



Cognitive Science 35 (2011) 381–398

Copyright © 2010 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/j.1551-6709.2010.01157.x

A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words

Max Louwerse,^a Louise Connell^b

^a*Department of Psychology, University of Memphis*

^b*School of Psychological Sciences, University of Manchester*

Received 14 October 2009; received in revised form 17 September 2010; accepted 19 September 2010

Abstract

Previous studies have shown that object properties are processed faster when they follow properties from the same perceptual modality than properties from different modalities. These findings suggest that language activates sensorimotor processes, which, according to those studies, can only be explained by a modal account of cognition. The current paper shows how a statistical linguistic approach of word co-occurrences can also reliably predict the category of perceptual modality a word belongs to (auditory, olfactory–gustatory, visual–haptic), even though the statistical linguistic approach is less precise than the modal approach (auditory, gustatory, haptic, olfactory, visual). Moreover, the statistical linguistic approach is compared with the modal embodied approach in an experiment in which participants verify properties that share or shift modalities. Response times suggest that fast responses can best be explained by the linguistic account, whereas slower responses can best be explained by the embodied account. These results provide further evidence for the theory that conceptual processing is both linguistic and embodied, whereby less precise linguistic processes precede precise simulation processes.

Keywords: Concepts; Embodied cognition; Linguistic Context; Modality-switch effect; Perceptual simulation; Property verification

Language processing elicits perceptual simulations. For instance, when people read the sentence “lemon can be sour,” they are faster if they have previously read the sentence “coffee can be bitter” than the sentence “stereo can be blaring.” In other words, verifying properties from different modalities (e.g., gustatory and auditory) produces switching costs. One reason for these switching costs is that the conceptual system is grounded in sensorimotor simulations.

Correspondence should be sent to Max M. Louwerse, Department of Psychology/Institute for Intelligent Systems, University of Memphis, Psychology Building, Memphis, TN 38152. E-mail: mlouwerse@memphis.edu

These findings and their conclusions are reported in Pecher, Zeelenberg, and Barsalou (2003, 2004) and provide evidence that cognition is embodied. The embodied cognition account claims that conceptual processing in cognitive tasks needs to be grounded in perceptual, perceptual, motor, and introspective states (Barsalou, 1999; Glenberg, 1997; Zwaan, 2004). According to these theories, meaning construction heavily relies on comprehenders' perceptual simulation of information communicated in language: When comprehenders read the word *sour*, they also "taste" the acid, and when we see the word *blaring* they also "hear" the noise (Pecher & Zwaan, 2005; Semin & Smith, 2008). Sensorimotor simulations during language comprehension have been found in a range of studies (see Louwerse & Jeuniaux, 2008; Pecher & Zwaan, 2005; Semin & Smith, 2008 for overviews). For instance, when participants read sentences about *ripe tomatoes* or *unripe tomatoes*, they were faster at Stroop color-naming when *tomato* appeared in the ink color implied by the sentence: red for ripe and green for unripe (Connell & Lynott, 2009; see also Connell, 2007). Even when told to expect words from a particular perceptual modality, participants needed more time to process touch-related words like *warm* or *itchy* than words relating to sight, sound, taste, or smell, reflecting a tactile disadvantage found in perceptual processing (Connell & Lynott, 2010).

The findings of modality-switching costs in Pecher et al. (2003, 2004) do not only support the embodied cognition account. Pecher et al. (2003, p. 123) also argued these findings provide evidence that the conceptual system cannot be modular and amodal. According to some amodal accounts of cognition, mental representations do not always have to be grounded, and can remain separate from sensorimotor experience, for instance because of the linguistic context (Fodor, 1975; Kintsch, 1998; Pylyshyn, 1984). That is, when comprehenders read the word *sour*, they also activate words like *lemon* and *acid* without any involvement of the gustatory cortex. Thus, useful information about the semantics of a word can arise from the association between words. Pecher and colleagues acknowledge that modality switching costs in conceptual processing could emerge from the distribution of connections between amodal representations, but they question the importance of such connections by arguing that "this idea is not supported by a number of studies that show a direct interaction between language and perception" (Van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008, p. 581).

Recently, various studies have explicitly acknowledged that conceptual processing is both linguistic and embodied (Barsalou, Santos, Simmons, & Wilson, 2008; Louwerse, 2008, 2010; Louwerse & Jeuniaux, 2008, 2010; Zwaan, 2008). That is, statistical linguistic factors and modal sensorimotor simulations interact with one another. For instance, Louwerse (2007, 2010) and Louwerse and Jeuniaux (2008, 2010) proposed the Symbol Interdependency Theory, arguing that linguistic forms are dependent on one another (linguistic context) while referring to the sensorimotor information (embodied representations). According to this theory, for shallow mental representations, the role of linguistic factors outweighs the role of embodiment factors, but for deeper mental representations the role of embodiment factors outweighs linguistic factors (Louwerse & Jeuniaux, 2010). A similar proposal can be found in Barsalou et al.'s (2008) Language and Situated Simulation (LASS) theory. According to LASS, the linguistic and simulation

systems are both engaged immediately, but the peak of linguistic activation precedes that of simulation.

Louwerse and Jeuniaux (2010) provided empirical evidence for the view that linguistic factors dominate in shallow processing and embodied factors dominate in deeper processing, by conducting four experiments in the line of Zwaan and Yaxley (2003). Word pairs were presented one underneath the other, where sometimes the word pairs had an iconic relation with the world (*attic* above *basement*) and sometimes a reverse-iconic relation (*basement* above *attic*). Louwerse and Jeuniaux identified linguistic factors (frequency of word order) and embodiment factors (iconicity ratings) and determined whether these factors differed in explaining response times under different conditions. Both factors predicted error rates and response times for both semantic and iconicity judgments of both words and pictures. However, these findings were modified by task, with the embodiment factor being strongest in iconicity judgments for pictures and the linguistic factor being strongest in semantic judgments for words. Congruent with the Symbol Interdependency and LASS theories, Louwerse and Jeuniaux's (2010) findings show that conceptual processing relies on both linguistic and embodied factors, dependent on tasks and stimuli. The results suggest that shallower processing relies more on the linguistic factor, whereas deeper processing relies more on embodiment factors. However, given that it is not possible to hold tasks and stimuli constant, it is difficult to draw conclusions on the time course of linguistic and embodiment activations in their study.

The current study had two goals. First, it aimed to investigate the extent to which statistical linguistic patterns can capture perceptual information, by testing whether the modality of a word can be predicted by linguistic context alone. Second, this study aimed to investigate whether the linguistic system precedes the simulation system during conceptual activation, as proposed by the Symbol Interdependency and LASS theories. We answered these questions using a computational linguistic algorithm on modality-specific words, and an experiment whereby participants respond to linguistic stimuli that share or shift modalities.

1. Corpus study

The aim of the first study was to determine whether a statistical linguistic approach allowed for accurate modality identification of a set of modality-specific adjectives. In other words, the research question we tested was whether it was possible to determine that *sour* refers to a gustatory modality and *blaring* to an auditory modality, solely based on the linguistic context in which these words occur. If the answer to this question is affirmative, it is difficult to argue that the statistical linguistic approach should be entirely dismissed in theories of cognition. After all, if language encodes modalities, humans might pick up on those statistical regularities in their continuous use of linguistic context in production and comprehension. On the other hand, if linguistic context does not allow for predicting the modality of a word, such a result would strengthen a purely embodied account of conceptual processing and afford dismissal of a statistical linguistic approach.

2. Method

All modality-specific words in Lynott and Connell's (2009) modality exclusivity norms were used. These norms comprise 423 adjectives, each describing an object property, where Lynott and Connell collected ratings of how strongly that property was experienced through each of five perceptual modalities: visual, haptic, auditory, olfactory, or gustatory.

Linguistic context was operationalized as the frequency of first-order co-occurrences of modality-specific words in the *Web 1T 5-gram* corpus (Brants & Franz, 2006). This corpus consists of 1 trillion word tokens (13,588,391 word types) from 95,119,665,584 sentences. The volume of the corpus allows for an extensive analysis of patterns in the English language. The frequency of co-occurrences of the 423 adjectives was computed in bigrams, trigrams, 4-grams, and 5-grams. For instance, the frequency of the words (*sour*, *bitter*) was determined by considering these words next to one another (*sour*, *bitter*), with one word in between (*sour w1 bitter*), and with two (*sour w1 w2 bitter*) or three intervening words (*sour w1 w2 w3 bitter*). This method is identical to the one used in Louwerse (2008) and Louwerse and Jeuniaux (2010).

The result of these computations was a 423×423 matrix of raw frequencies of co-occurrences, from which log frequencies were obtained. This matrix was submitted to a Principal Component Analysis with varimax rotation and Kaiser normalization. Rotation converged in 65 iterations, with 81 components being extracted. The first three components explained 31.42% of the variance, with additional components explaining less than 3% of the variance each. We therefore decided to run a factor analysis extracting only three (instead of 81) components. A total of 38.39% of the variance was explained by these three components, with the rotation converging in six iterations.

3. Comparison with human data

If a statistical linguistic approach using only co-occurrence frequencies is able to predict the modality of a particular word, then the factor loadings of the 423 words on the three components are expected to correlate with modality ratings from participants, as obtained by Lynott and Connell (2009). This was indeed the case. The higher the loading of a word on one of the three components, the higher participants rated the word as belonging to the modality corresponding to that component (Table 1).

Table 1 shows that for Component 1 both the visual and haptic modalities significantly correlate, and for Component 2 both olfactory and gustatory modalities correlate. In other words, the statistical linguistic approach is not able to distinguish between visual and haptic modalities (Component 1), nor between olfactory and gustatory modalities (Component 2). This is a weakness of the linguistic account in predicting modalities. However, there is systematicity in the inability to distinguish between these modalities. Any object that can be touched can be seen, and any object that has a taste also has a smell, so the inability to distinguish between visual and haptic modalities, and between gustatory and olfactory modalities, does not seem to be random. Moreover, when the statistical linguistic data are

Table 1
Correlations of factor loadings and participant ratings of each word's modality strength

Modality	Component 1	Component 2	Component 3
Visual	.522**	.041	-.130
Haptic	.259**	-.018	-.016
Auditory	-.240	-.488	.568**
Olfactory	-.243	.458**	-.370
Gustatory	-.233	.489**	-.393

Note. Only positive correlations are of relevance. ** $p < .01$.

compared with modal human data (Lynott & Connell, 2009), a very similar pattern emerges. In the human data the only modalities that have a positive correlation are visual and haptic, and olfactory and gustatory (see Table 2).

In other words, the statistical linguistic approach correlates with a modal simulation approach, except that the former lacks the precision of the modal approach. Whereas humans are able to distinguish between words related to visual and haptic modalities, and olfactory and gustatory modalities, the statistical linguistic approach cannot.

4. Predicting modality membership

Given that the three components correlate with three categories of modalities (visual–haptic, auditory, olfactory–gustatory), it is plausible that the factor loadings of each word on each of the three components allows for predicting which of those modality categories that word belongs to. For instance, if a word has the highest factor loading on Component 3, the prediction can be made that this word belongs to the auditory modality. If these predictions are accurate, it will show the dismissal of a statistical linguistic approach to be premature, and that its relevance should at least be reconsidered (Pecher et al., 2003, 2004).

In order to determine whether the statistical linguistic approach can distinguish between accuracy and bias, each component's performance was analyzed according to the methods of signal detection theory. In signal detection theory there are four possible outcomes for

Table 2
Correlations of participant ratings of each word's modality strength (from Lynott & Connell, 2009)

Modality	Visual	Haptic	Auditory	Olfactory	Gustatory
Visual		.38**	-.34	-.27	-.25
Haptic	.38**		-.24	-.23	-.09
Auditory	-.34	-.24		-.36	-.35
Olfactory	-.27	-.23	-.36		.78**
Gustatory	-.25	-.09	-.35	.78**	

Note. ** $p < .001$.

the detection of a signal: A hit constitutes a correct identification of the signal (e.g., system considers *sour* as being olfactory–gustatory), a false alarm constitutes a false identification of the signal (e.g., system considers *blaring* as being olfactory–gustatory), a miss constitutes an incorrect rejection of the signal (e.g., system does not consider *sour* as being olfactory–gustatory), and a correct rejection constitutes a correct rejection of the signal (e.g., system does not consider *blaring* as being olfactory–gustatory). We considered the “correct” category of each word as its dominant modality in Lynott and Connell’s (2009) norms.

From the distribution of these outcomes, the probabilities of correct [P(hit)] and of incorrect detection of the signal [P(false alarm)] can be calculated as:

1. $P(\textit{hit}) = \frac{\# \textit{ of hits}}{\# \textit{ of hits} + \# \textit{ of misses}}$
2. $P(\textit{false alarm}) = \frac{\# \textit{ of false alarms}}{\# \textit{ of false alarms} + \# \textit{ of rejections}}$

These hit and false alarm rates can then be used to determine *d'*, the discriminability performance:

3. *d'* = z-score of the hit-rate – z-score of the false-alarm-rate.

Table 3 shows that the hit rate was high, the false alarm rate was low, and therefore the system’s ability to discriminate between modalities was high. By comparison, a Monte Carlo simulation with 100 random component loadings of the 423 words revealed that the probability of random loadings having a *d'* being greater than 0.02 (75 times less than the obtained scores) was less than 1 in 100. These results provide further evidence that a statistical linguistic approach can discriminate between the three modalities. Moreover, it shows that the category of perceptual modality can be reliably predicted using linguistic context alone (see Appendix).

In addition, we can look at whether higher factor loadings for a word yields higher discriminability scores. This would then be informative for the predictability of a linguistic system and the certainty of these predictions. Factor loadings were categorized in 10 percentiles, and the *d'* for each of these 10 groups was computed. Percentile group and *d'* correlated strongly for visual–haptic ($r = .74, p < .001, n = 10$), auditory ($r = .92, p < .001, n = 10$), and olfactory–gustatory ($r = .89, p < .001, n = 10$), showing higher factor loadings yielded higher discriminability scores. In other words, not only is the statistical linguistic approach able to predict the modality of a word, it is also able to give the certainty of that prediction, with higher loadings being better predictors.

Table 3
Signal detection analysis of detection of the modality of 423 words

Modality	Hit Rate	False Alarm Rate	<i>d'</i>
Visual–Haptic	.67	.17	1.45
Auditory	.84	.10	2.29
Olfactory–Gustatory	.74	.17	1.58

5. Discussion

In a nutshell, the statistical linguistic approach was not able to distinguish visual from haptic modalities, or olfactory from gustatory modalities. However, it was able to reliably discriminate between three linguistic modality categories (visual–haptic, auditory, olfactory–gustatory), which can be seen as supersets of the five perceptual modalities (visual, haptic, auditory, olfactory, gustatory).

The finding that categories of modality can be reliably predicted through linguistic co-occurrences alone at least provides initial evidence that a statistical linguistic approach should not be entirely dismissed. Moreover, this finding supports the Symbol Interdependency Theory, which states that modality words do not become meaningful through the grounding of linguistic symbols to their referents alone but also through the linguistic context of those symbols. That is, with limited grounding, meaning induction can spread through the linguistic system.

The fact that the statistical linguistic approach was not as precise as the modal approach further supports predictions of the Symbol Interdependency Theory. It suggests that the linguistic system provides fuzzy representations, which the simulation system can in turn specify in greater detail. This then allows for the hypothesis that the linguistic system best predicts shorter response times in semantic tasks, whereas the simulation system predicts longer response times in those tasks (see also Barsalou et al., 2008; Louwerse & Jeuniaux, 2008, 2010), a hypothesis we tested in the following modality-shifting experiment.

6. Experiment

In their original study, Pecher et al. (2003: Experiment 2) attempted to address the possibility that their observed modality switching costs might emerge from associations between linguistic symbols. They argued that if associativity explained their data, then using highly associated properties in successive trials (e.g., *sheet can be spotless* → *air can be clean*) should lead to faster responses on the second item than unassociated properties (e.g., *sheet can be spotless* → *meal can be cheap*). When no such effect was found, Pecher et al. concluded that an associative explanation had been ruled out. However, their associativity measures came from free association norms (Nelson, McEvoy, & Schreiber, 1998), which are unlikely to capture the full complexity of the kind of linguistic-modality clustering that we demonstrated in our corpus study. In this experiment, therefore, we use the factor loadings from our corpus study to examine whether modality switching costs can be explained by linguistic associations as well as modality-specific embodied representations.

During processing of word-based stimuli, we argue that both linguistic and simulation representations are simultaneously activated, but that the linguistic system reaches peak activation before the simulation system (Barsalou et al., 2008; Louwerse, 2007, 2010; Louwerse & Jeuniaux, 2010). Since our corpus study shows that language alone clusters words into “linguistic modalities” (visual–haptic, olfactory–gustatory, auditory), the linguistic system therefore has useful information available to inform responses in property

Table 4

Contrasts of modality shift predictions for linguistic and embodied factors, where a modality shift represents a processing cost

Modality Transition	Linguistic Factor	Embodied Factor
Visual→Visual; Haptic→Haptic; Auditory→Auditory; Olfactory→Olfactory; Gustatory→Gustatory	Non-shift	Non-shift
Visual→Haptic; Haptic→Visual; Olfactory→Gustatory; Gustatory→Olfactory	Non-shift	Shift
Visual→Auditory; Visual→Olfactory; Visual→Gustatory; Haptic→Auditory; Haptic→Olfactory; Haptic→Gustatory; Haptic→Haptic; Auditory→Visual; Auditory→Haptic; Auditory→Olfactory; Auditory→Gustatory; Olfactory→Visual; Olfactory→Haptic; Olfactory→Auditory; Gustatory→Visual; Gustatory→Haptic; Gustatory→Auditory	Shift	Shift

verification tasks. In other words, not every word needs to be fully grounded (i.e., fully perceptually simulated) in order to produce a speeded response when a target is in the same modality as its cue. Some of the time, the linguistic relationship between modality-specific words will facilitate the target response.

Of course, the real test of our proposition comes from the mismatch between the precise simulation of five perceptual modalities and the fuzzy heuristic of three linguistic modalities. If we are correct, and some responses in property verification tasks emerge from the linguistic system rather than the simulation system, then these responses should show modality switching costs only between the three linguistic modalities and not between the five perceptual modalities (see Table 4). Statistically, since the linguistic system peaks in activation before the simulation system, most of these linguistic system responses will fall in the faster end of the distribution of property verification times. Linguistic factors (i.e., shifts between linguistic modalities) should therefore be a better predictor of fast responses than embodied factors, but embodied factors (i.e., shifts between perceptual modalities) should take over as a predictor as responses slow down and the simulation system peaks.

The data from an experiment reported in Lynott and Connell (2009) as an added evaluation study was used for the analysis of linguistic and embodied factors. We re-describe the method section of this study for clarification purposes.

7. Method

7.1. Participants

Twenty-five native speakers of English participated in the experiment for course credit. One participant was excluded for responding correctly to less than 70% of test items.

7.2. Materials

Forty strongly unimodal words were selected per modality and were attached to relevant objects (e.g., *moth: speckled*, *keys: jingling*). Two independent raters verified the appropriateness of these attributions. The pairing of each target item with its preceding modality was counterbalanced; for example, a visual item would be presented following another visual item (the same-modality condition), as well as following haptic, auditory, olfactory, and gustatory items (the different-modality conditions). Each participant saw every item, but in only one of these five possible pairs (Appendix, first column). In addition, a list of 300 object–property fillers was created, 250 false and 50 true, to provide an overall balance of 50:50 true:false responses per participant. Most of the false fillers were associated in Nelson et al.’s (1998) word association norms (e.g., *oven: baked*, *coffin: dead*) in order to ensure that the participants could not verify each property merely by word association (Solomon & Barsalou, 2004).

7.3. Procedure

The participants read instructions that asked them to press the button labeled “true” (the comma key) if the property was usually true of the concept but to press the button labeled “false” (the period key) if not. Each trial began with a fixation cross for 200 ms, followed by the item in the form “object can be property” (e.g., *moth can be speckled*), which stayed onscreen until the participant responded. The participants received immediate feedback if they responded incorrectly or too slowly (more than 2,000 ms), and each trial ended with a 200-ms blank screen. A practice session of 24 items, half true and half false, preceded the main experiment. Critical pairs and fillers appeared in a random order, with a self-paced break every 100 trials.

7.4. Design and analysis

Both the embodied and linguistic factors described modality shifts: If two consecutive properties did not share a modality, it was marked as a shift, otherwise as a non-shift. However, the factors differed in what constituted a modality. The embodied factor was operationalized as in Lynott and Connell (2009) and Pecher et al. (2003, 2004) as shifts between five perceptual modalities (visual, haptic, auditory, olfactory, gustatory). The linguistic factor was operationalized, according to the results of the corpus study, as shifts between three linguistic modalities (visual–haptic, auditory, olfactory–gustatory). The factors were therefore distinguished by certain transitions (e.g., haptic *marble can be cool* → visual *moth can be speckled*: see Table 4) being characterized as modality shifts in the embodied factor but non-shifts in the linguistic factor. Because conditions were counterbalanced by fully rotating each target item across all five possible cue modalities, the shift versus no-shift comparison remained within-items for both factors.

To investigate this further and to compare the time course of the linguistic and embodied factors, response times for the participants were divided into groups. To ensure that all 24 participants were represented in each group with on average at least six items per participant, three equal response time groups were used: fast, medium, and slow response times. The three-way split allowed the largest number of groups for examining trends of each factor while retaining sufficient data points per participant to detect a modality switch and test the time course hypotheses. For each of these RT groups there were on average 30 data points per participant, fast: $M = 30.54$, $SD = 13.11$, medium: $M = 30.5$, $SD = 7.54$, slow: $M = 29.79$, $SD = 12.82$. Because of the original 1:4 ratio of same and different modality item combinations, this meant on average 24 different modality pairs and 6 same modality pairs per participant, fast-same: $M = 23.96$, $SD = 11.07$, fast-different: $M = 6.58$, $SD = 2.76$, medium-same: $M = 24.54$, $SD = 6.55$, medium-different: $M = 5.96$, $SD = 1.85$, slow-same: $M = 24.67$, $SD = 11.71$, slow-different: $M = 5.85$, $SD = 2.23$.

For each of these groups, a mixed-effect regression model analysis was conducted on RTs with either linguistic-shifts or embodied-shifts as the fixed factor and participants and items as random factors (Baayen, Davidson, & Bates, 2008). The model was fitted using the restricted maximum likelihood estimation (REML) for the continuous variable (RT). *F*-test denominator degrees of freedom were estimated using the Kenward–Roger’s degrees of freedom adjustment to reduce the chances of Type I error (Littell, Stroup, & Freund, 2002).

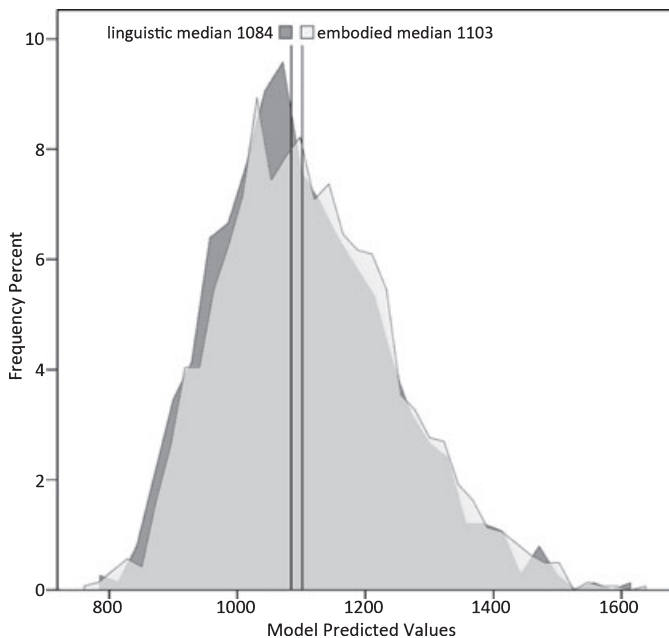


Fig. 1. Histogram of predicted RT values when a target requires a linguistic shift (dark gray) or a perceptual shift (light gray).

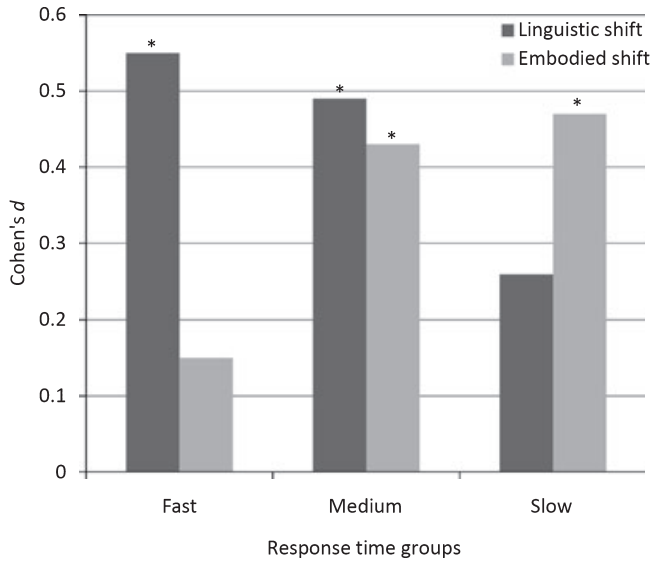


Fig. 2. Effect sizes for the linguistic and embodiment factors in each of the three response time groups. Asterisks mark significant differences ($p < .05$) in the mixed effect model analysis.

8. Results

Any target trials more than three standard deviations away from a participant's mean were removed as outliers (1.5% of the data). Only correct responses were analyzed in fast ($M = 860.18$, $SD = 72.31$, $n = 733$), medium ($M = 1,056.94$, $SD = 58.56$, $n = 732$), and slow ($M = 1,410.11$, $SD = 219.13$, $n = 733$) groups. All participants and all modalities were represented in each group. Each participant had an average of 30.32 data points ($SD = 11.64$) in each of the three response time groups. Analysis of the full dataset found an interaction between the embodied shifts and the response time groups, $F(5, 2040.85) = 1039.22$, $p < .001$, $R^2 = .72$, the linguistic shifts and the response time groups, $F(5, 2071.21) = 1035.53$, $p < .001$, $R^2 = .71$, as well as between the embodied shifts, linguistic shifts, and the response time groups, $F(11, 2092.67) = 471.39$, $p < .001$, $R^2 = .71$.

For the fast response times group, linguistic-shifts significantly predicted RTs, $F(1, 709.23) = 3.08$, $p = .03$, $d = .55$, with shifts yielding longer RTs ($M = 869.07$, $SD = 41.47$) than non-shifts ($M = 858.04$, $SD = 47.41$). No such effect was found for embodied-shifts, $F(1, 712.47) = .25$, $p = .27$, $d = .15$ (shift: $M = 865.57$, $SD = 30.70$ and non-shift: $M = 861.55$, $SD = 70.56$).

For the medium response times group, significant effects were obtained for the linguistic-shifts, $F(1, 721.44) = 3.83$, $p = .01$, $d = .49$ with shifts ($M = 1058.85$, $SD = 42.05$) yielding longer RTs than non-shifts ($M = 1050.04$, $SD = 37.43$). The same was true for the embodied-shifts, $F(1, 706.35) = 4.86$, $p = .04$, $d = .43$, with shifts ($M = 1061.11$, $SD = 21.44$) yielding longer RTs than non-shifts ($M = 1047.78$, $SD = 77.72$).

For the slow response times group, linguistic shifts did not predict RTs, $F(1, 651.98) = .81$, $p = .19$, $d = .26$ (shifts $M = 1389.52$, $SD = 131.51$; non-shifts $M = 1374.02$, $SD = 139.84$). However, embodied shifts did, $F(1, 709.37) = 3.54$, $p = .03$, $d = .47$, with modality shifts leading to longer RTs ($M = 1401.33$, $SD = 96.97$) than non-shifts ($M = 1362.21$, $SD = .216.01$).¹

What becomes clear from these findings is that the linguistic factor best explains fast RTs, whereas the embodiment factor best explains slow RTs, with the predicted RT when a target requires a shift in linguistic modalities peaking earlier than the predicted RT when the target requiring a shift in perceptual modalities (see Fig. 1). In fact, the linguistic and embodiment effects are the inverse for the fast and slow RT groups, as illustrated in the Cohen's d effect sizes in Fig. 2. When responses are made quickly, moving between haptic and visual stimuli or olfactory and gustatory stimuli does not incur a processing cost (i.e., it does not register as a shift) because these items share a common linguistic category. Only when more time is taken to make a response does a processing cost appear between such stimuli (i.e., it now registers as a shift) because different perceptual modalities are engaged during the simulation process.

In short, fuzzy regularities in linguistic context are more likely to be used for quick decisions, whereas precise perceptual simulations are more likely to be used in slower decisions.

9. General discussion

The current study shows that the category of perceptual modality a word belongs to can be predicted using linguistic context alone, even though these predictions are less precise than human predictions. When the predictions based on co-occurrences and predictions based on human ratings are compared in a modality-switching experiment, results suggest that the linguistic factor best explains shorter response times, and the embodiment factor best explains longer response times.

The findings of this study are fully in the line with the Symbol Interdependency Theory proposed elsewhere (Louwerse, 2007, 2010; Louwerse & Jeuniaux, 2008, 2010) and the LASS theory (Barsalou et al., 2008). Limited grounding of a word to its referent allows for distributing meaning across a network of linguistically related words. Each of these words could be individually grounded, but it does not have to be when semantic approximations suffice. These semantic approximations provide good-enough representations (Ferreira, Ferraro, & Bailey, 2002). For a more precise conceptual representation of a word, perceptual simulations are needed. This account would suggest that linguistic information will dominate early processing, because peak activations of the linguistic system precede those of the perceptual simulation systems. The results of response times presented in this paper show that linguistic factors best explain shorter response times, whereas embodied factors best explain longer response times.

Most replications of the modality switch effect during conceptual processing (Lynott & Connell, 2009; Marques, 2006; Pecher et al., 2003, 2004) are therefore a mix of linguistic and embodied effects: People are slower when verifying a property from a different modality to the previous trial because the linguistic relationship between properties is weaker and because attention must be decoupled from one modality-specific system and coupled to another.

An exception to the set of replications of the modality switch effect is a study by Van Dantzig et al. (2008), who, rather than presenting the cue trial in words (e.g., *leopard can be spotted* followed by *broccoli can be green*), presented it as a perceptual detection trial (e.g., visual light flash followed by *broccoli can be green*). Van Dantzig et al.'s modality switch effect is thus based on the embodied factor alone: People must re-allocate attention from one modality-specific perceptual simulation to another. Similar embodied-only findings were reported by Vermeulen, Corneille, and Niedenthal (2008), who found that a sensory memory load in a particular modality (e.g., a visual picture) interfered with verifying perceptual properties from that modality (e.g., visual *lemon can be yellow*). In both cases, combining perceptual and linguistic stimuli leaves no room for a linguistic shortcut to suffice. On the other hand, conceptual processing tasks that present all stimuli in words will always be subject to both linguistic and embodied effects, as we have demonstrated (see also Connell & Lynott, in press; Louwerse, 2010).

But why should linguistic information reflect perceptual modalities in the first place? It is possible that the distributional statistics that build up during language use are also supported by the statistical regularities of experience. For example, much of human experience of manipulable objects combine modalities (shape, size and texture = visual–haptic) as does the experience of food (flavor = olfactory–gustatory), and so the use of language reflects this modality overlap. Experience of sounds, however, is relatively distinct and so its language use is distinct also. However, even though our corpus study shows that “linguistic modalities” emerge from language use, the linguistic system itself should still be characterized as an arbitrary form of representation (Christiansen & Chater, 2008) because the linguistic forms that populate the distribution are unrelated in meaning to their referents. For instance, the visually processed orthography of the written word *sour*, or the auditorially processed phonological units of its spoken form, do not imply that its concept is grounded in either the visual or auditory modalities; rather, *sour* is grounded in the gustatory modality. However, these linguistic forms do not “contain” meaning or knowledge in their own right. Rather, the linguistic system offers a “quick and dirty” shallow heuristic that can provide good enough performance in certain tasks without recourse to deeper conceptual processing in the simulation system. The concept to which a word refers is ultimately grounded in the simulation system; however, a word does not need to be fully grounded every time it is processed.

Pecher et al. (2003, 2004) and Van Dantzig et al. (2008) question the relevance of representation other than embodied simulations in conceptual processing. We have shown that these representations should not be dismissed, because they explain behavioral data that embodiment factors do not explain. In conclusion, we would like to

advocate that a statistical linguistic account per se is incomplete, but so is a strictly modal account that explains conceptual processing. Language comprehension is both linguistic and embodied.

Note

1. Note that these findings are not affected by a potentially disproportionate number of observations in each group because of theoretical, methodological, and empirical reasons. First, slower and faster participants would still warrant the claim that fast responses are mostly linguistically driven and slow responses are mostly simulation driven. This position is also consistent with evidence of individual differences in processing meaning—those participants with poor WM tend to be more reliant on shallow lexical-level associations while those with good WM made greater use of deeper representations (i.e., simulation) of sentential context (Van Petten, Weckerly, McIsaac, & Kutas, 1997; see also Madden & Zwaan, 2006). Furthermore, fast participants and slow participants were defined as those participants with 2 *SD* above or below the mean number of cases per participants in one of the three groups. Three participants matched this criterion. When they were removed from the analysis results remained the same. The linguistic factor explained RT in the fast response group, $F(1, 647.94) = 3.41, p = .03$, but the embodiment factor did not; and the embodiment factor explained RT in the slow response group, $F(1, 579.87) = 3.89, p = .03$, but the linguistic factor did not.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavior and Brain Sciences*, *22*, 577–660.
- Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). *Language and simulation in conceptual processing*. In M. De Vega, A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press.
- Brants, T., & Franz, A. (2006). *Web 1T 5-gram Version 1*. Philadelphia: Linguistic Data Consortium.
- Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral & Brain Sciences*, *31*, 489–558.
- Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*, 476–485.
- Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic Bulletin and Review*, *16*, 573–577.
- Connell, L., & Lynott, D. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*, 1–9.
- Connell, L., & Lynott, D. (in press). Modality switching costs emerge in concept creation as well as retrieval. *Cognitive Science*.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–55.

- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Littell, R. C., Stroup, W. W., & Freund, R. J. (2002). *SAS for linear models* (4th ed.). Cary, NC: SAS Publishing.
- Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 107–120). Mahwah, NJ: Erlbaum.
- Louwerse, M. M. (2008). Embodied representations are encoded in language. *Psychonomic Bulletin and Review*, *15*, 838–844.
- Louwerse, M. M. (2010). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science (TopiCS)*. DOI: 10.1111/j.1756-8765.2010.01106.x.
- Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. de Vega, A. Glenberg & A. C. Graesser (Eds.), *Embodiment and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press.
- Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*, 96–104.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*, 558–564.
- Madden, C. J., & Zwaan, R. A. (2006). Perceptual representation as a mechanism of lexical ambiguity resolution: An investigation of span and processing time. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 1291–1303.
- Marques, J. F. (2006). Specialization and semantic organization: Evidence for multiple-semantics linked to sensory modalities. *Memory & Cognition*, *34*, 60–67.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*, 119–124.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin and Review*, *11*, 164–167.
- Pecher, D., & Zwaan, R. A. (Eds.) (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.
- Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Semin, G. R., & Smith, E. R. (Eds.) (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. New York: Cambridge University Press.
- Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*, 244–259.
- Van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, *32*, 579–590.
- Van Petten, C. M., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*, 238–242.
- Vermeulen, N., Corneille, O., & Niedenthal, P. M. (2008). Sensory load incurs conceptual processing costs. *Cognition*, *109*, 287–294.
- Zwaan, R. A. (2004). The immersed experience: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of language and motivation*, 44 (pp. 35–62). New York: Academic Press.
- Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 165–180). Oxford, England: Oxford University Press.
- Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic-relatedness judgments. *Psychonomic Bulletin and Review*, *10*, 954–958.

Appendix

Target stimuli used in the Lynott and Connell (2009) experiment with the predicted modality according to the embodiment account and the linguistic account.

Stimulus	Perceptual Modality	Linguistic Modality
Audience can be laughing	Auditory	Auditory
Baby can be gurgling	Auditory	Auditory
Barbecue can be smoky	Olfactory	Olfactory–Gustatory
Bathroom can be spotless	Visual	Visual–Haptic
Beard can be bristly	Haptic	Olfactory–Gustatory
Biscuit can be brittle	Haptic	Visual–Haptic
Bottle can be translucent	Visual	Visual–Haptic
Bouquet can be fragrant	Olfactory	Olfactory–Gustatory
Box can be rectangular	Visual	Visual–Haptic
Breath can be beery	Olfactory	Auditory
Brick can be solid	Haptic	Visual–Haptic
Brownie can be chocolatey	Gustatory	Olfactory–Gustatory
Butterscotch can be caramelized	Gustatory	Olfactory–Gustatory
Cake can be	Gustatory	Olfactory–Gustatory
Candle can be scented	Olfactory	Olfactory–Gustatory
Caramel can be gooey	Haptic	Olfactory–Gustatory
Child can be giggling	Auditory	Auditory
Chopping-boa can be onion	Olfactory	Olfactory–Gustatory
Coathanger can be bent	Visual	Visual–Haptic
Cocktail can be alcoholic	Gustatory	Visual–Haptic
Cola can be sweet	Gustatory	Visual–Haptic
Cookie can be nutty	Gustatory	Olfactory–Gustatory
Cotton-wool can be fluffy	Haptic	Visual–Haptic
Curtains can be patterned	Visual	Visual–Haptic
Dew can be glistening	Visual	Visual–Haptic
Dog can be smelly	Olfactory	Visual–Haptic
Dog can be snarling	Auditory	Auditory
Doughnut can be jammy	Gustatory	Olfactory–Gustatory
Drums can be rhythmic	Auditory	Auditory
Fire can be crackling	Auditory	Auditory
Fireworks can be sparkly	Visual	Visual–Haptic
Fish can be slippery	Haptic	Visual–Haptic
Floorboard can be creaking	Auditory	Auditory
Forest can be fresh	Olfactory	Visual–Haptic
Fridge can be chilly	Haptic	Visual–Haptic
Hand can be clammy	Haptic	Visual–Haptic
Hill can be steep	Visual	Visual–Haptic
Holly can be prickly	Haptic	Olfactory–Gustatory
Hook can be curved	Visual	Visual–Haptic
Ice-cream can be delicious	Gustatory	Olfactory–Gustatory
Incense can be flowery	Olfactory	Olfactory–Gustatory
Jaffa cake can be orangey	Gustatory	Olfactory–Gustatory
Juice can be fruity	Gustatory	Olfactory–Gustatory
Keys can be jingling	Auditory	Auditory
Kitten can be meowing	Auditory	Auditory

Appendix
(Continued)

Stimulus	Perceptual Modality	Linguistic Modality
Knife can be sharp	Haptic	Visual–Haptic
Laugh can be snorting	Auditory	Auditory
Laundry can be fresh	Olfactory	Visual–Haptic
Leaves can be rustling	Auditory	Auditory
Leopard can be spotted	Visual	Visual–Haptic
Lion can be growling	Auditory	Auditory
Man can be tall	Visual	Visual–Haptic
Manure can be reeking	Olfactory	Olfactory–Gustatory
Mattress can be lumpy	Haptic	Visual–Haptic
Meat can be rancid	Olfactory	Olfactory–Gustatory
Moisturiser can be scentless	Olfactory	Auditory
Moth can be speckled	Visual	Visual–Haptic
Moustache can be curly	Visual	Visual–Haptic
Overcoat can be damp	Haptic	Visual–Haptic
Perfume can be floral	Olfactory	Visual–Haptic
Piano can be tinkling	Auditory	Auditory
Pickles can be acidic	Gustatory	Olfactory–Gustatory
Pigsty can be stinky	Olfactory	Olfactory–Gustatory
Pizza can be cheesy	Gustatory	Olfactory–Gustatory
Pond can be murky	Visual	Olfactory–Gustatory
Raspberry can be ripe	Gustatory	Olfactory–Gustatory
Ring can be circular	Visual	Visual–Haptic
Rubbish can be rotten	Olfactory	Visual–Haptic
Sand can be gritty	Haptic	Visual–Haptic
Sandpaper can be abrasive	Haptic	Visual–Haptic
Satin can be smooth	Haptic	Visual–Haptic
Sausage can be fatty	Gustatory	Visual–Haptic
Scarf can be scratchy	Haptic	Auditory
Schnapps can be peachy	Gustatory	Olfactory–Gustatory
Seaweed can be slimy	Haptic	Visual–Haptic
Sewer can be stenchy	Olfactory	Olfactory–Gustatory
Shampoo can be lemony	Olfactory	Olfactory–Gustatory
Shout can be hoarse	Auditory	Auditory
Siren can be wailing	Auditory	Auditory
Skin can be blotchy	Visual	Olfactory–Gustatory
Snowball can be freezing	Haptic	Visual–Haptic
Soap can be lemony	Olfactory	Olfactory–Gustatory
Soup can be mushy	Gustatory	Olfactory–Gustatory
Steak can be tough	Haptic	Visual–Haptic
Stew can be meaty	Gustatory	Olfactory–Gustatory
Sugarpuffs can be honeyed	Gustatory	Olfactory–Gustatory
Sulfur can be eggy	Olfactory	Olfactory–Gustatory
Table can be sturdy	Haptic	Visual–Haptic
Thunder can be booming	Auditory	Auditory
Tide can be foamy	Visual	Olfactory–Gustatory
Tinsel can be glittery	Visual	Visual–Haptic
Toast can be burning	Olfactory	Visual–Haptic
Toffee can be chewy	Gustatory	Olfactory–Gustatory

Appendix
(Continued)

Stimulus	Perceptual Modality	Linguistic Modality
Truck can be rumbling	Auditory	Auditory
Voice can be husky	Auditory	Visual-Haptic
Wasp can be buzzing	Auditory	Auditory
Window can be misty	Visual	Visual-Haptic
Wolf can be howling	Auditory	Auditory
Woman can be petite	Visual	Visual-Haptic
Yoghurt can be creamy	Gustatory	Olfactory-Gustatory