

Grounding Language in Sensorimotor Experience: Mapping Word Embeddings to Perceptual and Motor Dimensions Through Computational and Behavioral Methods

Abhinav Gupta

University of Southern California
abhinavg@usc.edu

Jesse Thomason

University of Southern California
jessetho@usc.edu

Toben H. Mintz

University of Southern California
tmintz@usc.edu

Abstract

This research investigates whether sensorimotor grounding is encoded in distributional word embeddings through computational modeling and human behavioral experiments. We developed SENSE (Sensorimotor Embedding Norm Scoring Engine), a neural network projection model that maps word embeddings from Word2Vec, GloVe, and BERT onto 11 sensorimotor dimensions from the Lancaster Sensorimotor Norms. The model achieved mean squared errors below 0.017 across all embedding types, demonstrating that perceptual and motor information can be reliably extracted from purely text-based representations. To investigate whether this grounding operates at the sublexical level through phonesthemic associations, we conducted a behavioral study with 265 USC participants using computationally generated pseudowords. Results revealed significant correlations between model predictions and human judgments for interoceptive ($r = 0.75, p < .001$), auditory ($r = 0.70, p < .001$), visual ($r = 0.57, p = .001$), and gustatory ($r = 0.54, p = .003$) modalities. Sublexical analysis identified phoneme patterns (e.g., “er” and “in” for auditory, “atio” and “ni” for interoceptive) systematically associated with specific sensorimotor experiences. These findings bridge computational linguistics, cognitive science, and embodied cognition theory, demonstrating that sensorimotor grounding emerges from both semantic co-occurrence patterns and sublexical phonesthemic cues in language.

1 Introduction

Human cognition relies fundamentally on sensory and motor experiences. When we understand the word “apple,” we don’t merely access an abstract symbol; we activate sensorimotor representations associated with its visual appearance, tactile qualities, gustatory properties, and the motor actions involved in grasping and biting. This embodied view

of cognition, supported by extensive neuroimaging and behavioral research, suggests that conceptual knowledge is intrinsically linked to the perceptual and motor systems that enable our interactions with the world (??).

Despite this embodied foundation of human cognition, modern natural language processing (NLP) systems rely primarily on distributional semantics, learning word representations from patterns of co-occurrence in large text corpora without direct perceptual experience. Models like Word2Vec (?), GloVe (?), and BERT (?) have achieved remarkable success in capturing semantic relationships, raising a fundamental question: can purely text-based representations capture the sensorimotor grounding that characterizes human conceptual knowledge?

This research addresses this question through two complementary approaches. First, we investigate whether sensorimotor information can be systematically extracted from distributional word embeddings using computational modeling. Second, we examine whether such grounding operates not only at the semantic level but also at the sublexical level through phonesthemes, systematic sound-meaning correspondences that may provide iconic cues to sensorimotor properties even in the absence of semantic knowledge. Understanding these mechanisms has implications for both cognitive science theories of embodied cognition and practical applications in developing more grounded artificial intelligence systems.

2 Background

2.1 Embodied Cognition and Sensorimotor Grounding

The embodied cognition framework proposes that cognitive processes are deeply rooted in the body’s interactions with the world (?). Rather than viewing the mind as an abstract symbol manipulator, embodied theories suggest that perception, action,

and conceptual knowledge are fundamentally intertwined. Neuroimaging studies have shown that processing words for manipulable objects activates motor cortex regions (?), and behavioral experiments demonstrate action-sentence compatibility effects (?). Sensorimotor grounding specifically refers to the idea that concepts are partially constituted by reactivations of the perceptual and motor experiences associated with their referents (?).

2.2 The Lancaster Sensorimotor Norms

The Lancaster Sensorimotor Norms database quantifies sensorimotor grounding across multiple modalities (Lynott et al., 2019). Collected through crowdsourcing with over 3,500 participants, the norms provide ratings for 39,707 words across 11 dimensions: six perceptual modalities (auditory, gustatory, haptic, interoceptive, olfactory, and visual) and five action effectors (foot/leg, hand/arm, head excluding mouth/throat, mouth/throat, and torso). Participants rated each word on a scale from 0 to 5 based on the extent to which they experienced it through each modality. This multidimensional characterization enables nuanced investigations of grounding beyond simple concrete versus abstract distinctions.

2.3 Distributional Semantics and Word Embeddings

Distributional semantic models learn word representations from co-occurrence patterns in text corpora. Word2Vec uses neural networks to predict context words, GloVe leverages global co-occurrence statistics, and BERT generates contextualized representations through bidirectional transformers. While these models capture semantic relationships effectively, their relationship to perceptual grounding remains underexplored. This research addresses whether perceptual and motor information can be systematically extracted from different types of embeddings.

2.4 Phonesthemes and Sound Symbolism

Although sound-meaning relationships are largely arbitrary in language, systematic exceptions exist as phonesthemes: sublexical units carrying consistent semantic or sensory associations across words (?). Classic examples include “gl-” associated with light and vision (glitter, gleam, glow) and “-ash” associated with sudden, violent motion (crash, smash, bash). Phonesthemes provide direct, non-arbitrary links between phonological form and perceptual

experience. Previous computational work has identified phonesthemes using semantic clustering (Sangati et al., 2013), but whether neural language models naturally encode such associations remains unclear.

3 Experiments

3.1 SENSE: Sensorimotor Embedding Norm Scoring Engine

3.1.1 Dataset Construction

We created a parallel corpus aligning Lancaster Sensorimotor Norms with three embedding types: Word2Vec (300 dimensions), GloVe (100 dimensions), and BERT CLS token representations (768 dimensions). We selected only words present in both lexical embedding vocabularies and the Lancaster Norms. For multi-word phrases, we averaged constituent word vectors for Word2Vec and GloVe, and extracted BERT CLS tokens from “The word is [phrase].” This yielded 34,110 entries, each with an 11-dimensional sensorimotor vector and corresponding embeddings. We split this into training (70%), development (15%), and test (15%) sets, stratifying by maximum sensorimotor dimension.

3.1.2 Model Architecture

We compared three approaches: (1) a baseline predicting mean training sensorimotor vectors, (2) K-nearest neighbors ($K=5$) using cosine similarity with weighted averaging, and (3) a feed-forward neural network with one hidden layer (64-128 neurons, tuned on development set), ReLU activation, and 11-dimensional output. The neural network was trained using Adam optimization (learning rate 0.001) for 10 epochs with batch size 128, using mean squared error (MSE) as the loss function.

3.1.3 Results

Table 1 presents average test set MSE across models and embedding types. Both KNN and neural network approaches substantially outperformed the baseline, demonstrating that sensorimotor information is recoverable from distributional embeddings. The neural network showed advantages for BERT embeddings (MSE = 0.0160 vs. 0.0210 for KNN), likely due to its ability to learn non-linear transformations. Based on this superior performance, we selected the neural network as the final SENSE model.

Figure 1 shows per-dimension performance for the SENSE model across embedding types. Despite

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: Mean Squared Error for sensorimotor prediction across embedding types.

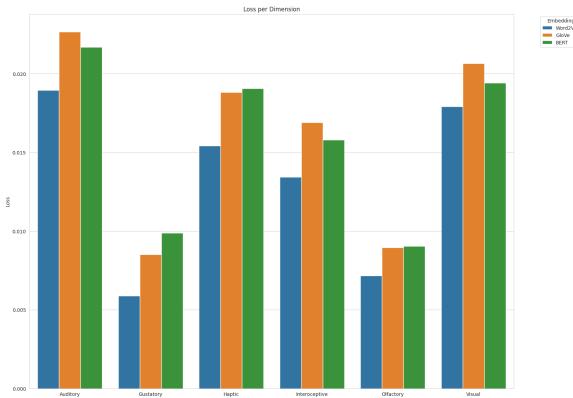


Figure 1: Mean squared error for each sensorimotor dimension across embedding types.

architectural differences, all embeddings showed similar patterns across modalities. Gustatory information achieved the lowest error, indicating taste-related concepts are well-captured by co-occurrence patterns. Visual and auditory dimensions showed higher errors, suggesting these perceptual modalities are less readily encoded. Interoceptive and haptic dimensions showed intermediate performance.

3.2 Human Behavioral Validation

3.2.1 Pseudoword Generation

To test whether SENSE captures sublexical phonosthetic associations, we used pseudowords: pronounceable nonwords following English phonotactics but lacking established meanings. We generated pseudowords using the Wuggy Pseudoword Generator (Keuleers and Brysbaert, 2010), which preserves subsyllabic structure and transition frequencies. For each word in an English dictionary, we generated 10 pseudoword candidates and removed any within one character edit distance of real words to ensure no semantic associations.

We filtered out pseudowords appearing in BERT’s vocabulary or having English homophones. Using SENSE, we predicted sensorimotor vectors for all remaining pseudowords and selected 12 per modality with scores above 0.50 for the target di-

Modality	r	p
Interoceptive	0.75	< .001
Auditory	0.70	< .001
Visual	0.57	.001
Gustatory	0.54	.003
Torso	0.50	.007
Hand/Arm	0.50	.007
Foot/Leg	0.20	.31
Olfactory	0.17	.38
Haptic	0.13	.50
Head	0.13	.51
Mouth/Throat	-0.13	.51

Table 2: Correlations between human selection rates and SENSE predictions.

dimension, creating a set of 132 pseudowords (12 per modality) predicted to strongly evoke specific sensorimotor dimensions based purely on sublexical cues.

3.2.2 Procedure

We recruited 296 participants from the USC subject pool who received course credit for participation. After excluding participants who did not complete the survey or took longer than 100 minutes, the final sample consisted of 265 participants. We designed a forced-choice survey with 4 question sets per modality (44 questions total across 11 dimensions). Each question presented 7 pseudowords and asked: “Which 3 of the following 7 words most strongly evoke [modality]?” For each question, 3 options were pseudowords SENSE rated above 0.50 for the target modality, while 4 were rated below 0.10. Each participant was randomly assigned 2 questions per modality to complete. The survey took approximately 15 minutes to complete.

3.2.3 Results

For each pseudoword, we computed the human selection rate (proportion of participants selecting it) and SENSE’s probability prediction for the target modality. Table 2 presents Pearson correlations between these measures for each modality. Significant positive correlations emerged for six modalities, with strongest effects for interoceptive ($r = 0.75, p < .001$), auditory ($r = 0.70, p < .001$), visual ($r = 0.57, p = .001$), gustatory ($r = 0.54, p = .003$), torso ($r = 0.50, p = .007$), and hand/arm ($r = 0.50, p = .007$) dimensions.

These results provide strong evidence that

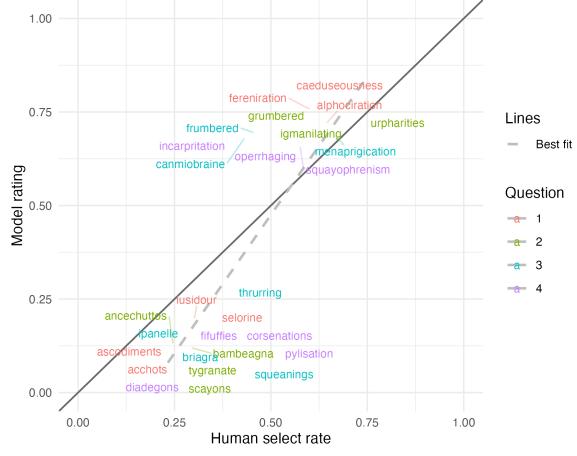


Figure 2: Correlation between human selection rate and model rating for interoceptive dimension. Each point represents a pseudoword, colored by question number. The best-fit line shows strong positive correlation ($r = 0.75$).

SENSE captures phonesthemic associations rather than purely semantic relationships. Figure 2 illustrates the correlation for the interoceptive dimension, showing that pseudowords SENSE rated highly were also frequently selected by human participants, while low-rated pseudowords were rarely selected. The diagonal pattern indicates systematic alignment between model predictions and human intuitions about sound-meaning correspondences.

3.3 Sublexical Pattern Analysis

To understand which phonological patterns drove the correlations, we analyzed sublexical units (phonemes and character sequences) overrepresented in highly-rated pseudowords. Table 3 presents example sublexical units systematically associated with specific modalities. For the auditory dimension, units “er” and “in” appeared frequently in pseudowords with high human selection rates. For the interoceptive dimension, “atio” and “ni” showed consistent associations. These patterns suggest that distributional embeddings encode systematic phonesthemic cues linking sound to sensorimotor experience.

4 Discussion

This research demonstrates that sensorimotor grounding is encoded in distributional word embeddings through both semantic co-occurrence patterns and sublexical phonesthemic associations. The SENSE model successfully extracted perceptual and motor information from text-based representa-

Sublexical Unit	Modality	Example Pseudowords
er	Auditory	frumbered, grumbered
in	Auditory	igmanilating, squeanings
atio	Interoceptive	alphodration, fereniration
ni	Interoceptive	igmanilating, menaprigication

Table 3: Example sublexical units associated with specific sensorimotor modalities.

tions, while pseudoword experiments revealed that this grounding operates at the phonological level independent of semantic content.

4.1 Implications and Applications

For cognitive science, these findings demonstrate that sensorimotor grounding can emerge from distributional statistics, suggesting perceptual experience and linguistic co-occurrence are tightly coupled. For artificial intelligence, SENSE provides a method for augmenting language models with explicit sensorimotor representations, potentially improving performance on tasks requiring grounded understanding, such as robot language understanding and multimodal AI systems. The phonesthemic patterns could also be leveraged for creative language generation tasks like brand naming or poetic language generation.

5 Conclusion

This research reveals that language encodes sensorimotor information through multiple complementary mechanisms operating at both semantic and sublexical levels. By demonstrating that distributional embeddings capture both semantic co-occurrence patterns and phonesthemic associations, these findings bridge computational linguistics, cognitive science, and embodied cognition theory. The work advances understanding of how symbolic language maintains connections to the perceptual experiences grounding human cognition and provides practical tools for developing more cognitively aligned AI systems.

Limitations

Several limitations warrant discussion. First, the Lancaster Norms represent subjective ratings from English speakers, and sensorimotor associations may vary across languages and cultures. Second, while pseudowords lack established meanings, participants may generate ad hoc semantic interpre-

tations based on similarity to real words. Third, SENSE operates as a static mapping without capturing dynamic aspects of sensorimotor simulation important for human cognition. Future work could explore context-sensitive approaches and extend the analysis to compositional semantics at the phrase and sentence level.

Acknowledgments

We thank the USC subject pool participants for their contributions to the behavioral study.

References

- 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 phonestemic senses.