When words get physical: evidence for proficiency-modulated somatotopic motor interference during second language comprehension

Nikola Vukovic (nv254@cam.ac.uk)
Department of Theoretical and Applied Linguistics, University of Cambridge

Abstract
New theories of cognition posit an intimate link between higher cognitive processes and the sensorimotor areas of the brain. In a reaction time-based translation task, second language (L2) speakers responded to action verbs using a microphone or a response pad. A significant interaction among Response Modality, Verb Type, and Proficiency indicated that more proficient L2 speakers took significantly longer to respond with their hands to previously seen hand-related verbs, but not mouth-related ones. Conversely, responding using a microphone led to slower latencies in the case of mouth-verbs, but not hand-verbs. Amidst virtually exclusively monolingual research on embodied cognition, the current study provides evidence that reading L2 action verbs selectively interferes with subsequently performed manual or verbal responses, suggesting that semantic representations of these verbs are distributed over neural substrates underlying action execution. The role of proficiency and experience in language comprehension is discussed.

Keywords: embodied cognition, second language, semantics

Introduction
Ever since the cognitive turn in psychology, the human mind has been likened to a computer, and the essence of our mental life envisaged as the outcome of complex calculation over abstract symbolic elements. In this traditional framework, human cognition is defined as modular, with distinct components operating on information independently and autonomously (Fodor 1998). In contrast to this perspective, recent years have witnessed the strengthening of theoretical paradigms which posit thought as being grounded in experience and sensation. These have become known as Embodied Cognition theories. According to Embodied Cognition, human concepts are not amodal transductions of sensory data, but are instead grounded in sensory-motor processing itself. Much evidence for embodied cognition comes from studies of language processing. Within this new framework, language comprehension is thought of as grounded in, and intimately linked to, neural resources used in action, perception, and emotion (Barsalou 2008).

By now there exists a wealth of behavioural research which supports the claim that language comprehension and action execution are subserved by common neural resources. For example, Glenberg and Kaschak (2002) demonstrated an Action-Sentence Compatibility effect, where judging the sensibility of sentences which implied a movement towards or away from the body (You gave Andy the pizza vs. Andy gave you the pizza) facilitated congruent arm movements. In a similar sentence sensibility judgement task, Zwaan and Taylor (2006) found that participants were significantly faster to perform manual rotation of a response knob when the rotation (clockwise or anti-clockwise) was congruent with the meaning of previously presented sentential material (turn the volume up/down). Bergen et al. (2003) asked participants to manually verify names of pictures representing actions, and found that response times were slower when they had to reject actions performed with the same (vs. different) effector. In a different study, Buccino et al. (2005) had participants listen to and judge the concreteness of sentences using hand or feet responses. They observed interference effects pointing toward the conclusion that verbally presented action sentences activate the motor system. These studies suggest that language and action are highly interconnected and that processing action language functionally involves activation of motor representations in the brain. Moreover, this interaction is differentially articulated as facilitation or inhibition, based on the temporal relationship between stimulus and response (Boulenger et al. 2006).

The findings from behavioural studies outlined above find additional support in the neurosciences, where experiments have shown interdependence between cognition and simulation of motor and perceptual states. There have now been numerous studies which demonstrate that semantic processing of a word activates distributed and diverse networks of sensory and motor information (Farah, McClelland, 1991; Damasio 1990; Caramazza et al., 1990). For example, processing the name of an action engages the motor area which is active during performing that same action (Martin et al., 1995 p. 649-652.). Hearing a word activates auditory associations (Pulvermüller et al. 2006), and action-related words elicit cortical activation in the motor areas of the brain, even when the participants are not aware of hearing the word (Pulvermüller et al. 2005). Intriguingly, comprehension of action words does not only reliably activate the motor cortex, but does so in an effectorspecific i.e. somatotopic manner: face, arm, or leg words activate the corresponding parts of the motor system in the central and precentral cortex which control face, arm, or leg movements (Buccino et al. 2005, Hauk et al. 2004).

Taken together, behavioural and neuroimaging data strongly support predictions and claims of embodied approaches to language and cognition. Semantic representations of words are not amodal or entirely symbolic, but seem to utilise the same neural resources involved in action execution, and it is these strong links which are made apparent in the interactions outlined in the studies above.

Some proponents of disembodied and symbolic approaches to cognition, however, have raised concerns
about the functional relevance of sensorimotor processes in language comprehension (Hickok, 2010; Mahon and Caramazza, 2008). For example, Mahon and Caramazza (2008) have argued that semantic motor activation, as it is described above, could also be incorporated into traditional theories. On this view, the motor cortex becomes activated as an epiphenomenon of, and not part of, semantic retrieval. In other words, semantic motor activation is a result of induced imagery of action, and is as such a downstream consequence of comprehension. However, a number of recent studies presented evidence for automaticity and causality of sensorimotor processes in language comprehension. For example, Pulvermüller et al. (2005) have demonstrated activity in the motor cortex as early as 100-200 ms following word recognition - speeds consistent with the idea that these processes are crucial to semantic retrieval (see also Shtyrov et al., 2004). Similarly, Liuzzi et al. (2010) present compelling evidence that the motor cortex is causally involved in learning and processing action words. In their study, transcranial direct current stimulation (tDCS) was used to temporarily disturb the functioning of the motor cortex, which in turn led participants to perform significantly worse in an action word acquisition task, compared to controls. The study indicates that the motor cortex is vital, and even necessary, for word processing, as is also suggested by other TMS, electrical stimulation, and behavioural experiments (Pulvermüller, 2005; Fischer and Zwaan, 2008; Glenberg et al., 2008). Results to date thus favour embodied approaches to language comprehension, and make symbolic interpretations of motor semantics rather difficult to maintain.

The Current Study

Motor and language processes in the brain seem to form a dynamic and highly interconnected system, with interactions appearing at very early stages. However, a major shortcoming of investigations performed so far is the fact that they are virtually exclusively based on monolingual data, and data obtained from native speakers of a language. Surely, however, any theory that seeks to explain linguistic processes cannot call itself complete without at the same time accounting for how these operate in the majority of the world’s population (i.e. bilinguals: Gordon, 2005). With the realisation that over half of the world’s population speaks more than one language, this study aims to test the predictions of embodied cognition on people other than monolinguals. The primary question in the experiment reported here is whether in second-language (L2) speakers action-word semantics are distributed over neural substrates involved in action execution. Will L2 speakers, just like native ones, exhibit sensorimotor effects in lexical processing? Do we, in other words, see evidence for interaction between motor and linguistic processes or, alternatively, is there evidence that in their case linguistic processes operate completely independently of sensorimotor ones? How are these affected by proficiency and experience? Clearly, the extensive work in embodied language processing is in need of specification in terms of how and when grounding takes place in the course of language development.

The current study employed a reaction time-based translation task, which methodologically synthesizes experimental and analytical tools drawn from second language (Altarriba and Mathis 1997, Duyck and Brysbaert 2004) and embodied cognition research (Shebani & Pulvermüller 2012, Bergen et al 2003, Marino BFM et al 2011). In every trial, participants were presented with a mouth, arm, or leg related verb in their native language (Serbo-Croatian), after which they would see a verb in English. The task was to quickly indicate whether the second verb was a good translation of the first verb, using a button box in one, and a microphone in the other half of the experiment. In half of the trials the English verb was a translation of the Serbo-Croatian one, and in the other half it was not. The critical trials in this experiment were those where the verbs were translations, and were split into two conditions: in one case, the English verbs denoted actions performed with the same body part or effector as the one used for the experimental response (mouth or hand, depending on which experimental response was required), whereas in the other case the verbs indicated actions performed with a different effector.

If we assume a neurobiological model of language in which lexicosemantic circuits are embodied (Pulvermüller, 1999), then verbs describing actions should be realised not only in perisylvian cortical regions traditionally associated with language, but also as circuits in the motor cortex, which is used for executing the actions themselves. If it is true that understanding an action verb produces motor activation, then introducing a concurrent task (the user response) which makes use of that same part of the brain should produce interference, reflected in slower reaction times. In addition, the semantic somatotopy model (Pulvermüller, 1999; 2001) predicts that this interference should be highly specific: processing mouth-related verbs should most strongly interfere with concurrent verbal responses, but not manual ones. Similarly, processing a hand-related verb should lead to much slower latencies when responding with the hand, but not with the mouth.

In the current context of second language speakers, we could expect several possible outcomes: 1. It might be that non-native speakers process language in a completely different way to that of their native counterparts. In other words, their semantic representations might not be distributed across sensorimotor neural substrates - a distribution that might, therefore, be a distinctive hallmark of native speakers. 2. It could be the case that L2 speakers show identical patterns to those of native speakers. 3. Finally, it is possible that L2 speakers show differing amounts of sensorimotor embodiment, as modulated by proficiency and experience.
Experiment

Participants

Twenty-four right-handed native speakers of Serbo-Croatian (13 female; mean age = 25.63 years, SD = 3.54) studying at the University of Cambridge took part in the experiment. All participants had normal or corrected-to-normal vision and reported no history of neurological, psychiatric, or language disorders. Participants varied in terms of L2 proficiency, as revealed by the language background questionnaire (see below). All gave their informed consent prior to participation. No participants were aware of the purpose of the experiment.

Stimuli

A total of 72 lexical stimuli were used in the experiment: 36 Serbo-Croatian (SC) verbs, and 36 English ones. For each language there were 12 mouth-related (e.g., bite, kiss, sing), 12 hand-related (e.g., peel, take, write), and 12 leg-related (e.g., dance, jump, walk) action verbs. In addition, 18 verb pairs were constructed for the practice trials: 9 practice trials per response modality (mouth/microphone vs. hand/button box). All critical lexical stimuli were matched for psycholinguistic variables such as number of letters, number of phonemes, lexical frequency, and letter bigram frequency. All SC verbs appeared inflected for first person singular present as, for example, in the verb pišem (write), where the suffix –m attaches to the base form piše-.

Procedure

The experiment took place in a sound-attenuated and dimly lit room. Participants sat comfortably in front of a computer screen at a distance of about 60 cm. Stimuli were presented centrally on the screen in lowercase Arial font (size = 20). Each trial started with a fixation cross presented in the centre of the screen for 1000 ms, followed by a SC verb displayed for another 1000 ms. After a 500 ms inter-stimulus interval (ISI), an English verb appeared and stayed on the screen until a response was given. Participants were instructed to respond, as quickly and as accurately as possible, whether the English verb was a translation of the previously seen SC verb. They did so by pressing “yes” or “no” on a button box, in one half of the experiment, and by saying “yes” or “no” into a microphone, in the other half (the order of “mouth response” and “hand response” blocks was counterbalanced across participants). After they gave a response (correct or incorrect), a blank screen was shown for 500 ms, after which a new trial started. Accuracy feedback was displayed only during the practice block. Stimulus presentation and response time collection were performed using the SuperLab Version 4.5 software package (Cedrus Corporation). The experiment consisted of a practice block, and two experimental blocks: one requiring hand, and the other mouth responses. Participants therefore saw each target verb twice, once in each half of the experiment. Each target verb would appear in both the matching and mismatching condition (actions conveyed by L2 verbs shared/did not share the effector with the experimental response). Items were rotated around two presentation lists. If an English target verb was in the same-effector condition on one list it was in the different-effector condition on the other list, and vice versa. There were equal numbers of same- and different-effector pairs on each list, and equal numbers of participants were tested on each list.

Language Background Questionnaire

All participants completed a language background questionnaire. The sample was homogeneous, with all participants being native speakers of Serbo-Croatian, raised in the ex-Yugoslavia territory, and having started learning their L2 at approximately the same time (Mean AoA: 7.86 years, SD = 2.4). Overall, participants rated their proficiency in English relatively high, though there were still differences, with some participants having just arrived to England for their undergraduate and graduate courses, and others having been in the country for longer as part of their PhD or postdoctoral research. A simple median split was performed on the proficiency scores, thus creating two groups, the lower and the higher proficiency group, with 12 people in each. The difference in L2 proficiency ratings between these two groups was statistically reliable, t(22) = -7.30, p = .001. These and other proficiency measures from the language background questionnaire which, importantly, were found to correlate with participants’ self-ratings, are summarized in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Proficiency rating a</th>
<th>Age of Acquisition (years)</th>
<th>Time spent in UK (months)</th>
<th>IELTS score b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Lower Proficiency Group</td>
<td>5.87</td>
<td>0.23</td>
<td>7.58</td>
<td>2.96</td>
</tr>
<tr>
<td>Higher Proficiency Group</td>
<td>6.71</td>
<td>0.30</td>
<td>8.08</td>
<td>1.78</td>
</tr>
</tbody>
</table>

Table 1. Mean data for participants’ (n = 24) language history and self-assessed proficiency ratings

a based on a scale from 1 – 7
b Since the majority of scores were IELTS scores, TOEFL and Cambridge Exam scores were also converted to the IELTS scale using the standard Equivalency Table.
Results

First, trials with erroneous responses were removed. Given that errors were very rare (0.02% of total trials), they were not analysed separately. No participant performed at less than 90% accuracy. Response latencies for correct trials larger than 2SD ± mean RT were excluded from statistical analysis as outliers. In total, no more than 0.5% of data was lost (Ulrich and Miller 1994).

For the remaining trials, mean RT values for each participant in each condition and block were calculated (see overall means in Figure 1) and entered into a repeated-measures analysis of variance (ANOVA) with two within-subject variables: Response Modality (microphone vs. button box), Verb Type (mouth vs. hand vs. leg), and Proficiency (lower vs. higher) as a between-subject variable. Response latencies were longer when using the microphone than with the button box, and correspondingly the ANOVA revealed a main effect of Response Modality, F(1, 22) = 44.27, p = .001; partial η² = .668. No other main effect was significant (p>.05). A two-way significant interaction among Response Modality and Verb Type was found: F(1, 22) = 12.58, p = .002; partial η² = .364. In addition, there was a three-way significant interaction among Response Modality, Verb Type, and Proficiency: F(1, 22) = 15.22, p = .001; partial η² = .409.

Most importantly, the significant three-way interaction directly addresses the main questions which motivate this study. The interaction indicates that second language speakers of different proficiencies differentially exhibit somatotopic interference on action execution during verb processing. The group of lower proficient speakers showed no significant differences in response times in any of the blocks and conditions. However, more proficient L2 speakers responding with the microphone were slower on mouth verbs than on hand verbs (1394 and 1206 ms respectively, t(11) = 3.096, p = .010). Conversely, their hand responses were slower when processing hand, but not mouth verbs (678 and 620 ms, respectively, t(11) = 2.80, p = .017). Responses to leg verbs always patterned with other non-response modality verbs.

Overall then, for the more proficient L2 speakers in the sample, the results reveal a double-dissociation pattern of response interference.

Discussion

This study reveals differential interference for mouth and hand responses, brought about as a consequence of processing lexical semantics of action verbs involving different parts of the body, namely the mouth and the hands. Interestingly, this selective effect was modulated by speaker L2 proficiency. In both participant groups, response latencies given with the microphone were longer than those given with the button box. However, our finding that verbal responses are slower than manual ones in L2 speakers is quite in line with results and latencies obtained in previous second language studies (see, for example Kroll et al., 2002). Apart from this main effect, participants in the lower proficiency group demonstrated no significant differences in response times across blocks and conditions. Interestingly, slower responses with hands were observed when higher proficiency participants processed hand verbs, but not mouth or leg verbs. The reverse effect, slower responses during mouth (but not hand or leg) verb processing, was seen when participants used a microphone to respond. These results follow a double dissociation pattern (Shallice 1988; Jones 1983), which allows for much more reliable inferences about the causal and interactive status of the systems and processes involved than was possible in some previous, conceptually similar studies, which only used a single response modality, or used pictorial stimuli where priming from visual features could not be reliably dissociated from true motor interference (see for example Bergen et al 2003; Marino BFM et al 2011). Specifically, this study demonstrates that in more proficient participants processing resources located in specific parts of the motor cortex are shared between action execution and lexico-semantic representations of related verbs. These data, although behavioural in nature, directly bear upon and

![Figure 1. Mean reaction times for both participant groups in all conditions (H=hand; M=mouth; L=leg). Significant differences in response latencies (p<.05) are marked with an asterisk.](image-url)
support predictions made by psychological and neural models of embodied language processing. Apart from providing novel and strong evidence for embodied semantics, the present study is the first to demonstrate double dissociated and effector specific shared neural resources between action execution and language in non-native speakers. In addition, it has implications for theories of how grounding takes place, as it suggests that speakers increasingly come to activate non-linguistic systems during word processing as a function of proficiency and experience.

The findings in this study can be interpreted with regard to neurobiological models of language learning and processing. These do not view lexicosemantic representations as static, but adopt an approach in which variation primarily comes not from the strength of form-meaning connections, but the distribution of semantic representations themselves. In other words, it might be the case that there is less meaning associated with L2 lexical forms (for similar proposals see Williams & Cheung 2011, and Duyck and de Houwer 2008). This explanation makes intuitive sense if we think about the amount of real-life experience with L2 words. For example, L2 words are associated with a much smaller range of senses than L1 words (Finkbeiner and Nicol 2003), and are in the majority of cases learned in artificial classroom environments, often through the use of crude lexical translation. This account is consistent with the literature on cortical learning, where words can be thought of as functional cell assemblies in the brain, formed through Hebbian processes (Pulvermuller 1999). Hebbian learning (“what fires together, wires together”) would thus predict that cortical distributions of L2 word semantics are much more restrictive than those of L1 words, due to different (and fewer) learning and usage experiences. If this is the case, then L2 words, learned through translation and with no or limited real-life usage, would be strongly left lateralized and distributed mainly over perisylvian cortices. We should, therefore, as was the case presently, find little activation outside of primary linguistic brain areas. However, if speakers start using L2 words in real-life embodied contexts, such as when studying in a foreign country, then these words would increasingly come to be co-activated with extra-linguistic neural substrates (including the motor cortex), and would “wire together” into a new assembly. The present results are therefore consistent with the idea that the amount of experience and real-life usage leads to changes in the way we semantically represent words. Crucially, there were differences in the amount of time both groups of speakers spent in an English-speaking country (see Table 1), and this difference was statistically reliable ($t(12) = -2.60, p = .023$). In fact, there is evidence for this conclusion from studies testing the linguistic performance of L2 students learning their second language in a classroom vs. an immersed, study abroad setting. Using experimental paradigms similar to the one employed presently, Linck et al. (2009) and Talamas et al (1999) demonstrate that as their proficiency increases, L2 speakers move from exhibiting primarily form-level effects, to showing increased semantic access. This could explain why in the results reported here the participants who spent less time in the UK showed no motor semantic interference during lexical processing, whereas the other group with a significantly longer residence in the country showed reliable double-dissociated interference when processing action verbs. However, while this proposal seems plausible and consistent with neurobiological models of language, it must presently remain at the level of speculation. More work is thus needed to further clarify the questions of how and when grounding takes place in the context of second language acquisition and learning.

**Conclusion**

It has been suggested that embodied sensorimotor systems are an integral part of language comprehension. The current study, for the first time in second language speakers, demonstrates that they increasingly come to incorporate into their action-word semantics the same processing resources used for effector-specific action execution. The resulting somatotopic interference effects observed in proficient L2 speakers present evidence in favour of embodied approaches to language.

**Acknowledgments**

I would like to thank my supervisor, Dr. John N. Williams, for his valuable comments and statistical advice. This study was supported by a CISS graduate scholarship from the University of Cambridge.

**References**


