version of the dual coding theory which maintains a strict separation between modality-specific subsystems (e.g. Baddeley, 1990) and the embodied cognition theory which supports the direct integration of the two modalities. Remember that in our results beat gestures without prosodic prominence (a mismatching or incongruent combination) did not have a beneficial effect on memory. As mentioned before, naturally produced beat gestures are almost invariably linked to prosodic prominence in speech (Shattuck-Hufnagel et al., 2016; Yasinnik et al., 2004; see Wagner, Malisz, & Kopp, 2014 and Jannedy & Mendoza-Denton, 2005 for a review). As described in the Introduction, these two conceptions lead to two different predictions regarding the effects of this incongruent audiovisual presentation. While the dual coding theory predicts a positive effect of beat gestures across conditions, regardless of whether they were matched or mismatched with prosodic prominence, this is not the case for embodied cognition theories. Thus, embodied cognition approaches correctly predicted that only matched cross-modal interactions has a reinforcing effect and leaves stronger memory traces. The results of our study thus seem to confirm the predictions based on embodied cognition paradigm, and show that not only gestures that convey semantic meaning have positive effects on learning processes.

However, it is important to also point out at this stage that more recent approaches to working memory (e.g. Baddeley, 2000) no longer maintain a strict separation between modality-specific subsystems. Baddeley (2000) proposed an extension to the working memory model presented in Baddeley (1990) by introducing a component called episodic buffer which includes a temporary storage of information that is presented through
The results of an ERP study by Wang and Chu (2013) seem to highlight the independent role of hand gestures relative to prosodic prominence. In this experiment, participants were asked to watch videos of a person speaking and gesturing. Target videos contained critical words, which were presented under six experimental conditions containing a combination of two factors, namely hand movement and pitch accent (accented and unaccented). The results showed that both beat gestures and pitch accentuation elicited smaller negativities in the N400 time window. These results suggest that prominence in gesture, like prominence in speech, triggers the attentional system separately for semantic processing. Thus Wang and Chu’s results seem to contradict the results of our experiment, showing that beat gestures have positive effects only when accompanied by prosodic prominence. Interestingly, aside from differences between the tasks performed in Wang and Chu (2013) and those in the present study, in Wang and Chu’s experiments participants could not see the mouth of the speaker. Further research is needed to disentangle this issue, as the authors’ decision to hide lip movements might have had a negative effect on the natural integration of prominence coming from speech and both articulatory and hand gestures.

In conclusion, our results are in line with recent work which has suggested that the use of hand movements (e.g. pointing gestures) can substantially benefit cognitive processing and enhance the learning of ideas, whether textual or diagrammatic (Hu, Ginns, & Bobis, 2015). The results of our study have implications for instructional practices in foreign language teaching because they suggest that the teachers’ use of prosodic and gestural prominences may help students to acquire novel words in a second language. Vocabulary is a core part of learning a new language. Vocabulary is obviously indispensable for comprehension and producing speech with appropriate meanings as well as generating syntactic, morphological, and phonological structures. In the second language classroom special attention is paid to vocabulary learning and especially to the ways novel vocabulary can be presented and learnt. In the current study we have used a translation procedure to introduce novel vocabulary in a foreign language. The results of the study showed that accompanying target words with prosodic and gestural prominence aids their learning. However, it is necessary to note that learning a novel word is a complex process. Lexical knowledge includes knowing a series of characteristics associated with the word beyond the mere notion of its semantic meaning (e.g. Richards, 1976; Nation, 1990). Thus our study only addresses the initial stage of the vocabulary learning process; the effects of prosodic and gestural prominence on L2 vocabulary learning could also be tested in a longitudinal design and with words embedded in a more natural pragmatic context. Still, the results of our study have direct implications in the foreign language.
classroom, where it is common to see a fair amount of gesture use by language teachers (e.g. Smotrova & Lantolf, 2013). The use of prosodic and gestural prominence together might constitute a good teaching strategy to cue relevant information in the foreign language classroom and also a promising approach for teaching vocabulary in computer-mediated education environments.

Notes
1. In this article, the term ‘second language learning’ is used as a cover term that refers to the process of learning another language after the native or dominant one. This is a common strategy in the field, which uses this term to refer to the learning of a third or a fourth language (Gass, 2013).
2. The Discourse Completion Task is an inductive method which has been applied for many years in research on pragmatics and sociolinguistics, and also recently on prosody, with good results (e.g. Prieto & Roseano, 2010).
3. The decision to work with nouns was primarily due to the fact that this allowed us to have better control over the number of syllables, syllable types, and stress positions within the target Russian words.
4. Participants had to memorise 2 rows of 9 numbers shown briefly on a PowerPoint slide and then write them down on a sheet of paper.

Acknowledgments
We would like to thank the students at the Universitat Pompeu Fabra who participated in the experimental tasks as well as Discourse Completion Task recordings. Many thanks to Carmen Pérez Vidal and Joan Borràs-Comes, who allowed us to contact the students in their classes and external groups, and to Joan also for his help with the statistical analysis. We are grateful to Anna Denissenko for her assistance with the recording of stimuli. Finally, this research would not have been possible without funding from the Spanish Ministry of Science and Innovation grant FFI2015-66533-P (“Intonational and gestural meaning in language”), and a grant awarded by the Generalitat de Catalunya (2014SGR-925) to the Prosodic Studies Group.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This research would not have been possible without funding from the Spanish Ministry of Science and Innovation grant FFI2015-66533-P (“Intonational and gestural meaning in language”), and a grant awarded by the Generalitat de Catalunya (2014SGR-925) to the Prosodic Studies Group. Agència de Gestió d’Ajuts Universitaris i de Recerca.

References
