Mental Simulation in the Processing of Literal and Metaphorical Motion Language: an Eye Movement Study

Emilia Castaño[a] and Gareth Carrol[b]

[a]*Department of Modern Languages and Literatures and English Studies, University of Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain.*
https://orcid.org/0000-0003-1276-4334

[b]*Department of English Language and Linguistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom.*
https://orcid.org/0000-0002-1886-4881

Corresponding author:

Emilia Castaño
Department of Modern Languages and Literatures and English Studies, University of
Barcelona, Gran Via de les Corts Catalanes, 585, 08007, Barcelona, Spain
e.castano@ub.edu

Word count: 10433

Mental Simulation in the Processing of Literal and Metaphorical Motion Language: an Eye Movement Study

An eye-tracking while listening study based on the blank screen paradigm was conducted to investigate the processing of literal and metaphorical verbs of motion. The study was based on two assumptions from the literature: that language comprehension by default engages mental simulation, and that looking behavior (measured through patterns of eye-movements) can provide a window into ongoing cognitive processes. This study specifically compared the comprehension of sentences that depicted actual physical motion (*the curtain is rising*) and sentences that described changes in quantity or emotional states in terms of vertical motion (*prices are rising*). Results showed that eye-movements were selectively biased upward or downward in accordance with the direction implied by the verb, regardless of the context (literal or metaphorical) in which they appeared, and in the absence of any visual stimuli or explicit task. Thus, these findings suggest that literal and metaphorical language drive spontaneous, direction-specific mental simulations captured by eye-movements, and that at least in the case of verbs presented in the present progressive, which emphasizes the ongoing nature of actions, visual biases along the vertical axis may start during the verb itself.

Introduction

Embodied approaches to language comprehension argue that meaning access and retrieval entail a process of mental simulation (Barsalou, 1999, 2008; Bergen, 2012; Gallese & Lakoff 2005; Zwaan, 2004). In this view, capturing the meaning of an utterance involves the construction of a mental simulation of its content, which draws on the automatic and unconscious activation of traces of earlier sensorimotor experiences stemming from language users' recurrent embodied interactions with the world. The re-enactment of such experiences during language comprehension is argued to engage, at least partially, brain areas responsible for actual perception and action despite the absence of appropriate perceptual stimuli or motor actions (Barsalou, Simmons, Barbey & Wilson, 2003; Barsalou 2008).

Sensorimotor Resonance in Literal Language Processing

In recent years, a growing body of empirical research ranging from behavioral to brain imaging studies has emerged in support of the involvement of sensorimotor systems in language processing. Thus, numerous studies have provided evidence for motor resonance during language comprehension. Understanding sentences depicting a physical action has been shown to display response compatibility effects when participants' responses involve motor actions that match the content of the sentences. So, for instance, judging whether a sentence such as *he closed the drawer* makes sense by pressing a button located further way from the body has been found to lead to faster responses than answering with an arm movement that is opposite to that indicated in the sentence (i.e., pressing a button located close to the body) (Glenberg & Kaschak, 2002). Likewise, conducting a categorization task in which the words to be categorized denote objects that either require precision grip (e.g., pencil, needle) or power grip (e.g., hammer, knife) by mimicking a precision or power grip, respectively, has also been shown to speed up participant's responses (Tucker & Ellis, 2004). In the same vein, fMRI studies have proven that passive processing of language about motor actions (e.g., to lick, pick, or kick) triggers the activation of brain areas that are adjacent to or overlap with those that activate during the actual movements involved in the execution of such actions (Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005). Neurological evidence also suggests that the level of activation of the brain regions involved in action processing varies as a function of the specificity of the motor action denoted by a word, i.e., activation is higher when processing a word such as *wipe* (specific action) compared to *clean* (Van Dam, Rueschemeyer & Bekkering, 2010), and that mental simulation is body-specific. Handedness, for example, has been shown to influence brain activity during manual action verb processing (Willems, Hagoort & Casasanto, 2010).

There also exists a rich seam of evidence that in word and sentence comprehension tasks participants perform mental simulations of implicit perceptual information on the intrinsic properties of an object such as shape (de Koning et al., 2017b; Zwaan & Pecher, 2012; Zwaan, Stanfield, & Yaxley, 2002); size (de Koning et al., 2017a, 2017b), and color1 (Connell, 2007; Hoeben Mannaert, Dijkstra & Zwaan, 2017; Simmons et al., 2007; Zwaan & Pecher, 2012). Evidence for these types of simulations is provided by sentence-picture verification tasks. In this tasks, participants are presented with sentences that provide implicit rather than explicit information on the perceptual properties of an object (their color, shape or size) and are asked to judge whether the object in the subsequently presented picture was mentioned in the previous sentence. Results confirm that faster responses are obtained whenever pictures or sounds match the properties implied by the previous sentence (e.g., verifying the image of a green traffic light after reading the sentence: *the driving instructor told Bob to go at the traffic lights*).

In the case of properties such as sound (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008), smell[2] (González et al., 2006) and taste (Barrós-Loscertales et al., 2012), evidence for such simulations mainly translates into the activation of modality-specific perceptual regions in the brain: the posterior and middle temporal gyrus when reading words with strong sound associations (e.g. *telephone*) and the primary and secondary olfactory and gustatory cortices

---

1 Interference effects in sentence-picture verification tasks have been reported by Connell (2007). Connell obtained faster responses to pictures that mismatched the color implied by the sentences.

2 Speed and Majid (2018) did not find conclusive evidence for the mental simulation of odor in a behavioural experiment.

after reading words like *cinnamon* or *cheese*. Finally, as for the simulation of properties related to touch (Brunyé et al., 2012), research has shown that processing sentences that implicitly convey tactile information induces biases in perceptual judgments (e.g., rating the texture of a fabric as smoother after processing a sentence that implicitly conveys congruent tactile information).

Likewise, the picture verification paradigm has also provided support for the simulation of an object's extrinsic properties such as orientation (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012) and distance (Winter & Bergen, 2012). In the former case, there is evidence that when the orientation of the object in the picture matches that implied by the sentence, reaction times are shorter. In the latter case, the implied distance of an object from the protagonist of the sentence has been shown to influence verification judgments in which the size of the object in a picture is manipulated to represent spatial distance (e.g., reaction times are shorter when verifying a large picture of a bottle of milk after reading the sentence: *You are looking at the milk bottle in the fridge* than after reading the sentence: *You are looking at the milk bottle across the supermarket*). There also exists evidence that speed is also simulated (Lindsay, Scheepers & Kamide, 2013; Speed & Vigliocco, 2014; van Dam et al., 2017). Eye-tracking and fMRI data support that the speed of the actions described in a sentence influences gaze patterns and modulates activity in brain areas involved in motion and action perception. Finally, results from eye-tracking, visual discrimination and picture verification tasks have shown that the direction or axis of motion of an entity is also simulated during language comprehension, which can influence gaze patterns and interfere with visual perception (Bergen, Lindsay, Matlock & Narayanan, 2007; Kamide, Lindsay, Scheepers & Kukona, 2016; Kaschak et al., 2005; Meteyard, Bahrami, & Vigliocco, 2007; Richardson, Spivey, McRae & Barsalou, 2003; Zwaan, Madden, Yaxley & Aveyard, 2004) – see Section 2 for more details.

Sensorimotor Resonance in Metaphorical Language Processing

Beyond literal language, behavioral and neurological evidence also seems to lend support to the hypothesis that abstract concepts are grounded in experience via metaphorical mappings, even though results do not always provide a consistent picture. Research has found motor resonance in response to sentences that denote both literal and abstract transfer, e.g., to *give something to someone* vs. to *delegate responsibilities to someone* (Glemberg et al., 2008). Likewise, the left anterior inferior parietal lobe, which is involved in action planning, has been shown to activate when action-verbs are used both literally and metaphorically, as in *the general grasped the diamond* vs. *the class grasped the solution* (Desai et al., 2011). Moreover, evidence has also been gathered that the areas of the cortex that plan actual hand movements are recruited when literal and figurative sentences involving hand actions are understood – e.g., *She handed me the apple* vs. *She handed me the theory* (Rohrer, 2001, 2005). Likewise, it has also been shown that performing upward or downward arm movements while reading words literally or metaphorically related to vertical space (e.g., ascend vs. defeat) elicits different ERP waveforms. A larger positivity in the waveform was found when word processing was accompanied by incongruent movements in both conditions (literal and metaphorical). However, effects of movement congruency appeared at different rates, between 200 and 300ms post-word onset in the case of literal words and 500ms post-word onset in the case of metaphorical words, which has been argued to support a weak version of the embodied metaphor hypothesis (Bardolph & Coulson, 2014).

On the sensory side of simulation processes, several studies have reported evidence that supports the involvement of the auditory, olfactory and gustatory systems in metaphorical language processing. Thus, for example, comprehending sentences such as *his voice was silky* or *the wind is sharp*, where texture-related language in used figuratively, has been found to activate texture-sensitive areas of the somatosensory cortex (Lacey, Stilla, & Sathian, 2012).

In the same line, taste and smell-related figurative language has been reported to activate primary and/or secondary gustatory and olfactory brain areas (Citron & Goldberg, 2014; Pomp et al., 2018). Finally, evidence for the involvement of the visual system in the processing of metaphorical language has also been provided by several eye-tracking studies of fictive motion. These studies have observed that the visual representations that are activated during language processing are sensitive to fictive motion as shown by the fact that people tend to look longer at the path region of a picture when it is described in fictive motion terms – e.g., *the road goes through the desert* – than when it is described literally – e.g., *the road is in the desert* (Richardson & Matlock, 2007; Mishra & Singh, 2010).

Mental Simulation in Literal and Metaphorical Motion Language

Although not exhaustive, the literature review provided above shows that there is mounting evidence for the involvement of the sensorimotor system in language processing in a wide variety of ways. Nevertheless, important questions concerning the way metal simulation can be influenced by factors such as (a) the linguistic context (i.e., literal or metaphorical) in which words are used, (b) the degree of conventionality of metaphorical expressions, (c) the type of task that subjects need to perform or (d) the visual support that is available during the task remain to be answered. In the case of spatial and action language results are mixed for sensorimotor activation in metaphorical language processing.[3] For example, while behavioral evidence suggests that valence and power-related notions such as *good*, *bad* or *high-born* are grounded in metaphorical projections that recruit the vertical axis as their source domain, (i.e., "GOOD IS UP" and "POWER IS UP" metaphors (Meier & Robinson, 2004; Schubert, 2005)), functional neuroimaging data challenge this interpretation. Research using fMRI suggests that the spatial representations activated during the processing of valenced words do not necessarily

---

[3] For a review see Cacciari, et al., 2011.

entail embodied simulation as there is no evidence for the activation of modality-specific brain areas (i.e., areas specifically involved in processing visual spatial information). Instead, data indicate that, for the most part, source domain representations engage brain regions involved in processing spatial information across multiple modalities. In the case of power-related terms, a null result has been obtained (Quadflieg et al., 2011). Several studies have also reported that action-related language only elicits the activation of the premotor and motor cortices when used literally. In this respect, it has been shown that the processing of action verbs related to different effectors (e.g., *grasp*, *kick,*) engages motor areas when verbs are presented either in isolation (Raposo, Moss, Stamatakis & Tyler, 2009; Rueschemeyer, Brass & Friederici, 2007) or in sentences that describe literal actions – e.g., *grasp the pen* – (Desai, Binder, Conant, Mano & Seidenberg, 2011; Desai, Conant, Binder, Park & Seidenberg, 2013; Aziz-Zadeh, Wilson, Rizzolatti & Iacoboni, 2006; Raposo, Moss, Stamatakis & Tyler, 2009).

In other cases, the involvement of the motor system in action-language processing is argued to be conditioned by factors such as the morphological context in which words are encountered, and the novelty/familiarity of the non-literal expressions in which those words are used. As far as the influence of the morphological context on simulation is concerned, although research has provided evidence for motor simulation during the comprehension of simple action verbs, no evidence for a similar pattern of activation exists for abstract derivative verbs formed from more simple action verbs –e.g., the German verb *begreifen* "to understand" made up of the prefix *ge-* and the verb stem *greifen* "to grasp". In fact, no differences in processing have been observed between these verbs and other morphologically complex verbs with abstract stems, e.g., *bedenken* "to consider", a prefixed form of *denken* "to think" (Rueschemeyer, Brass & Friederici, 2007). As for metaphorical conventionality, results also indicate that when action verbs are used in non-literal contexts, no activation of motor-related brain areas is found if verbs are encountered in highly conventional metaphorical contexts

(Aziz-Zadeh,Wilson, Rizzolatti & Iacoboni, 2006) or in idiomatic expressions, e.g. *bite the bullet* or *kick the bucket* (Desai, Conant, Binder, Park & Seidenberg, 2013; Raposo, Moss, Stamatakis & Tyler, 2009). By contrast, as conventionality/familiarity decreases, the involvement of the motor system in non-literal language understanding has been found to increase. Secondary and primary motor areas have been shown to activate in response to familiar but non-conventionalized figurative expressions and unfamiliar action metaphors, respectively (Desai, Binder, Conant, Mano & Seidenberg, 2011; Desai, Conant, Binder, Park & Seidenberg, 2013). These studies suggest that sensorimotor grounding exhibits graded sensitivity to meaning abstraction and familiarity, which involves progressive disembodiment as we move from literal to abstract language and from unfamiliar to familiar expressions (Desai, Binder, Conant, Mano & Seidenberg, 2011; Chatterjee, 2010; Zwaan, 2014).

In behavioral studies, as shown above, evidence for the involvement of the vision system in the simulation of metaphorical action-language (specifically motion-related language) has been provided by eye-tracking studies on fictive motion, however in visual categorization tasks, where participants are asked to categorize an object presented on the screen in locations that either match or mismatch the orientation implied by a previously presented sentence, only literal language has been shown to interfere with actual perception – i.e., participants took longer to categorize objects that appeared at the top of the screen after sentences that implied literal upward motion while the opposite was true for downward motion sentences (Bergen, 2012; Bergen et al., 2007). These results have been argued to suggest that metaphorical simulations are less detailed than simulations for literal language, or that they follow a different time course (Bergen et al., 2007). The strength of sensorimotor engagement has been argued to be modulated by context (literal or metaphorical) and familiarity with relatively detailed simulations for literal and non-conventionalized metaphorical expressions and less detailed simulations for entrenched metaphors (Desai et al., 2011; Cardillo, Watson, Schmidt, Kranjec & Chatterjee, 2012).

Concerning the time course of simulations, the analysis of the temporal dynamics of motor simulation, in particular the action compatibility effect (ACE, i.e. facilitation in executing a task that is congruent with a sentence describing that action such as movement toward or away from a participant), has yielded results that suggest that motor simulation involves two steps (Borreggine & Kaschak, 2006). First, motor information relevant to the action that is being described (e.g., direction of motion) is activated in preparation for mental simulation. The activation of this information seems to occur as soon as the verb is processed. Later, upon sentence completion, once all the relevant active features are bound together, a full mental simulation of the sentence takes place. Kaschak and Borreggine (2008) compared two accounts of how this leads to the ACE. The feature binding account suggests that as soon as this binding occurs (at or near the end of the sentence) the ACE is attenuated or prevented, since the motor features recruited for simulation are no longer available to influence motor responses. An alternative explanation is the feature activation account, which holds that motor features activate during online sentence processing and remain active until sentence processing has been completed. In this view, feature activation should lead to an ACE as soon as a motor response is executed (i.e. when the linguistic input induces this) and the effect is maintained as long as motor features are active. These accounts therefore make different predictions about the length of time that relevant motor features remain active during sentence processing, and Kaschak and Borreggine (2008) found evidence in favor of the feature binding account, whereby the ACE was not present consistently throughout the processing of a sentence featuring a motoric cue, but was most prominent immediately following the linguistic content that triggered a motor response and disappeared soon after (albeit re-emerging later in the sentence).

With regard to motion events, which are the focus of this paper, the possibility that metaphorical simulations are less detailed than literal simulations raises the question of how fine-grained parameters of motion such as the path implied by an utterance are incorporated in the mental representations elicited by literal and metaphorical language. In the case of literal language, visual categorization tasks have shown that mental simulation of motion-related language can either encode general spatial biases along the vertical axis (Richardson, Spivey, McRae & Barsalou, 2003) or spatial biases specific to the direction of motion – up or down – implied by the linguistic item (Bergen et al., 2007); this results in a decrease in the ability to perceive either visual stimuli presented at any point along the vertical axis (Richardson et al., 2003), or only those in locations that match the direction of motion implied by the linguistic items, i.e., at the top for up words and at the bottom for down items (Bergen et al, 2007). More recently, an eye-tracking study utilizing the visual world paradigm showed that eye-movements during motion language processing reveal the dynamic simulation of specific paths through space rather than global spatial biases (Kamide et al., 2016). In this study, participants' gaze was monitored while they listened to sentences that suggested a motion path in an upward (*jump*) or downward (*crawl*) direction and inspected visual scenes that contained an agent and a goal but did not depict the path of motion of the agent to the goal. Results showed verb-congruent visual attention biases with vertical gaze position varying as a function of the direction of motion implied by the verbs only in the scene region between the agent and the goal (i.e., the implied path region). This suggests that participants' eye-movements were not influenced by global spatial biases induced by the verbs but by more localized path-specific visual attention biases. In sum, these studies pose the question of whether the specificity of the spatial biases induced by language depends not only on the linguistic context (literal or metaphorical) but also on the type of task subjects are performing and the visual context that is available during the process. Finally, in reference to the time course of simulations, as discussed above, motor simulation studies suggest that subjects activate information relevant to sentence understanding as soon as the incoming linguistic input provides enough detail to understand the nature of the action that is being described. It still remains to be determined whether the same pattern applies in the case of metaphorical language.

Given these open questions, this study aims to investigate comprehenders' mental representations of literal and metaphorical motion by monitoring their eye-movements on a blank screen while listening to sentences that imply literal or metaphorical upward and downward motion. The main goals of this study are: Firstly, to compare eye-movement patterns for literal and metaphorical motion in order to determine to what extent context constrains mental simulation, that is, whether eye-movements are equally influenced by the semantics of sentence stimuli regardless of whether they depict actual physical motion or abstract metaphorical motion. Secondly, to explore whether the literal vs. metaphorical use of motion verbs modulates the time course of visual imagery and its specificity (in order to identify potential similarities or differences in the timing of literal and metaphorical language-induced mental simulation and to clarify whether space is globally or selectively recruited during simulations). Thirdly, to investigate the link between language processing and vision in a context where neither the presence of a relevant visual scene nor task demands prompt simulation.

Method

Participants

Thirty undergraduate students (18 women, M_{age} 20.3 ± 3.3) from the University of Birmingham participated in the experiment. All were native English speakers and had normal uncorrected vision or wore soft contact lenses or glasses. None reported any auditory impairments.

Stimuli

Experimental items were 40 spoken sentences implying literal or metaphorical motion in an upward or downward direction (e.g., *the curtain is rising* / *the amount is rising* vs. *the elevator is descending* / *the temperature is descending*). Metaphorical items were underpinned by the conceptual metaphors "MORE IS UP" and "LESS IS DOWN". Sentences always consisted of a two word noun phrase and a two word verb phrase in the progressive form. A total of 20 verbs were used, half of them denoting upward movement and the rest suggesting a motion path in a downward direction[4]. To ensure that the concrete and abstract nouns that functioned as subjects of the sentences did not trigger spatial biases along the vertical axis themselves, 30 native English undergraduate students (14 women, M_{age} 20.7 ± 5.3) at the CEA study abroad program in Barcelona rated them as to how strongly their meanings were associated with up or down: 1 (down associated) to 7 (up associated) (Bergen et al., 2007). Those nouns that had strong up or down associations, such as *balloon* (6.3) or *grass* (1.8), were discarded. The nouns chosen for the current study were selected from those rated between 3.5 and 4.5 (M = 3.83 ± 1.17). Forty filler sentences were also created. The filler items also contained motion verbs but these did not imply any specific direction (e.g., *to flee*, *to circulate*, *to wander*). Twenty fillers denoted literal motion and the rest metaphorical motion (see appendix for further details).

Sentences were recorded using a computer generated female American-English voice, Sally, implemented in Ivona Reader software (http://www.ivona.com/en/reader). We chose to use a text to speech software to record the stimuli because it ensures control over variables difficult to control in human voices (i.e. word stress and intonation), thus each sentence had a steady tone and intonation. The mean length of the sentences was 2,638ms and the mean onset time of the verb was 1,959ms. Finally, a visual display was created using MS Paint. This

---

[4] Verbs were selected from studies that had already normed verbs for upness and downess associations or that had studied motion simulation along the vertical axis (e.g., Berge et al., 2007; Kamide, Lindsay, Scheepers & Kukona, 2016).

display consisted of a white square (1,024 x 768 pixels) divided into four quarters by two black lines.

Procedure

The experiment was programmed using Experiment Builder (SR Research) and was structured as follows: following initial calibration and validation for each participant, each trial began with the presentation of a stationary black dot to allow for drift checking throughout the experiment. This was followed by a black fixation cross (+) presented at the center of a white screen for 1000ms. Participants were then presented with a blank display divided into four quarters, which remained on screen for 4500ms. At the same time as this visual display appeared, the sound file began playing. The monitor was then cleared and the next trial began. Some trials (32/80, or 40%) were followed by a yes-no comprehension question, to encourage attention to the spoken sentences throughout. Participants were instructed to listen carefully to each sentence while looking at the screen and to answer the comprehension questions displayed on the screen by pressing ‘A’ if the answer was “yes” and ‘L’ if the answer was “no”. Overall mean accuracy on the comprehension questions was 81% (± 39), suggesting that all participants paid close attention to the task. The experiment began with a 6-trial practice block whose stimuli were different from those used in the main experiment.

All data were collected using an Eyelink Portable Duo eye-tracker from SR Research. The experiment was run on an iMac, and the eye-tracker was positioned underneath the monitor on a desk-mounted tripod. Data was sampled from a single eye (left, unless this was problematic during setup in which case the right eye was used) at a rate of 500Hz.

Results

Data were extracted using the Eyelink DataViewer software to indicate the average gaze position throughout each trial. We divided each trial into 20ms bins, in order to analyze the change in gaze position over time. As the sample rate was 500Hz, this meant that each bin contained 10 samples, and for each bin we calculated the average y-position during that 20ms period. Bins where no samples were recorded were removed (8.9% of the data). We also removed any fixations that were recorded as being outside of the visible area of the monitor (a further 0.5% of the data). Analysis considered the average y-coordinates as the sentence unfolded to give an indication of vertical gaze position. Since the Eyelink software by default considers the top left of the screen to be the origin, values were transformed so that 0 = the mid-point of the screen; higher y-values therefore correspond to looks higher up the screen and negative values correspond to looks lower the screen. We conducted three analyses, corresponding to three distinct parts of the trial: the pre-verb region, from the start of the spoken sentence to the onset of the verb; the during-verb region, from the onset of the verb to the offset of the verb; and the post-verb region, extending for 1000ms immediately following the offset of the verb phrase. Table 1 summarizes average y-position for each section of analysis, and the overall development of average y-position (aggregated across all participants and trials) is shown in Figure 1. [5]

Table 1. Overall mean y-position (SD in brackets) for pre-verb, during-verb and post-verb sections, for literal and metaphorical verbs denoting upward and downward motion.

Figure 1. Average y-position over time (in periods of 20ms) for the pre-verb (left), during-verb (middle) and post-verb (right) sections of analysis. Top panels show literal uses of verbs while bottom panels show metaphorical uses.

---

[5] We additionally checked average x-position to determine whether there was any unexpected pattern in horizontal movement over the course of each trial. Gaze was consistently centered around x = -10, with an average change over the course of the trial of less than 4 pixels. In comparison, the average change in y-position over the course of the trial was more than 30 pixels.

Time Course Analysis of y-position

Data were analyzed using R (version 3.5.3; R Core Team, 2013) and R Studio (version 1.3.959), and the packages lme4 (version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (version 3.1-0; Kuznetsova & Brockhoff, 2017). For each section we constructed linear mixed effects models with corrected y-coordinate as the dependent variable and the interaction of Bin Index (numbered sequentially, hence bin 1 = the first 20ms, bin 2 = the next 20ms, etc.), and the sum-coded variables Direction of Motion (Up vs. Down) and Sentence Type (Literal vs. Metaphorical) as fixed effects. In order to account for the fact that consecutive bins are not independent samples, we created an autoregressive y-lag variable (the y-coordinate of the previous bin) and removed the first bin in each trial (e.g. Cho, Brown-Schmidt & Lee, 2018). Our initial model included verb length, log-transformed verb frequency (both centered) and trial order as covariates in all models. We used the drop-1 function in the lme4 package to determine that none of these made any improvement to the model for any region, so all were removed. We included random intercepts for subject, item and trial, by-subject random slopes for the interaction of and Direction and Sentence Type, and by-item random slopes for the effect of Sentence Type.

Pre-verb Region

For the region up to the onset of the verb there was a marginal main effect of Bin Index ($\beta = -0.01$, $t = -1.79$, $p = .073$) but not of Direction or Sentence Type. None of the two-way interactions of Bin Index and Direction, Bin Index and Sentence Type or Direction and Sentence Type were significant, nor was the three-way interaction of Bin Index, Direction and Sentence Type. Figure 1 (left panel) suggests that, as expected, there was little movement from the center of the screen for either literal or metaphorical verbs during this portion of the sentence.

During-verb Region

During the verb itself, there were no main effects of Bin Index, Direction or Sentence Type. There was a significant two-way interaction of Bin Index and Direction ($\beta = -0.04$, $t = -6.19$, $p < .001$) but no other two or three-way interactions were significant. Figure 1 (middle panel) shows that during this period the average gaze position moved in the direction indicated by the verb (toward the top of the screen for up verbs; toward the bottom for down verbs), with comparable effects for literal and metaphorical verbs.

Post-verb Region

In the 1000ms immediately following the offset of the verb there was a main effect of Direction ($\beta = -1.52$, $t = -2.90$, $p = .006$) but no other significant effects or interactions. Figure 1 (right panel) shows that following the offset of the verb, average gaze position continued to reflect the direction indicated by the verb for both literal and metaphorical trials.

This presents a clear picture of the development of gaze position over time, but inspection of the quantile-quantile plots for all models suggested that the residuals were not normally distributed which violates the assumptions of this kind of analysis. Although linear models are generally considered to be fairly robust against violations, we ran a second analysis using generalized linear models to better understand the pattern.[6]

Analysis of Fixation Position by Region

We created a binary variable according to whether a bin contained a fixation above or below the mid-point of the screen, hence a value of 1 corresponded to a fixation above the mid-point and a value of 0 corresponded to a fixation below the mid-point. Since Figure 1 suggests that average starting gaze position was several pixels below 0, we created a corrected mid-point

---

[6] Wiley (2020) suggests that mixed effects models with more than 6000 observations tend to be fairly robust to violations. Similarly, Minitab (2014) suggested that for any study with a sample size of at least 15, results are reliable even with residuals that are substantially different to the normal distribution.

according to the average y-position for each sentence type and direction during the 1000ms preceding each trial when a fixation cross appeared on the screen (i.e. when participants were asked to look directly at the center of the screen), hence the starting point for literal up sentences was y = -7.43, for literal down sentences was y = -7.11, for metaphorical up sentences was y = -8.19 and for metaphorical down sentences was y = -8.62. We then split each region into an early and late section in order to reflect any change in position during the region (e.g. the pattern seen in the middle panel of Figure 1 where the change in gaze begins during the verb region). Early and late were defined on a per-item basis, in order to account for variation in length during the pre-verb and verb regions.

We fitted generalized linear models with a binomial distribution to assess whether a fixation was more likely above or below the corrected center of the screen. We included the interaction of Direction of Motion (Up vs. Down) and Sentence Type (Literal vs. Metaphorical), which were both sum-coded, as fixed effects, and also included an autoregressive parameter (whether the previous bin featured a fixation above or below the corrected mid-point). We also removed the first bin in each trial, as in the time course analysis. We began with fixed effects of verb length, verb frequency and trial order and retained trial order in models where this was significant. Verb length and frequency were not significant so were removed from all models. We adopted the maximal random effects structure that did not lead to any convergence issues on a by-model basis.

Pre-verb Region

There were no significant main effects of Direction or Sentence Type, and no interactions, for the early part of the pre-verb region, nor during the later part of this region. In line with the time course analysis, and as reflected in Figure 1, gaze remained fairly constant in the center of the screen prior to the verb being heard, with fixations equally likely above or below this point.

Verb Region

There were no main effects of Direction or Sentence Type, and no interaction, during either the early or late portions of the verb region. Contrary to the pattern seen in the middle panels of Figure 1, likelihood of gaze position occurring above or below the center was not significantly affected by the direction of the verb. The lack of any interaction between Direction and Sentence Type suggested that this was true for both literal and metaphorical verbs.

Post-verb Region

During the first 500ms of the post-verb region there were main effects of Direction ($\beta = -0.27$, $z = -2.51$, $p = .012$) and Sentence Type ($\beta = -0.10$, $z = -2.45$, $p = .014$) but no interaction. The same pattern was seen in the later part of this region (500ms-1000ms post-verb), where there were main effects of Direction ($\beta = -0.28$, $z = -2.41$, $p = .016$) and Sentence Type ($\beta = -0.11$, $z = -2.50$, $p = .013$) but no interaction between the two.

Our second analysis confirms that eye-movements were congruent with the direction of the verb, with comparable effects for literal and metaphorical items, but this was only consistently present during the post-verb region. The main effects of sentence type during the post-verb region may suggest some overall difference in the magnitude of any effect for literal and metaphorical verbs, although visual inspection of the right-hand panel of Figure 1 suggests that this is not a particularly marked pattern.

Discussion

Drawing on embodied approaches to language comprehension and considering that looking behavior provides a window into ongoing cognitive processes such as language-induced mental imagery (e.g., Altman, 2004; Spivey & Geng, 2001; Huette, Winter, Matlock, Ardell & Spivey, 2014), in this study we investigated whether, even in the absence of a task or visual scene promoting mental simulation, the online processing of literal and metaphorical motion along the vertical axis influences overt spatial attention. We examined the specificity of mental simulations regarding the recruitment of vertical space (i.e., whether vertical space is used globally during mental simulations or in a way consistent with the direction of motion implied by the sentence stimuli) and their time course.

Results showed that eye-movements were biased upward and downward in accordance with the direction implied by the verb regardless of the linguistic context – literal or metaphorical – in which they appeared, and in the absence of any visual support. The average y-position increased over time when participants processed both literal and metaphorical up sentences and decreased over time when the sentences denoted literal and metaphorical down motion, with such spatial biases starting during the verb time window. Further consideration of the distribution of the data allowed us to understand this pattern more fully. Eye-movements are made up of a series of discrete fixations separated by saccades (movements from one point to another), meaning that effectively eye gaze "jumps" from one point to another. Inspection of the average y-position in each region suggested that the data had kurtosis larger than a normal distribution: values were mainly clustered around the mean (i.e. most of the time fixations were at the center point of the screen), with "heavy" tails, or a higher than normal probability of extreme values, and this pattern was most pronounced during the post-verb region. In other words, people spent most of their time looking at the center, but at some point during the verb (and more consistently in the 1000ms that followed) their gaze moved to the top or bottom of the screen, depending on the direction of motion implied by the verb, before returning to fixate on the center again. The representation in Figure 1 implies a smooth development over time, whereas a more accurate picture is that participants had one (or more) long fixations in the center portion of the screen, followed by a saccade (and subsequent fixation) to the top or bottom of the screen. Although there is some indication that this begins during the verb itself, the data suggests that this is variable across participants, and although some initiated movement during the verb, generally this saccade took place during the 1000ms following the verb. Accounting for individual differences in both the time course and extent of the mental simulation that is assumed to underpin this movement may be an important aspect of future investigations.

Our findings build on the previous literature on mental simulation in several ways. First, this study provides new evidence that motion-related literal language triggers spatial biases in visual attention (Bergen et al., 2007; Richardson et al., 2003), and that such biases are specific to the semantics of verbs, that is, the semantic processing of sentences denoting upward and downward motion does not recruit the vertical axis globally but selectively. Eye-movements bias as a function of the direction of motion encoded by the verb, i.e., up or down (Bergen, 2012; Bergen et al., 2007; Kamide et al., 2016). This fact provides further support to the claim that language understanding involves detailed mental simulations of described entities and events (Bergen, 2007; Brunyé et al., 2010; Speed & Vigliocco, 2014, among others). A consistent assumption in eye-tracking while reading is that what is being looked at is a reflection of what is being processed (what is sometimes called the eye-mind hypothesis - Just and Carpenter, 1980). Our results broadly support a picture whereby this is true not only for reading and visual processing, but that eye-movements additionally reflect biases that are triggered by auditory stimuli and the associated processes that these induce, such as mental simulation of motion events. (See also Hartmann, Mast and Fischer, 2015, for a comparable link between eye-movements and mental processes in the context of mental arithmetic, and Spivey and Geng, 2001 for eye-movements reflecting organizational and memory operations in the construction of a mental image.) The rapid generation of a saccade toward the top or bottom of the screen suggests that eye-movements are consistent with the simulation of the motion verb, even when participants are fully aware that there is nothing on the screen to look at.

Second, this study also adds to the literature by providing the first evidence, to the best of our knowledge, beyond fictive motion (e.g., Mishra & Singh, 2010; Richardson & Matlock, 2007), that metaphorical language also induces embodied simulation of upward and downward motion. Previous research has shown that processing sentences that depict actual physical motion along the vertical axis (up or down) triggers visual imagery that interferes with categorizing objects in the same part of the visual field; on the contrary, processing sentences that make metaphorical rather than literal use of motion verbs does not yield significant interference effects (Bergen et al., 2007). In our study, however, the spatial coordinates of participants' eye-movements suggest that the semantic processing of motion is supported by the oculomotor system whether the verbs convey actual physical motion or metaphorical motion, and the significant effect of phrase type in the post-verb region suggests that, if anything, this effect may be even more pronounced for metaphorical compared to literal examples (although Figure 1 suggests that any differences here they are relatively minor). A plausible explanation for our results (in comparison to studies that have found no effects for metaphorical language) might be that eye movements are a more sensitive indicator for automatic perceptual simulation in literal and metaphorical contexts than other experimental paradigms (see Altman, 2011, for a discussion of the benefits of using eye-tracking to examine the visual and cognitive processes underpinning spoken language comprehension). This would lend support to the hypothesis that failure to find interference or facilitation in a visual categorization task where the entities to be categorized appear in locations compatible with the direction implied by a previously processed metaphorical sentence does not entail that metaphorical language does not yield mental simulation. Instead, this might only suggest that the simulations of metaphorical events are less vivid (Bergen, 2012; Bergen et al., 2007; Bergen, 2005) and therefore more difficult to be observed through experimental paradigms that, as opposed to eye-tracking, do not allow for real-time monitoring of simulation during language comprehension and require additional tasks to test simulation.

An alternative explanation for the oculomotor behavior reported above is the use of the progressive aspect, which has been shown to increase the vividness of mental simulations and draw attention to the internal structure of events and the ongoing motion of described actions (Anderson et al., 2013; Bergen & Wheeler, 2010; Carreiras et al., 1997; Ferretti, Kutas & McRae, 2007; Huette et al., 2014; Madden & Zwaan, 2003). The use of such forms may therefore have induced a focus on the dynamics of motion and modulated the part of the event that participants simulated in the greatest detail, making metaphorical motion more vivid and salient in our study. In other words, grammatical structure (i.e., progressive aspect) might have influenced the extent to which the path of motion of the sentence was profiled during language processing in the present study. This might help to explain why previous studies on the interaction of language comprehension and perception found results consistent with participants mentally simulating motion in response to present tense and past tense sentences depicting actual motion events (Bergen et al., 2007; Kamide et al., 2016; Lindsay et al., 2013; Speed & Vigliocco, 2014), but failed to find similar effects for past tense sentences conveying metaphorical motion (Bergen et al., 2007).

Third, this paper also provides further evidence for spontaneous processing (i.e., "behavior and processing that occurs in the absence of an explicit task or concurrent visual referents to spoken words" – Huette et al., 2014: 2). In the present study, we tracked participants' eye movements on a blank screen while they merely listened to the sentence stimuli. The passive requirements of our experimental paradigm as well as the absence of supporting, task-relevant visual scenes rule out the possibility that participants' looking behaviors were constrained by the pictures used in the experiment or the task they were asked to perform. These experimental conditions, however, support the claim that visual simulations activate automatically while processing literal and metaphorical motion-related language. These results are in line with previous research on grammatical processing (Huette et al., 2014) and the spatial representation of time (Stocker, Hartmann, Martarelli & Mast, 2016, among others), but differ from those on fictive motion, given that no evidence of motion simulation for fictive motion sentences, as reflected in eye-movement measures, has been found in the blank screen paradigm (Mishra & Singh, 2010).

Finally, our findings also address the time course of visual simulations in literal and metaphorical contexts. In the case of literal motion, previous research has shown that verb-driven spatial biases along the vertical axis span verb-offset, when full bottom up phonological information about the verb has been gathered (Kamide et al., 2016). In the case of metaphorical language, although studies on fictive motion have made an important contribution to the analysis of visual processing in the context of figurative language (Matlock & Richardson, 2004, Richardson & Matlock, 2007; Singh & Mishra,

2010) by showing that fictive motion also induces mental simulation of physical movement, only some of these offer incrementally sampled measures (Mishra & Singh, 2010). This situation leaves open questions with respect to the time course of mental simulation. Our data, being time-locked to individual words in the target sentences, provide a window into the word-by-word processing of literal and metaphorical motion, showing that participants initiated a saccade to move their gaze in the direction indicated by the verb either during the verb itself or very soon afterward, during the 1000ms immediately following the offset of the verb.

With relation to the feature binding and feature activation accounts discussed by Kaschak and Borreggine (2008), our results confirm that a motor response can be induced as soon as a linguistic stimulus has been encountered (consistent with both accounts), but the presence of an effect for at least the next 1000ms might be taken as evidence against the feature activation account (since linguistic processing is presumed to have stopped by this point). The data are not entirely consistent with the feature binding account, however, since our assumption is that the conclusion of the sentence is also the point at which features are bound together and mental simulation begins. Kaschak and Borregine (2008) point out important task demands that may explain different patterns, so further investigation with a range of tasks, stimuli and response methods (e.g. direct comparison of eye-tracking on a blank screen vs. action responses) may be required to fully understand the unfolding processes at play during mental simulation.

Given the fact that there is evidence that directional biases in visual imagery depend on the global interpretation of the sentence rather than on lexical effects attributed to individual words (i.e., verbs or nouns with strong spatial associations) (Bergen et al., 2007; Bergen, 2012) and that mental simulations update as the text unfolds to integrate incoming information (Mannaert, Dijkstra & Zwaan, 2019), it can be argued that the findings reported here cannot be read in isolation from the context, literal or metaphorical, in which motion verbs appeared, but are the result of understanding processes affecting the sentences as a whole. Our results extend a growing body of evidence demonstrating that language comprehension generates detailed mental simulations of literal and metaphorical motion even in the absence of a relevant visual scene or a task that promotes mental imagery. This leads us to conclude that time bound shifts in visual attention during the course of language processing are not necessarily dependent on a concurrent visual scene, but on a mental simulation of the event, and that, although literal and metaphorical language may exhibit differences in the vividness of the embodied simulations that comprehenders construct (Bergen et al., 2007), a link between low-level sensorimotor features (initiation of a saccade to move the position of eye-gaze) and higher level language processing exists in both contexts.

Acknowledgements

This work was supported by the Obra Social la Caixa and University of Barcelona under grant [number LCF/PR/PR06/11010001]

We would like to thank the editor and one anonymous reviewer for helpful feedback on an earlier draft of this article. We would also like to thank Bodo Winter for his advice on aspects of data analysis.

Appendix A

Critical sentences for the study

Literal	Metaphorical
The curtain is rising	“*The amount is rising*”
The ball is soaring	“*The debt is soaring*”
The deer is climbing	“*The pound is climbing*”
The dolphins are jumping	“*The prices are jumping*”
The smoke is ascending	“*The cost is ascending*”
The bullet is rocketing	“*The business is rocketing*”
The platform is lifting	“*The mood is lifting*”
The waves are surging	“*The profits are surging*”
The bubble is levitating	“*The market is levitating*”
The fish is surfacing	“*The truth is surfacing*”
The leaves are falling	“*The sales are falling*”
The apples are dropping	“*The shares are dropping*”
The elevator is descending	“*The temperature is descending*”
The bridge is collapsing	“*The firm is collapsing*”
The building is slumping	“*The euro is slumping*”
The shelves are toppling	“*The rates are toppling*”
The ship is sinking	“*The brand is sinking*”
The walls are tumbling	“*The stocks are tumbling*”
The boat is dipping	“*The numbers are dipping*”
The water is lowering	“*The ratio is lowering*”

Appendix B

Filler sentences for the study

Literal	Metaphorical
The bus is leaving	"*The rumor is circulating*"
The soldiers are withdrawing	"*The holidays are approaching*"
The coin is rolling	"*The system is swaying*"
The tourists are embarking	"*The negotiation is progressing*"
The horse is galloping	"*The mind is wandering*"
The boy is walking	"*The economy is lumbering*"
The thief is fleeing	"*The weather is oscillating*"
The girl is swimming	"*The weekend is arriving*"
The baby is crawling	"*The time is dragging*"
The rabbit is bolting	"*The crisis is receding*"
The cyclist is sprinting	"*The days are passing*"
The snow is drifting	"*The conversation is flowing*"
The man is running	"*The discussion is meandering*"
The pony is trotting	"*The project is lurching*"
The scarf is fluttering	"*The feeling is spreading*"
The train is departing	"*The date is coming*"
The child is swinging	"*The anxiety is returning*"
The vessel is sailing	"*The threat is retreating*"
The car is zigzagging	"*The message is penetrating*"
The woman is skiing	"*The events are accelerating*"

References

Anderson, S. E., Matlock, T., and Spivey, M. J. (2013). Grammatical aspect and temporal distance in motion descriptions. *Frontiers in Psychology*. *4*:337. http://doi.org/10.3389/fpsyg.2013.00337.

Altmann, G. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, *93*(2), B79–B87.

Altmann, G. T. M. (2011). The mediation of eye movements by spoken language. In S. P. Liversedge, I. D. Gilchrsit, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 979–1003). Oxford, England: Oxford University Press.

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, 1818–1823

Bardolph, M., & Coulson, S. (2014). How vertical hand movements impact brain activity elicited by literally and metaphorically related words: An ERP study of embodied metaphors. *Frontiers in Human Neurosicence*, *8*, 1031.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J.C., Costumero, V., Parcet, M.A., and Ávila, C. (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex*, *22* (11), 2554–2563.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660. http://dx.doi.org/10.1017/s0140525x99002149.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91. http://dx.doi.org/10.1016/S1364-6613(02)00029-3.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01.

Bergen B. (2005) Mental simulation in spatial language processing. In Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.

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

Bergen, B. & Wheeler, K. 2010. Grammatical Aspect and Mental Simulation. *Brain and Language, 112*(3). 150–158.

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.

Borreggine K.L. & Kaschak, M.P. (2006). The action-sentence compatibility effect: It’s all in the timing. *Cognitive Science, 30,* 1097–1112.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Walters, E. K., & Taylor, H. A. (2010). You heard it here first: Readers mentally simulate described sounds. *Acta Psychologica*, *135*, 209– 215.

Brunyé T. T., Walters E. K., Ditman T., Gagnon S. A., Mahoney C. R., Taylor H. A. (2012). The fabric of thought: priming tactile properties during reading influences direct tactile perception. *Cognitive Science*, *36*, 1449–1467.

Cacciari, C., Bolognini, N., Senna, I., Pellicciari, M. C., Miniussi, C., & Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: A TMS study. *Brain and Language*, *119*, 149–157.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Ktanjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage,* 59, 3212–3221.

Carreiras, M., Carriedo, N., Alonso, M. A., and Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during. *Memory & Cognition*, *25*, 438–446.

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition,* 2(1), 79–116.

Cho, S., Brown-Schmidt, S. & Lee, W. (2018). Autoregressive Generalized Linear Mixed Effect Models With Crossed Random Effects: An Application to Intensive Binary Time Series Eye-Tracking Data. *Psychometrika, 83*(3), 751-771.

Citron, F. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience*, *26*, 2585–2595.

Connell, L. (2007) Representing object colour in language comprehension. *Cognition*, *102*, 476–485.

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., and Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. J. *Cognitive Neuroscience*, *23*, 2376–2386.

Desai, R. H., Conant, L. L., Binder, J. R., Park, H., & Seidenberg, M. S. (2013). A piece of the action: Modulation of sensory– motor regions by action idioms and metaphors. *Neuroimage*, 83, 862–869.

Ferretti, T. R., Kutas, M., and McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*, 182–196. https://doi.org/10.1037/0278-7393.33.1.182.

Fox, J. & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd edition). Sage, Thousand Oaks CA. http://tinyurl.com/carbook.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, *22*, 455–479.

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

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palombo, D., Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly journal of experimental psychology*, *61*, 905–919.

González, J., Barros-Loscertales, A., Pulvermuller, F., Meseguer, V., Sanjuan, A., Belloch, V., et al. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*, 906–912.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in the motor and premotor cortex. *Neuron*, *41*, 301–307.

Hartmann, M., Mast, F. W., Fischer, M. H. (2015). Spatial biases during mental arithmetic: Evidence from eye movements on a blank screen. *Frontiers in Psychology*, *6*, 12.

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, *45*(6), 1–9. https://doi.org/10.3758/s13421-017-0708-1.

Huette, S., Winter, B., Matlock, T., Ardell, D. H., & Spivey, M. (2014). Eye movements during listening reveal spontaneous grammatical processing. *Frontiers in Psychology*, *5*, 410.

Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review, 87*, 329-354.

Kamide Y, Lindsay S, Scheepers C, Kukona A. (2016)Event Processing in the Visual World: Projected Motion Paths During Spoken Sentence Comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(5),804–812.

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

Kaschak, M. P., & Borreggine, K. L. (2008). Temporal dynamics of the action-sentence compatibility effect. *The Quarterly Journal of Experimental Psychology*, 61, 883–895.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J. & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*,12224–12230.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017a). Size does matter: The implied object size is represented during language comprehension. *Discourse Processes*, *54*(7), 493-503.

de Koning, B., Wassenburg, S., Bos, L., & van der Schoot, M. (2017b). Mental simulation of four visual object properties: Similarities and differences assessed by the sentence-picture verification task. *Journal of Cognitive Psychology*, *29*(4), 420–432. https:// doi.org/10.1080/20445911.2017.1281283.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: 10.18637/jss.v082.i13.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: comprehending textural metaphors activates somatosensory cortex. *Brain and Language*, *120*, 416–421.

Lindsay, S., Scheepers, C., Kamide Y. (2013).To dash or to dawdle: verb-associated speed of motion influences eye movements during spoken sentence comprehension. *PLOS ONE*, *8*(6), e67187.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory and Cognition*, *31*, 663–672.

Mannaert, L. N., Dijkstra, K., Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory & Cognition*, *47*, 1201–1214. https:// doi.org/10.3758/s13421-019-00928-2.

Matlock, T., & Richardson, D. C. (2004). Do eye movements go with fictive motion? Paper presented at the 26th Annual Cognitive Science Society, Chicago.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Sci*ence, 15, 243–247.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs Language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013.

Minitab (2014, October 16). How Important Are Normal Residuals in Regression Analysis? [Blog post]. Retrieved from: https://blog.minitab.com/blog/adventures-in-statistics-2/how-important-are-normal-residuals-in-regression-analysis.

Mishra, R.K., & Singh, N. (2010). Online Fictive Motion Understanding: An Eye-Movement Study With Hindi. *Metaphor and Symbol*, *25*, 144–161.

Pomp, J., Bestgen, A. K., Schulze, P., Müller, C. J., Citron, F. M., Suchan, B., & Kuchinke, L. (2018). Lexical olfaction recruits olfactory orbitofrontal cortex in metaphorical and literal contexts. *Brain and Language*, *179*, 11–21.

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. *Journal of Cognitive Neuroscience*, 23, 2636–2649.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://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.

Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767–780.

Richardson, D. C., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, *102*, 129– 138.

Rohrer, T. (2001) Understanding through the Body: fMRI and of ERP studies of metaphoric and literal language. Paper presented at the 7th International Cognitive Linguistics Association Conference.

Rohrer, T. (2005) Image Schemata in the Brain. In: B. Hampe (Ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp.165–196). Berlin: Mouton de Gruyter.

Rueschemeyer, S. A., Brass, M. & Friederici, A. D. (2007). Comprehending prehending: Neural correlates of processing verbs with motor stems. *Journal of Cognitive Neuroscience* 19(5). 855–865.

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology* 89(1). 1–21.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802– 2810.

Singh, N., & Mishra, R. K. (2010). Simulating Motion in Figurative Language Comprehension. *The Open Neuroimaging Journal*, *4*, 46–52. https://doi.org/10.2174/1874440001004010046.

Speed, L. J. & Vigliocco, G. (2014) Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science, 38*(2), 367–382.

Speed, L. J., & Majid, A. (2018). An Exception to Mental Simulation: No Evidence for Embodied Odor Language. *Cognitive Science*, *42*(4), 1146–1178.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65, 235–241.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156.

Stocker, K., Hartmann, M., Martarelli, C., & Mast, F. W. (2016). Eye movements reveal mental looking through time. *Cognitive Science*, *40*, 1648–1670.

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. Neuroscice*. *17*, 273–28.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*, 185–203.

van Dam, W. O., Rueschemeyer, S. A. & Bekkering, H. (2010).How specifically are action verbs represented in the neural motor system: an fMRI study. *Neuroimage*, *53*, 1318–1325.

van Dam W.O., Speed L.J., Lai V. T., Vigliocco G., Desai R. H. (2017). Effects of motion speed in action representations. *Brain and Language*, *168*, 47–56.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wiley, J. (2020, February 25). Multilevel Models using lmer [Blog post]. Retrieved from: https://cran.r-project.org/web/packages/multilevelTools/vignettes/lmer-vignette.html.

Willems, R. M., Hagoort, P. & Casasanto, D. (2010). Body-specific representations of action verbs neural evidence from right-and left-handers. *Psychological Science*, *21*, 67–74.

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

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

Zwaan, R.A., (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Eds.), *The Psychology of Learning and Motivation* (pp. 35–62). New York: Academic Press.

Zwaan, R.A., Madden C. J., Yaxley R. H., Aveyard M.E., (2004). Moving words: dynamic mental representations in language comprehension. *Cognitive Science*, *28*, 611–619.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*(12), e51382. doi:10.1371/journal.pone.0051382.

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. Trends in Cognitive. Sciences. 18, 229–234. doi: 10.1016/j.tics.2014.02.008