Inferring Sensorimotor Meaning from Sublexical Form: A Computational and Human Study

Abhinav Gupta, Toben H. Mintz and Jesse Thomason
University of Southern California

Abstract

Drawing on previous work developing SENSE, a computational model that extracts sensorimotor information across 11 perceptual and motor modalities from word embeddings, we investigated whether SENSE captures sensorimotor associations in the same way humans do. We created pronounceable nonwords that lack semantic meaning to isolate sublexical processing. Over 250 participants rated which pseudowords evoked specific sensorimotor experiences (e.g., auditory, interoceptive, visual), and we compared human selection rates to SENSE's probability predictions. Results revealed striking variation across modalities: very high correlations for auditory ($r = 0.70$, $p < 0.001$) and interoceptive ($r = 0.75$, $p < 0.001$) dimensions, but no significant correlation for others. Sublexical analysis showed ..... (These findings suggest that text-based language models can capture certain dimensions of embodied meaning through distributional statistics but lack the systematic sound-symbolism humans use when processing novel words.)

1 Introduction

A central question in cognitive science is whether language understanding requires embodied, sensorimotor experience or whether meaning can be derived from linguistic co-occurrence patterns alone (need source). Recent work suggests that distributional language models can capture aspects of sensorimotor grounding despite being trained only on text (Lynott et al., 2019). Building on this, we previously developed SENSE (Sensorimotor Embedding Norm Scoring Engine), a projection model that successfully maps word embeddings onto eleven perceptual and motor dimensions from the Lancaster Sensorimotor Norms. However, a critical question remains: does SENSE capture sensorimotor associations through the same cognitive mechanisms humans use?

To address this, we conducted a behavioral study using pseudowords (pronounceable nonwords like "squeping" and "grumbered" that lack semantic meaning). Pseudowords allow us to isolate sublexical processing, the cognitive mechanism by which humans break down unfamiliar words into phonemic and orthographic units (Keuleers and Brysbaert, 2010)

2 Background

2.1 Lancaster Study

The Lancaster Sensorimotor Norms database, retrieved from the Lancaster university's website (Lynott et al., 2019) consists of norms of sensorimotor strength across six perceptual modalities (touch, hearing, smell, taste, vision, and interoception) and five action effectors (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and torso). The data was gathered by surveying 3,500 participants, who rated the degree to which each of the 39,707 words in the collection was associated with specific perceptual and action modalities, using a scale from 0 to 5.

how we use lancaster norms:

2.2 Related Work

2.2.1 Lancaster Sensorimotor Norms

The Lancaster Sensorimotor Norms database (Lynott et al., 2019)) consists of norms of sensorimotor strength across six perceptual modalities (touch, hearing, smell, taste, vision, and interoception) and five action effectors (mouth/throat, hand/arm, foot/leg, head excluding mouth/throat, and torso). The data was gathered by surveying 3,500 participants, who rated the degree to which each of the 39,707 words in the collection was associated with specific perceptual and action modalities, using a scale from 0 to 5. This dataset provides a comprehensive mapping of how English words evoke different sensory and motor experiences, making it an ideal resource for investigating whether computational models can capture embodied semantic information.

2.2.2 Phonesthemes and Sound-Meaning Correspondences

Phonesthemes are sublexical units (phonemes, morphemes, or character sequences) that carry systematic sound-meaning associations across multiple words. Classic examples include ’gl-’ in English words related to light and vision (glimmer, gleam, glisten) and ’sn-’ in words related to the nose (sniff, sneeze, snout). Previous work has attempted to develop computational methods to detect and validate phonesthemes (Sangati et al., 2013), demonstrating that these sound-symbolic patterns exist across languages and may play a role in word learning and processing. Understanding whether computational language models capture phonesthemic information could be crucial for assessing their alignment with human cognitive processing of novel words.

3 Experiments

3.1 Sensorimotor Embedding Norm Scoring Engine

The Lancaster Sensorimotor Norms dataset (Lynott et al., 2019) was used to train and evaluate the Sensorimotor Embedding Norm Scoring Engine (SENSE), a projection model that maps lexical and contextual embeddings onto the eleven perceptual and motor dimensions of the Lancaster Sensorimotor Norms.

3.1.1 Dataset Preparation

We created a parallel corpus by aligning Lancaster sensorimotor vectors with **Word2Vec**, **GloVe**, and **BERT CLS** embeddings. Since the Word2Vec and GloVe vocabularies did not completely overlap, we selected only words and phrases present in both lexical embedding vocabularies and the Lancaster Norms. For multi-word phrases, we averaged the constituent word vectors. BERT representations were extracted using the CLS token for each phrase. This yielded 34,110 aligned entries, each consisting of a Lancaster sensorimotor vector with corresponding Word2Vec, GloVe, and BERT embeddings.

We split this dataset into training, development, and test sets (70-15-15), ensuring representative coverage of sensorimotor associations across splits.

3.1.2 Model Training

We trained separate models for each embedding type (Word2Vec, GloVe, BERT). As a baseline, we predicted the mean training set sensorimotor vector for all inputs.

We then compared two models: (1) a **K-Nearest Neighbors** model (K=5) using cosine similarity with weighted averaging, and (2) a **feed-forward neural network** with one hidden layer (64 or 128 neurons, tuned on development set), ReLU activation, and 11-dimensional output.

The neural network architecture adapted to each embedding’s dimensionality (Word2Vec: 300, GloVe: 100, BERT: 768). Training used Adam optimization (learning rate 0.001) for 10 epochs with batch size 128. All models were evaluated using Mean Squared Error (MSE).

3.1.3 Results

Table 1 presents average test MSE across embedding types. Both models substantially outperformed the baseline, demonstrating that sensorimotor information is recoverable from distributional embeddings.

Performance varied by embedding type. However, the neural network showed clear advantages on BERT embeddings (0.0160 vs 0.0210), likely due to its ability to learn non-linear transformations of contextual representations. Given the neural network’s superior performance on contextual embeddings and competitive results on lexical embeddings, we chose this this model for the rest of our work and refer to it as SENSE (Sensorimotor Embedding Norm Scoring Engine) in the subsequent sections.

Notably, even static lexical embeddings (Word2Vec, GloVe) achieved low error rates, suggesting that sensorimotor grounding is encoded in distributional semantics despite these models being trained solely on co-occurrence patterns without explicit perceptual supervision.

Figure 1 shows per-dimension performance across perceptual modalities for SENSE. Despite differences in embedding architectures, all embeddings showed a similar trends across modalities, suggesting consistent patterns in how sensorimotor information is encoded in distributional semantics. Gustatory information achieved the lowest error across all embeddings, indicating that taste-related concepts are well-captured by word co-occurrence patterns. In contrast, visual and auditory dimensions showed the highest errors, sug-

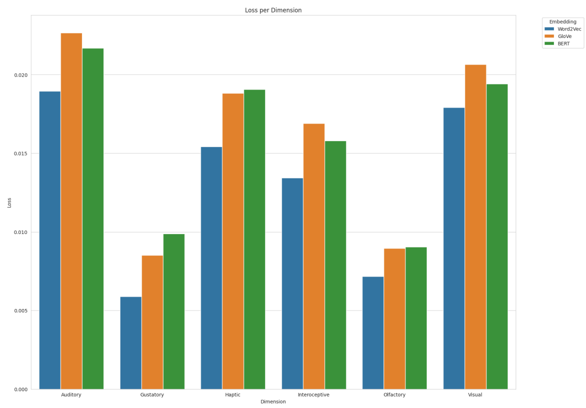

Figure 1: Mean squared error for each perceptual dimension across embedding types. Word2Vec shows consistent performance across modalities, while GloVe and BERT exhibit greater variability.

Embedding	Baseline	KNN	Neural Net
Word2Vec	0.0280	0.0138	0.0140
GloVe	0.0270	0.0170	0.0170
BERT CLS	0.0270	0.0210	0.0160

Table 1: Average MSE across models predicting sensorimotor information from different word embeddings

gesting that these perceptual modalities are less readily encoded. Haptic and interoceptive dimensions showed intermediate performance, while olfactory predictions were relatively accurate despite the abstract nature of smell descriptors.

3.2 Human Experiment

3.2.1 Method

After developing SENSE, we investigated whether the model captures sensorimotor associations through sublexical iconicity (systematic sound-meaning correspondences at the phoneme or morpheme level) rather than purely through semantic meaning learned from word co-occurrence patterns. This distinction is important because sublexical processing, the cognitive mechanism for breaking down words into phonemic or orthographic units, plays a crucial role in processing unfamiliar words, proper nouns, and second language acquisition (Ubellacker and Hillis, 2022; Indefrey and Davidson, 2009).

We hypothesized that if learned word vectors encode not only whether a word is concrete versus abstract, but what kind of concrete experience it evokes (e.g., auditory vs. visual), this information should be reflected in systematic sublexical patterns. To test this, we used pseudowords: pronounceable nonwords following English phonotactics but lacking established meaning. Since pseudowords have no semantic content and do not appear in model training data, any sensorimotor associations must arise from sublexical cues alone, allowing us to isolate phonesthemic processing from semantic knowledge.

Pseudoword Dataset Creation We generated pseudowords using the Wuggy Pseudoword Generator (Keuleers and Brysbaert, 2010), which creates pronounceable nonwords by preserving the subsyllabic structure and transition frequencies of real English words. [BRIEF WUGGY DESCRIPTION]. We used the following parameters: [PARAMETER DETAILS] applied to seed words from [SOURCE DICTIONARY].

To ensure participants had no prior semantic associations, we filtered out any pseudowords appearing in BERT's vocabulary. We also removed any pseudoword that had an ENligh homophone or a word stem that is English word. We then used SENSE to predict sensorimotor vectors for all remaining pseudowords and selected the top 12 per modality based on probability predictions (minimum 0.5 for the target modality). [INSERT TABLE OF SELECTED PSEUDOWORDS WITH EXAMPLES].

[insert table of pseudo words or atleast a brief version]

Study Procedure We designed a forced-choice survey with 4 question sets per modality (44 questions total across 11 sensorimotor dimensions). Each question presented 7 pseudowords and asked:

Which 3 of the following 7 words evoke [modality]?

For each question, 3 options were pseudowords SENSE rated highly for the target modality (probability >0.5), while the remaining 4 were pseudowords SENSE rated highly for other modalities but low for the target (probability 0.4). This design allowed us to test whether human judgments aligned with SENSE's predictions when only sublexical cues were available. We recruited [150/200?] participants through [RECRUITMENT METHOD] to complete the survey.

3.2.2 Results

We calculated the correlation between human selection rates (the proportion of participants who selected each pseudoword for a given modality) and SENSE's probability predictions. The results

Modality	**R value**	**P value**
Interoceptive	0.75	$3.75e-06$
Auditory	0.70	$3.55e-05$
Visual	0.57	$1.49e-03$
Gustatory	0.54	$2.70e-03$
Torso	0.50	$6.61-03$
Hand/Arm	0.50	$7.29e-03$
Foot/Leg	0.20	$3.11e-01$
Olfactory	0.17	$3.83e-01$
Haptic	0.13	$4.99e-01$
Head	0.13	$5.11e-01$
Mouth	-0.13	$5.09-01$

Table 2: Correlations between the human selection rate and the model's probability prediction of pseudowords for each modality, sorted in the decreasing order of R value. The p value is also noted.

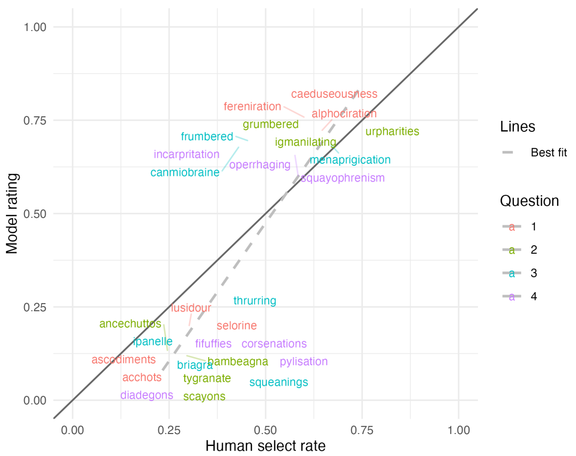

Figure 2: Graph showing the correlation between the rate of human selection vs Model's rating for words shown to the humans under the Interoceptive category

revealed striking variation across modalities in how well the model aligned with human judgments.

Table 2 presents the Pearson correlation coefficients and associated p-values for each modality, sorted in decreasing order of correlation strength. We found very high correlations ($r >= 0.65$) for interoceptive ($r = 0.75$, $p < 0.001$) and auditory ($r = 0.70$, $p < 0.001$) modalities, indicating strong alignment between SENSE's predictions and human sensorimotor associations for these dimensions. Moderate correlations emerged for visual ($r = 0.57$), gustatory ($r = 0.54$), torso ($r = 0.50$), and hand/arm ($r = 0.50$) modalities. However, we observed no significant correlation for olfactory, haptic, head, mouth, and foot/leg modalities, with correlations ranging from -0.13 to 0.20.

4 Discussion and Analysis

Our findings reveal important insights into the relationship between distributional language models and human cognitive processing of sensorimotor information. The very high correlations for interoceptive and auditory modalities ($r >= 0.65$) demonstrate that text-based models like SENSE can capture certain dimensions of embodied experience remarkably well.

References

P. Indefrey and D.J. Davidson. 2009. Second language acquisition. In Larry R. Squire, editor, *Encyclopedia of Neuroscience*, pages 517–523. Academic Press, Oxford.

Emmanuel Keuleers and Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3):627–633.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, 52(3):1271–1291.

Ekaterina Sangati, Raquel Fernández, and Federico Sangati. 2013. Automatic labeling of phonesthemic senses.

Delaney M. Ubellacker and Argye E. Hillis. 2022. Chapter 12 - the neural underpinnings of word comprehension and production: The critical roles of the temporal lobes. In Gabriele Miceli, Paolo Bartolomeo, and Vincent Navarro, editors, *The Temporal Lobe*, volume 187 of *Handbook of Clinical Neurology*, pages 211–220. Elsevier.