Acta Psychologica 145 (2014) 98–103

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals

Nikola Vukovic *, John N. Williams

Department of Theoretical and Applied Linguistics, University of Cambridge, UK

ARTICLE INFO

Article history:
Received 18 September 2013
Received in revised form 6 November 2013
Accepted 8 November 2013
Available online 10 December 2013

PsycINFO classification:
2720 - Linguistics & Language & Speech
2340 - Cognitive Processes

Keywords:
Embodied Cognition
Bilingual
Semantics

ABSTRACT

Research supports the claim that, when understanding language, people perform mental simulation using those parts of the brain which support sensation, action, and emotion. A major criticism of the findings quoted as evidence for embodied simulation, however, is that they could be a result of conscious image generation strategies. Here we exploit the well-known fact that bilinguals routinely and automatically activate both their languages during comprehension to test whether this automatic process is, in turn, modulated by embodied simulatory processes. Dutch participants heard English sentences containing interlingual homophones and implying specific distance relations, and had to subsequently respond to pictures of objects matching or mismatching this implied distance. Participants were significantly slower to reject critical items when their perceptual features matched said distance relationship. These results suggest that bilinguals not only activate task-irrelevant meanings of interlingual homophones, but also automatically simulate these meanings in a detailed perceptual fashion. Our study supports the claim that embodied simulation is not due to participants' conscious strategies, but is an automatic component of meaning construction.

1. Introduction

The nature of meaning – how we represent and transmit it – stands as one of the foundational problems in cognitive science. The ability to communicate using language is a distinctively human trait, yet the mechanisms describing how exactly the linguistic code becomes meaningful are much debated. Understanding a sentence, on traditional accounts, consists of computing a propositional representation which, in essence, specifies necessary and sufficient conditions needed for the sentential content to be true (e.g., Fodor, 1998). However, any semantic theory in which meaning is the exclusive domain of amodal computation runs into the serious problem of symbol grounding (Harnad, 1990; Searle, 1980). An alternative approach to this issue has recently been presented in theories whose aim is to explain the operation of the evolutionarily recent linguistic system as built upon, and grounded in, phylogenetically much older sensorimotor brain structures. On this embodied interpretation of cognition, language comprehension is achieved by recruiting the very same resources which are used for action, perception, and emotion. Linguistic meaning, in other words, arises through performing mental simulations of sentential content, during which we reactivate fragments of experience formed during past perception and action (Barsalou, 1999).

Evidence from behavioural and neuroimaging studies supports the claim that, when understanding language, people perform mental simulation, and that this simulation is embodied. Studies indicate that we mentally represent perceptual and visual information described in comprehended sentences (Glenberg & Kaschak, 2002; Horton & Rapp, 2003; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). For example, Zwaan et al. (2002) had participants read sentences implying certain actions (e.g., *The ranger saw the eagle in the sky*), and subsequently verify pictures that were congruent or incongruent with the implied event (e.g., a picture of an eagle with outstretched wings vs. an eagle in a nest). They found that people were significantly faster to respond to pictures that were consistent with the implied sentential content, suggesting that they were simulating the shape of objects and animals involved. Similarly, behavioural studies have confirmed that comprehenders simulate a range of other perceptual features, such as orientation (Stanfield & Zwaan, 2001), location (Bergen, Lindsay, Matlock, & Narayanan, 2007), visibility conditions (Horton & Rapp, 2003), and motion (Kaschak et al., 2005). Interestingly, mental simulation also includes movement direction (Glenberg & Kaschak, 2002), as well as other properties such as action duration. In Matlock (2004), for example, participants read the sentence "*Road 49 crosses the desert*" much slower if they were previously told that the desert was 400 miles in diameter, than when they were told that the desert was only 30 miles in diameter.

Neuroimaging studies also support the view that language comprehension crucially involves simulation of sensory, motor, and emotional content. Pulvermüller (2005) found that reading action words such as "kick" or "run" reliably activates areas in the motor cortex used for performing arm or leg movements. Similarly, auditory perception has overlapping neural correlates with auditory imagination (Halpern,

* Corresponding author at: Department of Theoretical and Applied Linguistics, English Faculty Building, University of Cambridge, 9 West Road, Cambridge CB3 9DP, UK.
E-mail address: nv254@cam.ac.uk (N. Vukovic).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.11.002

Zatorre, Bouffard, & Johnson, 2004). Speer, Reynolds, Swallow, and Zacks (2009) found a congruent simulation pattern, not only during individual word presentation, but also during sentence or story reading. They conclude that understanding a story produces brain activation very similar to that during performing, imagining, or observing relevant actions/events in the world.

In sum, then, both behavioural and neuroimaging data suggest that language comprehension is more than the kind of purely symbolic process assumed by traditional theories. Comprehension seems to involve activating sensorimotor neural resources to perform detailed, multimodal, and dynamic simulation of perceptual and motor states, through which the linguistic codes we use in communication become meaningful.

The studies outlined above have sparked an active discussion in the field, highlighting the need for additional inquiry into the automaticity and universality of perceptual simulation processes, especially across diverse linguistic populations. Concerns have been raised that the reported findings might not reflect a process that is basic to language comprehension, but is rather the result of participants' conscious decision to imagine a described scene after they have already understood the meaning (Kiefer & Pulvermüller, 2012; Mahon & Caramazza, 2009). The aim of the present study was to address this question by exploiting a known fact about bilingual lexical processing — automatic, and unconscious, activation of lexical representations in the non-current language (usually the L1 during an L2 processing task).

Previous research has demonstrated that bilinguals cannot switch off one of their languages during comprehension (De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001; de Groot, Delmaar, & Lupker, 2000; De Groot & Nas, 1991; Schwartz & Kroll, 2006; Van Assche, Duyck, & Hartsuiker, 2012; Van Heuven, Dijkstra, & Grainger, 1998). This non-selective activation of both L1 and L2 words at the same time has been found to persist irrespective of task demands and participants' cognitive strategies (Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & Van Hell, 2003). For example, research in bilingual language processing has demonstrated that bilinguals non-selectively activate both meanings of a homophone word (e.g., Lagrou, Hartsuiker, & Duyck, 2011). These task-irrelevant L1 meanings become active not only in single word reading, but also in a sentential and semantic context (Lagrou, Hartsuiker, & Duyck, 2012). In conclusion then, language non-selective and automatic lexical access is a robust finding confirmed by decades of research in bilingual language processing.

Our study aims to exploit this known fact about L2 lexical processing to investigate a crucial issue in the literature: the automaticity of embodied simulatory processes in language comprehension. The bilingual mind seems to be the perfect testing ground for this hypothesis derived from monolingual research. We know, as previously described, that people perform detailed perceptual simulation during language processing; however, it is debated whether such simulation is consciously initiated and maintained, or if it proceeds automatically during comprehension.

Here, we adapt the experimental design used in Winter and Bergen (2012), which was originally introduced by Stanfield and Zwaan (2001). Winter and Bergen (2012) had participants read sentences about objects, where some sentences implied visual distance (*You are looking at the microphone on the other side of the stage*) and some implied closeness (*You are looking at the microphone in your hand*). Afterwards, participants were presented with pictures of mentioned objects which, crucially, varied in size. Their task was to make a judgment on whether the object in the picture was mentioned in the previous sentence. The reasoning was that if readers mentally simulate perceptual features, including distance, then they should be faster to verify images congruent with sentence-implied distance. Indeed, participants were found to respond faster to small images of mentioned objects after reading sentences implying visual distance, and to big images following sentences implying closeness, suggesting that they visually simulated the sentence meaning. No such effect was obtained in control trials, when participants responded to pictures of objects not mentioned in the sentences.

In the present study the above paradigm was modified so as to test the main hypothesis on a bilingual population and, exploiting the well-known fact about non-selective bilingual language processing, ascertain more closely the nature of embodied simulation processes. Dutch-English bilinguals heard sentences in their L2 (English), after which they saw images which varied in size, and had to make a judgment on whether the depicted objects were mentioned in the previously heard sentences. In the Winter and Bergen (2012) study, participants would read sentences on the screen and would manually indicate when they have read them, thus advancing to the next trial. This procedure may be problematic, because it gives the participants the freedom and time for slow (re-)reading, and might possibly favour strategies wherein participants purposefully imagine the sentence content in anticipation of the picture. Because of this, we decided to use auditory presentation of sentences, which is a more rapid procedure which minimises variation in stimulus presentation time. Additionally, Winter and Bergen (2012) used sentences within which the position of the target object-word varied. Here, all critical targets were object-words located uniformly across sentences (always in sentence-final position). Most importantly, however, the current experiment is more likely to uncover automatic processes because of a novel modification of the classic design: all the critical object-words were interlingual English-Dutch homophones (words which sound similar in both languages, but denote something different in each).

If, following second language research, bilinguals unconsciously activate L1 meanings when processing L2 sentences containing interlingual homophones, and if it is true, as Embodied Cognition suggests, that meaning activation includes performing perceptual simulation, then we should see evidence of simulation processes even on task-irrelevant L1 meanings of words. Alternatively, if perceptual simulation is a process initiated intentionally by participants, then we might expect that unconsciously activated semantic material will not participate in a top-down driven simulation. If the former account is true, we would expect to see slower latencies in the matching homophone condition (where image size matches sentence implied distance), and faster ones in the mismatching condition (image size does not match sentence implied distance). The reasoning behind this prediction is as follows. Winter and Bergen (2012) saw facilitation of congruent responses in monolinguals when the pictures matched the implied sentence distance. Here, however, participants have to reject, not accept, the critical pictures, because they show task irrelevant representations of L1 homophone meaning. For example, after hearing the English sentence *"On the plate in front of you, you can see a bone,"* participants would see a picture showing beans — the word for which in Dutch is "boon"/bo:n/. The picture shown varied in size between conditions, such that it could match or mismatch the implied sentence distance (for an example of sentence–picture pairs, see Table 2). Critically, if bilinguals end up perceptually simulating both the task-relevant L2 meaning and the irrelevant L1 meaning, it should be harder for them to reject the critical pictures. This is because, even though task-irrelevant, the pictures would match the participants' mental simulation on one additional dimension — visual distance. The main RT differences to look out for, then, will be those between congruent and incongruent homophone target pictures.

The results of the current experiment could have important implications for debates both in Embodied Cognition and bilingual processing. As we have seen, Embodied Cognition theory predicts mental simulation of meaning that is automatic and central to language comprehension. Crucially, on this account, this simulation would be one that is multimodal, and embodying a specific perspective — namely, one that mirrors actual perception. Therefore, understanding sentences in our experiment should lead to simulation of objects at specific distances, and whose size is consistent with the first-person perspective of an immersed observer (Barsalou, 2002). If we find evidence of such simulation even for homophone meanings from the task irrelevant language (Dutch), then this would present strong support for the automatic and cross-linguistic nature of embodied simulation processes. In addition to testing the predictions made by Barsalou (2008) and other Embodied Cognition supporters, the results of this study could inform debates in linguistics about the bilingual lexicon, and the kind of representations activated during bilingual language processing.

2. Experiment

2.1. Participants

The participants in this study were 24 proficient Dutch-English bilinguals, at the time living and working/studying in Cambridge, UK. The data of 3 participants had to be excluded due to computer malfunction (experimental software crashing and/or not recording user responses). In addition, data of one participant had to be removed because they reported being "mildly dyslexic" in the post-experiment questionnaire. Thus, the data are reported for the remaining 20 participants (10 females, mean age = 29.35 years, SD = 8.67). All participants were quite proficient in their L2, due to regular exposure to English in school, media, as well as doing university courses and living in the UK. After the experiment, participants were asked whether they noticed that the size of the pictorial stimuli was manipulated in a meaningful way. None reported awareness of this manipulation, indicating that they were not aware of the perceptual simulation nature of the experiment. Additionally, participants were asked to provide a self-rating of their L2 proficiency (in both reading and speaking) on a 7-point Likert scale. We also recorded the age at which they started learning English, as well as the number of years spent living in an English-speaking country. Mean ratings are reported in Table 1. No participant reported any history of neurological, psychiatric, or language disorders, and had normal or corrected-to-normal vision. Prior to participation, all gave their written, informed consent.

2.2. Materials

For this experiment, 24 critical, contextually low-constraint sentences in English were constructed, which used an interlingual English-Dutch homophone (e.g., "bone," which in Dutch sounds like the word "boon" [beans]/bo:n/; "cook", which in Dutch sounds like the word "koek" [cake]/kuk/) occurring in the final position. The L2 homophone nouns and their psycholinguistic characteristics were obtained from Lagrou et al. (2011, 2012) and from the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993). This previous research has demonstrated that our selection of homophones reliably activates both L1 and L2 meanings both in isolation and in a sentence context, and irrespective of sub-phonemic cues related to one's native language. The critical sentences were organised in pairs, where one member of the pair implied closeness (e.g., *On the plate in front of you, you can see a bone*) and the other distance (e.g., *On the plate at the far end of the table, you can see a bone*). Across conditions, each participant would hear only one sentence out of each pair (Near vs. Far). To accompany these homophone sentences, 24 colour image pairs were obtained using Getty Images (see http://www.gettyimages.co.uk), a provider of stock digital media. Each pair consisted of a big and a small image (produced by resizing an image to occupy 500 px on the longest dimension, in the former, and 200 px, in the latter case). Critically for this design, all images required a "no" response, since they represented the task irrelevant, i.e. Dutch, meaning of the homophone word.

In addition, using the same criteria, an equal number of sentence and picture pairs were created as controls for the homophone trials. The target words in critical and control sentences were matched for the number of phonemes and lexical frequency. The only difference was that control sentence-image pairs were completely unrelated. Thus, each participant would be exposed to 12 critical and 12 control sentences i.e. 24 sentence picture pairs requiring a "no" response. Finally, seventy-two sentence–picture pairs were created to act as fillers. Twenty-four of these filler pairs required a "no," and forty-eight a "yes" response. To distract participants from the fact that the experiment investigated sensory simulation of distance, two thirds of the filler sentences referred to spatial position using left/right (e.g., "You put the umbrella to the left of the door" or "You place your book on the right side of the desk"). In addition, all pictures used in filler trials varied widely in size, ranging from 150 to 700 px on the longest dimension. All sentences were spoken by a male native speaker of American English, and digitally recorded and edited to finish at the offset of the last word using the Audacity 2.0 software package (see http://audacity.sourceforge.net/). The length of sentences varied approximately from 1800 ms to 3500 ms.

In total, then, each participant saw 48 trials requiring a "yes" response, and 48 trials requiring a "no" response. Half of these trials included sentences implying distance relations, and the other half used left/right spatial position.

2.3. Design and procedure

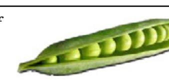
All items and conditions were counterbalanced, yielding four lists. Each participant would see only one of these lists, and between lists, each homophone-object would appear in one out of four possible versions (Near/Far Sentence × Big/Small Picture). Table 2 shows examples of critical and control sentence–picture pairs. There were equal number of matching and mismatching pairs on each list, and on each an equal number of participants was tested.

Participants would sit in a quiet and dimly illuminated experiment room, at a distance of around 60 cm from the computer screen. The presentation of stimuli was managed using the SuperLab software package (v4.5, Cedrus Corporation). The experiment started with an

Table 1
Mean age of L2 (English) acquisition, self-reported L2 proficiency ratings (7-point Likert scale), and length of stay in an English-speaking country.

	Mean	SD
Age of L2 acquisition	8.65	3.32
L2 proficiency: written	6.4	0.59
L2 proficiency: spoken	6.3	0.65
Years in L2-speaking country	3.85	5.37

Table 2
Examples of critical (1, 2) and control (3, 4) sentence–picture pairs used in the experiment. In this example, the English words "bone" and "cook," sound like the Dutch words for "bean" and "cookie".

Sentence [Near/Far]	Picture [Large/Small]	
1. *On the plate [in front of you/at the far end of the table], you can see a BONE.*		
2. *[Right next to you in the kitchen/At the far end of the kitchen], you can see the COOK.*		
3. *[In front of you/In the distance], you can see a WHEELCHAIR.*		
4. *[In the cage in front of you/On the top branch of the tree], you can see an EAGLE.*		

introduction screen, informing participants that their task is to quickly judge whether presented pictures depicted something mentioned in a previously heard sentence. They responded by pressing "yes" or "no" on a button box, with button location counterbalanced across participants. To ensure that the participants paid attention to the entire sentence, on a third of the trials they were asked to repeat the sentence after giving a response. The trials started with a blank screen, displayed for 2000 ms. Then, a red fixation cross would appear and stay centrally on the screen, while the sentence sound file was played. At audio offset, the fixation cross would turn green for 150 ms, after which it was immediately replaced by a picture. The picture would remain on the screen until a response was given, after which the next trial would start. Accuracy feedback was not displayed, except during the practice block, which consisted of 8 sentence–picture pairs.

3. Results

Due to disproportionate error/outlier responses for one homophone item across participants, reaction times for this item were excluded and replaced by the condition mean. In addition, response latencies 2SD longer than the average RT over each subject's data, in each condition, were marked as outliers and removed. The combined removal of erroneous and outlier data resulted in the loss of 4.3% (SD = 1.38) of the total data.

The remaining mean RT values were entered into a 2 (Sentence: Near vs. Far) × 2 (Picture: Big vs. Small) × 2 (Target Word: Homophone vs. Control) × 4 (list) repeated measures analysis of variance (ANOVA). Sentence, Picture, and Target Word were within-subject variables, while List was a between-subjects variable. In addition, we performed an identical ANOVA analysis using error percentage as the dependent variable. Consistent with the results of previous studies, and the claim that bilinguals activate multiple languages during comprehension, the analysis of reaction times revealed a main effect of Target Word: $F_1(1, 16) = 70.31$, $p < .001$ partial $\eta2 = 0.815$, indicating that participants, on average, responded significantly slower to homophone targets (914 ms) than to control targets (699 ms). This effect was mirrored in our analysis of response accuracy, which returned only one significant effect — namely, a main effect of Target Word: $F_1(1, 16) = 32.66$, $p < .001$ partial $\eta2 = 0.632$ (participants made errors in homophone trials 2.9% of the time, and no errors in control trials). Crucially, however, the ANOVA testing for reaction time differences revealed a significant two-way interaction between Sentence and Picture: $F_1(1, 16) = 12.64$, $p = .003$, partial $\eta2 = 0.44$; as well as a significant three-way interaction between Sentence, Picture, and Target Word: $F_1(1, 16) = 9.62$, $p = .007$, partial $\eta2 = 0.376$. The by-items 2 (Sentence: Near vs. Far) × 2 (Picture: Big vs. Small) × 2 (Target Word: Homophone vs. Control) ANOVA confirmed the main effect of Target Word: $F_2(1,11) = 138$, $p < .001$, partial $\eta2 = 0.92$; as well as the Sentence × Picture × Target Word interaction: $F_2(1,11) = 5.34$ $p = .041$, partial $\eta2 = 0.32$.

Supporting the starting hypothesis, pairwise comparisons revealed that participants were significantly slower to reject matching pictures over mismatching ones for Far homophone sentences, $t(19) = 2.68$, $p = .015$. This difference between matching and mismatching conditions was also significant for Near homophone sentences, $t(19) = 2.71$, $p = .014$. On the other hand, there were no significant differences in participants' responses to matching and mismatching pictures after hearing control sentences ($p > 0.05$). RT means for all critical conditions can be found in Table 3.

Table 3
Mean response times (ms) for all conditions: sentence (Near vs. Far) × picture (Match vs. Mismatch) × target word (Homophone vs. Control). Standard deviations are given in parentheses.

Sentence	Picture	Homophone trials RT	Control trials RT
Near	Match	966.07 (250.36)	702.05 (171.49)
	Mismatch	859.85 (205.23)	679.24 (143.59)
Far	Match	968.31 (258.60)	709.51 (148.36)
	Mismatch	864.87 (225.68)	705.41 (173.18)

3.1. Discussion

Combining experimental and analytical tools drawn from linguistic and Embodied Cognition research, this study sought out to test the automaticity of perceptual simulation by capitalising on the unique nature of bilingual comprehension. As outlined in the Introduction section, it is known that bilinguals non-selectively activate both their languages during processing. Unsurprisingly, this claim is supported by the presence of a main effect of Target Word (Homophone vs. Control) in our data. Participants were slower to respond to target stimuli in the homophone condition, suggesting they activated both languages which, in turn, led to greater competition between alternative meanings. Our finding of a significant three-way interaction between Sentence, Picture, and Target Word variables has valuable implications for theories of Embodied Cognition and bilingual language research in general. After hearing interlingual homophone sentences implying visual distance, participants were much slower to reject small target pictures than large ones. For example, on hearing the sentence "*Across the field, you can see a lake*" Dutch-English bilinguals are much slower to indicate that a picture of a corpse ("lijk" [corpse]/lɛ k/) was not mentioned in the sentence when it was small (congruent with sentence distance) than when it was large (incongruent with implied distance). Conversely, after hearing homophone sentences implying visual closeness, such as "*On your friend's desk, you can see a big mess,*" they responded slower to large pictures, than to small ones (in this case, the picture of a knife, which in Dutch is "mes" [knife]/mes/). This suggests that after hearing English sentences containing a homophone word, participants in parallel activated and perceptually simulated the homophone's L1 meaning. For example, after hearing the English noun "lake," participants would perceptually simulate the L1 meaning i.e. "corpse." Similarly, hearing the English words "mess" or "rock" would lead to activations of Dutch same sounding words for "knife" and "skirt," respectively.

The crossover pattern observed here lends strong support in favour of the hypothesis that participants mentally simulated the sentential content and that, moreover, these simulations were perceptual and fine-detailed enough to include visual features such as object distance.

The fact that bilinguals cannot but activate both L1 and L2 words when comprehending homophones is by now an unsurprising finding. Such activations were found in numerous previous studies, and can be argued to be analogous to the multiple activation of homophones within languages (such as *bank* in English), which is also evident in slower reading times (e.g., Rayner & Pacht, 1994). What is surprising and highly novel, however, is the current finding that this automatic process, in turn, is modulated by perceptual simulatory processes entailed by embodied approaches to cognition. The fact that participants in this study not only activated task-irrelevant L1 items, but also simulated them in enough detail to include visual distance, presents striking evidence in support of the idea that embodied simulation is an automatic building block of meaning construction. Indeed, the central role of perceptual simulation in comprehension is not predicted a priori by amodal theories of meaning (such as that by Mahon & Caramazza, 2009), and could only be reconciled with them post hoc, whereas such effects are readily explainable and directly predicted by embodied approaches (for an extensive discussion of this issue, see Pulvermüller, 2013).

One issue that might be raised with respect to current findings is that the effect may not be due to a meaning-specific process. In other words, it could be the case that a congruency effect will be observed with any object whose size matches the perspective implied in the previously heard sentence. In such a scenario, hearing a sentence implying distance/closeness would indiscriminately lead to facilitatory processing of small/big target objects. However, the present data does not support such a conclusion. Responses to control items in our experiment do not follow the pattern one would expect if meaning simulation were item non-specific. Rather, the latencies are similar to those found in previous research and show no statistical difference between distance-congruent and distance-incongruent conditions. The pattern of responses to control items, therefore, supports the hypothesis that embodied simulation is a semantic and contextually-driven process. In previous studies, concerns have also been raised about the external validity of simulation findings. It has been suggested that such findings might have been due to task demands which would encourage participants to actively perform visual simulation. In the study reported here, we used an equal number of matching and mismatching trials, so that adopting an active imagination strategy would not lead to better performance in participants. Moreover, the task only asked participants whether the depicted object was mentioned in the previously heard sentence — a judgment which does not require the use of fine-grained perceptual representations. In addition, as was the case in Winter and Bergen (2012), most sentences mentioned more than one object, so that participants could not know which one would be presented in subsequent pictures. The fact that our design used auditory sentence presentation also meant that participants could not freely (re-)read the material and try to imagine the described scenarios, as was the case in previous studies (e.g., Stanfield & Zwaan, 2001; Winter & Bergen, 2012; Yaxley & Zwaan, 2007; Zwaan et al., 2002). Finally, one further argument that the presently reported simulation effects are not due to a consciously driven process is one following from the unique experimental manipulation introduced into this study, informed by our knowledge of bilingual lexical access. There is no straightforward reason to believe that participants would actively try to include unconsciously activated L1 meanings into a mental simulation, especially in the context of an experiment where both the language and the perceptual feature of distance were task irrelevant. For these reasons, the results of the current study are more likely due to automatic perceptual simulation processes, rather than conscious image generation strategies. Moreover, this conclusion is consonant with the findings of several previous studies investigating the task specificity of embodied mental simulation. The studies of Pecher, van Dantzig, Zwaan, and Zeelenberg (2009), Wassenburg and Zwaan (2010), and Ditman, Brunyé, Mahoney, and Taylor (2010), indicate the presence of sentence–picture matching effects as well as improved memory performance even with testing delays ranging between several minutes to several days, thus going against the argument of task-driven mental image generation.

In addition to informing Embodied Cognition research, this study has relevance for second language research and research on bilingualism. It supports the view that bilingual language access is highly non-selective, and this non-selectivity persists even in a context where only one language (in this case English) is task relevant. With regard to models of the bilingual lexicon this study shows that during comprehension in an L2 linguistic task bilinguals activate not only low level lexical forms, but also automatically simulate context-driven L1 meaning in a detailed perceptual format. In this respect, the study nicely fits with findings showing early and parallel semantic and contextual integration processes (Penolazzi, Hauk, & Pulvermüller, 2007). We should note that having the present task be performed entirely in English had unique benefits: namely, it ascertained that any semantic effects related to L1 word meanings were automatic, and not due to the task itself promoting a mixed or dual-language experimental setting. With this in mind, it would be very interesting for future research to investigate the effects of embodied simulation in an L1 context, i.e. when reading L1 homophone sentences, and responding to pictures of L2 word meanings.

4. Conclusion

Theories of Embodied Cognition posit that language, action, and perception form a highly interdependent system supported by common processing and neural resources (Barsalou, 1999, 2008). An important question frequently debated is whether embodied simulation is a conscious strategy on behalf of the participants. Modifying a classic sentence–picture matching task, the current study presents novel evidence that simulated mental representations arising during language comprehension are indeed automatic, and qualitatively analog to modal states involved in actual vision and perception. Moreover, this study is the first to successfully test the predictions of theories of non-selective meaning access in bilinguals through methods developed in embodied simulation research. Embodied cognitive science has started to question the basic assumptions underlying most of linguistics, and especially the science of second language learning and processing: the current study demonstrates that there is much to be gained by combining these two fields of inquiry.

Acknowledgements

This study was supported by a Cambridge International Scholarship Scheme graduate scholarship from the University of Cambridge.

References

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]*. Philadelphia, PA: University of Pennsylvania Linguistic Data Consortium.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, *22*, 577–660.

Barsalou, L. W. (2002). Being there conceptually: Simulating categories in preparation for situated action. In N. L. Stein, P. J. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 1–15). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*, 733–764.

De Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. (2001). Language context effects on interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition*, *4*, 155–168.

de Groot, A.M. B., Delmaar, P., & Lupker, S. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology*, *53A*, 397–428.

De Groot, A.M. B., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*, 90–123.

Dijkstra, A., & Van Hell, J. (2003). Testing the language mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism*, *6*, 2–16.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, *42*, 445–464.

Ditman, T., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition*, *115*(1), 172–178.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Cambridge, MA: Harvard University Press.

Glenberg, A.M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565.

Halpern, A.R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined timbre. *Neuropsychologia*, *42*, 1281–1292.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*(1), 104–110.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79–B89.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*, 805–825.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952–965.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2012). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, *16*(03), 508–517.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition*, *32*, 1389–1400.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1108–1114.

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 374–388.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576–582.

Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, *127*, 86–103.

Rayner, K., & Pacht, J. M. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, *33*, 527–544.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417–424.

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, *20*(8), 989–999.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *121*, 153–156.

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, *3*, 174.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458–483.

Wassenburg, S. I., & Zwaan, R. A. (2010). Readers routinely represent implied object rotation: The role of visual experience. *Quarterly Journal of Experimental Psychology*, *63*, 1665–1670.

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*(1), 1–16.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*(1), 229–236.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, *13*, 168–171.