

Embodiment of action-related language in the native and a late foreign language – An fMRI-study

E. Monaco^{a,1}, M. Mouthon^{a,1}, J. Britz^a, S. Sato^b, I. Stefanos-Yakoub^a, J.M. Annoni^a, L.B. Jost^{a,*}

^a Department of Neuroscience and Movement Science, Faculty of Science and Medicine, University of Fribourg, Switzerland

^b Department of Psychology, University of Fribourg, Switzerland

ARTICLE INFO

Keywords:

fMRI
Embodiment
Bilingualism
Semantic processing

ABSTRACT

Theories of embodied cognition postulate that language processing activates similar sensory-motor structures as when interacting with the environment. Only little is known about the neural substrate of embodiment in a foreign language (L2) as compared to the mother tongue (L1). In this fMRI study, we investigated embodiment of motor and non-motor action verbs in L1 and L2 including 31 late bilinguals. Half had German as L1 and French as L2, and the other half vice-versa. We collapsed across languages to avoid the confound between language and order of language acquisition. Region of interest analyses showed stronger activation in motor regions during L2 than during L1 processing, independently of the motor-relatedness of the verbs. Moreover, a stronger involvement of motor regions for motor-related as compared to non-motor-related verbs, similarly for L1 and L2, was found. Overall, the similarity between L1 and L2 embodiment seems to depend on individual and contextual factors.

1. Introduction

“Embodied” theories claim that cognition is grounded in the body’s interaction with the environment, and the involvement of the motor system in language processing has received particular attention. According to the embodied cognition account, sensory-motor systems are involved in several aspects of cognition, including language processing (Barsalou, 2008; Gallese & Lakoff, 2005; Glenberg & Kaschak, 2002; Jirak et al., 2010; Meteyard et al., 2012; Pulvermüller et al., 2005). This is supported by electrophysiological (Alemanno et al., 2012) and non-invasive brain stimulation studies (Buccino et al., 2005; Gianelli & Dalla Volta, 2015; Johari et al., 2022; Tremblay et al., 2012; Vukovic et al., 2017) showing that nouns, verbs and sentences referring to action activate cortical sensory-motor structures. Such action-language coupling has also been reported in several studies using functional magnetic resonance imaging (fMRI), revealing that processing action-related language stimuli partly recruits motor brain areas engaged in the execution of the corresponding movement (Aziz-Zadeh et al., 2006; Buccino et al., 2001; Boulenger et al., 2009; Hauk et al., 2004;

Kemmerer et al., 2008; Raposo et al., 2009; Tettamanti et al., 2005), also discussed as “semantic somatotopy” (Pulvermüller, 2005): reading action verbs recruits the sensory motor areas involved in the execution of that action (e.g., “kick” activates the motor cortex for the foot, “lick” for the mouth and “pick” for the hand). These areas primarily include the primary motor cortex (M1), the ventral and dorsal premotor cortex (vPM, dPM) as well as the supplementary and pre-supplementary motor areas (SMA, preSMA) (Courson et al., 2017; Courson & Tremblay, 2020; Jirak et al., 2010, see also Bidet-Ildei et al., 2020). However, it seems that effector specificity is more present in premotor areas (Jirak et al., 2010), and it has also been suggested that action language, at least in terms of word and sentence-level processing, is more similar to action observation than to action imagery and execution, and also more similar to action than movement processing (e.g., Courson & Tremblay, 2020).

While the above-mentioned studies investigated such motor involvement in the first language (L1), less is known about how and to what degree a foreign language (L2) is embodied. An early-acquired L1 is grounded in heteromodal and modality specific cortical regions based on the body’s rich experiences and interactions with the environment, as

* Corresponding author at: Laboratory for Cognitive and Neurological Sciences, Neurology Unit, Department of Neuroscience and Movement Science, Faculty of Science and Medicine, University of Fribourg, Ch. du Musée 5, CH-1700 Fribourg, Switzerland.

E-mail address: lea.jost@unifr.ch (L.B. Jost).

¹ Shared first authorship.

well as social learning. A late acquired L2 (after age 7) is typically acquired through formal instruction, translation and rote memory, and less through interaction with the environment (Adams, 2016; Dudschig et al., 2014; Li & Jeong, 2020; Macedonia et al., 2019; Yang & Li, 2012).

Numerous models have been applied to bilingual speakers to explain first and second language processing and how age, proficiency and immersion influence the interplay between the L1 and an L2, proposing either a shared, partially shared or common semantic storage (e.g. RHM; Kroll & Stewart, 1994; Inhibitory Control Model, Green, 1998; Bilingual Interactive Activation BIA + Model, Dijkstra & van Heuven, 2002; Conceptual Selection Model, CSM, Bloem & La Heij, 2003; Sense Model, Finkbeiner et al., 2004; Silverberg & Samuel, 2004; Multilink, Dijkstra et al., 2019). A more recent model postulates a shared semantic space between the different languages, while at the same time allowing for semantic representations of translation equivalents to be associated with different features (Blanco-Elorrieta & Caramazza, 2021). Independent of which of the above-mentioned models explain L2 lexico-semantic language processing best, two possibilities seem conceivable regarding the embodiment of a late L2. The first one is that a late acquired L2 is less embodied than the L1 as less semantic senses might be linked to L2 words and the semantic representations might be less “rich” in their features. The second possibility is that a late acquired L2 is similarly embodied as the L1 given that it shares its semantic space with the L1 or given that it requires mediation via the L1 translation equivalent.

Previous studies investigating the role of the sensorimotor system in L2 semantic processing using behavioral measures such as reaction times in word-image matching tasks, go/no-go tasks or sentence judgement tasks implying perceptual simulation concluded that L2 is embodied (Bergen et al., 2010; Buccino et al., 2017; Vukovic & Williams, 2014), but that this embodiment differs from the one in L1, being lower in magnitude or presenting itself in a different pattern (Ahlberg et al., 2018; Qian, 2016; Sheikh & Titone, 2016; for a review see Monaco et al., 2019). In line with the behavioral evidence, studies using temporally sensitive methods such as electroencephalography (EEG) and electromyography (EMG), all point towards a similar result, namely that L2 is embodied, but differently than L1 (Baumeister et al., 2017; Feroni, 2015; Vukovic & Williams, 2014; Xue et al., 2015). Similarly, a recent TMS study suggests differences in terms of timing and magnitude between L2 and L1 embodiment (Monaco et al., 2021): the motor cortex showed a larger amount of excitability at 275 ms vs. 125 ms post-stimulation for L2, a difference that was not present for L1. Moreover, a stronger modulation of reaction times in response to L1 motor verbs compared to L1 non motor verbs and compared to L2 motor verbs was found only at an early latency of lexico-semantic processing.

In contrast to these homogenous results, the few studies investigating motor-language coupling in L1 and L2 using fMRI yield a diverse picture. De Grauwe et al. (2014) showed that processing motor-verbs leads to stronger activation in motor and somatosensory brain areas for both L1 and L2 in highly proficient late Dutch-German bilinguals. However, another recent fMRI study using hand- or arm- related verbs embedded in literal, metaphorical and abstract L1/L2 sentences, showed overall larger activations of motor regions of interest (ROIs) during L2 compared to L1 processing in Mandarin-English bilinguals (Tian et al., 2020). The authors ruled out a higher degree of embodiment for L2 based on their result showing that abstract language involving no action-related meaning also induced a higher degree of motor activation in L2. Tian et al. (2020) concluded that this generally higher motor response is not necessarily linked to action-related semantic simulation, but rather to increased demands of cognitive control (i.e. memory retrieval, execution control, information integration) during the processing of a less automatized language. Another recent fMRI study (Zhang et al., 2020) showed that participants performing a semantic judgment task in L1 generally recruited a larger network (encompassing frontal and temporal areas) as compared to a group of other participants performing the same task in a late L2. Verbs presented in L1 and L2 elicited similar magnitudes of activation in the primary motor cortex and precentral

gyrus, albeit less connectivity between the middle temporal gyrus (MTG) as part of the semantic integration hub and sensorimotor regions for the processing of L2 compared to L1 verbs (Zhang et al., 2020).

In the present fMRI study, we investigated the neural substrate of embodiment of motor and non-motor related action verbs in L1 and a late L2. To this end, we examined the effects of motor-relatedness (motor vs. non-motor verbs) and language (L1 vs. L2) using a within-subject design including a total of 31 late bilinguals with varying levels of L2 proficiency.

We intended to 1) replicate previous findings and show stronger involvement of brain regions related to motor processing for motor-related as compared to non-motor-related verbs and 2) investigate whether this effect was stronger for L1 as compared to L2 verbs. We focused on M1, PM and SMA since previous studies reported verb-induced activation peaks in these areas (Kemmerer, 2015) and in order to be more selective. Moreover, we investigated if motor involvement during language processing would be associated with higher L2 proficiency. During the experimental task, subjects were asked to silently read hand-related motor and non-motor verbs in L1 and L2. To ensure that participants paid attention to the stimuli and accessed their meaning, once in a while they had to perform a semantic judgment task (i.e., subjects had to indicate whether the currently presented noun was semantically related to the preceding verb). This experimental task was chosen over a semantic or lexical decision task to avoid a motor response in each trial.

Importantly, we wanted to take previous results one step further by implementing several methodological aspects we believe to be of interest. A) We pooled together French L1/German L2 and German L1/French L2 to have each language equally represented in L1 and L2 and to compensate for language effects between L1 and L2. This first point is particularly important, as previous studies on embodiment in L1 and L2 confounded the order of language acquisition (L1 vs. L2) with the actual language. Collapsing the two languages avoids the wrong attribution of differences in the languages (L1 vs. L2) to differences in embodiment as a function of order of acquisition (L1 vs. late acquired L2). B) Valence, emotional experience, involvement of the body and motor-relatedness were assessed and matched for the L1 and L2 stimuli. C) Only those L2 stimuli that were known by the participants were included in the analysis to assure semantic processing, which in turn would lead to activation of motor information (e.g. Meteyard et al., 2012).

2. Materials and methods

2.1. Stimulus selection for the experimental task

Two surveys were conducted prior to the study to create the stimulus material. None of the subjects who participated to the surveys took part in the experiment.

The first survey aimed at assessing valence, emotional experience, involvement of the body and motor-relatedness of 361 verbs in French and in German. This list was randomly divided into two lists, so that each participant had to evaluate only one list of 181 verbs. A total of 31 native French speakers (age $M = 27$ years, Standard Deviation (SD) = 6 years, 10 males) and 41 native German speakers (age $M = 25$ years, SD = 5 years, 7 males) took part in this survey. Participants were asked to indicate for each verb presented in their L1: 1) to which degree that verb involved the body on a Likert scale from 1 (not at all) to 7 (completely); 2) if that verb represented a movement of the body or not; 3) which body part was associated with that verb (head, face/mouth, one hand/arm, both hands/arms, one foot/leg, both feet/legs, the entire body, none); 4) to which extent that verb evoked an emotional experience on a Likert scale from 1 (not at all) to 7 (completely); 5) the valence, indicating if that verb was rather positive or negative on a Likert scale from 1 (very negative) to 7 (very positive). Based on questions 2 and 3, words that were considered as too ambiguous to be classified as either motor-related or non-motor-related were excluded. Based on question 1, 4

and 5, words rated below or above $SD = 2$ from the mean for embodiment, emotional experience and valence were excluded. Since the embodiment ratings were generally higher in French than in German for the unambiguous verbs (French motor verbs: mean = 5.02, $SD = 0.74$; French non-motor verbs: mean = 1.52, $SD = 0.47$; German motor verbs: mean = 4.53, $SD = 0.76$; German non-motor verbs: mean = 1.73, $SD = 0.43$), these values were transformed into z-scores.

The second survey aimed at choosing verbs that would likely be familiar to the participants in their L2. This was important, as we had decided to include only verbs in the analyses of the fMRI data that are known to the participants, to make sure that the words were semantically processed. Subjects who filled out this survey had acquired the L2 after the age of 6 and judged their L2 written comprehension level > 3 on a 1-to-7 scale. Participants had to evaluate the total of 361 L2 verbs, indicating to which degree they were familiar with the verbs presented on a Likert scale from 1 (completely unfamiliar) to 7 (completely familiar). Data were collected from 36 French native speakers (age $M = 22$ years, $SD = 3$ years, 5 males) for the German verbs and from 26 German native speakers (age $M = 22$ years, $SD = 2$ years, 2 males) for the French verbs. L2 words with median and mode below 4 were excluded, resulting in 258 remaining French verbs and 226 remaining German verbs.

Taking into consideration the results of both surveys, a final list of 200 verbs was created, i.e. 50 verbs per condition (L1 motor, L1 non-motor, L2 motor, L2 non-motor). No cognates, defined as $> 80\%$ of grapheme overlap, were included. Translational equivalents were similarly distributed (16 French/German motor verbs and 13 French/German non-motor verbs plus 1 verb that was rated as motor in French but as non-motor in German). Verbs did not differ regarding frequency ($F = 0.120$, $p = 0.729$), number of letters ($F = 2.160$, $p = 0.143$), number of syllables ($F = 0.351$, $p = 0.555$), and valence ($F = 0.423$, $p = 0.516$) across the four lists. Embodiment ratings were not different across languages ($F = 1.974$, $p = 0.162$), but differed between motor-related and non-motor-related verbs ($F = 1141.208$, $p < 0.001$). The same holds true for emotional experience (language: $F = 1.775$, $p = 0.184$; motor-relatedness: $F = 50.479$, $p < 0.001$). This can be explained by the concreteness of the verbs: non-motor verbs often represent more abstract verbs, which are generally rated as more emotional than concrete words (Küenecke et al., 2015). Using Zipf-scores for the frequency evaluation allowed for direct comparisons of the subtitle-based frequency of French (database: Lexique 3.81) and German (database: Subtlex-DE) verbs (Brysaert et al., 2011).

2.2. Participants

An a-priori power analysis indicated that a 2X2 repeated measures ANOVA with the parameters: $\alpha = 0.05$, $f = 0.25$, $1-\beta = 0.9$, $r = 0.5$ correlation among repeated measures, $\epsilon = 1$, requires a total sample size of $n = 30$ (G*Power). Moreover, the two most recent fMRI studies investigating semantic representations in first and second languages report data of 30 and 29 participants (Zhang et al., 2020; Tian et al., 2020). As such, we collected data of 34 participants, out of which the data of three participants had to be excluded due to signal loss or excessive moving during fMRI. The final group consisted of 31 healthy late bilingual adults (13 men) aged 19–45 years (mean = 24.5 years, $SD = 6.0$ years), all recruited on the campus of a German-French bilingual university. 17 of the subjects were L1-French speakers who had learned German as L2, and 14 subjects were L1-German speakers who had learned French as an L2. Participants with French and German mother tongue were pooled in one group. Their L2 was acquired at a mean age of 9.8 (± 1.7) years, with varying proficiency levels (see section “2.3 Language evaluation” for details). None of the subjects reported any history of neurological or psychiatric disorders, and all participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal or corrected-to-normal vision. The study protocol was in accordance with the Declaration of Helsinki approved by the

local ethics committee. Participants gave written informed consent and were compensated with 25 CHF/hour for their participation.

2.3. Language evaluation

The Language Experience and Proficiency Questionnaire (LEAPQ; Marian et al., 2007) was used to assess self-reported L2 proficiency, age of acquisition (AoA) and past and present L2 exposure. In addition, each participant performed a sub-test from the computer-based DIALANG language diagnosis system (Zhang & Thompson, 2004), in order to obtain an estimation of L2 receptive vocabulary. After the fMRI experiment, participants also performed a translation recognition task for the L2-words previously presented during the experimental task (see Section 2.4.1 for details). Accuracy scores from this task gave us an additional L2-proficiency measure specific to the stimuli used in the experiment.

2.4. Tasks and procedure

Each participant took part in one session of approximately three hours, which took place at the local hospital (HFR Fribourg - Hôpital Cantonal). After filling out questionnaires assessing language background, handedness and general health, participants performed a silent reading task with an embedded semantic judgment task in the fMRI scanner. After completion of the experimental task, subjects were asked to complete a translation recognition task for the L2-words previously presented during the experimental task and another rating task to estimate the AoA of the L1 verbs presented during the experimental task.

2.4.1. Experimental task

During the fMRI experiment, participants performed a silent reading task, during which a total of 340 words were presented. In addition to the 200 verbs of interest (i.e., 50 L1 motor, 50 L1 non-motor, 50 L2 motor, 50 L2 non-motor), 50 filler nouns in L1 and 50 filler nouns in L2 randomly distributed among the verbs to distract participants from the fact that the experimental stimuli were verbs. Occasionally, participants were asked to perform a semantic judgment task on 40 supplementary filler items (verb + noun), which corresponds to 11.8 % of all items (Fig. 1). This embedded semantic judgment task was added for two reasons: a) to ensure that participants were paying attention to the stimuli and b) to promote semantic processing of the experimental verbs. Filler and semantic judgment trials were not included in the analyses of the fMRI data but were matched for frequency and length with the target verbs and used as regressors without interest.

We used a mixed block/event design (Petersen & Dubis, 2012). Language blocks were composed of motor, non-motors or fillers words. E-Prime 3.0 (Psychology Tools, Inc., Pittsburgh, PA, USA) was used for controlling stimulus presentation and response collection. Each word appeared on a screen in black with a white background for 800 ms and was followed by a pseudorandom inter-stimulus interval (ISI) between 2200 and 4260 ms which ensures the design efficiency for the event part of the design (jittering and sampling frequency). Words were presented in two separate lists, which allowed participants to have a short break, and to counterbalance the order of lists between participants. Each list consisted of 20 blocks alternating for language (L1/L2). In each block, there were 8 stimuli. In 10 out of 20 blocks, the blocks pseudo-randomly ended with an active semantic judgment task. In this case, the last word of the block was a filler verb followed by a filler noun. Simultaneously with the presentation of the noun, the question “associated?” was shown on the screen, indicating to the participant that he/she needed to indicate if the noun currently presented on the screen was semantically related to the verb previously shown. To give their response, participants had to press either the right or the left mouse button. Participants were allowed a maximum of 5.8 s to respond. Immediately after their response, a white cross was shown on the screen. The white cross duration depended on the response time of the participant in order to have a minimal duration (6 s) between the beginning of the active task

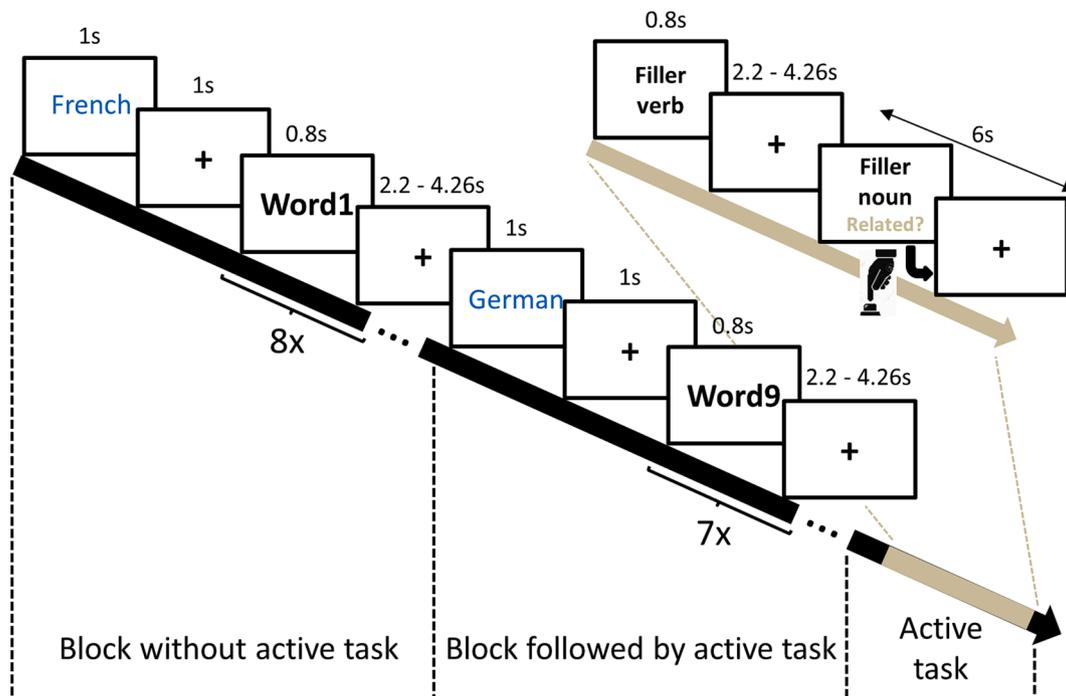


Fig. 1. Schematization of the silent reading task with the embedded semantic judgment task (active task).

and the start of the next block (Fig. 1). Each block's duration was around 30 s and was designed to preserve the signal at task-related frequencies from the noise reduction high-pass filter (see below). All stimuli within a block were of the same language (L1 or L2) and the language context alternated between L1 and L2 blocks to avoid switching between languages within a block. Motor-relatedness condition (motor or non-motor) was randomized within blocks as events. In each experimental run, two rest periods of 25 s were included in the middle and at the end. The total stimulus presentation time lasted approximately 25 min.

To familiarize participants with the task, they trained on 32 items prior to the fMRI experiment. These items were not part of the stimulus material of the experimental task.

2.4.2. Translation recognition task for the L2 words

After fMRI data acquisition, each participant performed a translation recognition task for the L2 -verbs presented during the silent reading task to evaluate the knowledge of the L2 words. Each L2 -verb appeared on a screen with four choices of potential translations and the option "I don't know" to avoid guessing the correct answer. Participants were asked to either choose the correct translation or to indicate that they did not know the word. Only the verbs for which participants chose the correct translation were included for fMRI analyses. The purpose was to make sure that only verbs that were known to the participants – that could be semantically processed – would be included in the analyses. Moreover, accuracy scores from this task were used as an additional L2 -proficiency measure specific to the stimuli used in the experiment.

2.4.3. AoA rating task for the L1 verbs

After fMRI data acquisition and the translation recognition task, participants estimated the age at which they had acquired the meaning of the L1 verbs. Each L1 -verb appeared on a screen with seven potential age range answers, 0–2 years old, 3–4 years old, 5–6, 7–8, 9–10, 11–12, 13 and older. Each range was then converted to a seven-point scale, with lower ages corresponding to lower points on the scale, following previous studies (e.g., Bird et al., 2001).

2.5. Data acquisition, preprocessing and analyses

2.5.1. MRI data acquisition

Data were acquired with a 3 T MRI scanner (Discovery MR750; GE Healthcare, Waukesha, Wisconsin) equipped with a 32-channel standard head coil. To reduce movements, the head of the participants was maintained by a sound-attenuating memory foam. The task was displayed on an LCD screen at 60 Hz, with a 1920 × 1080 resolution (32" NNL LCD monitor, NordicNeuroLab, Bergen, Norway). Participants viewed the screen positioned at their feet through a mobile mirror system. Words subtended a horizontal visual angle between 1.86° to 6.75° and a vertical angle of 0.51°. A high-resolution T1-weighted anatomical scan was recorded in the coronal plane with 270 slices, and a voxel size of 0.86 × 0.86 × 1 mm (acquisition parameters: matrix size: 256 × 256, Field Of View (FOV) = 220 mm, TR = 7.3 ms, TE = 2.8 ms, Flip angle = 9°, Prep Time = 900 ms, parallel imaging acceleration factor (PIAF) = 1.5, intensity correction: PURE). Functional T2*-weighted echo planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast were acquired with 417 dynamic volumes for the experimental silent reading task and 115 dynamic volumes for the motor-localizer task (not discussed in the present paper) with axial contiguous ascending acquisitions (voxel size: 2.3 × 2.3 × 3 mm, inter slice spacing (ISS) = 0.5 mm, acquired matrix size: 96 × 96, FOV = 220 mm, number of slices: 36, TR = 2000 ms, TE = 30 ms, Flip angle = 85°, PIAF = 2). Each scan run started with 8 s of dummy scans to assure a steady-state magnetization of the tissues. To correct the distortion of the static magnetic field during post-processing, a fieldmap was acquired after the experimental task. It required two FAST SPGR sequences with distinct Echo Time and the same space coverage as the functional EPI (TR = 50 ms, TE1 = 4.9 ms, TE2 = 7.3 ms, flip angle = 45°).

2.5.2. MRI data preprocessing

MRI data preprocessing and analyses were conducted with SPM12 (the Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London) running on MATLAB R2016b (MathWorks, <https://www.mathworks.com>, MA). Functional images were pre-processed with the following pipeline: setting the origin on the anterior commissure, slice timing, computation of the Voxel Displacement map

(VDM) (using the FieldMap2.1 toolbox, <https://www.fil.ion.ucl.ac.uk/spm/software>), spatial realignment and unwarping (using VDM previously created), normalization to the Montreal Neurological Institute (MNI) coordinate system with a voxel size of $3 \times 3 \times 3$ mm based on the Unified-Segmentation procedure of the co-registered T1-weighted anatomical image to fMRI images, smoothing with a Gaussian kernel of 8-mm full-width-at-half-maximum (FWHM). The ArtRepair toolbox was used on the preprocessed images only to detect volumes with fast motion (>0.5 mm/TR) without interpolating them. Two participants were excluded from the data analysis because $>10\%$ of the fMRI images suffered from these rapid movements during a run. In order to correct these motion artefacts, the realignment parameters of each experimental run were included as covariate of non-interest in individual models (see next paragraph).

The resulting preprocessed images were analyzed at the individual subject level using a general linear model (GLM). In the case of the silent reading task, each experimental trial was modeled as event of 800 ms of duration and convolved with the canonical hemodynamic response function (HRF). Filler trials, semantic judgment trials, instruction periods and trials including words that were not correctly identified by the participants during the translation recognition task were included as separate regressors of no interest. Movement parameters were introduced as covariate of non-interest and the rest period as model baseline. A high-pass filter with a $1/128$ Hz threshold was applied at time series from each voxel to remove low frequency noise and signal drifts. To correct for temporal correlations between neighboring voxels in the whole brain, an autoregressive function (AR(1)) was implemented.

2.5.3. Region of interest (ROI) statistical analysis

For the Region of Interest (ROI) analysis, the positive contrasts against the baseline, previously computed at the individual subject level, were sent to a random effect analysis. We performed a 2×2 repeated measures factorial design (second-level statistical analysis) with the within-subject factors *Motor-relatedness* (motor vs. non-motor verbs) and *Language* (L1 vs. L2). Both directions of each main effect and the interaction between the factors were studied with T-contrasts. This voxel-wise analysis was restricted by an explicit mask defined using the Human Motor Area Template (HMAT) atlas (Mayka et al., 2006). Based on previous studies investigating the involvement of motor areas during action-language processing (Courson et al., 2017; Courson & Tremblay, 2020; Jirak et al., 2010, see also Bidet-Ildei et al., 2020), we chose the following ROIs to be included: the primary motor cortex (M1), the ventral and dorsal part of the premotor cortex (PM), and the supplementary motor area (SMA), which includes also the pre-supplementary motor area. A statistical threshold of $p < 0.05$ family-wise error (FWE) corrected for multiple comparisons at the peak level was used. The MNI space was used to localize the clusters' maxima and figures are displayed in Neurological convention.

In a next step, we computed a linear mixed effects (LME) model, which, among the statistical advantages over classical approaches (e.g., Chen et al., 2013), allows to better take into account individual differences, playing an important role in second-language research (Linck & Cunnings, 2015). To this end, we used the extracted pondered activation mean (eigenvariate) values for each individual participant and for each M1, PM and SMA mask (HMAT atlas). The extracted individual contrast values per experimental conditions were used as dependent variable in the mixed modelling approach. The mixed models were computed in R (R Core Team, 2018, 2018) version 4.0.3, using the package lme4 (Bates et al., 2014), lmerTest with Satterthwaite method for denominator degrees of freedom (Kuznetsova et al., 2017), tidyverse (Wickham et al., 2019), RePsychLing (Bates et al., 2018) and the restricted maximum likelihood estimation method. We chose to model the data from the three ROIs separately, since both the literature (e.g., Courson & Tremblay, 2020; Jirak et al., 2010) and our data suggest that the three motor regions behave differently. The factors *Language*, *Motor-relatedness* and their interaction were modeled as fixed predictors. Random

slopes were required to account for the non-independence of repeated measures and to minimize type I error rates (Linck & Cunnings, 2015). In the maximal model possible by design (Barr et al., 2013), the random structure included intercepts for subjects and by-subject random slope for both the factors *Language* and *Motorrelatedness*. This model was however overparameterized and risked reducing the statistical power (Bates et al., 2018). Hence, we simplified the model (Barr, 2013; Brauer & Curtin, 2018; Matuschek et al., 2017) and we made sure that the procedure did not reduce the goodness of fit (differences in Akaike information criterion, AIC, according to Burnham & Anderson, 2004). The final random structure included by-subject random slopes for *Language* only. *Language*, and *Motorrelatedness* predictors were centered with contrast coding ($-0.5, 0.5$). Given that the three ROIs were analyzed separately, we corrected for multiple comparisons using the Bonferroni method. As such, only p-values of $p < 0.05/3 = 0.017$ were considered significant.

As previously discussed, our research question focuses on the difference of motor activation between L1 and L2, and not between two specific languages. However, we decided to further control if the results are influenced by the language per se. We added the *LanguageGroup* predictor (L1 French vs. L1 German) as a fixed factor centered with contrast coding ($-0.5, 0.5$) and tested if this element was redundant via the likelihood ratio test (Baayen et al., 2008). This seemed even more important given that the L1 German and L1 French participants differed regarding L2 AoA and L2 exposure. We also added the participant's age as a factor to test the goodness of the model.

In order to test if proficiency played a role in modulating the results, we also added these measures as fixed predictors, after centering their values to the group mean (Judd et al., 2017).

The data used for this manuscript is shared on the open repository 'Zenodo' and can be accessed via the following link: DOI: <https://doi.org/10.5281/zenodo.4761370>.

2.5.4. Whole brain statistical analysis

In order to better understand our results, we also performed a whole brain analysis, with a similar voxel-wise 2×2 repeated measures factorial design as in the ROI analysis. As we wanted to illustrate the general effect of the task, results are shown using a more sensitive statistical threshold than for the ROI analysis. This threshold was of $p < 0.001$ uncorrected at the voxel level and $p < 0.05$ family-wise error (FWE) corrected for the cluster size (min cluster size = 99 voxels). Anatomical locations of brain activations were checked with the neuromorphometrics probabilistic atlas provided by SPM12.²

3. Results

3.1. Behavioral results

Results of the translation recognition task showed that on average, participants correctly recognized 78.4 % of the L2 verbs previously presented during the experimental silent reading task. Moreover, the L2 proficiency scores of the DIALANG correlated ($r = 0.81, p < 0.001$) with the number of correct responses in the translation recognition task. Levene's test indicated unequal variances between the German and French participants for L2 AoA ($F = 32.522, p < 0.001$) and L2 exposure ($F = 38.555, p < 0.001$). L2 proficiency measures (i.e. DIALANG scores, number of correct responses in the translation recognition task as well as reading, writing and comprehension reported in the LEAP-Q) did not differ significantly between the L1 French and L1 German participants (all $p > 0.12$). See Table 1 for more detailed information on the L2 background and proficiency of the participants.

² MRI scans as originating from the OASIS project (<https://www.oasis-brains.org/>) and the labeled data as "provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>) under academic subscription".

Table 1
L2 background and proficiency measures of participants (n = 31).

Variable	L1 French participants		L1 German participants	
	Mean	SD	Mean	SD
Age of L2 acquisition (years)	9.24	1.1	10.50	2.1
Current L2 exposure (% of the day)	11.29	8.2	26.29	14.8
Subjective measure of proficiency: self-evaluation (0–10 scale)				
Speaking	5.41	1.7	6.36	1.7
Comprehension	6.59	1.7	6.57	1.9
Reading	6.59	1.7	6.93	1.2
Objective measure of proficiency L2 Vocabulary				
DIALANG (ranging from 0 to 1000)	411	250	552	231
Translation recognition (% correct)	75	15	82	13

According to the LEAP-Q, all L1 French participants reported French to be their most dominant language and all L1 German participants reported German to be their most dominant language. Since proficiency is a component of the dominance construct too, we also subtracted for each participant the average score across the three self-reported proficiency measures “speaking”, “comprehension” and “reading” reported in the LEAP-Q of L2 from L1. A Mann-Whitney-U-Test did not reveal a significant difference for this L1 vs. L2 proficiency score across the two groups (Mann-Whitney U = 104, $n_1 = 17$, $n_2 = 14$, $P > 0.05$ two-tailed). The DIALANG confirmed these self-report data: for L1 French participants, it showed an average score of 931 (SD 68) for L1 and 411 (SD 243) for L2. For L1 German participants, it showed a score of 916 (SD 69) for L1 and of 552 (SD 222) for L2.

Results of the semantic judgement task showed that on average, participants correctly responded to 96 % (SD = 0.7) of the L1 items and 82 % (SD = 1.3) of the L2 items, indicating that the participants were attentive to the task.

Results of the AoA rating task showed that motor verbs were acquired earlier than non-motor verbs (main effect of *Motor-relatedness*: $F = 331.47$, $p < 0.001$) and post hoc analysis using the Tukey criterion for significance confirmed this result for both languages ($p < 0.001$). Additionally, AoA of verbs correlated with their embodiment values (Spearman’s $r = -0.330$, $p < 0.001$), showing that early acquired verbs are also more embodied.

3.2. fMRI results

The ROI analysis showed a main effect of *Motor-relatedness* in the contrast “motor vs. non-motor verbs” with stronger activation for motor than non-motor verbs in the left dorsal premotor (dPM) and M1 cortex (Fig. 2 A). This cluster has a size of $K_E = 8$ voxels and its highest peak is located at $-33 -10 56$ with a Z-score of 4.22. The opposite contrast “non-motor vs. motor verbs” contrast did not show significant results. Moreover, it showed a main effect of *Language* in the contrast “L2 vs. L1” reflected in stronger activation for L2 verbs irrespective of their motor-relatedness primarily in the M1, vPM/dPM (Fig. 2 B1) and the SMA/pre-SMA (Fig. 2 B2). The highest peak for the cluster found in the M1, vPM/dPM ($K_E = 473$) is located at $-51 5 26$ with a Z-score of 7.02. The peak for the cluster found in the SMA/pre-SMA ($K_E = 123$) is located at $-6 17 50$ with a Z-score of 5.63. The opposite “L1 vs. L2” contrast did not show significant results. Importantly, the interaction between the factors *Language* (L1 vs. L2) and *Motor-relatedness* (motor vs. non-motor verbs) was not significant.

To investigate the relation between L2-proficiency and the strength of activation in the ROIs in response to L2 motor-related action verbs, we performed Pearson-correlations with the L2 proficiency score of the DIALANG as well as the number of correct responses in the translation recognition task for each ROI (M1, PM, SMA). None of the correlations was significant (both $p > 0.017$).

As shown in Table 2 and depicted in Fig. 3, the mixed effect model showed a main effect of *Language* in the PM ($\beta = 1.089$, $t = 6.499$, $p < 0.001$) and in the SMA ($\beta = 0.621$, $t = 4.348$, $p < 0.001$), at a corrected threshold of $p < 0.017$, reflecting stronger activation for L2 compared to L1.

Adding the *LanguageGroup* predictor as a fixed factor did not improve the final models ($p > 0.05$) and the results didn’t change. Adding the age of the participants as a fixed factor did not improve the final models ($p > 0.2$), nor did it change the results. Moreover, also adding the different measures of proficiency as fixed factors did not improve the final models ($p > 0.05$) and the results didn’t change.

To gain a better understanding of these findings, and particularly to see whether the increased activation of motor areas during L2 compared to L1 processing extends to other brain areas (regardless of the type of verbs), we report the results of the whole-brain analysis. There was a main effect of *Motor-relatedness* showing a higher activation for non-

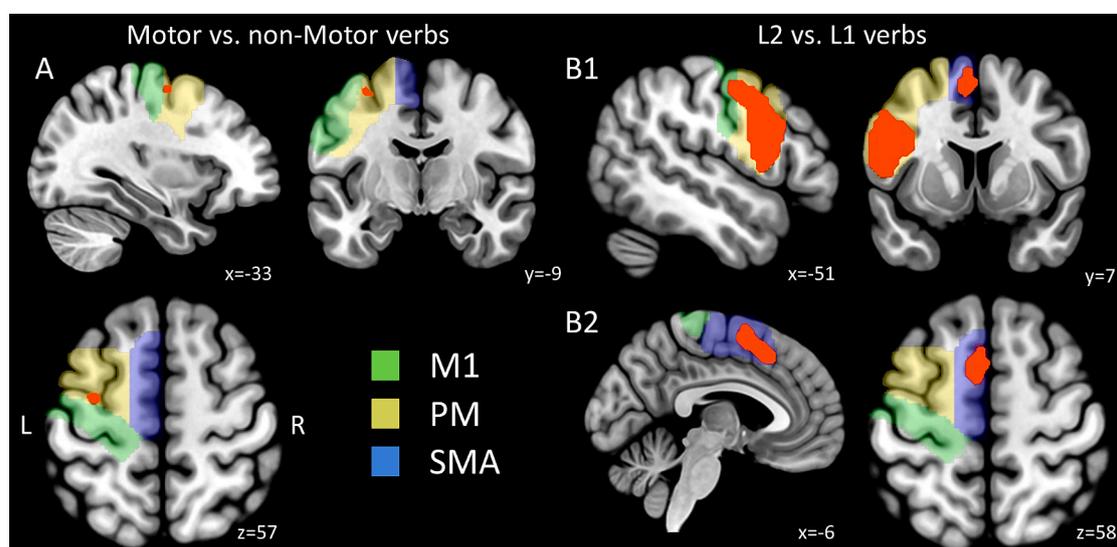


Fig. 2. Illustration of the significant activation blobs in orange with a voxelwise corrected $p_{FWE} < 0.05$ statistical threshold for the ROI analysis on the left hemisphere. A. Contrast motor vs. non-motor verb reading on MNI: $-33 -9 57$ position. B. Contrast foreign language (L2) vs. mother tongue (L1) verb reading. B1 is centered on MNI: $-51 7 26$ and B2 on MNI: $-6 17 58$. All results are overlain on the left primary motor cortex (green), left pre-motor cortex (yellow), left supplementary and pre-supplementary motor area (blue) used as region of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Linear mixed effect model for each region of interest. CI = confidence interval.

Predictors	Model M1			Model PM			Model SMA					
	Estimates	CI (95 %)	Statistic	P-Value	Estimates	CI (95 %)	Statistic	P-Value	Estimates	CI (95 %)	Statistic	P-Value
Intercept	0.491	[0.013,0.968]	2.015	0.053	1.074	[0.687,1.461]	5.438	<0.001	0.835	[0.538,1.131]	5.514	<0.001
Language	0.238	[-0.066,0.542]	1.534	0.129	1.089	[0.760,1.417]	6.499	<0.001	0.621	[0.341,0.901]	4.348	<0.001
Motorrelatedness	-0.206	[-0.503,0.090]	-1.363	0.176	-0.078	[-0.315,0.160]	-0.641	0.524	0.085	[-0.183,0.352]	0.62	0.537
Language:Motorrelatedness	-0.705	[-1.298,-0.112]	-2.329	0.022	0.396	[-0.078,0.870]	1.637	0.107	-0.063	[-0.597,0.472]	-0.229	0.819
N	31 Subjects				31 Subjects				31 Subjects			
Observations	124				124				124			
R model equation:	BOLD_ROI2 ~ Language * Motorrelatedness + (Language Subjects)											

motor vs. motor-related verbs in the left middle/superior temporal gyrus (Fig. 4 A). The cluster had a size of $K_E = 99$ voxels. Its highest peak (located at the coordinate $-48 -40 -1$) had a Z-Score of 4.76. The results of the main effect of *Language* are shown in Fig. 4 B-C and detailed in the Table 3. No significant result was found for the interaction between the two factors *Language* and *Motor-relatedness*.

4. Discussion

In this study we investigated whether action-verbs are embodied differently in L1 than in L2. In line with previous research (for a review see Kogan et al., 2020; Monaco et al., 2019), we show a main effect of motor-relatedness, with stronger activation of motor and premotor areas during the processing of motor as compared to non-motor action verbs. We also found a main effect of language in motor-related regions, with stronger activation of PM and SMA for L2 than L1 verbs. However, the two factors did not interact: neither motor cortex, nor premotor and supplementary motor cortex were recruited differently for motor vs. non-motor verbs in L1 and L2, which could have been expected based on behavioral (Ahlberg et al., 2018; Qian, 2016; Sheikh & Titone, 2016) and EEG/EMG studies (Baumeister et al., 2017; Foroni et al., 2015; Vukovic & Shtyrov, 2014; Xue et al., 2015). Moreover, no association between the strength of activation during action language processing and L2 proficiency was found.

Our first result showing higher activation in motor-related ROIs for motor than non-motor verbs, similar for L1 and L2, corroborates the fMRI study of De Grauwe et al. (2014) and Zhang et al. (2020). This result confirms an involvement of motor systems in semantics, and it indicates that L2 representations are similarly rich as L1 representations, as also postulated by De Grauwe et al. (2014). L1 and L2 semantic representations can be (partly) shared and stored in a common system, as postulated by several long-standing and also more recent models of bilingual language representations (Blanco-Elorrieta & Caramazza, 2021; Dijkstra & van Heuven, 2002; Finkbeiner et al., 2004; Kroll & Stewart, 1994; Van Hell & De Groot, 1998). Moreover, since the concept of a concrete action is usually acquired through interaction with the environment, the content is likely the same for both languages (e.g. Van Hell & De Groot, 1998). Nevertheless, it should be noted that the effect of *Motor-relatedness* in the present study was only found in the voxel-based analysis but not in the LME models. This could likely be due to the high interindividual variability of bilingual subjects. The whole brain analysis showed a *Motor-relatedness* effect as well, but only as higher activation for the non-motor vs. motor related verbs in the middle/superior temporal gyrus. According to Binder (2007), both concrete and abstract word processing involve the left lateral temporal cortex. However, when the semantic task is more challenging, the left lateral temporal cortex is more activated by abstract concepts, to which the non-motor verbs used in the present study can be compared.

Results from empirical studies suggesting that L2 is differently or less embodied than L1 appear at first glance to contradict our results too and the above-mentioned models of bilingual language representation that assume a common semantic store for both languages. However, this is not necessarily the case: it could also be that an L2 word points to the same semantic features as its L1 translation equivalent, yet the degree to which these features are activated is less salient or strong. Moreover, also the most recent model of multilingual language processing postulating a shared semantic system across languages (Blanco-Elorrieta & Caramazza, 2021), allows for semantic representations of L1 and L2 words to be associated with distinct features, therefore leading to differences in embodiment.

An important methodological difference between previous studies and the current one is that previous studies collapsed across L2 words irrespective of whether they were known to the subjects or not, therefore including also words that were not necessarily semantically processed. By pre-testing the stimulus material such to include words that are likely to be known by many participants and only including those words that

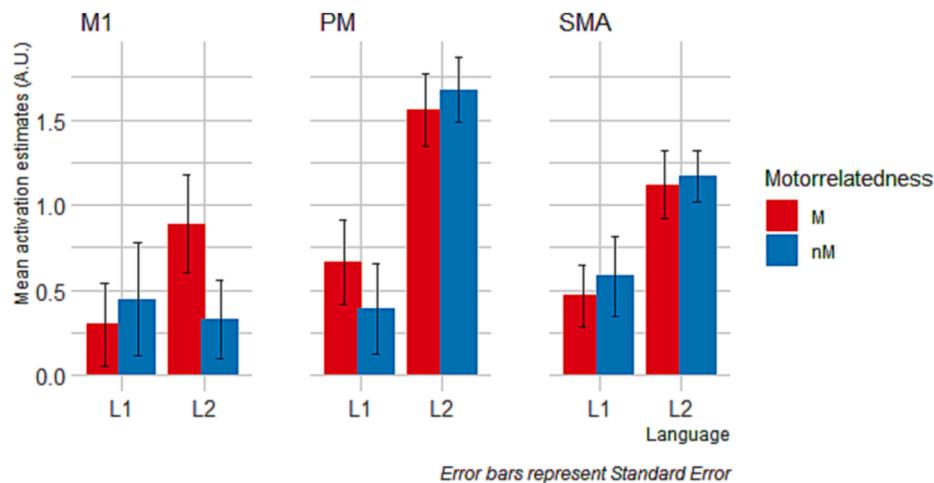


Fig. 3. Illustration of mean activation estimates across the experimental condition and ROIs. A.U. = arbitrary unit.

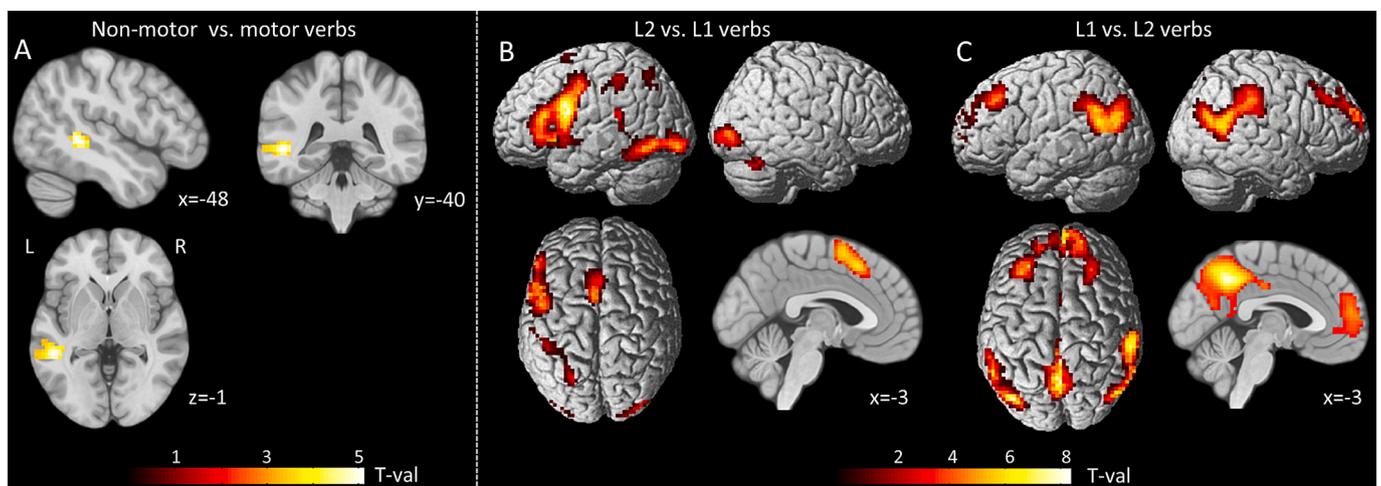


Fig. 4. Whole brain results for the 2×2 factorial design. **A.** The T-contrast non-motor vs. motor verbs shows a significant difference in the left middle/superior temporal gyrus. The opposite contrast does not show any significant result at the whole brain level. **B. and C.** Language factor with T-contrast L2 vs. L1 verb reading (**B**) and T-contrast L1 vs. L2 verb reading (**C**). In both cases, the three first images show a 3D-projection of fMRI results on brain surface (render) and the fourth image is a section to illustrate brain activations in brain center. More details about activity location can be found in Table 3. Results are shown with a statistical threshold of $p < 0.001$ uncorrected at the voxel level and $pFWE < 0.05$ corrected at the cluster level (min cluster size = 99 voxels). Color maps represent the T-value of the comparison. Coordinates (x, y, z) are in MNI space and results are displayed in neurological convention (left hemisphere on the left side[L]).

the individual participants knew, we assume that the stimuli were understood and lexically and semantically processed by the participants. This also explains why we did not find a relation between the strength of activation in the ROIs in response to L2 motor-related action verbs and L2 proficiency. As such, the measure of L2 proficiency predicts well how many words were recognized during the translation recognition task, but is not reflected in the results depicting the brains' response to the known words. Another possibility is that L2 embodiment is mediated by L1 embodiment because of a language-nonselective lexical access, as for example suggested by the Bilingual Interactive Activation Plus model (Dijkstra & van Heuven, 2002) or because of L1 transfer in meaning (Pavlenko, 2009).

Our second result, showing stronger activation for L2 as compared to L1 verbs in motor-related regions, was consistent across both analyses (voxel based, and LME models). Although this result might be interpreted as stronger embodiment for L2 as compared to L1 at first glance, it is important to note that this effect was independent of whether the verbs were motor related or not. Considering that motor activation is not always category specific, given its activation in pseudoword processing and abstract meanings (e.g., Aravena et al., 2014; Dreyer &

Pulvermüller, 2018; Tomasino et al., 2010), it is plausible that the network differences between L2 and L1 processing do not stem from a single category of verbs. Other studies have also found stronger activation of motor-related areas during L2 compared to L1 processing that was not specifically related to the processing of motor-related language information (Monaco et al., 2021; Rüschemeyer et al., 2006; Tian et al., 2020). This is also in line with a recently published review paper and meta-analysis by Sulpizio et al. (2020) investigating which brain areas are involved in bilingual processing and in the control of different types of linguistic knowledge such as for example lexico-semantics. Results showed stronger activation in late bilinguals for L2 compared to L1 in the precentral gyrus, despite stimulus material that was not motor-related, therefore suggesting an alternative role of motor areas during language processing other than simulation of the content of the meaning. For example, Tian et al. (2020) interpreted their result as reflecting increased demands for cognitive control, such as memory retrieval, execution control and information integration in a less proficient language. The stimuli used in the study by Tian et al. (2020) were verb phrases such as “catch the ball” instead of single verbs as were used in the current study. However, the meta-analytic study by Liu and Cao

Table 3

Whole brain activation details for the main effect of *Language*. Anatomical locations were retrieved from the Neuromorphometric atlas from SPM12.

Cluster coverage	Coordinates of the highest peak [MNI] (x y z)			Z-max	Cluster size
A. L2 vs. L1					
Left inferior occipital/temporal gyrus, fusiform gyrus	-45	-61	-10	7.64	616
Left precentral gyrus, inferior/middle frontal, anterior insula, putamen, caudate	-51	5	26	7.58	1475
Bilateral supplementary motor area (with a left dominance)	-6	17	50	6.26	253
Right cerebellum	27	-64	-25	6	145
Right inferior occipital, fusiform gyrus	21	-94	-4	5.06	214
Left superior parietal lobule, supramarginal gyrus, angular gyrus	-24	-64	44	4.98	252
B. L1 vs. L2					
Bilateral precuneus, posterior/middle cingulate gyrus	-6	-52	41	7.15	1873
Left angular gyrus, middle occipital gyrus, supramarginal, middle/superior temporal gyrus	-42	-79	26	6.16	707
Right angular gyrus, middle occipital gyrus, supramarginal, middle temporal gyrus	48	-73	26	5.89	764
Left fusiform gyrus, lingual gyrus, parahippocampal gyrus	-30	-37	-16	5.3	122
Bilateral medial frontal gyrus, middle/superior frontal gyrus, anterior cingulate gyrus	-36	32	41	5.21	1294

(2016) including a wide range of tasks and stimuli also showed additional involvement of brain regions during L2 as compared to L1 processing. In this study, Liu and Cao (2016) investigated which factors play a role in the similarities and differences between L1 and L2 networks and found that several regions (including the left frontal and precentral gyrus and the left insula) were more involved in L2 than in L1 processing. Interestingly, they also found that more additional regions were involved during L2 processing in late as compared to early bilinguals, suggesting higher cognitive demands for late than early bilinguals during L2 processing. Following this interpretation, the higher demand for more cognitive resources for L2 processing would call for the support of motor areas, with a negative correlation between L2 proficiency and motor involvement. The lack of correlation with proficiency in the present study, likely due to the fact that the participants' general L2 proficiency was not reflected in the results as only known words were included in the analyses, does not support this hypothesis. Future studies could include fine measurements of task-related proficiency (e.g., processing speed) to better investigate Tian and colleagues' hypothesis.

To further explore whether the increased activation of motor areas during L2 compared to L1 processing extends to other brain areas, a whole brain analysis was conducted. This analysis showed partial activation of the control system, namely left caudate, for the contrast L2 vs. L1. The left caudate has been shown to play a role in the control of verbal interference (e.g., Abutalebi et al., 2008; Ali et al., 2010; Branzi et al., 2016). However, no difference in the classical language control areas such as the anterior cingulate cortex (e.g., Green & Abutalebi, 2013) was found. This is not very surprising, since most of the studies on bilingual language control have focused on the control network involved while switching from one language to another. The code switching, however, was not the focus of our task. For the contrast L1 vs. L2, we found an increased activation in the posterior cingulate cortex, bilateral precuneus, left angular gyrus and medial prefrontal cortex, suggesting stronger implication of the Default Mode Network (e.g., Buckner et al., 2008). We suggest that the L2-L1 differences may reflect both control activation in L2, as well as resting state activation within the reading network, shown to be correlated to reading abilities in L1 and L2 (M.

Zhang et al., 2014). Another potential explanation for the stronger implication of motor areas in L2 can be found in the language statistics hypothesis by de Zubicaray et al. (2021), according to which we could speculate that L2 and L1 show different involvement of the motor cortex not because of a language-specific motor simulation of the semantic content, but because of differences in the cognitive demands when processing probabilistic *ortho*-phonological regularities. Another potential explanation for the stronger implication of motor areas in L2 can be found in the language statistics hypothesis by de Zubicaray et al. (2021), according to which we could speculate that L2 and L1 show different involvement of the motor cortex not because of a language-specific motor simulation of the semantic content, but because of differences in the cognitive demands when processing probabilistic *ortho*-phonological regularities. Accordingly, since L2 orthographic forms usually occur less frequently than L1 ones, they are also more demanding. Such cost is shown in the whole brain analysis, where L2 words activate the left fusiform gyrus to a larger extent than L1 words. This is in line with previous studies that have linked a higher activation of this area with an increased processing demand for unfamiliar word forms as well (Binder et al., 2006; de Zubicaray et al., 2021). It is likely then that perceptual information such as *ortho*-phonological occurrences, and not only semantic information, plays a role in L2 and L1 processing in the motor areas.

Considering the recent literature, it seems important that future studies explore not only the differences in activation between L2 and L1 processing in motor and language areas, but also investigate more closely the different connectivity patterns between key-language, sensory-motor and semantic areas) during L1 and L2 processing (Zhang et al., 2020) with the use of functional and structural connectivity analysis. Further confirmation of X. Zhang et al.'s finding of an increased L2 network activation without strong connections within the network may favor the view of the motor system as having a supporting role during L2 processing. Moreover, recent data also suggest that the language effect in motor areas may differ in terms of temporal processing and thus tapping into different stages of processing, which fMRI cannot resolve (Monaco et al., 2021).

In light of the diverse results across behavioral (Ahlberg et al., 2018; Qian, 2016; Sheikh & Titone, 2016; for a review see Monaco et al., 2019), EEG/EMG (Baumeister et al., 2017; Foroni, 2015; Vukovic & Williams, 2014; Xue et al., 2015) and fMRI (De Grauwe et al., 2014; Tian et al., 2020; Zhang et al., 2020) studies, the question arises to what extent these results specifically reflect the degree of sensorimotor involvement during L1 and L2 processing, or simply mirror methodological aspects of the studies. Apart from the weak temporal resolution of fMRI in contrast to TMS and EEG, that might have hidden eventual differences in L1 and L2 processing of motor-related words, we believe that several other methodological aspects such as 1) language distance between L1 and L2, 2) the choice of within vs. between subject design, 3) the choice of action related language stimuli that could potentially be more or less embodied depending on the task and context, and that are related to different effectors (e.g., mouth, hand, legs) play a crucial role in the above-mentioned results. In the following, we will discuss these issues and how they might play a role in the results of the present study, but also in future studies investigating embodiment in L1 and L2.

Linguistic distance has previously been shown to modulate the involvement of different brain regions and therefore to play a role in bilingual language processing (e.g. Ghazi-Saidi & Ansaldo, 2017; Kim et al., 2016). As such, results of previous studies investigating embodiment in an L1 and L2 could have been influenced by the fact of comparing embodiment in L1 and L2 in two Germanic languages (Foroni, 2015; Vukovic & Shtyrov, 2014), in a Germanic and a Romance language (Baumeister et al., 2017; Sheikh & Titone, 2016) or even in languages from two more distant language families (Germanic & Sino-Tibetan) (Qian, 2016; Tian et al., 2020). This seems even more crucial in light of the first large-scale quantitative examinations of semantic structure across languages conducted by Thompson et al. (2018), who

investigated whether “all languages convey semantic knowledge in the same way” (p.1). This study revealed that common actions (e.g. “wash”) show only intermediate levels of semantic alignment, indicating that the meaning of words, even if seemingly straightforward, varies across languages. Several studies indicate that languages and cultures co-adapt to shape knowledge and semantic representations that can vary across populations (Ghandhari et al., 2020; Majid et al., 2015). In the present study, we can rule out that the languages themselves or the distance between L1 and L2 groups can account for the differences because we pooled together a Romance and a Germanic language to avoid the confound between language and order of acquisition. When language (or language family) and order of acquisition are confounded (e.g. L1: Chinese, L2: English), differences in motor activation for action verbs in L1 and L2 cannot be uniquely attributed to the motor content of the material, but they can likewise arise from differences in motor involvement between the two languages. In our design we contrasted differences in L1 vs. L2 collapsed across the languages of L1 (German vs. French) and we can hence argue that the differences in motor activation can be attributed to L1 vs. L2 and not to French or German. We acknowledge the fact that pooling two different groups of bilinguals constitutes a challenge, especially considering the unequal variances in L2 AoA and exposure. Nevertheless, all participants were recruited from the same bilingual environment and had a similar variance in the proficiency profile. Most importantly, the inter-subject variability was taken into account in the LME analysis, which further showed that being French or German did not influence the results.

A confound usually present in studies with bilinguals, is the differences in cultural context. The generally higher embodiment ratings in L1 French than in L1 German speakers in the pretest indicates that the judgment of body involvement indeed differs between languages and language families. Such an interpretation has previously been put forward by Ghandhari et al. (2020), proposing differences in the gesture-language system between Persian and Italian speaking participants accounting for differences in the relationship between language and action. When possible, future studies should rule out the potential crucial confound of cross-cultural differences in embodiment studies in L1 and L2. A similar difference in motor relatedness judgment of the L2 vs. L1 stimuli was reported e.g. by Tian et al. (2020). By adopting a within-subject design, differences in semantic representations that could potentially vary across populations could be reduced. The within-design also minimizes the impact of inter-individual sociolinguistic differences, which seems of primary importance considering the role of personal learning history and experience on embodiment (Bidet-Ildei et al., 2020; Pulvermüller, 2013).

Another methodological aspect that we would like to point out is the choice of stimuli. The above-mentioned differences in semantic knowledge across languages highlight the importance of making clear distinctions between motor and non-motor related word stimuli in each language separately. In the present study we conducted a pilot survey with L1 speakers of both languages that allowed us choose stimuli that could be attributed to either the (hand) motor or non-motor category in a more objective way. This seems even more important in light of the fact that the verbs were presented in isolation and not embedded in a sentence, which might otherwise have left room for activation of different semantic associations (e.g. Aravena et al., 2014). Moreover, the pilot study also allowed controlling for emotional experience and valence, variables that are often a confound between languages in studies investigating embodiment in L1 and L2 (e.g. Pavlenko, 2017; Sheikh & Titone, 2016).

We decided not to include imageability in the survey for practical and theoretical reasons. First, the survey was already lengthy, therefore we had to choose what would have been more relevant to our study while at the same time ensuring cooperation and data quality. Second, embodiment and imageability ratings are highly positively correlated (Sidhu et al., 2014). Embodiment ratings, though, seem to capture something more than imageability in visual verb recognition in Sidhu

et al. (2014), thus we decided to discard imageability and follow the authors' guidelines to collect embodiment ratings. Third, since the stimuli selected referred only to familiar words, the likelihood of having verbs with low imageability and higher embodiment should be very low.

The embodiment ratings collected during the survey also correlated with the AoA values related to the verbs, confirming that early acquired words are indeed considered to be more grounded in the body (Pexman, 2019). Following the stages of child development, verbs related to the mouth are indeed acquired first, followed by those related to the hands and arms, and finally by those not related to the body (Maouene et al., 2008).

To conclude, the present study revealed stronger involvement of motor, premotor and supplementary motor regions during L2 than during L1 processing, independently of the motor-relatedness of the verbs, indicating that these regions play a role in L2 processing independently of the motor features of the stimuli. Moreover, activation in motor regions was stronger for motor-related as compared to non-motor-related verbs, an effect that did not differ between L1 and L2. Finally, motor involvement was not stronger in higher proficient L2 bilinguals. Our study not only adds another part to our understanding of embodiment of language, but also takes up on key methodological aspects that seem to be particularly important in this line of research and that could account for the heterogeneity in the results reported in studies so far.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Swiss National Science Foundation (No. 325130_182594).

References

- Abutalebi, J., Annoni, J. M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., Lazeyras, F., Cappa, S. F., & Khateb, A. (2008). Language control and lexical competition in bilinguals: An event-related fMRI study. *Cerebral Cortex*, 18(7), 1496–1505. <https://doi.org/10.1093/cercor/bhm182>
- Adams, A. M. (2016). How language is embodied in bilinguals and children with specific language impairment. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01209>
- Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozzyk, J. V. (2018). Grounded cognition: Comparing Language × Space interactions in first language and second language. *Applied Psycholinguistics*, 39(2), 437–459. <https://doi.org/10.1017/S014271641700042X>
- 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. <https://doi.org/10.1016/j.brainres.2012.09.030>
- Ali, N., Green, D. W., Kherif, F., Devlin, J. T., & Price, C. J. (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience*, 22(10), 2369–2386. doi: 10.1162/jocn.2009.21352.
- Aravena, P., Courson, M., Frak, V., Cheylus, A., Paulignan, Y., Deprez, V., & Nazir, T. A. (2014). Action relevance in linguistic context drives word-induced motor activity. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00163>
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16(18), 1818–1823. <https://doi.org/10.1016/j.cub.2006.07.060>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00328>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3). <https://doi.org/10.1016/j.jml.2012.11.001>
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. <http://arxiv.org/abs/1406.5823>.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. <http://arxiv.org/abs/1506.04967>.
- Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., & Winkielman, P. (2017). Embodiment and emotional memory in first vs second language. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00394>
- Bergen, B., Lau, T.-T.-C., Narayan, S., Stojanovic, D., & Wheeler, K. (2010). Body part representations in verbal semantics. *Memory & Cognition*, 38(7), 969–981. <https://doi.org/10.3758/MC.38.7.969>
- Bidet-Ildi, C., Beauprez, S.-A., & Badets, A. (2020). A review of literature on the link between action observation and action language: Advancing a shared semantic theory. *New Ideas in Psychology*, 58, Article 100777. <https://doi.org/10.1016/j.newideapsych.2019.100777>
- Binder, J. R. (2007). Effects of word imageability on semantic access: Neuroimaging studies. In J. Hart, & M. A. Kraut (Eds.), *Neural basis of semantic memory* (pp. 149–181). Cambridge University Press. <https://doi.org/10.1017/CBO9780511544965.007>.
- Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage*, 33(2), 739–748. doi: 10.1016/j.neuroimage.2006.06.053.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1), 73–79. <https://doi.org/10.3758/BF03195349>
- Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, 213, Article 104625. <https://doi.org/10.1016/j.cognition.2021.104625>
- Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, 48(3), 468–488.
- Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral cortex*, 19(8), 1905–1914. <https://doi.org/10.1093/cercor/bhn217>
- Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex*, 26(6), 2367–2380. <https://doi.org/10.1093/cercor/bhw052>
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389–411. <https://doi.org/10.1037/met0000159>
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Böhle, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13(2), 400–404. <https://doi.org/10.1111/j.1460-9568.2001.01385.x>
- Buccino, G., Marino, B. F., Bulgarelli, C., & Mezzadri, M. (2017). Fluent speakers of a second language process graspable nouns expressed in L2 Like in their native language. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01306>
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24(3), 355–363. <https://doi.org/10.1016/j.cogbrainres.2005.02.020>
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38. <https://doi.org/10.1196/annals.1440.011>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S., & Cox, R. W. (2013). Linear mixed-effects modeling approach to fMRI group analysis. *NeuroImage*, 73, 176–190. <https://doi.org/10.1016/j.neuroimage.2013.01.047>
- Courson, M., Macoir, J., & Tremblay, P. (2017). Role of medial premotor areas in action language processing in relation to motor skills. *Cortex*, 95, 77–91. <https://doi.org/10.1016/j.cortex.2017.08.002>
- Courson, M., & Tremblay, P. (2020). Neural correlates of manual action language: Comparative review, ALE meta-analysis and ROI meta-analysis. *Neuroscience & Biobehavioral Reviews*, 116, 221–238. <https://doi.org/10.1016/j.neubiorev.2020.06.025>
- De Grauwe, S., Willems, R. M., Rueschemeyer, S.-A., Lemhöfer, K., & Schriefers, H. (2014). Embodied language in first- and second-language speakers: Neural correlates of processing motor verbs. *Neuropsychologia*, 56, 334–349. <https://doi.org/10.1016/j.neuropsychologia.2014.02.003>
- de Zubicaray, G. I., McMahon, K. L., & Arciuli, J. (2021). A sound explanation for motor cortex engagement during action word comprehension. *Journal of Cognitive Neuroscience*, 33(1), 129–145. https://doi.org/10.1162/jocn_a.01640
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(03). <https://doi.org/10.1017/S1366728902003012>
- Dijkstra, T. O. N., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 22(4), 657–679.
- Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, 100, 52–70. <https://doi.org/10.1016/j.cortex.2017.10.021>
- Dudschig, C., de la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain and Language*, 132, 14–21. <https://doi.org/10.1016/j.bandl.2014.02.002>
- Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, 51(1), 1–22. <https://doi.org/10.1016/j.jml.2004.01.004>
- Foroni, F. (2015). Do we embody second language? Evidence for ‘partial’ simulation during processing of a second language. *Brain and Cognition*, 99, 8–16. <https://doi.org/10.1016/j.bandc.2015.06.006>
- Gallese, V., & Lakoff, G. (2005). The Brain’s concepts: The role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3–4), 455–479. doi: 10.1080/02643290442000310.
- Ghandhari, M., Fini, C., Da Rold, F., & Borghi, A. M. (2020). Different kinds of embodied language: A comparison between Italian and Persian languages. *Brain and Cognition*, 142, Article 105581. <https://doi.org/10.1016/j.bandc.2020.105581>
- Ghazi-Saidi, L., & Ansaldo, A. I. (2017). The neural correlates of semantic and phonological transfer effects: Language distance matters. *Bilingualism: Language and Cognition*, 20(5), 1080–1094. <https://doi.org/10.1017/S136672891600064X>
- Gianelli, C., & Dalla Volta, R. (2015). Does listening to action-related sentences modulate the activity of the motor system? Replication of a combined TMS and behavioral study. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01511>
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565. <https://doi.org/10.3758/BF03196313>
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1(2), 67–81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301–307. [https://doi.org/10.1016/S0896-6273\(03\)00838-9](https://doi.org/10.1016/S0896-6273(03)00838-9)
- Jirak, D., Menz, M. B., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language – A short story on embodiment. *Consciousness and Cognition*, 19(3), 711–720. <https://doi.org/10.1016/j.concog.2010.06.020>
- Johari, K., Riccardi, N., Malyutina, S., Modi, M., & Desai, R. H. (2022). HD-tDCS of primary and higher-order motor cortex affects action word processing. *Frontiers in Human Neuroscience*, 16, Article 959455. <https://doi.org/10.3389/fnhum.2022.959455>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68(1), 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Kemmerer, D. (2015). Does the motor system contribute to the perception and understanding of actions? Reflections on Gregory Hickok’s The myth of mirror neurons: The real neuroscience of communication and cognition. *Language and Cognition*, 7(3), 450–475. <https://doi.org/10.1017/langcog.2014.36>
- Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, 107(1), 16–43. <https://doi.org/10.1016/j.bandl.2007.09.003>
- Kim, S. Y., Qi, T., Feng, X., Ding, G., Liu, L., & Cao, F. (2016). How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean–Chinese–English trilinguals. *NeuroImage*, 129, 25–39. <https://doi.org/10.1016/j.neuroimage.2015.11.068>
- Kogan, B., Muñoz, E., Ibáñez, A., & García, A. M. (2020). Too late to be grounded? Motor resonance for action words acquired after middle childhood. *Brain and Cognition*, 138, Article 105509. <https://doi.org/10.1016/j.bandc.2019.105509>
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149–174. <https://doi.org/10.1006/jmla.1994.1008>
- Künecke, J., Sommer, W., Schacht, A., & Palazova, M. (2015). Embodied simulation of emotional valence: Facial muscle responses to abstract and concrete words: Facial responses to abstract and concrete words. *Psychophysiology*, 52(12), 1590–1598. <https://doi.org/10.1111/psyp.12555>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(1), 1. doi: 10.18637/jss.v082.i13.
- Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *Npj Science of Learning*, 5(1), 1. doi: 10.1038/s41539-020-0068-7.
- Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185–207. <https://doi.org/10.1111/lang.12117>
- Liu, H., & Cao, F. (2016). L1 and L2 processing in the bilingual brain: A meta-analysis of neuroimaging studies. *Brain and language*, 159, 60–73. <https://doi.org/10.1016/j.bandl.2016.05.013>
- Macedonia, M., Repetto, C., Ischebeck, A., & Mueller, K. (2019). Depth of encoding through observed gestures in foreign language word learning. *Frontiers in Psychology*, 10, Article 33. <https://doi.org/10.3389/fpsyg.2019.00033>
- Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, 49, 1–18. <https://doi.org/10.1016/j.langsci.2014.11.002>

- Maouene, J., Hidaka, S., & Smith, L. B. (2008). Body parts and early-learned verbs. *Cognitive Science*, 32(7), 1200–1216. <https://doi.org/10.1080/03640210802019997>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067)
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: A meta-analysis. *NeuroImage*, 31(4), 1453–1474. <https://doi.org/10.1016/j.neuroimage.2006.02.004>
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7), 788–804. <https://doi.org/10.1016/j.cortex.2010.11.002>
- Monaco, E., Jost, L. B., Gygax, P. M., & Annoni, J.-M. (2019). Embodied semantics in a second language: Critical review and clinical implications. *Frontiers in Human Neuroscience*, 13, Article 110. <https://doi.org/10.3389/fnhum.2019.00110>
- Monaco, E., Jost, L. B., Lancheros, M., Harquel, S., Schmidlin, E., & Annoni, J.-M. (2021). First and second language at hand: A chronometric transcranial-magnetic stimulation study on semantic and motor resonance. *Journal of Cognitive Neuroscience*, 1–18. https://doi.org/10.1162/jocn_a.01736
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, 17, 74–78. <https://doi.org/10.1016/j.copsyc.2017.06.005>
- Pavlenko, A. (2009). 6. Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon* (pp. 125–160). Multilingual Matters. doi: 10.21832/9781847691262-008.
- Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage*, 62(2), 1177–1184. <https://doi.org/10.1016/j.neuroimage.2011.09.084>
- Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition and Neuroscience*, 34(10), 1274–1283. <https://doi.org/10.1080/23273798.2017.1303522>
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582. <https://doi.org/10.1038/nrn1706>
- Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, 127(1), 86–103. <https://doi.org/10.1016/j.bandl.2013.05.015>
- Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems: Functional links between motor and language systems. *European Journal of Neuroscience*, 21(3), 793–797. <https://doi.org/10.1111/j.1460-9568.2005.03900.x>
- Qian, W. (2016). Embodied cognition processing and representation of power words by second language learners with different proficiency levels. *Chinese Journal of Applied Linguistics*, 39(4). <https://doi.org/10.1515/cjal-2016-0030>
- R Core Team. (2018). *R: A language and environment for statistical computing* (3.5.1). R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, 47(2), 388–396. <https://doi.org/10.1016/j.neuropsychologia.2008.09.017>
- Rüschmeyer, S.-A., Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage*, 31(1), 354–365. <https://doi.org/10.1016/j.neuroimage.2005.11.047>
- Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion*, 30(3), 488–500. <https://doi.org/10.1080/02699931.2015.1018144>
- Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica*, 149, 32–39. <https://doi.org/10.1016/j.actpsy.2014.02.009>
- Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, 51(3), 381–398.
- Sulpizio, S., Del Maschio, N., Fedeli, D., & Abutaleb, J. (2020). Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 108, 834–853. <https://doi.org/10.1016/j.neubiorev.2019.12.014>
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17(2), 273–281. <https://doi.org/10.1162/0898929053124965>
- Thompson, B., Roberts, S., & Lupyan, G. (2018). *Quantifying semantic similarity across languages*.
- Tian, L., Chen, H., Zhao, W., Wu, J., Zhang, Q., De, A., Leppänen, P., Cong, F., & Parviainen, T. (2020). The role of motor system in action-related language comprehension in L1 and L2: An fMRI study. *Brain and Language*, 201, Article 104714. <https://doi.org/10.1016/j.bandl.2019.104714>
- Tomasino, B., Weiss, P. H., & Fink, G. R. (2010). To move or not to move: Imperatives modulate action-related verb processing in the motor system. *Neuroscience*, 169(1), 246–258. <https://doi.org/10.1016/j.neuroscience.2010.04.039>
- Tremblay, P., Sato, M., & Small, S. L. (2012). TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia*, 50(2), 319–326. <https://doi.org/10.1016/j.neuropsychologia.2011.12.002>
- Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193–211. <https://doi.org/10.1017/S1366728998000352>
- Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, 96, 222–229. <https://doi.org/10.1016/j.neuropsychologia.2017.01.025>
- Vukovic, N., & Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: Evidence from rapid mu-rhythm desynchronization. *NeuroImage*, 102, 695–703. <https://doi.org/10.1016/j.neuroimage.2014.08.039>
- Vukovic, N., & Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica*, 145, 98–103. <https://doi.org/10.1016/j.actpsy.2013.11.002>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. doi: 10.21105/joss.01686.
- Xue, J., Marmolejo-Ramos, F., & Pei, X. (2015). The linguistic context effects on the processing of body-object interaction words: An ERP study on second language learners. *Brain Research*, 1613, 37–48. <https://doi.org/10.1016/j.brainres.2015.03.050>
- Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PLoS ONE*, 7(8), Article e42993. <https://doi.org/10.1371/journal.pone.0042993>
- Zhang, M., Li, J., Chen, C., Xue, G., Lu, Z., Mei, L., Xue, H., Xue, F., He, Q., Chen, C., Wei, M., & Dong, Q. (2014). Resting-state functional connectivity and reading abilities in first and second languages. *NeuroImage*, 84, 546–553. <https://doi.org/10.1016/j.neuroimage.2013.09.006>
- Zhang, S., & Thompson, N. (2004). DIALANG: A diagnostic language assessment system (review). *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, 61(2), 290–293. <https://doi.org/10.1353/cml.2005.0011>
- Zhang, X., Yang, J., Wang, R., & Li, P. (2020). A neuroimaging study of semantic representation in first and second languages. *Language, Cognition and Neuroscience*, 35(10), 1223–1238. <https://doi.org/10.1080/23273798.2020.1738509>