C O P *i* L CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **14(1)**, Article 6: pp. 180–199, 2022 | ISSN: 2050-5949

Embodiment Effects in Idiom Processing*

T.R. Williamson
University of Cambridge

Abstract In recent research, there have been mixed results as to whether idioms with meanings related to motor action are subject to embodiment effects during semantic processing, and no behavioural study to date has illustrated any kind of differential processing for action-based idioms over non-action-based idioms. In this paper, the processing of idioms is investigated from the perspective of embodied semantics. Native English speakers with no history of dyslexia completed a semantic decision task and a semantic congruency task with stimuli containing action idioms, non-action idioms, and literal expressions. All stimulus groups were meticulously controlled and matched for a number of linguistic and psycholinguistic properties. Results indicate the influence of embodiment in idiom processing, with the semantic congruency task in particular presenting a significant positive correlation between mean response times and action-based processing facilitation.

1 Introduction

Since the foundational philosophical contribution to cognitive science of Lakoff & Johnson (1980), embodied cognition has expanded as a discipline into many fields. It is the idea that cognitive processes and representations (e.g., those concerned with language) are inextricably linked with the function and inputs of perceptual, affective, motor, introspective, interoceptive, proprioceptive, i.e., non-cognitive, areas of the brain. In other words, this school of thought proposes that there is a fundamental connection between the neurological underpinnings of human bodily function and cognition. In linguistics, the relevance of embodiment has been in understanding the neurological dynamics of language processing. No truer is this than in determining the extent of embodiment effects in the processing of semantics. Much evidence has accrued (Glenberg & Kaschak 2002, Hauk, Johnsrude & Pulvermüller 2004, Pulvermüller, Härle & Hummel 2001, Pulvermüller, Hauk, Nikulin & Ilmoniemi 2005a), albeit with some resistance (Buccino, Riggio, Melli, Binkofski, Gallese & Rizzolatti 2005, Meteyard, Bahrami & Vigliocco 2007), that sensorimotor systems have a functional role in semantic processing.

This evidence, though, has almost exclusively been for sensorimotor involvement in *literal* semantic processing. The picture is less clear for *metaphorical* semantic processing (Cacciari & Pesciarelli 2013, Desai, Binder, Conant, Mano & Seidenberg

* I give thanks to Dr John Williams, whose supervision of the MPhil thesis inspiring this paper was consistently diligent, thoughtful, and kind. The thesis' examiners, Dr Napoleon Katsos and Dr Chris Cummins, are also the recipients of my gratitude for their insightful and useful comments.

2011, Lauro, Mattavelli, Papagno & Tettamanti 2013). Opaquer still is the literature on the involvement of the sensorimotor system in *idiomatic* semantic processing. While a body of work already exists in psycholinguistics on the processing of idioms outside of an embodied context (e.g., with definitions from Kiparsky 1976 and Geoffrey Nunberg 1994; whose preferential treatment in semantic processing over literal phrases was shown by Swinney & Cutler 1979), few have approached the question from the embodied perspective.

Amongst those that do, many papers to date have not found that the processing of idioms involved embodied neurological systems (Cacciari & Pesciarelli 2013, Cuccio, Ambrosecchia, Ferri, Carapezza, Piparo, Fogassi & Gallese 2014, Desai et al. 2011, Lauro et al. 2013, Raposo, Moss, Stamatakis & Tyler 2009), suggesting that the semantic content of idioms is too abstract or disconnected from the syntactic form (or that their meanings are literally *untransparent*; Gibbs & Nayak 1989; Gibbs, Nayak & Cutting 1989) for the sensorimotor system, for instance, to be triggered in semantic processing.

However, some fMRI (functional magnetic resonance imaging) and MEG (magnetoencephalography) evidence has been found to indicate that the semantic processing of idioms is at least influenced by the sensorimotor system. Boulenger, Hauk & Pulvermuller (2009) used fMRI with literal and idiomatic sentences and both arm- and leg-related action semantics in a passive reading task that presented stimuli word-by-word. The researchers chose to take fMRI images in both early and late stages in phrasal comprehension; the former was at the onset of the final phrasal constituent and the latter occurred three seconds later. Target lexical stimuli were controlled for word frequency, lemma frequency, number of letters, number of syllables, bigram frequency, trigram frequency, and number of orthographic neighbours. Sentential stimuli were controlled for number of words, syntactic structure, and cloze probability. The findings illustrate that idiomatic sentences activated the motor cortex more than literal sentences, with both arm- and leg-related idioms also indicating differential activation patterns corresponding to respective motor-action-specific regions. Motor cortex activation increased from the early- to the late-stage imaging window across areas of the brain, including further areas of the motor cortex, for idioms.

Adding to this, Boulenger, Shtyrov & Pulvermüller (2012) used MEG to better understand the time-course effects picked up in Boulenger et al. (2009), with the same stimuli, task and procedures, only controlling for more features. Results indicate that activations of the motor cortex for idioms and literal expressions were equally early upon the onset of the target word (at around 150–200ms), irrespective of both conditions: action-based semantics and sentence type. As early as 150–180ms, differential activation patterns were also noted for the body-part-specific semantics of idioms, corresponding to arm- and leg-related motor areas respectively.

A methodological lesson from the apparent conflict in the literature on the embodied processing of idioms appears to concern the selection and controlling of stimuli features and properties. Problems include ratings' reliability from inter-study differences (Cacciari & Pesciarelli 2013), little to no controlling of stimuli properties (Cuccio et al. 2014), and no actual evidence of the employment of idioms (Raposo et al. 2009). In contrast, the stimuli used in Boulenger et al. (2009) and reused in Boulenger et al. (2012) are extensively controlled and tested. While one might argue that the need for such extensive stimuli selection and controlling procedures dampens results' impact qua applicability to the otherwise noisy and uncontrolled real world of ordinary language use, one must remember that a laboratory environment requires homogeneity for reliability and reproducibility, especially when neuroscientific methods are employed.

This paper attempts to contribute to this debate and the otherwise nascent literature on the involvement of the sensorimotor system in idiom processing using an online, behavioural paradigm (given SARS-CoV-2 restrictions). The central hypothesis is that, following careful stimuli controlling, action idioms will display a processing effect, in the form of response time (RT) facilitation, within a behavioural task relative to non-action idioms. It is predicted that this facilitation will arise due to the employment of the hands as participants' effectors in tandem with the hand-action-based semantics of idiomatic stimuli. Following a norming study to gain ratings for idioms' transparency and familiarity, a semantic decision task (judging the meaningfulness of a string of English words) and a semantic congruency task (judging the congruency of meaning between a two-sentence vignette and a one-sentence target) will be employed to gauge the extent of the processing effects for hand-action semantics. Results present a positive correlation between participants' mean RTs and action-based processing facilitation, suggesting that embodiment effects are present in the processing of idiomatic semantics at time windows indicating the comprehension of sentential meaning has occurred.

2 Methodology

2.1 Aims and hypotheses

Three studies were designed for this paper. The first was a norming study, aiming to gain ratings for idioms' transparency and familiarity prior to their inclusion within the context of a processing study. The second was a basic semantic decision task with the aim of assessing embodied semantic processing on a phrasal level and of assuring the reliability of stimuli via replicating Swinney & Cutler (1979). This has three key hypotheses: that RTs to judge semanticity for action idioms will be quicker than those for non-action idioms; that RTs for idioms will be quicker than literal equivalents; and that accuracy across stimuli groups would not be significantly different. The third was a semantic congruency task, aiming to test participants' ability at determining the semantic connection between a vignette and a target with respect to the presence of an action-semantic idiomatic phrase at the sentence-final position of the target. This also had three hypotheses: that action idiom target sentences would receive quicker congruency judgements than non-action idiom target sentences; that idiom target sentences would have quicker RTs than non-idiom target sentences; and that no accuracy difference across target sentence groups would be observed.

2.2 Participants

Fifty-seven individuals participated in the norming study, recruited from the researchers' personal network and online, via Prolific (`https://www.prolific.co`). Each was a native speaker of English with no reported history of dyslexia – no other personal data was taken by the researcher. They all expressed consent to participate. Participants from Prolific were paid at a rate of £7.12 per hour.

For the semantic decision and semantic congruency tasks, forty-four participants, who were native English speakers with no reported history of dyslexia, were recruited from the researcher's professional network. Before beginning, individuals were given information about the study at hand and expressed their consent to participate. Information about excluding participants whose RTs or accuracy represented outliers will be explained for each experiment individually. Both studies received ethical approval from the Research Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge, United Kingdom.

2.3 Stimuli selection and controlling

Stimuli were selected by gathering lexical-level linguistic and psycholinguistic features for action-related verbs from two sources: The English Lexicon Project (ELP; Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman 2007) and Gijssels & Casasanto (2020). As very little data on phrasal-level features is available, characteristics of stimuli were taken for the head verbs of the idiomatic verb phrase. This follows precedent in the literature (Barber, Otten, Kousta & Vigliocco 2013, Boulenger et al. 2009, Pulvermüller, Shtyrov & Ilmoniemi 2005b).

First, a list of 12,357 lexical items was generated from the ELP for seven linguistic and psycholinguistic properties: the number of letters, phonemes, syllables, morphemes, and log frequency (linguistic, though the latter is not linguistic-form-specific), plus concreteness and body-object interaction (BOI), which is a variable that measures a person's ease of interaction with an object in everyday situations (compare chair vs. ceiling; see Pexman, Muraki, Sidhu, Siakaluk & Yap 2018) (psycholinguistic).

This initial list was then cross-compared with a small corpus compiled by Gijssels & Casasanto (2020). Their contribution was to question the reliability of explicit, self-reported, ratings of phenomena like action-relatedness (e.g., with BOI) by designing a pantomime task where participants were instructed to act out the action underpinning one of 250 hand-action verbs in English. Video recordings of these pantomimes were taken and a different group of participants rated how much the hands were used by each participant during the pantomiming action per word. In combining these with explicit ratings themselves, they find that an action's relatedness to the hands is more nuanced and represents a continuous variable. Importantly, they present the mean scores on both the pantomime task (implicit manuality) and the ratings task (explicit manuality) per English verb. Scores for implicit and explicit manuality were reported on a -5 to +5 scale. For the purposes of the experiments in this paper, a 'strength of action representation' (SAR) score was created by averaging the implicit and explicit ratings to provide a useful way of encapsulating both measures. It was decided that a score of over 2.5 in SAR indicated that an item was suitable for selection, as a value over this score would symbolise an action representation strength closer to unimanuality (which indicates a more pronounced/obvious hand-relatedness in the action) than bimanuality (which would be a score of 0 on implicit manuality, meaning that no specific, single-hand action was implicated).

From this cross-comparison, 202 lexical items remained; that is, there were 202 hand-action-related verbs present in both the 250 rated in Gijssels & Casasanto (2020) and the 12,357 from the ELP. With items removed that had under a 2.5 score on SAR, 97 remained to be used to search for idiomatic phrases.

These 97 hand-action related verbs were entered into the Farlex Dictionary of Idioms (`https://idioms.thefreedictionary.com/`). From this list of 97, a total of 113 potential idioms were recorded. Of these, 72 were verb-initial; to keep the stimuli syntactically consistent, all idiomatic expressions started with a verb and were then either followed by a prepositional or noun phrase. To ensure the same verb or noun did not occur more than once in the action idiom stimuli group, 32 were excluded, leaving 40 action idioms in total.

Producing a list of 40 non-action idioms began with revisiting the original list of 12,357 lexical items from the ELP and all those with a BOI higher than 3.0 were excluded, leaving 1,295 lexical items. This was done to ensure that non-action verbs would have semantics with as low interaction with motor action as possible, and given that the Gijssels & Casasanto (2020) corpus did not survey non-action related words, there was not another quantifiable resource from which to draw. All of the 1,295 non-action items from the ELP that could be a verb in at least one sense (according to the researcher's intuitions) were inputted into the Farlex Dictionary of Idioms. From this, a list of 40 non-action idioms was generated.

To ensure that these two sets of 40 idioms were sufficiently homogenous, two-tailed two-sample t-tests assuming unequal variance were conducted on the scores for each idiom per feature. Idioms from action and non-action groups were not significantly different in any feature apart from BOI and concreteness; which actually covary significantly: as body-object interaction increases, concreteness also increases (β = 1.284, $t(78)$ = 9.912, p = $<$ 0.001).

2.3.1 Stimuli for the semantic decision task

Two additional stimuli groups were created for the semantic decision task: literal phrases and nonsense phrases. To ensure that control stimuli groups across the semantic decision task were homogenous, literal phrases were constructed out of the idioms by altering either the verb or (one of) the noun(s) in the idiomatic expression to make it literal (inspired by Swinney & Cutler 1979). This resulted in two kinds of literal phrase stimuli groups: 'de-action de-idiom', where the action verb from action idioms was replaced with a new verb; and 'non-action de-idiom', where the nouns from the prepositional or noun phrases of non-action idioms were replaced with new nouns. This entailed the creation of 80 additional literal phrases to act as controls for the action and non-action idiom groups.

All 80 of these items were subject to meticulous feature controlling, and results indicated no significant difference between any literal phrase group and its corresponding idiom group, apart from in the case of BOI for de-action de-idiom phrases. This suggests that the action-based semantic element from the action idioms was sufficiently removed during their creation. Moreover, no significant difference in BOI was found between non-action idioms and the non-action de-idiom stimuli, indicating that these were representative of a similar removal of action-based semantics.

To provide stimuli for the semantic decision task designed to elicit 'no' responses, nonsense phrases were created and controlled with idiomatic stimuli. These consisted of 80 meaningless strings of English words pseudorandomly generated using the script of the Bee Movie in 'Level 5'[1] of the nonsense phrase generator 'Gibberish Generator' (`https://thinkzone.wlonk.com/Gibber/GibGen.htm`). These 80 were matched with the 80 idiomatic phrases on five key linguistic features (numbers of words, syllables, phonemes, morphemes, and letters), between which there were no significant differences.

2.3.2 Stimuli for the semantic congruency task

In the semantic congruency task, two kinds of sentential stimuli were presented to participants: vignettes, which contained two sentences, and targets, which contained one. Every vignette was paired to a target sentence such that the meaning of the target was related with the meaning of the vignette (i.e., their meanings were congruent). There were four groups of target sentence: action idiom sentences, non-action idiom sentences, non-idiom sentences, and filler idiom sentences. These were defined by the construction that came at the end of the sentence; for example, if a sentence ended in an idiom with action-based semantics, then it was an action idiom sentence. There were 40 vignette-target pairs for action and non-action idioms each, 80 pairs for non-idiom phrases, and 80 target sentences with the filler idioms (there was no vignette for filler idioms because they were only ever meant as 'no' responses; they followed vignettes from the test and non-idiom groups). These were all controlled for relevant linguistic and psycholinguistic properties, as explained below.

Several rules were kept in mind when writing these stimuli: all vignettes were two sentences and all target sentences were only one sentence; all verbs were presented in a form of the past tense; all test items (action and non-action idioms, non-idiom phrases, and filler idioms) occurred in sentence-final position in the target sentences; no other idiomatic expressions occurred in either the vignettes or

[1] The program generates a random output from an input sample of text that maintains the same statistical distribution of a certain sized character cluster in the output as in the input. This certain size is determined by the Level number; using Level n on a specific input will yield an output with the same statistical distribution of n-sized character clusters (e.g., single characters at Level 1, character pairs at Level 2, etc.) as the input.

the target sentences; all sentences were in the declarative; no commas were used; and no negation of the idiom was included.

To ensure that the target sentences did not all contain idiomatic expressions at sentence-final positions, 80 non-idiom phrases were constructed to populate control sentences. As with action and non-action idioms, these were selected using the ELP in order to acquire feature ratings for the verb in log frequency and BOI. The non-idiom phrases were all phrasal-level constructions with no figurativeness of meaning involved. As they were never presented on their own, only in the context of their corresponding non-idiom sentences, they were not controlled for any of the linguistic properties by themselves. No significant difference between action and non-action idioms against non-idiom phrases was found for log frequency, and a significant difference between action idioms and non-idiom phrases for BOI was found ($p = < 0.001$) according to two-sample two-tailed *t*-tests assuming unequal variance. Filler idioms were also constructed; these were included in a target sentence that would be incongruous with a previously-presented vignette from either the action idiom, non-action idiom, or non-idiom phrase groups. They comprised idioms from English that were not used in either the action or non-action idiom groups. All filler idioms were matched with the action and non-action idioms for the five linguistic characteristics, with no significant differences found according to two-sample two-tailed *t*-tests assuming unequal variance. All the vignettes and all the target sentences were also all controlled amongst each other for these same five linguistic characteristics with two-sample, two-tailed *t*-tests assuming unequal variance, with no significant differences found between any.

Materials in the semantic congruency task were divided into four groups for separate presentation to participants. Each group contained 10 action idiom vignette-target pairs, 10 non-action idiom vignette-target pairs, and 20 non-idiom vignette-target pairs. Twenty additional non-idiom target sentences, taken from a group of stimuli that the particular participant would not be seeing, were included in each group, alongside 20 filler idiom sentences, to balance 'yes' and 'no' responses to vignettes.[2] This entailed that, for each group, participants saw 20 congruous target sentences and 20 incongruous target sentences. All participants were presented with two groups' worth of items, with a self-paced break in between groups.[3]

2.4 Procedure

Both the norming and the processing studies were completed online using personal computers, although participants had the option of using a mobile device for the former. Before each task, participants were given comprehensive explanations of the concepts at hand (e.g., definitions of transparency, familiarity, meaningfulness and semantic congruency) and afforded the opportunity to complete practice trials.

[2] In this way, there were two kinds of non-idiom target sentence: one that participants would see as congruous, and one that participants would see as incongruous. Henceforth, they shall be referred to as 'congruous non-idiom sentences' and 'incongruous non-idiom sentences'.

[3] All materials are available upon a request made via email to the author.

During the norming study, idioms were presented one-by-one, page-by-page, and participants had to manually click a separate button to progress onto the next item. Stimuli were presented in a random order, alternating transparency and familiarity questions per idiom between participants. The action and non-action idiom sets were split randomly into two groups of forty items. Participants were pseudorandomly assigned to either group so that all groups contained an equal number of responses, and thus meaning that each individual answered forty questions; for each of the twenty action and twenty non-action idioms they saw, they answered either the transparency or the familiarity question. It was designed and distributed using Qualtrics software (Qualtrics; `https://www.qualtrics.com`). Results were analysed using data analysis functions in Microsoft Excel.

Both the semantic decision and congruency tasks were distributed online to participants, who completed it external to a laboratory environment and on personal computers. Gorilla Experiment Builder (`www.gorilla.sc`) was used to create and host the experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed 2020).

For the semantic decision task, procedurally, it was important to avoid potentially confounding familiarity effects in participants seeing the same verb or noun twice in both an idiomatic and the literal expression derived from it. To rectify this, these literal phrases were pseudorandomly sorted into groups with action and non-action idioms, along with the nonsense phrases, to present to participants such that they never saw both the idiom and the literal phrase from which it was created.

To start, participants were introduced to the tasks they were about to undertake. The task required judging whether a string of words had a meaning; it was explained, for example, that 'stack the books' has a meaning, whereas 'just a on slowly' does not. Following this, information about the context of the study was given and consent was requested to participate. Instructions were given for participants to place their left forefinger on the 'f' key and their right forefinger on the 'j' key to give 'yes' and 'no' responses respectively. No overt request for participants to respond to each stimulus as quickly as they could was given.

During the task, participants were shown two groups of stimuli. RT and accuracy were both recorded. Between the presentation of each phrase, a fixation cross appeared on the screen to refocus attention, which lasted 250ms with a 100ms pause before and after, meaning stimuli were presented with a 450ms gap in total. Halfway through between stimuli groups, participants were given a break that was self-paced.

At the start of the semantic congruency task, it was explained that they would first see a 'context' (the vignette) with a length of two sentences and then a 'target' (the target sentence) with only one sentence. Their instructions were to indicate whether the meaning of the target sentence was related to, followed on from, or was otherwise connected with the vignette. It was also elaborated that making such a judgement would entail deciding if the target sentence was semantically congruous with the vignette. Participants were requested to make their responses by using keys 'f' and 'j', with left and right forefingers respectively, to make affirmative or negative judgements about congruency.

The presentation of congruous/incongruous target sentences with vignettes was pseudorandomised with the help of Gorilla Experiment Builder. Specifically, participants were directed into one of two options for each of the semantic congruency task groupings. These two options (A or B) were designed to ensure an even spread of congruous and incongruous responses to each vignette. For example, for a given vignette in groups 1 or 2, if the target sentence in the (A) option was congruous, then the target sentence in the (B) option was incongruous.

Before the presentation of each vignette, a fixation cross appeared on the screen for 250ms. A 100ms blank-screen pause came before and after each fixation. Progression past the vignette was self-paced by a press of the spacebar, following which participants were immediately presented with either a congruent or incongruent target sentence. RTs to vignettes and target sentences were both recorded, along with accuracy scores, but only latencies for target sentences are relevant for data analysis. No indication was given that participants should maintain some time-sensitivity in their reading or judgements, nor was feedback about accuracy.

2.5 Data analysis

Results from the norming study were analysed using statistical tools in Microsoft Excel. Criteria for participant exclusion in the norming task was calculated on two bases: accuracy in the pre-test training section and mean rating across the study. For the former, participants were candidates for exclusion if their rating on one of the practice examples was above 2 $\times$ SD (standard deviation) of the mean rating for that example. For the latter, participants were candidates for exclusion if their mean rating (for all the items they were presented with in the study) was outside 2 $\times$ SD of the mean of all participants' ratings across transparency and familiarity. Only one participant was a candidate for exclusion in both criteria, and thus their response was not included. This left fifty-six eligible responses, which were selected for analysis.

Responses to stimuli in the semantic decision task were analysed from the perspective of the mean RTs and accuracy scores for each participant. Responses were mainly excluded on two bases: if it was inaccurate and if the RT exceeded 2.5 $\times$ SD of that participant's mean RT.

Data analysis from the semantic congruency task was also completed from two perspectives: for both RT and accuracy. Individual responses were excluded on two grounds: inaccuracy and RT above 2.5 $\times$ SD of all responses. Two participants were excluded for having a response accuracy below 2.5 $\times$ SD of mean accuracy.

3 Results

3.1 Norming study for transparency and familiarity

Action idioms (M = 4.39, SD = 1.07) and non-action idioms (M = 4.39, SD = 1.09) were not rated significantly differently across both features (p = 0.821) in the norming study, according to a two-sample t-test assuming equal variance. This was also true for comparisons for each feature; neither transparency (M = 4.56, SD = 1.06, p = 0.700) nor familiarity (*M* = 4.14, *SD* = 1.06, $p = 0.475$) were rated significantly differently across both idiom groups. An interesting line of analysis concerns the relation between transparency and familiarity ratings themselves. While a two-sampled *t*-test assuming equal variance indicated a significant difference between the two groups ($p = 0.011$), this does not necessarily mean the two are not correlated. To assess this, regressions were carried out on transparency and familiarity scores for both action and non-action idioms. Results indicated significant positive correlations between transparency and familiarity in action idioms ($\beta = 0.413$, $t(39) = 2.606$, $p = 0.013$) and in non-action idioms ($\beta = 0.403$, $t(39) = 2.912$, $p = 0.006$).

3.2 Semantic decision task

In the raw data, the mean RT to a given stimulus was 1306.84ms (*SD* = 1084.66ms) and mean accuracy was 0.87 (*SD* = 0.33). Mean accuracy for idioms (*M* = 0.91, *SD* = 0.28) was significantly higher than for literal phrases (*M* = 0.77, *SD* = 0.42; $p = < 0.001$). This finding corroborates older studies from psycholinguistics on the processing of idioms (Ortony, Schallert, Reynolds & Antos 1978, Swinney & Cutler 1979). In terms of an embodiment effect, though accuracy is not a window into semantic processing, it is noteworthy that mean accuracy for action idioms (*M* = 0.93, *SD* = 0.25) was significantly higher than for non-action idioms (*M* = 0.89, *SD* = 0.31; $p = 0.003$).

Outlier exclusion was then carried out. Correct responses amounted to 4,610 of 5,280 total responses, indicating that 12.69% of responses were inaccurate. Of these 4,610, the RTs of 106 were 2.5 SD slower than that participant's mean RT, leaving 4,504 for data analysis. Table 1 presents mean RTs for all stimuli groups after this; at which point overall mean RT was 1238.27ms (*SD* = 799.25ms)

	Mean response time (ms)
All Idioms (Action and Non-Action)	1015.22 *(456.46)*
Action Idioms	1007.01 *(466.01)*
Non-Action Idioms	1023.821 *(446.06)*
Literal Phrases	1335.53 *(921.74)*
Nonsense Phrases	1367.38 *(893.03)*

Table 1 Mean response times in milliseconds for all stimuli groups after the removal of outlier response latencies and inaccurate responses in the semantic decision task. Standard deviations are given in brackets.

Paired, two-tailed *t*-tests were carried out on this data. To start, RTs to idioms were significantly faster than literal phrases ($p = < 0.001$), adding further support to the older findings in the psycholinguistics of idioms. Results also indicated a strong trend towards significance for RTs to action idioms over non-action idioms on by-participant means (p = 0.068), presenting a case for the involvement of embodiment effects in idiom processing.

3.3 Semantic congruency task

Before outlier exclusion, the overall mean RT for making congruency judgements on target sentences was 2065.12ms (SD = 1601.68ms), with mean accuracy at 0.91 (SD = 0.29). Means for RTs and accuracy rates per item group are given in Table 2. Table 3 below shows the means for RTs per item group after the exclusion of outliers and inaccurate responses.

	Mean response time (ms)	**Mean accuracy**
All Idiom Sentences (Action and Non-Action)	2149.96 *(1862.72)*	0.92 *(0.23)*
Action Idiom Sentences	2173.24 *(1988.63)*	0.93 *(0.26)*
Non-Action Idiom Sentences	2126.69 *(1727.35)*	0.92 *(0.27)*
All Non-Idiom Sentences (Congruous and Incongruous)	2031.66 *(1490.75)*	0.92 *(0.28)*
Congruous Non-Idiom Sentences	2057.19 *(1397.09)*	0.94 *(0.24)*
Incongruous Non-Idiom Sentences	2006.11 *(1578.48)*	0.90 *(0.30)*
Filler Idiom Sentences	2047.39 *(1527.85)*	0.881 *(0.32)*

Table 2 Mean response times in milliseconds and accuracy scores for target sentences in the semantic congruency task. Standard deviations are given in brackets.

	Mean response time (ms)
All Idioms (Action and Non-Action)	1864.61 *(992.37)*
Action Idiom Sentences	1840.94 *(967.63)*
Non-Action Idiom Sentences	1888.34 *(1016.01)*
All Non-Idiom Sentences (Congruous and Incongruous)	1826.06 *(935.04)*
Congruous Non-Idiom Sentences	1871.74 *(987.26)*
Incongruous Non-Idiom Sentences	1778.44 *(874.75)*
Filler Idiom Sentences	1798.95 *(963.08)*

Table 3 Mean response times in milliseconds for target sentences in the semantic congruency task after outlier exclusion and the removal of incorrect responses. Standard deviations are given in brackets.

During outlier exclusion, inaccurate attempts and RTs above 2.5 SD of the mean RT for all judgements were excluded. Incorrect responses totalled 320 of 3520 total responses, which represents 9.09%. Of the 3,200 remaining, 91 response latencies exceeded 2.5 SD of the mean RT, leaving 3109 for data analysis. Additionally, two participants were removed for having accuracy scores below 2.5 SD of mean accuracy across all participants. Analysis of the remaining responses did not reveal any significant difference between groups. The differences in RT between action and non-action idiom sentences and between idiom and non-idiom sentences were not found to be significant.

However, noteworthy findings arise in a comparison between participants' RTs and the difference between their RTs to action sentences and non-action sentences. This non-action minus action RT difference presents a useful indication for whether a given participant may have experienced embodiment effects during processing. For example, if mean RT for action sentences is lower than that for non-action sentences, there may have been some measure of action-based semantic processing facilitation involved. A visualisation of this comparison can be found in **??**. The line of best fit indicates a positive correlation between mean RT and non-action sentence RT minus action sentence RT; in other words, **??** suggests that, as a participant's mean RT increases, action-based semantic facilitation arises. This is attested with a regression analysis of these two variables; results indicate that this correlation is statistically significant ($r(40) = 0.402$; $p = 0.008$).

This significant correlation presents a rationale for undertaking further investigation via a *post hoc* analysis. Namely, if there is evidence to suggest that the presence of embodiment effects on congruency judgements arises at later mean response latencies, it is useful to assess whether participants with greater mean RTs themselves display an action-idiom-sentence preference in semantic processing. It makes sense to suggest that participants with later mean latencies in the semantic congruency task may have experienced action-semantic processing facilitation, as opposed to those with earlier mean latencies. This is for two reasons. First, quite simply, sentences are longer than idioms and phrases (as in the semantic decision task); they take longer to read, and reading times differ from individual to individual. Second, semantic comprehension of sentences may incur additional processing costs, which one would expect to be reflected in RT within a task investigating semantic congruency judgements.

As a point of interest motivated by curiosity, a visual inspection of the data reveals that participants with mean RTs above 1500ms (see **??**) seem to demonstrate more, and more pronounced, embodiment effects – as illustrated by the difference in non-action minus action sentence mean response latencies. When taking only participants with a mean RT above 1500ms ($n = 30$), a paired, two-tailed, two-sample t-test on participants' mean RTs to action and non-action idiom sentences revealed a significant difference. This test showed that action idiom sentences ($M = 2114.67$ms, $SD = 473.06$ms) were judged to be semantically congruous with the vignette significantly more quickly ($p = 0.043$, $df = 28$) than non-action idiom sentences ($M = 2243.66$ms, $SD = 583.29$ms).

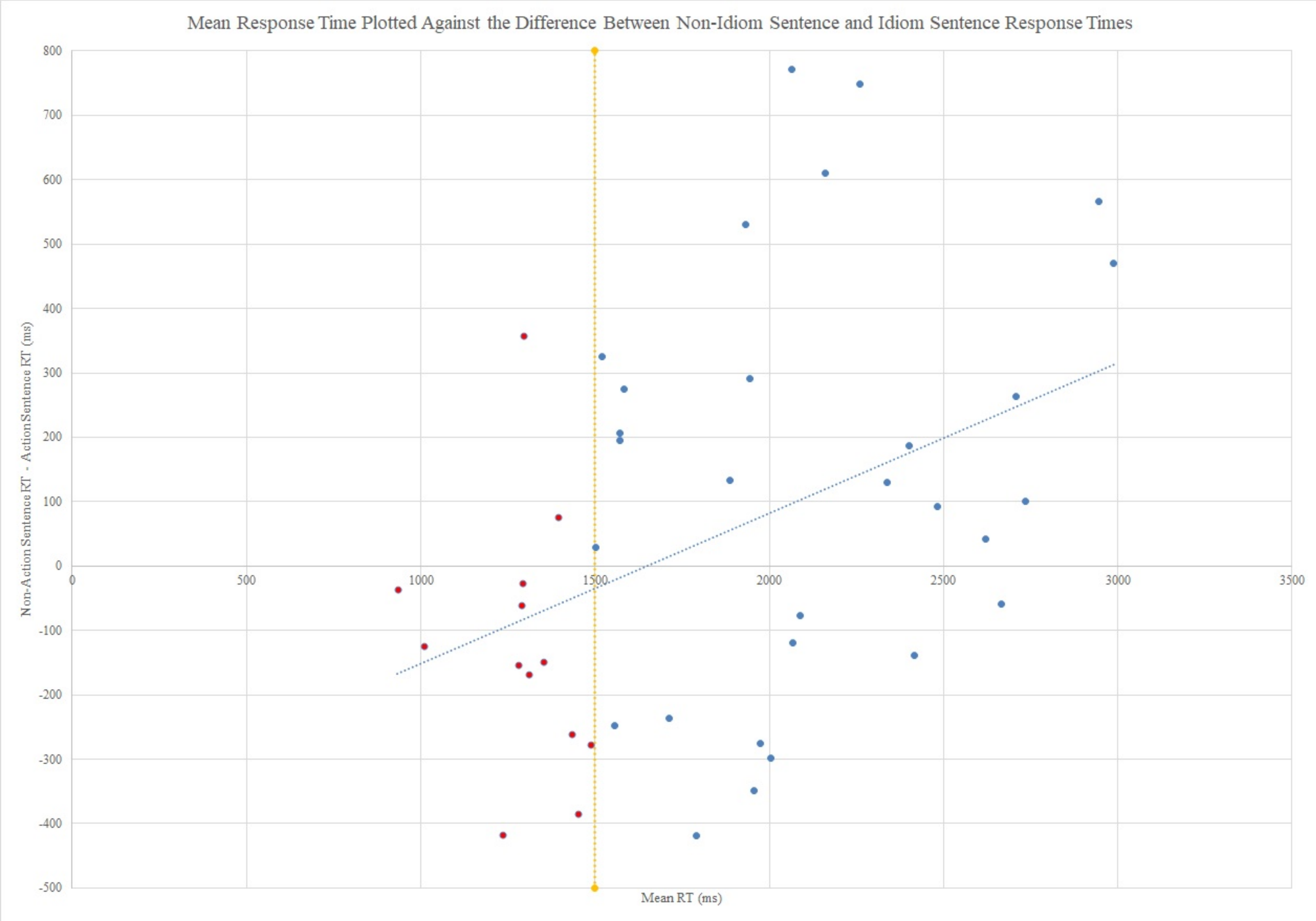

Figure 1 Mean response time of each participant plotted against their difference between non-action sentence mean response times and action sentence mean response times. The orange dotted line indicates where it is suggested that action-facilitated sentence comprehension may begin to be observed. Red dots indicate participants whose responses were excluded in the *post hoc* analysis; blue dots indicate participants whose responses were included.

4 Discussion

In this paper, three studies were carried out. The first comprised a norming study to gain feature ratings for 80 idioms in transparency and familiarity from 56 participants. Results indicated that transparency and familiarity ratings were not significantly different between action and non-action idiom groups.

In the second, a semantic decision task, participants were presented with four kinds of stimuli and requested to judge the meaningfulness of a given string of English words. Three hypotheses were attached to this study. First, that RTs to judge an action idiom as meaningful would be quicker than RTs to non-action idioms. Results indicated that RTs to action idioms trended towards being significantly quicker than non-action idioms. Second, that all idioms would receive quicker meaningfulness judgements than literal phrases; which was affirmed, and thus reproduced the findings of Swinney & Cutler (1979). Third, that participants' accuracy in responding to all item groups would not differ; it was found that the response accuracy difference was significant between idioms and non-idiom phrases and between action idioms and non-action idioms (with idioms, then action idioms, being significantly more accurate).

In the semantic congruency task, participants were presented with four kinds of vignette-target stimuli pairs and judged whether the target's meaning was related to that of the vignette. Three hypotheses were also associated with this task. First, that RTs would differ significantly for target sentences ending with an action idiom over a non-action idiom. No difference of this kind was found, but results indicate a significantly positive correlation between mean RT and an RT preference for action idiom sentences over non-action idiom sentences. *Post hoc* analyses affirmed the prediction that action idiom sentences would receive quicker judgements than non-action idiom sentences, though only amongst the three-quarters of participants whose mean RTs exceeded 1500ms. The second hypothesis for the semantic congruency task suggested that target sentences ending in any kind of idiom, action or non-action, would receive significantly quicker responses than those with literal, non-figurative meanings. This suggestion was not borne out in the data. The final hypothesis predicted that response accuracy for congruency judgements would not be significantly different between idiom and non-idiom sentences, which was confirmed.

4.1 Transparency and familiarity

Overall, there were no significant differences between action and non-action idioms for either transparency or familiarity. However, what remains curious is the correlation uncovered between ratings for both transparency and familiarity in both action and non-action idioms. At face value, it is not clear why this would be the case; transparency ratings fundamentally hinge upon a semantic decision whereas familiarity ratings are concerned with an individual's experiences. One reason could be that a participant may be more inclined to identify idioms as more or less transparent on the basis of their familiarity (or *vice versa*), which would explain why they correlate. However, this does not clarify the direction of the correlation; why was the correlation positive? One answer to this question might suggest that the more transparent an idiom is, the more salient its meaning within a speaker's mind, the more frequently it gets used, and thus the more familiar it seems. This would justify why the correlation was positive. Usage-based approaches to mental grammar might support the notion that psycholinguistic salience and frequency of use are interrelated (Bybee 2006, 2010, Bybee & Hopper 2001, Giora 1997, 2003, Ortony et al. 1978, Swinney & Cutler 1979, Williamson 2021), which supports to this explanation. However, how one gets from idiom transparency to psycholinguistic salience remains unexplained.

This explanation, though, assumes that it is transparency that ultimately affects familiarity judgements. The data do not suggest that this is necessarily the case; it is also plausible that someone's familiarity with a given idiom increases their leniency to judge that its constituents are more easily composed into its meaning. It may be that the more familiar an individual word (within an idiom) is to someone, the easier its meaning might be composable. This might be attested by the trend towards significance noted in covariance analyses between log frequency and familiarity in both action ($\beta = 0.192$, $t(39) = 1.647$, $p = 0.108$) and non-action ($\beta = 0.208$, $t(39) = 1.801$, $p = 0.08$) idioms – if familiarity is somewhat a function of log frequency, and transparency and familiarity covary, then perhaps log frequency does have a say in an item's transparency.

4.2 The embodiment correlation

In the semantic congruency task, a significantly positive correlation was observed between participants' processing preference for action idiom sentences and their overall mean response times. In other words, for a given participant, a stronger embodiment effect (represented as that individual's faster processing of action idiom target sentences over non-action idiom target sentences) was observed when their mean response time was greater. This finding is novel, even amongst the two studies reporting action-based facilitation in idiom processing (Boulenger et al. 2009, 2012).

Moreover, the result that latencies for congruency judgements were not significantly different between idiom and non-idiom sentences does not replicate many contemporary studies in the psycholinguistics of idiom processing in which sentences containing idioms are processed considerably faster than sentences with non-figurative language (Cacciari & Tabossi 1988, Carrol & Conklin 2017, Conklin & Schmitt 2008, Gibbs 1980, Rommers, Dijkstra & Bastiaansen 2013, Siyanova-Chanturia, Conklin & Schmitt 2011).

In attempting to explain why the significant correlation between mean RT and action-semantic facilitation emerged, recourse can be made to Boulenger et al. (2009). Using fMRI, Boulenger et al. (2009) observed that motor cortex activation for sentences containing action idioms actually increased from the onset of the verb of the idiom to a late-stage time window (3 seconds later) in a passive listening task. Their analysis suggests that this 3-second increase in metabolic activity was specifically a semantic, sentential comprehension effect. They argue that one word from a sentence is extremely unlikely to be processed in such depth in isolation as to dominate the brain's response to it, and as such the processing demands of a given target sentence continually recruit resources from the motor cortex throughout understanding the meaning of a sentence. The findings of Boulenger et al. (2009) are thus consistent with those in this paper; the longer a participant spent making a congruency judgement, the significantly more motor-facilitated processing resources they appeared to recruit.

At face value, a further result from the semantic congruency task seem to contradict arguments made thus far in this paper. The finding that RTs were not significantly different between action and non-action sentences seems to contribute to the literature observing no embodiment effects in idiom processing (Cacciari & Pesciarelli 2013, Cuccio et al. 2014, Desai, Conant, Binder, Park & Seidenberg 2013, Lauro et al. 2013, Raposo et al. 2009).

Crucially, however, results from the *post hoc* analysis in the semantic congruency task present findings that do illustrate embodiment effects. It was found that participants whose mean RTs to all target sentences exceeded 1500ms displayed evidence of motor-facilitated semantic processing via quicker RTs to action over non-action idiom sentences. If these results are generalisable, they represent a novel contribution to the literature on embodied semantics and idiom processing.

Indeed, there is evidence to suggest that this *post hoc* analysis is reliable. Titone & Libben (2014) found that, for idioms embedded in sentential contexts, activation of the idiom's meaning was maximal 1000ms after presentation. Thus, for idioms within sentential contexts, it may be plausible to expect embodiment effects in semantic processing after at least 1000ms plus however long it takes an individual to have read and understood the rest of the sentential content, especially given that participants were also required to make a judgment about congruency with a vignette. This, along with the fact that Boulenger et al. (2009) discovered increased activation in the motor cortices for action idioms over a 3-second window, helps to affirm the reliability of these findings.

Such a finding represents an interesting contribution to the field of embodied semantics. This is due to the presence of a common challenge (Mahon & Caramazza 2008) to the contemporary, neurocognitive programme pursued within the field. When faced with fMRI studies like those mentioned above (Boulenger et al. 2009, Hauk et al. 2004), whose stimuli frequently only reach lexical or phrasal levels of length, one can reasonably argue for an incongruity between the situations presented to participants in the laboratory and those faced in real-world discourse. As the critique goes, the latter often contains far longer (e.g., sentential-level), messier, less semantically-transparent linguistic materials, so how can fMRI studies reasonably purport to give us insight into real-world linguistic processing?

Findings of the present study, using sentence-level materials within the context of making discourse-level judgements, suggest that embodied semantics is applicable to understanding how language is processed on a level slightly closer to more natural discourse. This result implies that, even in making sentential-level semantic judgements (at time windows indicating that comprehension of sentential semantics has occurred), the motor cortex is involved and facilitates semantic processing.

Such an implication is made even more noteworthy when one considers that no behavioural study to date has yet found that action-based semantics facilitate semantic processing even on the phrasal-level of idiomatic stimuli, where there are necessarily fewer confounding variables and linguistic properties.

5 Conclusion

In this paper, evidence has been found that the processing of idioms is embodied. In responses to both semantic decision and congruency tasks, a processing facilitation was observed for action idioms via faster RTs when compared with non-action idioms. In the semantic decision task, participants judged action idioms as meaningful more quickly than non-action idioms on a trend approaching significance. In the semantic congruency task, not only was a significant correlation between participants' RTs and the extent of action-based processing facilitation observed, but participants whose mean RTs indicated semantic comprehension of the sentential-level stimuli also judged action idiom sentences as congruous with a preceding vignette significantly more quickly than non-action idiom sentences. These results represent novel findings in the field of embodied semantics for studies using behavioural methods.

A crucial orthogonal point concerns stimuli control and feature matching for studies in embodied semantics and psycholinguistics more generally. Confounding variables are highly prevalent within linguistic stimuli and can significantly influence RTs in behavioural tasks. When semantic processing is at issue, any variable that might influence this ought to be sufficiently controlled and matched for between stimuli groups to ensure results are reliable. It has been suggested that methodological flaws in stimuli controlling can cause studies not to show embodiment effects in the case of idiom processing (though this is generalisable to further explanantia and fields). The experimental design of this paper attempted to alleviate such flaws – contrasting many fMRI studies (Desai et al. 2013, Lauro et al. 2013, Raposo et al. 2009) whose methods are much finer-grain in the ability to image the motor cortex than those behavioural found here.

It would be useful to make several methodological improvements were the research repeated. First, if participants were explicitly prompted in their instructions to respond as quickly as possible, more automatic processing may be engaged in a repeat of the semantic decision task. Second, it would be useful to operationalise multiple effectors, other than the hands, for responding to hand-action idioms to ensure that it was embodiment, and not concreteness, that was responsible for the processing facilitation found here.

References

Anwyl-Irvine, A. L., J. Massonnié, A. Flitton, N. Kirkham & J. K. Evershed. 2020. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52(1). 388–407. doi:10.3758/s13428-019-01237-x.

Balota, D. A., M. J. Yap, K. A. Hutchison, M. J. Cortese, B. Kessler, B. Loftis, J. H. Neely, D. L. Nelson, G. B. Simpson & R. Treiman. 2007. The English Lexicon Project. *Behavior Research Methods* 39(3). 445–459. doi:10.3758/bf03193014.

Barber, H. A., L. J. Otten, S.-T. Kousta & G. Vigliocco. 2013. Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language* 125(1). 47–53. doi:10.1016/j.bandl.2013.01.005.

Boulenger, V., O. Hauk & F. Pulvermuller. 2009. Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. *Cerebral Cortex* 19(8). 1905–1914. doi:10.1093/cercor/bhn217.

Boulenger, V., Y. Shtyrov & F. Pulvermüller. 2012. When do you grasp the idea? MEG evidence for instantaneous idiom understanding. *NeuroImage* 59(4). 3502–3513. doi:10.1016/j.neuroimage.2011.11.011.

Buccino, G., L. Riggio, G. Melli, F. Binkofski, V. Gallese & G. Rizzolatti. 2005. Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research* 24(3). 355–363. doi:10.1016/j.cogbrainres.2005.02.020.

Bybee, J. 2006. From Usage to Grammar: The Mind’s Response to Repetition. *Language* 82(4). 711–733.

Bybee, J. 2010. *Language, use, and cognition.* Cambridge: Cambridge University Press.

Bybee, J. & P. Hopper. 2001. *Frequency and the Emergence of Linguistic Structure.* Amsterdam/Philadelphia: John Benjamins.

Cacciari, C. & F. Pesciarelli. 2013. Motor activation in literal and non-literal sentences: does time matter? *Frontiers in Human Neuroscience* 7. doi:10.3389/fnhum.2013.00202.

Cacciari, C. & P. Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27(6). 668–683. doi:10.1016/0749-596x(88)90014-9.

Carrol, G. & K. Conklin. 2017. Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea ‘has legs’. *Bilingualism: Language and Cognition* 20(2). 299–317. doi:10.1017/s1366728915000103.

Conklin, K. & N. Schmitt. 2008. Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics* 29(1). 72–89. doi:10.1093/applin/amm022.

Cuccio, V., M. Ambrosecchia, F. Ferri, M. Carapezza, F. L. Piparo, L. Fogassi & V. Gallese. 2014. How the Context Matters. Literal and Figurative Meaning in the Embodied Language Paradigm. *PLoS ONE* 9(12). e115381. doi:10.1371/journal.pone.0115381.

Desai, R. H., J. R. Binder, L. L. Conant, Q. R. Mano & M. S. Seidenberg. 2011. The Neural Career of Sensory-motor Metaphors. *Journal of Cognitive Neuroscience* 23(9). 2376–2386. doi:10.1162/jocn.2010.21596.

Desai, R. H., L. L. Conant, J. R. Binder, H. Park & M. S. Seidenberg. 2013. A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage* 83. 862–869. doi:10.1016/j.neuroimage.2013.07.044.

Geoffrey Nunberg, T. W., Ivan A. Sag. 1994. Idioms. *Language* 3(70). 491–538.

Gibbs, R. W. 1980. Spilling the beans on understanding and memory for idioms in conversation. *Memory Cognition* 8(2). 149–156.

Gibbs, R. W. & N. P. Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology* 21(1). 100–138. doi:10.1016/0010-0285(89)90004-2.

Gibbs, R. W., N. P. Nayak & C. Cutting. 1989. How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language* 28(5). 576–593. doi:10.1016/0749-596x(89)90014-4.

Gijssels, T. & D. Casasanto. 2020. Hand-use norms for Dutch and English manual action verbs: Implicit measures from a pantomime task. *Behavior Research Methods* 52(4). 1744–1767. doi:10.3758/s13428-020-01347-x.

Giora, R. 1997. Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics* 8(3). 183–206. doi:10.1515/cogl.1997.8.3.183.

Giora, R. 2003. *On Our Mind: Salience, Context, and Figurative Language*. Oxford: Oxford University Press.

Glenberg, A. M. & M. P. Kaschak. 2002. Grounding language in action. *Psychonomic Bulletin & Review* 9(3). 558–565. doi:10.3758/bf03196313.

Hauk, O., I. Johnsrude & F. Pulvermüller. 2004. Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron* 41(2). 301–307. doi:10.1016/s0896-6273(03)00838-9.

Kiparsky, P. 1976. Oral Poetry: Some Linguistic and Typological Considerations. In R. S. S. Benjamin A. Stolz (ed.), *Oral Literature and the Formula*, vol. 3, 71–106. Ann Arbor, MI: Centre for the Coordination of Ancient and Modern Studies.

Lakoff, G. & M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lauro, L. J. R., G. Mattavelli, C. Papagno & M. Tettamanti. 2013. She runs, the road runs, my mind runs, bad blood runs between us: Literal and figurative motion verbs: An fMRI study. *NeuroImage* 83. 361–371. doi:10.1016/j.neuroimage.2013.06.050.

Mahon, B. Z. & A. Caramazza. 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris* 102(1-3). 59–70. doi:10.1016/j.jphysparis.2008.03.004.

Meteyard, L., B. Bahrami & G. Vigliocco. 2007. Motion Detection and Motion Verbs. *Psychological Science* 18(11). 1007–1013. doi:10.1111/j.1467-9280.2007.02016.x.

Ortony, A., D. L. Schallert, R. E. Reynolds & S. J. Antos. 1978. Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior* 17(4). 465–477. doi:10.1016/s0022-5371(78)90283-9.

Pexman, P. M., E. Muraki, D. M. Sidhu, P. D. Siakaluk & M. J. Yap. 2018. Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods* 51(2). 453–466. doi:10.3758/s13428-018-1171-z.

Prolific. 2014 [2021]. Prolific (Version May 2021). Retrieved from `https://www.prolific.co`. Oxford, UK.

Pulvermüller, F., O. Hauk, V. V. Nikulin & R. J. Ilmoniemi. 2005a. Functional links between motor and language systems. *European Journal of Neuroscience* 21(3).

793–797.

Pulvermüller, F., M. Härle & F. Hummel. 2001. Walking or Talking?: Behavioral and Neurophysiological Correlates of Action Verb Processing. *Brain and Language* 78(2). 143–168. doi:10.1006/brln.2000.2390.

Pulvermüller, F., Y. Shtyrov & R. Ilmoniemi. 2005b. Brain Signatures of Meaning Access in Action Word Recognition. *Journal of Cognitive Neuroscience* 17(6). 884–892. doi:10.1162/0898929054021111.

Qualtrics. 2005 [2021]. Qualtrics (Version May 2021). Retrieved from `https://www.qualtrics.com`. Provo, Utah, USA.

Raposo, A., H. E. Moss, E. A. Stamatakis & L. K. Tyler. 2009. Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia* 47(2). 388–396. doi:10.1016/j.neuropsychologia.2008.09.017.

Rommers, J., T. Dijkstra & M. Bastiaansen. 2013. Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience* 25(5). 762–776. doi:10.1162/jocn_a_00337.

Siyanova-Chanturia, A., K. Conklin & N. Schmitt. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27(2). 251–272. doi:10.1177/0267658310382068.

Swinney, D. A. & A. Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18(5). 523–534. doi:10.1016/s0022-5371(79)90284-6.

Titone, D. & M. Libben. 2014. Time-dependent effects of decomposability, familiarity and literal plausibility on idiom meaning activation. *The Mental Lexicon* 9(3). 473–496. doi:10.1075/ml.9.3.05tit.

Williamson, T. R. 2021. The Graded Co-Salience Hypothesis for Polysemous Ambiguity. *Journal of the Undergraduate Linguistics Association of Britain* 1(1). 57–77.

T.R. Williamson
The University of Cambridge
`tom.r.williamson98@gmail.com`