



HAL
open science

The neural correlates of embodied L2 learning Does embodied L2 verb learning affect representation and retention?

Ana Zappa, Deirdre Bolger, Jean-Marie Pergandi, Raphael Fargier, Daniel Mestre, Cheryl French-Mestre

► To cite this version:

Ana Zappa, Deirdre Bolger, Jean-Marie Pergandi, Raphael Fargier, Daniel Mestre, et al.. The neural correlates of embodied L2 learning Does embodied L2 verb learning affect representation and retention?. 2022. hal-03657366v1

HAL Id: hal-03657366

<https://hal.science/hal-03657366v1>

Preprint submitted on 2 May 2022 (v1), last revised 5 Dec 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The neural correlates of embodied L2 learning

Does embodied L2 verb learning affect representation and retention?

Ana Zappa^{1, 2,6 *}, Deirdre Bolger^{1, 6}, Jean-Marie Pergandi^{1, 3,4}, Raphael Fargier^{1, 2,6},
Daniel Mestre^{1, 3,4,5,6}, & Cheryl Frenck-Mestre^{1, 2,5,6}

1Aix-Marseille Université, 2Laboratoire Parole et Langage, 3IMS, 4Centre de Réalité Virtuelle de la Méditerranée, 5Centre National de Recherche Scientifique, 6Institute of Language, Communication and the Brain

The present study aims to investigate how naturalistic actions in a highly immersive, multimodal, interactive 3D virtual reality (VR) environment may enhance word encoding by recording EEG in a pre/post-test learning paradigm. Both imaging and electrophysiological data have established motor activation during language processing, and behavioral data has shown that coupling word encoding with gestures enhances learning. However, the neural underpinnings of facilitated action language learning have yet to be elucidated. Herein, we couple EEG recording with virtual reality to examine whether “embodied learning”, or learning that occurs using specific physical movements that are coherent with the meaning of new verbs, creates linguistic representations that produce greater motor resonance (a decrease in power in the mu and beta frequency bands), due to stronger motor traces, compared to learning without accompanying specific gestures. We will also investigate whether greater motor resonance while listening to learned action verbs post-learning correlates with improved retention.

1. Introduction

Following decades of considering language processing as independent from motor processes, theories of embodied cognition are building an argument for reuniting the two (Barsalou, 2008), based on the assumption that cognition is grounded in multimodal representations originating in human experience (Pulvermüller, 2005). As concerns language, this means that modal representations replace amodal symbolic linguistic representations (Fodor, 1983), giving motor processes an essential role in language processing (Wilson & Golonka, 2013). Indeed, sensory and motor systems are recruited during lexical processing, both during development (James & Swain, 2011) and in adults (Hauk, Johnsrude & Pulvermüller, 2004; Pulvermüller, 1999, 2005; Thompson-Schill, 2003). Furthermore, neuroimaging has revealed an overlap in neural mechanisms for processing speech and hand movements (Nishitani, Schürmann, Amunts & Hari, 2005). Along the same lines, gestural studies have suggested that gesture and speech comprise an integrated system (Goldin-Meadow, 2011; Graziano & Gullberg, 2018). Language processing is facilitated by gesture such that language-action congruency that occurs early in sentence processing can facilitate lexical retrieval (Glenberg & Kaschak, 2002). Furthermore, the retrieval of stored semantic representations directly influences sensorimotor activation as indexed by greater motor preparation when congruent action language is presented prior to movement (Aravena et al., 2012). Importantly, incongruity between actions and meaning can cause interference in meaning retrieval (Aravena et al., 2010; Barsalou, 1999). These results suggest that motor representations are not simply reactivated by linguistic representations post-lexically (Mahon & Caramazza, 2009), but that they can play a role in representations. In the current investigation, we will examine the interaction between motor and semantic processes and how it may affect the mapping of novel action verbs to physical actions. Word encoding will be

coupled with compatible physical actions in an interactive virtual environment that enables pseudo-natural movements, to test whether motor activation enhances novel action word learning in a foreign language. We will examine whether action verbs learned with specific actions produce greater motor activation post-training, as revealed by a decrease in the beta and mu band power, compared to verbs learned without accompanying actions. Finally, we will examine whether our experimental manipulation leads to improved retention, due to a stronger motor trace in memory (Engelkamp & Krumnacker, 1980).

Encoding new words is an essential part of language learning and has been addressed in various learning studies that investigated cortical changes associated with learning, whether in the native language (L1) or in a second language (L2). In a seminal study, McLaughlin and colleagues (2004) found differences in L2 learners' cortical activity after around 14h of classroom instruction when processing newly learned L2 words compared to pseudo words, as indexed by an N400 effect (McLaughlin, Osterhout & Kim, 2004). Evidence for semantic encoding was only found following far more instruction. Similarly, differences in electrophysiological indices of word processing between the L1 and the L2 were seen via an increase in N400 amplitude for L2 words after one semester of learning (Soskey, Holcomb & Midgley, 2016). These studies, however, investigated extended L2 training and did not allow for the observation of cortical changes occurring during the very first stages of encoding. In this vein, rapid cortical changes have been observed as a result of contextual word learning (Borovsky, Elman & Kutas, 2012; Borovsky, Kutas & Elman, 2010; Mestres-Missé, Rodriguez-Fornells & Münte, 2007; Shtyrov, Nikulin & Pulvermüller, 2010). Changes in N400 amplitude, thus indicating meaning integration, have been reported as quickly as after three exposures to novel words in highly constraining sentential contexts in the L1 (Mestres-Misse et al., 2007). Borovsky et al. (2010) concluded from their ERP data that a single exposure to novel words in highly constraining contexts is sufficient to derive meaning, as

demonstrated by differences in the N400 amplitude to the verbs preceding these items in a subsequent test sentence. Note, however, that no N400 modulations were found at the noun itself. In a subsequent study, although the cloze manipulation did not produce significant modulations of the N400 for either known words or novel words, Borovsky et al. (2012) reported significant N400 modulations in a primed lexical decision task for both known and newly learned words. This effect was reported to be restricted to words learned in highly constrained contexts, as shown in independent pairwise comparisons. Bakker and colleagues (2015) found that ERPs only showed effects of lexicalization after a 24h period of consolidation (Bakker, Takashima, van Hell, Janzen & McQueen, 2015). From the above studies we can conclude that the neural response associated with semantic encoding can be modified following relatively little exposure. Nonetheless, the neural underpinnings of learning following a short training period have not yet been fully explored. The current study aims to help to fill this void by observing cortical activity after a two-day, explicit, word-learning training using physical movement.

The benefit of physical movement for language learning and memorization is well established (Moskowitz, 1976; Quinn-Allen, 1995). Behavioral studies dating back to the 1980s have shown that illustrative gestures support language retention better than other conditions (Engelkamp & Krumnacker, 1980; Engelkamp & Zimmer, 1984). For example, Engelkamp and Krumnacker (1980) showed that verb phrases such as “shuffle the cards” were better memorized when learners performed representative gestures during learning compared to either watching someone else perform the action, imagining the action or simply listening to the sentence. Outside of the language domain, a number of studies have shown that participants encode new information better when they perform gestures that are congruent with the new content. Physical activity facilitated the integration of sung melodies (Wakefield & James, 2011) as well as mathematical (Kontra, Lyons, Fischer & Beilock, 2015) and

scientific principles (Johnson-Glenberg & Megowan-Romanowicz, 2017; Johnson-Glenberg, Megowan-Romanowicz, Birchfield & Savio-Ramos, 2016) more than verbalization. This “enactment effect”, has been replicated in studies focusing on second language or artificial language word-learning studies. After 20-30 minutes of learning novel words by simply pointing to or touching the corresponding objects, participants showed associations between sensorimotor experiences (the location of an object in a vertical space) from training and the novel words (Öttl, Dudschig & Kaup, 2017).

It has been argued that truly embodied learning involves “self-performed” or “self-generated” action, as opposed to simply observing or imagining action (James & Bose, 2011; James & Swain, 2011; Johnson-Glenberg, 2017, 2018; Johnson-Glenberg & Megowan-Romanowicz, 2017). In other words, highly embodied learning generally implies that learners physically perform gestures or movements that are directly linked to the content they are learning (Johnson-Glenberg, 2018). Both L1 and L2 lexical encoding studies generally use representative or iconic gestures (McNeil, 1992), which illustrate and map onto meaning directly. Studies with both adults (de Nooijer, van Gog, Paas and Zwaan, 2013; Macedonia & Knösche, 2011) and children (Tellier, 2008) have shown that the production and recall of (L2) lexical items is enhanced by performing representative gestures.

The studies cited above indicate that action boosts memory performance and therefore supports language encoding. However, it is still unclear what cognitive processes underlie this facilitation. One explanation is that physical action relays and helps establish implicit knowledge. Indeed, we often express information without even realizing it through gestures (Church & Goldin-Meadow, 1986). According to Sun and colleagues (2001) what they describe as the “synergy” between explicit and implicit performance can aid in learning new skills (Sun, Merrill & Peterson, 2001). The theory of Hebbian associative learning claims that the synchronous activity of neurons forms neuronal assemblies (Hebb, 1949); hence when

lexical items are acquired along with action, cortical areas involved in language processing and those involved in action planning and execution quickly develop into shared neural circuits (Pulvermüller, 1999, 2005; Tomasello, Garagnani, Wennekers & Pulvermüller, 2018). To better understand how learning may be enhanced by movement, several studies have examined the neural underpinnings of lexical-motor interactions. In two functional magnetic resonance imaging (fMRI) studies, results showed that direct interaction with objects (James & Swain, 2011) and/or performing meaningful gestures (Macedonia, Muller & Friederici, 2011) were necessary to produce activation in the motor system during the subsequent auditory processing of newly learned lexical items. Moreover, performing iconic gestures during the learning of new lexical labels led to greater activation of the semantic network or “deeper semantic encoding” (Krönke, Mueller, Friederici & Obrig, 2013).

Despite the importance of the above studies, fMRI may not be the ideal tool to show motor to language effects or vice versa. Indeed, much debate surrounds the *role* of motor activation during language processing. One of the arguments against embodied semantics is that language-induced motor activations are post-lexical and not a necessary part of language processing (Mahon & Caramazza, 2008). High temporal resolution – an advantage of EEG compared to fMRI – is hence an important element in arguing for embodied language representations. One way of quantifying motor cortex activity is to use EEG to measure event-related synchronization/desynchronization (ERS/ERD) via stimulus-locked time-frequency analysis (Vukovic & Shtyrov, 2014). Decrease in alpha, mu (8-13 Hz) and beta-band (13-30 Hz) power, mostly over central or centro-parietal sites, has been associated with sensorimotor activation involved in movement preparation and execution (Pfurtscheller & Lopes da Silva, 1999; Nicolai et al., 2014; Pineda, 2005). A decrease in the alpha rhythm has likewise been linked to motor imagery (Höller et al., 2013). Recently, desynchronization in oscillations associated with motor processes has also been observed during action language

understanding. Reading sentences describing manual actions versus abstract sentences led to the suppression of mu rhythms at fronto-central sites (Alemanno et al., 2012; Moreno et al., 2015). This is not to state that a one-to-one mapping exists between power decreases in specific frequency bands and specific cognitive functions, but that there is an association between mu and beta oscillations over central and centro-parietal sites and motor/sensorimotor activity which can be used to index language-motor interactions (cf. Klepp et al., 2019). To our knowledge, the only study that has used time-frequency to measure motor activation during language processing pre and post-training was conducted by Fargier and colleagues (Fargier et al., 2012). They showed that learning novel words in association with specific self-performed actions led to greater mu desynchronization post-training, over centro-parietal sites, which they interpreted as motor activation, compared to learning in association with abstract animations. (Fargier et al., 2012). However, on the second day of training, a fronto-central distribution of the effect, as opposed to a typical central parietal mu distribution, lead the authors to conclude that it was confined to a convergence zone.

Embodied cognition binds social and physical contexts to cognition, and therefore the environment in which learning takes place could potentially play an important role in learning outcome (Black, Segal, Vitale & Fadjo, 2012). According to Atkinson (2010), learning is not just a mental process but one that occurs in environments made up of “bodies, cognitive tools, social practices and environmental features” and this multimodality calls for an experimental approach that is likewise multimodal. One caveat of experimental protocols that examine “embodied” learning is that, given the need for control, movement is generally reduced to minimal hand actions and training most often occurs in isolated and decontextualized environments (Peeters, 2019). This is especially true of studies that analyze the neural correlates of language processing and learning using techniques such as fMRI, MEG or EEG.

For instance, when interaction with objects has been made possible, it has been limited to pointing at or touching objects, hence making it impossible for participants to map specific actions to specific words. When one considers the importance of interlocutors, social context and physical cues on how language is understood in real life (Knoeferle, 2015), physical and environmental limitations likely affect how language is learned. Within the framework of embodied cognition, it is especially important to take a closer look at the gap between real-life language processing and that which takes place in an experimental environment (Tromp, Peeters, Meyer & Hagoort, 2018).

Virtual reality (VR) is an important tool for investigating embodied language learning. Numerous L2 studies have used VR paradigms involving varying degrees of immersion to investigate language learning. They have generally found facilitation for learning in immersed conditions compared to word-word or picture-word paired associations (Berns, Gonzalez-Pardo & Camacho, 2013; Lan, Fang, Legault & Li, 2015). Furthermore, participants who learned in a VR environment using avatars (*Second Life*) showed neural activations that were more distributed and associated with embodied networks compared to the control group (Lan et al., 2015). However, *Second Life* paradigms are limited when it comes to exploring truly interactive embodied learning (for a review of L2 video games, see Legault et al., 2019). To overcome this, Legault and colleagues taught participants a set of L2 words using an ecologically valid immersive virtual reality zoo or kitchen, using word-word paired association as a control. Participants — especially less successful learners — showed higher accuracy in the immersive VR condition (Legault et al., 2019).

Peeters (2019) claims that VR “shifts the theoretical focus toward the interplay between different modalities [...] in dynamic and communicative environments, complementing studies that focus on one modality in isolation.” (p.1, 2019). VR uses visual and auditory stimuli to create an immersive sensory experience, providing participants with

believable environments. In addition, participants' head and body movements are tracked by input tools (e.g. hand controls) and participants are given real-time feedback for their actions, which provides a sensation analogous to real life (Burdea & Coiffet, 2003). The fact that participants can interact with the environment by manipulating virtual objects and carrying out naturalistic actions gives them a sense of "agency" (Johnson-Glenberg, 2018). Compared to traditional experiments, this leads to the sensorimotor system being more implicated and responses and actions being closer to what occurs in real life (Bohil, Alicea, & Biocca, 2011). Finally, VR combines ecological validity with full control over the onset, location and duration of presentation of the multimodal stimuli. Very few studies have combined virtual reality and EEG to study language processing. In an exploratory EEG-VR experiment, participants listened to a sentence ("I just ordered this salmon") and saw a virtual object that either matched (salmon) or mismatched (pasta) the object in the sentence. An N400 effect was observed for mismatched versus matched pairs, and the authors interpreted this as proof of validity for combining VR and EEG to examine language processing (Tromp et al., 2018). However, participants did not actually manipulate objects and the involvement of the motor cortex was not examined. Recently, Zappa and colleagues measured motor-related EEG activity in an interactive virtual reality environment while participants performed a Go-Nogo task and listened to action verbs prior to executing the corresponding actions (Zappa et al., 2019). Motor activation was found via a decrease in power in the mu and beta bands during verb processing and prior to movement proper, providing compelling evidence in a naturalistic setting of how motor and linguistic processes interact. Moreover, greater ERD was found for Go trials, suggesting that motor preparation influenced semantic processing. These results provide the basis for the present study, investigating the association of new linguistic labels to motor actions.

Our study will use a combined EEG-VR methodology to explore the neural correlates

of embodied learning. EEG and VR will not be employed simultaneously, but EEG will be used for measuring learning pre and post-training and VR will be used to facilitate embodied and situated learning during the training. Using a head mounted VR system (Oculus Rift) and controller, participants will be exposed to an auditory L2 lexicon of action verbs associated with videos of congruent physical actions. Participants will be assigned to one of two groups, according to whether they are expected to subsequently perform a motor action that corresponds to the specific action (picking up an object and throwing it for the verb “throw”) (Specific action condition) or simply pointing to the object (Pointing condition). Both pre and post training, learners’ knowledge of the semantic meaning of the training verbs will be measured behaviorally and through EEG using a match-mismatch task. Motor resonance will also be measured using EEG while participants listen to the training verbs as well as a set of filler verbs that are never taught, both pre and post-training (See Table. 1). We expect motor resonance during auditory verb processing to vary as a function of learning condition during the post-training session. We hypothesize that representations of verbs learned with specific actions will carry a stronger motor trace and hence produce greater motor activation than verbs learned in the pointing condition. We also predict that embodied learning using specific self-performed congruent physical actions will lead to better learning outcomes post-training compared to the pointing condition.

2. Hypotheses

1. In accordance with the theory that learning lexical items along with action can form shared neural networks (Pulvermüller, 1999, 2005; Tomasello et al., 2018) and studies showing greater motor activation for object labels learned with direct object interactions (James & Swain, 2011) or specific self-performed actions (Fargier et al., 2012), we expect

to find a decrease in beta (13-30 Hz) and mu (8-13 Hz) band power (motor activation) post-training compared to pre-training during the processing of the training verbs (passive listening task). Given that only training verbs will have been associated to meaning, these effects are not expected to be observed for filler verbs, for which no variation pre-post training should occur.

2. Activity in the premotor context has been found when learners process verbs learned with iconic gestures but not those learned with meaningless gestures (Macedonia et al., 2011). We therefore expect to find greater motor resonance for verbs learned in the Specific action condition compared to the Pointing condition.
3. Studies have shown that learners associate a new word-form to semantic content after very little exposure (Borovsky et al., 2012; Mestres-Missé et al., 2007; Yum, Midgley, Holcomb & Grainger, 2014). During the match-mismatch task, we expect that pre-training, we will not find an N400 effect for match versus mismatch trials. Post-training, we expect to find greater N400 amplitude for mismatch versus match trials in both learning conditions, due to participants accessing the semantic meaning of newly learned verbs.
4. Along with studies in non-linguistic domains showing enhanced learning when gestures are used (Broaders, Cook, Mitchell & Goldin-Meadow, 2008; Johnson-Glenberg et al., 2016; Johnson-Glenberg & Megowan-Romanowicz, 2017; Kontra et al., 2015; Wakefield & James, 2011), both behavioral (Mayer, Yildiz, Macedonia & von Kriegstein, 2015; Tellier, 2008) and electrophysiological (Macedonia & Knösche, 2011; Macedonia et al., 2011; de Nooijer et al., 2013) evidence from L2 learning studies has revealed that congruent gestures support linguistic memory and encoding and improves performance. We therefore hypothesize that the N400 effect outlined in hypothesis 3 will be greater for the Specific action condition compared to the Pointing condition.
5. In accordance with hypotheses 3 and 4, we expect to find a positive correlation between

greater motor resonance during the passive listening task and a greater N400 amplitude for mismatch versus match trials in the match-mismatch task.

6. In accordance with hypothesis 4, we predict that our behavioral results will show greater accuracy for verbs learned in the Specific action condition compared to the Pointing condition.

3. Methods

In the current study we will manipulate the type of action performed (specific object manipulation vs pointing) during L2 learning in a VR environment. During learning, participants will visualize movements performed by a virtual hand. The Specific action group will reproduce the movement on a virtual object and the control group will point to the virtual object on which the action was performed. EEG will be recorded both pre and post-training.

3.1. Ethics. This research complies with all relevant ethical regulations and has been approved by the local university ethics committee.

3.2. Statistical power analysis. For hypothesis 1, a statistical power analysis was performed for sample size estimation using G*Power 3.1 (Faul, Erdfelder, Buchner & Lang, 2009). The analysis was based on data from a previous published study (Author et al., 2019) (N=20), comparing decrease in mu and beta band power for Nogo vs Go trials. The effect size (ES) in this study was .8, considered to be large using Cohen's (1988) criteria. However, given that large effect-sizes are often over-estimated and rare, we selected a medium effect-size (.5) for our power analysis using a Cohen's d. With an alpha = .05 and power = 0.80, the projected sample size needed is approximately N = 34 for this simplest within group comparison

(Hypothesis 1). Thus, our proposed sample size of 40 will be adequate for the three within-subjects comparisons (Hypotheses 1, 3 and 5) as they either concern a decrease in power in the mu/beta bands or the N400 effect, which has been shown to require a smaller sample than 34 (Fields & Kuperberg, 2019), and should also allow for attrition. We also expect that this sample size should be adequate for our other hypotheses (2, 4 and 6), although, given that these hypotheses require between-subjects comparisons there is a chance that we will not reach 0.80 power.

3.3. *Participants.* Eighty (40 per group) right-handed French native speakers (aged 20–26) with no previous knowledge of Serbian or related languages will participate in the study. Participants will be right-handed volunteers from the student population of the Aix-Marseille Université, having no history of neurological insult. All participants will give their written informed consent prior to the experiment. Participants will receive 40 euros for their participation.

3.4. *Stimuli.* Auditory stimuli consist of 12 imperative transitive verbs in Serbian that are not transparent with their translation equivalents in French, Spanish, Italian, Portuguese, German or English. Serbian is a South Slavic language that is linguistically distant from both Romance and Germanic languages such that transparency poses little threat. Auditory verbs will be recorded in a professional sound booth and produced by two trained female native speakers. The verbs denote actions that can be performed using one's hand and arm, and were previously validated in a VR environment (Author et al., 2019): /'gurni/[push], /zɑ'grebi:/ [scratch], /'pu:sti:/ [drop], /'batsi:/ [throw], /o'krɛni:/ [pivot], /'prɛmesti:/ [move], /'ku:tsni:/ [tap], /'uxvati:/ [catch], / podigni:/ [lift], /'lu:pi:/ [hit], /o'bori:/ [tip over], / pro'tresi:/ [shake].

Verbs will be recorded by two native female speakers of Serbian. Half of the participants in each group (Specific action vs. Pointing) will hear Speaker 1 during training and Speaker 2 during EEG testing, and the other half the opposite assignment of speakers. A set of 12 filler verbs denoting different actions will be recorded for the passive listening EEG task. Visual stimuli for learning will consist of an office environment containing a 3D 10-point star polygon and a CRT screen (*Figure 1.*)

Figure 1.

Animations of hand and arm movements corresponding to the training verbs, performed on the 3D 10-point star polygon, will be recorded. These animations will be used in both learning conditions to teach participants the movements that correspond to the verbs. They will also be used for the match-mismatch task pre and post-training.

3.5. Learning apparatus. An Oculus VR headset and controller will be used for training purposes. The Oculus headset visually immerses participants by presenting them with a 360-degree visual scene and 3D virtual objects. The controller allows participants to manipulate objects while motion capture is recorded online.

3.5.1. Software. During pre and post-tests, StimPres (Tufts University) will be used for stimulus presentation on a desktop computer and a 64-channel Biosemi system (Actiview) will be used for acquisition. UNITY software will control virtual object presentation during learning.

3.6. General Procedure. The experiment will take place over two days. On the first day, participants will undergo EEG and behavioral pre-tests followed by a VR learning session. On the second day, they will take part in a second VR learning session with the same materials as day 1, followed by a EEG and behavioral post-tests.

3.6.1. Learning procedure. Participants will be comfortably seated at a desk wearing a VR Oculus headset and holding a controller. Participants in both the Specific action and the Pointing conditions will be presented with an auditory verb and requested to overtly repeat the verb prior to observing an action on the virtual CRT screen within the VR environment. Following this, a virtual object will appear on the virtual desk. The Specific action group will manipulate the object, performing the action observed on the virtual CRT screen (*Figure 2.*). The Pointing group will point to the object.

Figure 2.

3.6.2. EEG procedure. EEG will be recorded during both pre and post-tests. Participants will be comfortably seated at a desk situated 60 cm away from a computer screen in an electrically shielded sound-attenuated booth.

3.6.2.1. Passive listening task. During the first task participants will be asked to listen to the list of verbs passively, with no associated task. They will hear the 12 verbs used for learning and 12 filler verbs, twice. A trial will begin with an ocular fixation cross displayed in the center of the computer monitor for 200 msec prior to and for the duration of the auditory

word, which will be presented via electrically shielded speakers. A visual “blink” prompt will be displayed immediately thereafter for 2 seconds. The experimental session will last roughly 10 minutes.

3.6.2.2. Match-Mismatch task. During the match-mismatch task the auditory verbs used in learning will be preceded by either the compatible (match) or an incompatible (mismatch) animation. A question mark will appear directly following the auditory verb. Participants will be required to answer yes or no on a response box. A visual “blink” prompt will be displayed immediately thereafter for 2 seconds. The experimental session will last roughly 25 minutes, including one break.

3.6.3. Behavioral procedure. Behavioral responses and response times will be recorded during the match-mismatch task. In an exploratory manner, word-retention will be tested behaviorally in two tasks, after each training session. In the first task, participants will be asked to name the actions depicted in the animations using the label in Serbian. In the second task, they will be asked to give the French translation for each auditory verb.

3.7. EEG data acquisition. During pre and post-tests, EEG activity will be recorded continuously from 64 scalp electrodes located at left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas by means of a 64-channel electrode cap mounted with silver-chloride active electrodes (BioSemi Active Two system AD box). During acquisition, the offset of the electrodes will be maintained within the $-/+20\text{mV}$ range, in line with common practice using active electrode set-ups and data will be sampled online at 512 Hz. Blinks and vertical eye movements will be monitored via an electrode placed under the right eye and horizontal eye movements will be monitored via an electrode placed at the outer

canthus of the left eye. One electrode will be placed over each mastoid. EEG will be recorded continuously during the experiment and periods spanning from -100 pre-stimulus onset to 1100 msec post-stimulus onset will be used post-recording for analyses.

3.8. EEG data processing. EEG data will be bandpass filtered between 0.4 and 40 Hz using a 1408-order FIR filter windowed (Kaiser) sinc filter. The filtered data will be re-referenced offline to the average of the two mastoids. To detect noisy electrodes, we will apply several approaches to take into account different noise sources such as muscle artifacts, electrode pops, ocular movements etc. First, based on the continuous data, we will identify those electrodes whose amplitudes exceed a pre-defined threshold of +/- 50mV and, for each electrode, the total above-threshold time will be calculated. In addition, we will determine those electrodes with extreme amplitudes by calculating the robust z-score, as described by Bigdely-Shamlo and colleagues (Bigdely-Shamlo, Mullen, Kothe, Su & Robbins, 2015). The robust z-score is calculated based on the median and the robust standard deviation ($z_{\text{robust}} = 0.17413 * \text{interquartile range}$) and those electrodes with a $z_{\text{robust}} > 5$ are marked as bad. We will also test the electrodes based on the noisiness criterion described by Bigdely-Shamlo and colleagues (Bigdely-Shamlo et al., 2015), which calculates the ratio of the power of high frequency signal components to the power of low frequency components. This will be complemented by visual examination of the power spectral density of each electrode to determine those with excessive low and high frequency activity or contaminated by line noise. We will perform a baseline correction using a 100 msec prestimulus period

Noisy electrodes marked for rejection will be removed. Before carrying out independent components analysis (ICA), to correct for ocular movements, sections of the EEG signal that are highly contaminated with noise will be removed from the dataset. ICA will be carried out on the continuous data of each participant. Principal component analysis

(PCA) will be applied prior to ICA computation to reduce the dimension of the data and speed up the ICA computation time. The number of PCA components will be estimated by calculating the explained variance of each principal component and conserving only those principal components explaining 99% of the variance. Those independent components corresponding to eye-blinks will be identified automatically and rejected. Once the ocular artifacts are corrected using ICA, the rejected electrodes will be interpolated using spherical spline interpolation. The data will be segmented and epochs will be visually inspected. Those contaminated by noise will be removed. The epoched data will then be divided into separate conditions for analysis. The number of trials per condition will be kept as uniform as possible across conditions. If the percentage of rejected trials exceeds 20% for any given participant, their data will be excluded.

For the time-frequency analysis, the continuous data will be filtered into the frequency bands of interest (mu (8-13 Hz) and beta (13-30 Hz)). Continuous rather than segmented data will be filtered to avoid the possibility of losing portions of each trial to distortions introduced by filtering. Bandpass filtering will be carried out using a 4th-order Butterworth filter. After segmentation of the continuous data, the instantaneous amplitude and phase of each narrow-band can be extracted from the analytic signal, calculated via a Hilbert transform.

3.9. EEG data analysis

3.9.1. ERPs. The ERP data will be modeled in linear mixed effect models for the mean voltage amplitudes in the established N400 window, between 300-600msec (Kutas & Federmeier, 2011; Tromp et al., 2018), time-locked to the onset of the verb. Analyses for the N400 component will be conducted on the data acquired at 35 electrodes, including 5 over midline (Fz, FCz Cz, CPz, Pz), and 30 lateral electrodes divided equally over the left (F1, F3,

F5, FC1, FC3, FC5, C1, C3, C5, CP1, CP3, CP5, P1, P3, P5) and right (F2, F4, F6, FC2, FC4, FC6, C2, C4, C6, CP2, CP4, CP6, P2, P4, P6) hemispheres. A generalized linear mixed effects model (glmer) including the fixed factors Group (Specific movement vs Pointing), Session (Pre vs Post), Condition (Match vs Mismatch), and ROI (Midline, Left and Right lateral electrodes) and their interactions will be performed. Participant and Item will both include random intercepts. Both will include random slopes for Session provided the model converges. The fixed factors will be sum-coded to allow for the interpretation of main effects. In an exploratory analysis to determine where significant differences between Match and Mismatch conditions emerged, a permutation test with false discovery rate (FDR) correction will be carried out on all time points of the post stimulus interval for each electrode. A significant difference will only be considered ($q \leq .05$) if its duration exceeds 10msec (~5 consecutive time samples for a sampling frequency of 512Hz).

3.9.2. Event-related Spectral Perturbation (ERSP). The event-related spectral perturbation (ERSP) will be calculated on the data from the passive listening task, time-locked to the onset of the verb, using the FieldTrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2010). Based on our previous study, analyses will include 9 fronto-central electrodes associated with motor processes (FC3, FC4, C3, C4, CP3, CP4, FCz, Cz and CPz). To compute the ERSP, time-frequency decomposition will be effectuated at the single trial level for each participant and each condition (Pre-training, Post-training, Training verbs, Filler verbs) by applying complex Morlet wavelets over the 4 Hz to 35 Hz frequency band; the number of wavelet cycles will be adjusted as a function of frequency to ensure optimal time and frequency tradeoff for the frequency bands of interest (principally mu and beta bands). The trial-level power will be averaged to yield a single-subject average time-frequency map for each participant. The grand-average post-stimulus power will be expressed in terms of decibel

change relative to the pre-stimulus interval (-200 msec – 0 msec prior to verb onset). The frequency of the peak alpha for each individual will be extracted from the resting state as a control, if possible (Haegens, Cousijn, Wallis, Harrison & Nobre, 2014).

A generalized linear mixed effects model (glme) including the fixed factors Group (Specific movement vs Pointing), Session (Pre vs Post), Verb (Learned vs Filler) and ROI (Midline (FCz, Cz, CPz), Left (FC3, C3, CP3), Right lateral (FC4, C4, CP4)) as well as their interactions, will be performed. Participant and Item will both include random intercepts. Both will include random slopes for Session provided the model converges. The fixed factors will be sum-coded to allow for the interpretation of main effects. Individual peak alpha frequency will be applied as a co-variate in the model, if it can be extracted.

3.10. Behavioral data analysis. A generalized linear mixed effects model (glmer) will be used to examine accuracy in the match-mismatch task. Fixed effects factors will include Condition (Match vs Mismatch) and Session (Pre vs Post-training) and their interaction. Participant and Item will both include random intercepts. Fixed factors will be sum-coded. For the post-training session, fixed effects factors will include Condition (Match vs Mismatch, Group (Specific action vs Pointing) and their interaction. Participant and Item will both include random intercepts. Fixed factors will be sum-coded. As an exploratory measure, we will also perform permutation tests with FDR correction on all time points of the post-stimulus interval

3.11. Exploratory analyses. The success of movement during training will be used to predict a decrease in power in the mu and beta bands (passive listening experiment), while the N400 effect size and behavioral learning success (match-mismatch experiment) will be entered as covariates in the models outlined above. In like manner, both time-frequency results (mu and

beta band power decrease) and ERP results (N400 effect) will be used to predict behavior in an exploratory manner.

3.12. Timeline. We predict that the study will take 7 to 8 months to complete with the following breakdown: 4 to 5 months for data collection, 2 months for analysis and 1 month for the write-up.

References

- Alemanno, F., Houdayer, E., Cursi, M., Velikova, S., Tettamanti, M., Comi, G., Cappa, S. F., & Leocani, L. (2012). Action-related semantic content and negation polarity modulate motor areas during sentence reading : An event-related desynchronization study. *Brain Research, 1484*, 39-49.
- Aravena, P., Delevoeye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., & Nazir, T. (2012). Grip Force Reveals the Context Sensitivity of Language-Induced Motor Activity during “Action Words” Processing : Evidence from Sentential Negation. *PLOS ONE, 7*(12), e50287.
- Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., & Ibáñez, A. (2010). Applauding with closed hands : Neural signature of action-sentence compatibility effects. *PloS One, 5*(7), e11751.
- Atkinson, D. (2010). Extended, Embodied Cognition and Second Language Acquisition. *Applied Linguistics, 31*(5),
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs : Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia, 79*, 33-41.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660.
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology, 59*(1), 617-645.
- Berns, A., Gonzalez-Pardo, A., & Camacho, D. (2013). Game-Like Language Learning in 3-D Virtual Environments. *Computers & Education, 60*(1), 210-220.
- Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline : Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics, 9*, 16.

- Black, J. B., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and enhancing learning and motivation. In *Theoretical Foundations of Learning Environments* (D. Jonassen and S. Land, p. 2015).
- Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752-762.
- Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is Enough : N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. *Language Learning and Development: The Official Journal of the Society for Language Development*, *8*(3), 278-302.
- Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words : Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289-296.
- Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539-550.
- Burdea, G. C., & Coiffet, P. (2003). *Virtual Reality Technology*. John Wiley & Sons.
- Church, R. B., & Goldin-Meadow, S. (1986). The Mismatch Between Gesture and Speech as an Index of Transitional Knowledge. *Cognition*, *23*(1), 43–71.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058.
- de Nooijer, J. A., van Gog, T., Paas, F., & Zwaan, R. A. (2013). When Left Is Not Right : Handedness Effects on Learning Object-Manipulation Words Using Pictures With Left- or Right-Handed First-Person Perspectives. *Psychological Science*, *24*(12), 2515-2521.
- Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. [Image- and motor-processes in the retention of verbal materials.]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*(4), 511-533.
- Engelkamp, J., & Zimmer, H. D. (1984). Motor programme information as a separable memory unit. *Psychological Research*, *46*(3), 283-299.
- Fargier, R., Paulignan, Y., Boulenger, V., Monaghan, P., Reboul, A., & Nazir, T. A. (2012). Learning to associate novel words with motor actions : Language-induced motor activity following short training. *Cortex*, *48*(7), 888-899.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1 : Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

- Fields, E.C., & Kuperberg, G.R. (2019). Having your cake and eating it too : Flexibility and power with mass univariate statistics for ERP data. *Psychophysiology*, 57(2), e13468-e13468.
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Science*, 2(6), 595-607.
- Graziano, M., & Gullberg, M. (2018). When Speech Stops, Gesture Stops : Evidence From Developmental and Crosslinguistic Comparisons. *Frontiers in Psychology*, 9.
- Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, 92, 46–55.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, 41(2), 301-307.
- Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory* (pp. xix, 335). Wiley.
- Höller, Y., Bergmann, J., Kronbichler, M., Crone, J. S., Schmid, E. V., Thomschewski, A., Butz, K., Schütze, V., Höller, P., & Trinka, E. (2013). Real movement vs. Motor imagery in healthy subjects. *International Journal of Psychophysiology*, 87(1), 35-41.
- James, K. H., & Bose, P. (2011). Self-generated actions during learning objects and sounds create sensori-motor systems in the developing brain. *Cognition, brain, behavior : an interdisciplinary journal*, 15(4), 485-503.
- James, K. H., & Swain, S. N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental Science*, 14(4), 673-678.
- Johnson-Glenberg, M. C. (2017). Embodied Education in Mixed and Mediated Realities. In D. Liu, C. Dede, R. Huang, & J. Richards (Éds.), *Virtual, Augmented, and Mixed Realities in Education* (p. 193-217). Springer.
- Johnson-Glenberg, M. C. (2018). Immersive VR and Education : Embodied Design Principles That Include Gesture and Hand Controls. *Frontiers in Robotics and AI*, 5.
- Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality : How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications*, 2(1), 24.
- Johnson-Glenberg, M. C., Megowan-Romanowicz, C., Birchfield, D. A., & Savio-Ramos, C. (2016). Effects of Embodied Learning and Digital Platform on the Retention of Physics Content : Centripetal Force. *Frontiers in Psychology*, 7.

- Klepp, A., Dijk, H. van, Niccolai, V., Schnitzler, A., & Biermann-Ruben, K. (2019). Action verb processing specifically modulates motor behaviour and sensorimotor neuronal oscillations. *Scientific Reports*, 9(1), 1–11.
- Knoeferle, P. (2015). Language comprehension in rich non-linguistic contexts: Combining eye tracking and event-related brain potentials. In R. M. Willems (Ed), *Cognitive neuroscience of natural language use* (pp. 77-100). Cambridge: Cambridge University Press.
- Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical Experience Enhances Science Learning: *Psychological Science*.
- Krönke, K.-M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(9), 2553-2568.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Lan, Y.-J., Fang, S.-Y., Legault, J., & Li, P. (2015). Second language acquisition of Mandarin Chinese vocabulary : Context of learning effects. *Educational Technology Research and Development*, 63(5), 671-690.
- Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. *Languages*, 4(1), 13.
- Macedonia, M., & Knösche, T. R. (2011). Body in Mind : How Gestures Empower Foreign Language Learning. *Mind, Brain, and Education*, 5(4), 196-211.
- Macedonia, M., Muller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, 32, 982-998.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1), 59-70.
- Mahon, B. Z., & Caramazza, A. (2009). Concepts and Categories : A Cognitive Neuropsychological Perspective. *Annual review of psychology*, 60, 27-51.
- Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology: CB*, 25(4), 530-535.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, 7(7), 703–704.
- McNeill, D. (1992). *Hand and mind : What gestures reveal about thought* (p. xi, 416). University of Chicago Press.

- Mestres-Missé, A., Rodríguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. *Cerebral Cortex*, 17(8), 1858-1866.
- Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST : An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2), 229-240.
- Moreno, I., Vega, M. D., León, I., Bastiaansen, M. C. M., Lewis, A. G., & Magyari, L. (2015). Brain dynamics in the comprehension of action-related language : A time frequency analysis of mu rhythms. *NeuroImage*, 109(April), 50-62.
- Moskowitz, G. (1976). The Classroom Interaction of Outstanding Foreign Language Teachers. *Foreign Language Annals*, 9(2), 135-143.
- Neuper, C., & Pfurtscheller, G. (2001). Event-related dynamics of cortical rhythms : Frequency-specific features and functional correlates. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 43(1), 41-58.
- Niccolai, V., Klepp, A., Weessler, H., Hoogenboom, N., Schnitzler, A., & Biermann-Rubén, K. (2014). Grasping hand verbs : Oscillatory beta and alpha correlates of action-word processing. *PloS One*, 9(9), e108059.
- Nishitani, N., Schürmann, M., Amunts, K., & Hari, R. (2005). Broca's region : From action to language. *Physiology (Bethesda, Md.)*, 20, 60-69.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010, décembre 23). *FieldTrip : Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data* [Research Article]. Computational Intelligence and Neuroscience; Hindawi.
- Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition*, 9(1), 156-171.
- Peeters, D. (2019). Virtual reality : A game-changing method for the language sciences. *Psychonomic Bulletin & Review*, 26(3), 894-900.
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization : Basic principles. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 110(11), 1842-1857.
- Pineda, J. A. (2005). The functional significance of mu rhythms : Translating « seeing » and « hearing » into « doing ». *Brain Research. Brain Research Reviews*, 50(1), 57-68.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22(2), 253-336.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews. Neuroscience*, 6(7), 576-582.

- Quinn-Allen L. (1995). The effects of emblematic gestures on the development and access of mental representations of french expressions. *Mod. Lang. J.* 79, 521–529
- Shtyrov, Y., Nikulin, V. V., & Pulvermüller, F. (2010). Rapid Cortical Plasticity Underlying Novel Word Learning. *Journal of Neuroscience*, 30(50), 16864-16867.
- Soskey, L., Holcomb, P. J., & Midgley, K. J. (2016). Language effects in second-language learners : A longitudinal electrophysiological study of spanish classroom learning. *Brain Research*, 1646, 44-52.
- Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge : A bottom-up model of skill learning. *Cognitive Science*, 25(2), 203-244.
- Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8(2), 219-235.
- Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory : Inferring “how” from “where”. *Neuropsychologia*, 41(3), 280-292.
- Tomasello, R., Garagnani, M., Wennekers, T., & Pulvermüller, F. (2018). A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity. *Frontiers in Computational Neuroscience*, 12.
- Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, 50(2), 862-869.
- Vukovic, N., & Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: Evidence from rapid mu-rhythm desynchronisation. *NeuroImage*, 102, 695–703.
- Wakefield, E. M., & James, K. H. (2011). Effects of sensori-motor learning on melody processing across development. *Cognition, brain, behavior : an interdisciplinary journal*, 15(4), 505-534.
- Wilson, A. D., & Golonka, S. (2013). Embodied Cognition is Not What you Think it is. *Frontiers in Psychology*, 4.
- Yum, Y. N., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2014). An ERP study on initial second language vocabulary learning. *Psychophysiology*, 51(4), 364-373.
- Zappa, A., Bolger, D., Pergandi, J.-M., Mallet, P., Dubarry, A.-S., Mestre, D., & Frenck-Mestre, C. (2019). Motor resonance during linguistic processing as shown by EEG in a naturalistic VR environment. *Brain and Cognition*, 134, 44-57.

ADD TABLE 1 WITH STIMULI