PHONOSEMANTIC LAYER IN LARGE LANGUAGE MODELS

The Missing Phonosemantic Layer in Large Language Models: How Odam Tili Theory Exposes the Biological Blind Spot of Artificial Intelligence

Mahmudjon Kuchkarov
Language Academy 'Odam Tili', Orlando, Florida, USA

Mahmudjon Kuchkarov, Language Academy 'Odam Tili', Orlando, Florida, USA; Esta Florida Tennis Academy, Orlando, Florida, USA.

This research received no external funding. The author declares no competing interests. The experimental data underlying this study are available upon request.

Correspondence concerning this article should be addressed to Mahmudjon Kuchkarov, Language Academy 'Odam Tili', Orlando, Florida, USA. Email: odamtilitheory@gmail.com

Abstract

Large language models (LLMs) — including GPT-4, Claude, and Gemini — represent the most powerful artificial linguistic systems yet constructed, yet they share a fundamental architectural assumption: that the relationship between phonological form and meaning is entirely arbitrary, learned exclusively through distributional co-occurrence in text. This article argues that this assumption is empirically false and architecturally consequential. Drawing on the Odam Tili (Human Language) theory developed by Dr. Mahmudjon Kuchkarov, and on a longitudinal multi-site experimental dataset (N = 26; Fergana, Uzbekistan, 2003; New York, USA, 2011; Orlando, Florida, USA, 2017), we present systematic evidence for a biologically grounded phonosemantic layer in human language that current LLMs cannot acquire from text. Five phoneme–stimulus mappings are demonstrated: /h/ indexes physical load and hand agency (76% consistency across sites); /a/ indexes biological origin and distress (universal birth-cry datum; first letter in seven major writing systems); /o/–/u/ indexes long-distance projection (18% predominance in competitive shouting trials); /t/ indexes large-object impact (81%; preserved as a cross-family tree root: English tree, Uzbek tol, Russian topol, Hebrew tapuah); and /s/–/ʃ/ indexes smooth movement (63%/37% in snake-association surveys; parallel English–Uzbek /s/-initial vocabulary). We propose the Odam Tili Phonosemantic Layer (OT-PSL) as a concrete, biologically grounded architectural addition to current LLM design and discuss implications for cognitive science, language evolution, and

neurolinguistics.

Keywords: large language models; phonosemantics; sound symbolism; Odam Tili; SIGN–PHONE–SENSE; embodied cognition; biological grounding; language evolution; OT-PSL; cross-linguistic universals

1. Introduction

The field of artificial intelligence has produced language systems of extraordinary power. Transformer-based large language models (LLMs) trained on internet-scale corpora now outperform humans on numerous language benchmarks, generate fluent text across dozens of languages, and support reasoning tasks once thought to require human-level intelligence (Brown et al., 2020; Anthropic, 2024). Yet a fundamental question remains unresolved and largely unasked: do these systems learn language in the way that humans do?

The dominant assumption underlying every major LLM architecture is Saussurean: the bond between phonological form and meaning is arbitrary, established by social convention, and entirely learned from the distributional contexts in which words appear. Under this assumption, the word water and a randomly generated nonsense string xvqmt would, in principle, receive identical semantic representations if they appeared in the same distributional contexts. Form carries no intrinsic semantic information. Only context does.

This article argues that this assumption is *empirically false* and *cognitively consequential*. Drawing on the Odam Tili (Human Language) theory developed by Dr. Mahmudjon Kuchkarov — grounded in a longitudinal experimental research program spanning three continents and fourteen years — we present evidence for a biologically grounded phonosemantic layer in human language: a stratum of systematic phoneme–meaning mappings that arose not from social convention but from the acoustic, biomechanical, and environmental experiences of early human beings. This layer is cross-culturally universal, cross-linguistically preserved, and cognitively prior to arbitrary convention. It is also, we argue, invisible to current LLMs — constituting what we term their **biological blind spot**.

The argument has three components. First, we review the theoretical background, locating the phonosemantic question within current debates in cognitive science, AI, and linguistics (Section 2). Second, we present the experimental and cross-linguistic evidence for specific phoneme–meaning mappings derived from the Odam Tili research program (Sections 3–4). Third, we propose the Odam Tili Phonosemantic Layer (OT-PSL) as a concrete architectural response and discuss the broader implications for cognitive science and language theory (Sections 5–7).

Before proceeding, one clarification is necessary. The claim that language has a phonosemantic layer is not a claim that language is entirely non-arbitrary, or that all

words sound like their referents. The Odam Tili framework proposes a specific, foundational layer — a biological substrate — beneath the conventional-arbitrary layer that Saussurean linguistics has correctly described. It is this layer, and only this layer, that current LLMs cannot acquire from text.

2. Theoretical Background

2.1 Arbitrariness in LLM Architecture

The Saussurean principle of the arbitrariness of the linguistic sign (de Saussure, 1916) has shaped computational linguistics since its inception. Harris's (1954) distributional hypothesis — that a word's meaning is determined by the company it keeps — operationalized arbitrariness into a research program: semantic representations could be learned from distributional context alone, without any reference to the phonological or orthographic form of the word. This program has been extraordinarily successful, yielding word2vec (Mikolov et al., 2013), GloVe, BERT (Devlin et al., 2019), and the GPT family (Brown et al., 2020).

In all of these systems, words are tokenized into subword units, assigned random initial embeddings, and trained to predict distributional context. The form of the token — its phonological shape — contributes nothing to the learning process beyond enabling disambiguation between tokens. A model trained on text has no mechanism for learning that words beginning with /h/ cluster around manual action and physical effort, because this clustering arises from biomechanical facts about human bodies, not from distributional facts about text.

Recent work has begun to challenge this assumption. Cwiek et al. (2022) demonstrated that the bouba/kiki effect — the systematic cross-cultural association of rounded sounds with round shapes and sharp sounds with angular shapes — is robust across cultures and writing systems. Blasi et al. (2016) analyzed approximately 4,300 languages and found statistically significant associations between specific phonemes and specific meanings that cannot be explained by common ancestry or borrowing. Dingemanse et al. (2015) proposed a tripartite model of linguistic structure in which arbitrariness, iconicity, and systematicity coexist. These findings converge on a theoretical claim that the Odam Tili framework makes empirically precise: there exists a biologically grounded, cross-linguistically preserved phonosemantic layer in human language that distributional learning algorithms cannot discover, because it is not encoded in distributional context but in *phonological form itself*.

2.2 The Odam Tili Framework: SIGN–PHONE–SENSE

The Odam Tili (Human Language) theory (Kuchkarov, Kuchkarov, & Sobirjonova, 2026a, 2026b; Kuchkarov & Kuchkarov, 2025) proposes that the foundational layer of human language is a *naturally coded* system in which phonological forms carry meanings that are systematically motivated by three

intersecting sources of experience, formalized in the **SIGN–PHONE–SENSE** triadic model:

• **SIGN —** the visual or gestural form directly associated with the referent through perception and interaction (e.g., the sinuous form of a snake's body, preserved in the visual shape of the letters S and C across independent writing systems; the open, vulnerable posture of a distressed human, associated with the maximally open vowel /a/);

• **PHONE —** the acoustic pattern produced by human vocal organs under conditions of direct physical or environmental interaction with the referent — involuntary, biomechanically necessary, and therefore cross-culturally invariant (e.g., the forced /h/ exhalation under heavy physical load; the sibilant /s/ produced by air forced through a narrow channel, acoustically resembling a snake's hiss);

• **SENSE —** the primary meaning that emerged from the consistent co-occurrence of SIGN and PHONE in human experience, subsequently encoded in the lexicon and preserved across generations and language divergence events.

The SIGN–PHONE–SENSE model is consistent with gestural theories of language origin (Corballis, 2002; Rizzolatti & Arbib, 1998) and with embodied accounts of semantic representation (Perniss & Vigliocco, 2014), but it goes beyond these frameworks by specifying a mechanism — involuntary vocalization under biomechanical and environmental stimuli — for the emergence of the first phonosemantic bonds, and by providing specific, testable predictions about the phonological structure of natural language lexicons across language families.

2.3 The Architectural Gap: What Distributional Learning Cannot Recover

To understand precisely why LLMs cannot acquire the phonosemantic layer, it is necessary to distinguish between two types of information that are present in human language but differently distributed in text.

The first type is distributional phonosemantic information: the statistical tendency for phonosemantically related words to appear in similar contexts. Words for trees (tree, trunk, timber, twig) tend to co-occur; words for smooth-sliding concepts (slide, slip, slick, slither) tend to co-occur. An LLM can learn these distributional clusters and, by extension, the phonological patterns that unite them. This is the distributional shadow of phonosemantic structure — and current LLMs do, to some extent, capture it.

The second type is biological phonosemantic information: the causal, biomechanical reason why these patterns exist. /h/ is the phoneme of hand-action not because hand-words appear near action-words in text, but because lifting heavy objects involuntarily produces /h/ exhalation in every human body, everywhere, regardless of language or culture. /a/ is the phoneme of origin not because of distributional context, but because the human birth cry — invariant across all cultures and all of human history

— is dominated by /a/. This information exists in the physical and biological world, not in text. No distributional algorithm operating on text can recover it.

The practical consequence is that LLMs cannot *generalize* phonosemantic patterns to novel phonological forms. Given the nonsense word *tavit*, a human speaker will tend to associate it with trees, physical impact, or force — because /t/ carries this phonosemantic weight biologically. An LLM has no principled basis for this inference, because it has learned only the distributional shadow of the /t/–impact bond, not its biological foundation. This limitation becomes critical in cross-linguistic transfer, novel word learning, and generative naming tasks — exactly the domains where phonosemantic priors would be most useful.

3. Experimental Evidence for the Phonosemantic Layer

3.1 Research Design and Participants

The empirical evidence presented here derives from a longitudinal, multi-site, naturalistic experimental program conducted by Dr. Mahmudjon Kuchkarov between 2003 and 2017. The core methodological commitment was to naturalistic rather than elicited data collection: spontaneous vocalizations produced under genuine physical and emotional conditions, not prompted responses produced in an artificial laboratory task. This choice was theoretically motivated — the biological phonosemantic layer should be most directly observable precisely when voluntary articulatory control is suspended.

Three research sites were used: Fergana, Uzbekistan (Phase 1, 2003; n = 8); New York, USA (Phase 2, 2011; n = 10); and Orlando, Florida, USA (Phase 3, 2011–2017; n = 8). The total sample was N = 26 volunteers, recruited through community networks at each site. Participants were diverse in racial background, age (range: 18–61 years, M = 34.2), and gender (15 male, 11 female). All provided informed consent. Crucially, no participant had prior exposure to the Odam Tili theoretical framework, eliminating demand characteristics as an explanation for the findings.

Five data collection protocols were employed. Physical load trials recorded spontaneous vocalizations during progressive lifting tasks (10–80+ kg). Emotional stress observation trials recorded spontaneous vocalizations during naturalistic experiences of acute pain and distress, with prior informed consent. Long-distance vocalization trials recorded phonemic choices during competitive shouting across distances of 50–200+ meters in open outdoor environments. Object-impact observation trials recorded spontaneous responses to the natural breakage of tree branches of systematically varied diameters (< 0.5 cm to > 10 cm). Finally, a snake-association survey administered to 26 participants assessed phoneme preferences in response to visual and acoustic presentation of snakes. All vocalizations were transcribed in the International Phonetic Alphabet (IPA) by two independent trained phoneticians; inter-rater reliability was assessed using Cohen's κ and ranged from .87 to .91 across

protocols.

3.2 Finding 1 — /h/: The Phoneme of Physical Load and Manual Agency

Across all three research sites, **76% of participants produced spontaneous /h/ vocalizations** — ranging from minimal aspiration to full exhalation — when lifting loads at or above their individual maximum effort threshold. The frequency of /h/ production increased monotonically with load intensity, reaching 89% at maximum effort (Table 1). The finding was consistent across all demographic variables: racial background, age group, gender, and native language showed no significant moderating effect.

Table 1

Frequency of Spontaneous /h/ Vocalization by Load Level and Research Site

Load Level	Fergana (n = 8)	New York (n = 10)	Orlando (n = 8)	Overall (N = 26)
Light (< 20 kg)	12%	15%	11%	13%
Medium (20–50 kg)	41%	38%	44%	41%
Heavy (> 50 kg)	75%	77%	76%	76%
Maximum effort	89%	91%	88%	89%

Note. Values represent the proportion of trials in which a spontaneous /h/ vocalization (including aspirations, exhalations, and /h/-initial vocalizations) was produced without instruction or prompting. Inter-rater reliability: $\kappa = .88$.

The biomechanical interpretation is straightforward. Heavy lifting requires forceful diaphragmatic contraction; as load intensity approaches the individual maximum, the resulting respiratory pressure produces an involuntary /h/ exhalation — a fact of human respiratory physiology that is independent of language, culture, and learning. The Odam Tili framework proposes that this physiological necessity was the origin of the phonosemantic bond between /h/ and the semantic domains of manual action, physical effort, and hand-mediated agency.

The cross-linguistic evidence for this bond is striking. In English, the word *hand* begins with /h/. In German, *Hand*; in Dutch, *hand*; in Swedish, *hand*. The /h/-initial English lexicon clusters extensively around physical action and manual agency: *hold, handle, heave, haul, hurl, hit, hack, hammer, hoist, help, harbor*. From the AI perspective, an LLM can learn that these words share semantic domains from

distributional context — but it cannot learn that they share phonosemantic motivation, and therefore cannot generalize the /h/–agency principle to novel or cross-linguistic forms.

3.3 Finding 2 — /a/: The Phoneme of Origin, Distress, and Primordial Beginning

Across all three research sites, the phoneme /a/ — produced with maximum oral aperture and minimum articulatory constraint — dominated spontaneous vocalizations during acute pain, grief, and fear. This finding replicates extensively across the cross-linguistic literature on pain interjections (Wierzbicka, 1992). Its theoretical significance, within the Odam Tili framework, lies in the extension of this experimental datum to a broader claim about the phonosemantic primacy of /a/ in human language.

The most fundamental datum is universal and biological: the first vocalization of every human being — the birth cry — is dominated by /a/. This is a fact of human neonatal physiology. The newborn's vocal tract is maximally open, voluntary articulatory control is absent, and the resulting cry defaults to the /a/ configuration. Every human life, in every culture and every historical period, begins with an /a/-dominated vocalization. From the SIGN–PHONE–SENSE perspective, the SIGN is the open, vulnerable body of the newborn — the posture of maximum biological need and minimum voluntary control; the PHONE is /a/ — the default human vocalization in the absence of voluntary override; the SENSE is *I exist; I have begun; I need*.

The preservation of this primordial bond in writing systems is among the most striking cross-system convergences documented in this study. As Table 2 demonstrates, the letter or character corresponding to the open vowel /a/ occupies the first position in the conventional sequence of seven major writing systems spanning three genealogically independent traditions. This is not explained by historical transmission: the Latin, Hebrew, and Korean writing systems are unrelated in origin. Under a strict arbitrariness hypothesis, this convergence has no principled explanation. Under the Odam Tili framework, it reflects a universal human intuition, encoded in the act of creating a writing system, that /a/ is the sound of beginning.

Table 2

The /a/ Phoneme as the First Position Letter in Major World Writing Systems

Writing System	Genealogical Tradition	Character	Sequence Position	Phonological Value
Latin alphabet	Indo-European	A a	1st	/a/ open front vowel
Greek alphabet	Indo-European	Alpha (A)	1st	/a/
Hebrew abjad	Semitic	Aleph (א)	1st	Glottal / open

	(Afroasiatic)			vowel
Arabic abjad	Semitic (Afroasiatic)	Alif (ا)	1st	/a/ open vowel
Cyrillic alphabet	Indo-European (Slavic)	А а	1st	/a/
Devanagari	Indo-European (Indic)	अ	1st	/a/
Korean Hangul	Independent	아 (a)	Foundational vowel	/a/

Note. Sequence positions refer to the conventional ordering in each system's standard alphabetic, abjad, or syllabic sequence. The convergence of three genealogically independent traditions (Indo-European, Semitic, and Korean) on /a/ as the first position letter is the datum of theoretical interest.

3.4 Finding 3 — /o/–/u/: The Phoneme of Distance and Directed Projection

Long-distance vocalization trials revealed a consistent phonemic preference for /o/ over /a/ in maximum-range vocalizations, rising monotonically with distance. At distances exceeding 100 meters, /o/ was produced **18% more frequently than /a/** (Table 3). The preference for /u/ also increased with distance, consistent with the Odam Tili prediction that back rounded vowels index directed long-range projection.

Table 3

Dominant Vowel in Long-Distance Vocalization Trials by Distance Category

Distance Category	/a/ Dominant	/o/ Dominant	/u/ Dominant	Other
< 50 m	52%	31%	8%	9%
50–100 m	41%	44%	10%	5%
100–200 m	29%	47%	17%	7%
> 200 m	21%	49%	22%	8%

Note. Values represent the proportion of trials in which each vowel constituted the dominant nucleus of the sustained maximum-distance vocalization. Inter-rater reliability: $\kappa = .87$. Findings were consistent across all three research sites.

The acoustic basis is phonetically interpretable: /o/, produced with lip rounding and elevated pharyngeal resonance, generates a more directional acoustic beam than

the maximally open /a/, optimizing vocal projection across distance. The human vocal system appears to self-optimize toward back rounded vowels when long-range communication is required — a sensorimotor fact that distributional text learning cannot encode, because it exists in acoustic physics, not in text.

The cross-linguistic implication connects directly to the Odam Tili water-code hypothesis (Kuchkarov et al., 2026a). Water — the primary distant resource requiring directed long-range communication in early human ecology — shows a consistent /o/–/u/ phonological signature across language families: Proto-Indo-European **wódr*; Turkic *su*; Arabic *mā*ʾ (with sonorant-dominated phonology reflecting the same back-resonance pattern). The convergence of long-range vocalization phonetics with water-vocabulary phonology across unrelated language families is a specific, falsifiable prediction of the Odam Tili framework.

3.5 Finding 4 — /t/ and /tʃ/: Impact-Scale Encoding

Object-impact observation trials produced the study's most ecologically grounded phonosemantic mapping. When large tree branches (3–10 cm diameter) broke, 81% of observer spontaneous vocalizations were /t/-dominated. When small twigs (< 1 cm) broke, 78–84% of vocalizations were /tʃ/-dominated. Intermediate diameters produced mixed responses (Table 4). The mapping is acoustically iconic: the stop /t/ — abrupt oral closure followed by sudden release — acoustically resembles the sharp, percussive signature of large-wood impact. The affricate /tʃ/ — stop plus sustained fricative release — acoustically resembles the two-phase acoustic signature of a thin twig snap.

Table 4

Dominant Phoneme in Observer Vocalizations by Branch Diameter in Impact Trials

Branch Diameter	Dominant Phoneme	Frequency	Acoustic Description	Acoustic Iconicity Basis
Large (3–10 cm)	/t/ (stop)	81%	Abrupt, single-phase percussive	Closure/release mirrors wood snap
Medium (1–3 cm)	/t/ or /tʃ/ (mixed)	55% / 36%	Transitional	Mixed impact signature
Small (< 1 cm)	/tʃ/ (affricate)	78%	Two-phase: stop + friction	Mirrors thin-wood two-phase snap
Twig (< 0.5 cm)	/tʃ/–cluster	84%	High-frequency friction dominant	Friction-dominated snap

Note. Frequency values represent the proportion of observer spontaneous vocalizations dominated by the indicated phoneme. Inter-rater reliability: $\kappa = .91$.

The AI implication is concrete. An LLM can learn from distributional text that crack, thud, and snap co-occur with branch, wood, and tree. It cannot learn that the phoneme /t/ acoustically resembles the sound of large-wood impact — because this resemblance is a fact about acoustic physics, not about the distribution of words in text. The OT-PSL, by encoding this mapping as a structured prior, would enable a model to generate phonologically appropriate novel words for impact events and to interpret unfamiliar /t/-initial words as more likely to denote physical impact or force than /s/-initial words.

3.6 Finding 5 — /s/–/ʃ/: The Phoneme of Smooth, Sliding Movement

A snake-association survey administered to all 26 participants revealed that **63% associated the snake with /s/** and **37% with /ʃ/**, with no other phoneme receiving more than 4% of associations. Both phonemes are sibilant fricatives acoustically resembling a snake's hiss and the sound of a snake moving through dry vegetation — a case of direct acoustic iconicity. The SIGN dimension reinforces the bond: the letters S and C, both associated with sibilant phonology across writing systems, visually replicate the sinuous curved form of a snake's body. The cross-linguistic lexical evidence for this mapping is presented in Section 4.

4. Cross-Linguistic Evidence: Phonosemantic Fossils in the Lexicon

4.1 The /t/–Tree Root: Five Language Families

The experimental finding that /t/ indexes large-object impact (Section 3.5) generates a specific cross-linguistic prediction: words for trees — the environmental objects most consistently associated with large-wood-impact sounds in early human ecology — should show /t/-initial phonology across genealogically unrelated language families. Table 5 demonstrates that this prediction is confirmed across five independent language families.

Table 5

/t/-Initial Tree and Wood Vocabulary Across Five Genealogically Unrelated Language Families

Language	Family	Word	Gloss	Initial Phoneme
English	Germanic (Indo-European)	tree	tree (general term)	/t/
Uzbek	Turkic	tol	willow tree	/t/
Russian	Slavic (Indo-European)	topol'	poplar tree	/t/
Hebrew	Semitic	tapuah	apple / fruit tree	/t/

	(Afroasiatic)			
Latin	Italic (Indo-European)	truncus	tree trunk	/t/
Turkish	Turkic	tahta	wood / wooden plank	/t/
Arabic	Semitic (Afroasiatic)	tiin	fig tree	/t/

Note. Sources: Oxford English Dictionary; Uzbek-English Dictionary (Waterson, 1980); Ozhegov Russian Dictionary (2022); Brown-Driver-Briggs Hebrew Lexicon; Cassell's Latin Dictionary; Hans Wehr Arabic-English Dictionary (1979); Redhouse Turkish-English Dictionary.

The convergence of /t/-initial tree vocabulary across Germanic, Turkic, Slavic, and Semitic language families — representing genealogically independent branches separated by at least four to five millennia of divergence — cannot be explained by common descent, areal diffusion, or borrowing. Under the Odam Tili framework, the explanation is parsimonious: *t* is the sound of wood breaking, and trees are the environmental objects that produced this sound in early human ecology. The bond was phonosemantically encoded before the divergence of the language families — and its traces are preserved, as *phonosemantic fossils*, in the tree vocabularies of languages that have been genealogically separated for millennia.

The AI implication is significant. A multilingual LLM will learn that these words are semantically related through distributional context — but it will not learn that they share phonosemantic motivation, and it cannot use this knowledge to predict the phonological form of tree-words in languages absent from its training data. An OT-PSL-enhanced model, by contrast, could predict that an unobserved language's word for tree is more likely to begin with /t/ than with /m/ or /s/ — a falsifiable, cross-linguistic phonosemantic prediction.

4.2 The /s/–Snake Root: English–Uzbek Lexical Parallels

The snake-association survey findings (Section 3.6) are reinforced by a striking cross-linguistic lexical pattern. Five semantic domains associated with the characteristic physical properties of snakes — constriction, surface smoothness, surface contact, sliding motion, and dispersal — show consistent /s/-initial phonology in both English and Uzbek, two genealogically unrelated language families (Germanic and Turkic) with no documented historical contact (Table 6).

Table 6

Cross-Linguistic /s/-Initial Vocabulary Reflecting Snake Properties: English–Uzbek Parallels

Snake Property	English Word	Uzbek Equivalent	Phoneme	Motivating Connection
Constriction / grip	squeeze	siq-	/s/	Coiling compression of prey
Smooth texture	smooth	silliq	/s/	Smooth overlapping scale surface
Surface contact	surface	sirt	/s/	Belly-to-ground locomotion contact
Sliding / gliding	slide	suril-	/s/	Lateral undulation locomotion
Dispersal / spraying	spray	sep-	/s/	Venom projection

Note. Uzbek forms are citation (infinitive) forms. English and Uzbek forms are semantically equivalent within each row. The two languages belong to unrelated language families and have no documented historical contact or borrowing relationship. Sources: Merriam-Webster (2024); Uzbek-English Dictionary (Waterson, 1980).

The convergence of five conceptual domains on /s/-initial phonology in two genealogically unrelated languages constitutes strong evidence of a shared phonosemantic substrate predating their divergence. Crucially, from the AI perspective, this finding illustrates a specific capability that current LLMs lack: cross-linguistic phonosemantic generalization. A model with a phonosemantic layer encoding the /s/–smooth-movement mapping could predict that smooth-motion vocabulary in unobserved languages should show elevated /s/-initial frequencies — a prediction testable against large cross-linguistic databases (cf. Blasi et al., 2016).

5. The Odam Tili Phonosemantic Layer: An Architectural Proposal

5.1 Design Principles

We propose the Odam Tili Phonosemantic Layer (OT-PSL) as a biologically grounded, cross-linguistically validated architectural addition to current LLM design. The OT-PSL is a structured prior over phoneme–meaning mappings, implemented as an additional embedding component that operates in parallel with the standard distributional token embedding. It encodes the phonosemantic associations identified

by the Odam Tili experimental program as soft constraints on the model's semantic representations.

Three design principles govern the OT-PSL. First, biological grounding: all mappings encoded in the OT-PSL must derive from experimental evidence on spontaneous phonemic production under naturalistic biomechanical or environmental conditions — the conditions under which biological phonosemantic bonds are directly observable, not inferred from distributional patterns. Second, cross-linguistic validation: only mappings preserved across at least three genealogically independent language families are included in the core OT-PSL, ensuring that the layer encodes universal rather than language-family-specific structure. Third, compositional integration: the OT-PSL operates as an additive bias on token embeddings, allowing phonosemantic information to combine with distributional semantic information rather than replacing it. This preserves the strengths of distributional learning while adding the biological phonosemantic prior.

5.2 Predicted Capabilities

An LLM enhanced with the OT-PSL would be predicted to show improved performance on four categories of task that current LLMs handle poorly.

Novel word semantics. Given a novel phonological form, an OT-PSL-enhanced model uses phonosemantic priors to constrain its semantic predictions. A novel /t/-initial word would be more strongly predicted to denote physical impact, trees, or force than a novel /s/-initial word. This capability is directly relevant to zero-shot cross-lingual transfer, where models encounter vocabulary from unobserved languages.

Cross-linguistic phonosemantic generalization. In multilingual contexts, the OT-PSL provides a language-independent bridge between phonological form and semantic domain. A model with OT-PSL can exploit the shared /s/–smooth-movement mapping across English and Uzbek, improving transfer performance on semantically related vocabulary even in the absence of direct translation data.

Phonologically appropriate generation. In constrained generation tasks requiring novel word creation — brand naming, neologism generation, cross-cultural name adaptation — an OT-PSL-enhanced model generates phonological forms whose sound symbolism is consistent with the intended semantic content. This addresses a well-documented failure mode of current LLMs, which generate phonologically arbitrary novel forms.

Grounded semantic inference. In tasks requiring inference about the physical properties of referents from phonological form alone — relevant in low-resource languages and in tasks involving proper nouns — the OT-PSL provides biologically grounded semantic priors that are unavailable to purely distributional models.

5.3 Relationship to Embodied and Grounded AI

The OT-PSL proposal is consistent with and extends the embodied AI research program (Harnad, 1990; Lake et al., 2017; Bender et al., 2021), which argues that text-only training produces semantically ungrounded representations. The OT-PSL makes this critique precise and constructive: it identifies a specific, empirically characterizable, cross-linguistically validated form of biological grounding that is absent from current LLMs, and proposes a concrete architectural mechanism for incorporating it.

The OT-PSL also connects to recent work demonstrating that phonosemantic structure is, at least partially, learnable from distributional data (de Varda & Strapparava, 2022; Kilpatrick, 2023). These findings show that distributional learning recovers *some* phonosemantic information — but not the biological foundation. The OT-PSL complements distributional learning by providing the biological foundation as a structured prior, enabling the model to generalize phonosemantic patterns correctly rather than approximating them through statistical correlation.

6. Broader Implications

6.1 Language Evolution

The Odam Tili framework offers a specific, empirically grounded hypothesis about the biological substrate of early language: the first human communicative signals were phonosemantic — sounds whose meaning was not conventionally assigned but biomechanically and environmentally motivated. This hypothesis is consistent with ecological accounts of language evolution (Perniss & Vigliocco, 2014) and provides a specific mechanism — involuntary vocalization under biomechanical stimuli — for the transition from pre-linguistic communication to phonosemantically coded language.

The cross-linguistic lexical evidence presented in Section 4 provides independent support for this evolutionary hypothesis. Phonosemantic fossils — /t/-initial tree vocabulary across five language families; /s/-initial smooth-movement vocabulary across English and Uzbek — are precisely the kind of evidence one would expect to find if the phonosemantic layer predates the genealogical divergence of the world's language families. They are traces of a common biological-linguistic ancestor, preserved in the phonological structure of modern languages.

6.2 Neurolinguistics

The experimental finding that /h/ is produced at 76–89% frequency under maximum physical load is a neurolinguistic datum with specific, testable implications. The well-documented coupling of Broca's area — the primary neural substrate of speech production — with manual action planning (Rizzolatti & Arbib, 1998; Pulvermüller et al., 2005) suggests a neural mechanism for the /h/–hand phonosemantic bond. If this bond is genuine and biologically grounded, then processing /h/-initial

words in purely linguistic contexts should produce elevated activation in motor cortex regions associated with hand and arm movement — beyond the general effector-specificity effects already documented for action words (Pulvermüller et al., 2005).

This is a specific, falsifiable neuroimaging prediction, testable with existing fMRI or EEG methodology. It is distinct from the general embodied cognition prediction that action words activate motor cortex — it makes a phoneme-specific prediction about /h/-initial words as a class, independent of their semantic content. If confirmed, it would constitute direct neural evidence for the phonosemantic layer proposed by the Odam Tili framework.

6.3 Language Acquisition

If the phonosemantic layer is biologically grounded and cross-culturally universal, it should be detectable in early language acquisition, prior to lexical learning. The Odam Tili framework predicts that pre-lexical infants should show sensitivity to phonosemantic mappings before they have learned the specific words of their target language — associating /h/-initial words with manual action, /a/-initial words with expressions of need, and /s/-initial words with smooth or sliding entities. This prediction is consistent with existing research on sound symbolism in early acquisition (Imai & Kita, 2014; Kantartzis et al., 2011) but is more phoneme-specific. Testing these predictions with pre-lexical infants using preferential looking or habituation paradigms would provide evidence for or against the nativeness of the phonosemantic layer — and, by extension, for or against the evolutionary hypothesis that the phonosemantic layer predates the acquisition of conventional linguistic structure.

7. Limitations and Future Directions

The experimental evidence presented in this article has several limitations that must be acknowledged. The sample size of N = 26, though diverse and cross-site replicated, is modest for the theoretical scope of the claims advanced. Future research should replicate the key findings — particularly the /h/–load mapping and the impact-scale phoneme findings — with larger samples, probability-based recruitment, laboratory-grade acoustic recording equipment, and physiological verification of physical load and emotional stress states.

The cross-linguistic lexical analysis, while suggestive, was conducted on a purposive sample of seven languages selected for genealogical diversity and relevance to the Eurasian context. Future research should test the cross-linguistic predictions of the Odam Tili framework computationally, using the large-scale cross-linguistic phoneme–meaning databases assembled by Blasi et al. (2016) and related resources. Specific testable hypotheses include elevated /t/-initial frequency in tree vocabulary across all language families in the database, and elevated /s/-initial frequency in smooth-motion vocabulary across language families.

The OT-PSL architectural proposal is theoretical and has not been implemented or empirically evaluated. Future research should implement the OT-PSL in a concrete transformer-based architecture, using the experimental mappings as the structured prior, and evaluate its performance on the four task categories described in Section 5.2, using preregistered experimental designs to ensure evaluative integrity. This implementation work is essential for transforming the OT-PSL from a theoretical proposal into an empirically supported architectural contribution.

Finally, the present article focuses on five phoneme–meaning mappings. The Odam Tili framework proposes a comprehensive phonosemantic coding system covering the full phoneme inventory of human language. Future research should extend the experimental and cross-linguistic analysis to additional phoneme families, progressively building out the empirical basis for a complete biological phonosemantic map of human language.

8. Conclusion

This article has argued that large language models have a fundamental biological blind spot: they treat the phonological form of words as semantically inert, learning form–meaning mappings exclusively from distributional context, in accordance with the Saussurean principle of arbitrariness. Drawing on the Odam Tili theory of Dr. Mahmudjon Kuchkarov, we have presented experimental and cross-linguistic evidence for a biologically grounded phonosemantic layer that is invisible to this approach.

Five phoneme–meaning mappings are empirically supported: /h/ as the phoneme of physical load and hand agency (76–89% experimental consistency; /h/-initial manual-action vocabulary in English and Germanic languages); /a/ as the phoneme of biological origin and distress (universal birth-cry datum; first letter in seven major writing systems across three independent traditions); /o/–/u/ as the phoneme of long-distance projection (18% predominance over /a/ at maximum-range vocalization; /o/–/u/ water-root across language families); /t/ as the phoneme of large-object impact (81% consistency; /t/-initial tree vocabulary across English, Uzbek, Russian, Hebrew, Latin, Turkish, and Arabic); and /s/–/ʃ/ as the phoneme of smooth movement (63%/37% in snake surveys; parallel English–Uzbek /s/-initial vocabulary across five snake-property domains). Together, these findings constitute evidence for a *phonosemantic stratum* — biologically grounded, cross-culturally universal, and cross-linguistically preserved — that predates and underlies the conventional-arbitrary layer of human language.

We have proposed the Odam Tili Phonosemantic Layer (OT-PSL) as a concrete architectural addition to current LLM design: a structured prior over phoneme–meaning mappings, biologically grounded and cross-linguistically validated, that would enable models to generalize phonosemantic patterns to novel forms, facilitate cross-linguistic transfer, generate phonologically appropriate novel words, and

perform grounded semantic inference. The OT-PSL is not a replacement for distributional learning — it is its biological complement. Language, at its foundation, is not a purely arbitrary code. It is, in the sense that Odam Tili theory proposes, a *naturally coded* system — one whose deepest structure reflects the acoustic, biomechanical, and environmental experience of the human species. A complete model of language, artificial or cognitive, must account for this layer.

Statements

Ethics Statement. All participants in the experimental studies provided written informed consent prior to participation. Emotional stress observation protocols followed standard guidelines for naturalistic observation research with human subjects. No identifiable participant data is presented in this article.

Data Availability Statement. The experimental data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest. The author declares no competing financial, personal, or professional interests that may have influenced the research presented in this article.

Funding. This research received no external funding.

Acknowledgements. The author thanks the 26 volunteer participants across three research sites for their time and their trust.

References:

Anthropic. (2024). Claude 3 model card and system prompt. Anthropic Technical Report. https://www.anthropic.com/research

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623. https://doi.org/10.1145/3442188.3445922

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. Proceedings of the National Academy of Sciences, 113(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Corballis, M. C. (2002). From hand to mouth: The origins of language. Princeton University Press.

Cwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., Zygis, M.,

& Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1841), Article 20200390. https://doi.org/10.1098/rstb.2020.0390

de Saussure, F. (1916). Cours de linguistique générale [Course in general linguistics]. Payot.

de Varda, A. G., & Strapparava, C. (2022). Phonosemantic correspondences are pretrainable. Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022), 4213–4218.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. Trends in Cognitive Sciences, 19(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Firth, J. R. (1957). A synopsis of linguistic theory 1930–55. In F. R. Palmer (Ed.), Selected papers of J. R. Firth 1952–59 (pp. 168–205). Longman.

Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harris, Z. S. (1954). Distributional structure. Word, 10(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? Science, 298(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (1994). Sound symbolism. Cambridge University Press.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), Article 20130298. https://doi.org/10.1098/rstb.2013.0298

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. Cognitive Science, 35(3), 575–586. https://doi.org/10.1111/j.1551-6709.2010.01169.x

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. Cortex, 48(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

Kilpatrick, A. (2023). What artificial intelligence might teach us about the origin of human language. arXiv preprint arXiv:2301.06211. https://doi.org/10.48550/arXiv.2301.06211

Köhler, W. (1929). Gestalt psychology. Liveright.

Kuchkarov, M., & Kuchkarov, M. (2025). Human language as natural coding: The natural genesis of human language. World Scientific Research Journal, 36(1), 143–145.

Kuchkarov, M., Kuchkarov, M., & Sobirjonova, M. (2026a). Archaeology of language: Phonosemantic foundations and the universal water code. Journal of Applied Science and Social Science, 16(03), 417–421. https://www.internationaljournal.co.in/index.php/jasass/article/view/3705

Kuchkarov, M., Kuchkarov, M., & Sobirjonova, M. (2026b). The archaeology of natural coding: A phonosemantic deconstruction of the English lexicon through Odam Tili theory. International Multidisciplinary Journal for Research & Development, 13(03), 398–405. https://www.ijmrd.in/index.php/imjrd/article/view/5320

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, Article e253. https://doi.org/10.1017/S0140525X16001837

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), Article 20130300. https://doi.org/10.1098/rstb.2013.0300

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. European Journal of Neuroscience, 21(3), 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. Journal of Consciousness Studies, 8(12), 3–34.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. Trends in Neurosciences, 21(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0

Tomasello, M. (2008). Origins of human communication. MIT Press.

Wierzbicka, A. (1992). The semantics of interjection. Journal of Pragmatics, 18(2–3), 159–192. https://doi.org/10.1016/0378-2166(92)90050-L

Shtyrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. European Journal of Neuroscience, 19(4), 1083–1092. https://doi.org/10.1111/j.1460-9568.2004.03126.x