Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Is There Meaning Without Anyone Meaning Anything? A Structuralist Proposal for Artificial Semantics

Francesco Maria Cianciaruso

francescocia.99@gmail.com

Research Article

Keywords: Deep neural networks , Meaning , Embodiment , Structuralism , Semantics, LLMs

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10230022/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Global Philosophy on November 5th, 2026. See the published version at https://doi.org/10.1007/s10516-026-09824-2.

Abstract

Recent developments in the field of deep neural networks (DNNs) and large language models (LLMs) have led to the spread of systems capable of producing coherent texts, drawing inferences and acting in a contextually situated manner. Such systems appear to exhibit semantically relevant behaviour, despite lacking features traditionally considered essential for meaning, such as sensory experience or human-like intentionality. This situation has revived the philosophical debate on the nature of meaning in intelligent machines. Seeking to contribute to this debate, this article proposes a change of perspective, linking semantic organisation in artificial systems not to symbols or embodiment, but to the concepts of *structure* and *network*. Through a trajectory that draws on the historical epistemology of AI, this paper explores a structural conception of semantic organisation in artificial systems. Our claim is that LLMs may exhibit a form of *structurally organised semantic competence* based on the internal relational organisation of networks rather than on symbols or bodily experience. Through the analysis of distributed representations and relational spaces embedded in LLMs, this paper aims to clarify in what sense such systems can exhibit forms of semantic organisation without possessing human understanding or phenomenal intentionality. This perspective has significant epistemological implications, suggesting a reconsideration of traditional concepts of representation, knowledge, and meaning. In this sense, this article aims to contribute to contemporary debates on the nature of semantic organisation in artificial systems.

1. Introduction. Is semantics a structural property?

Since the earliest attempts to implement intelligent behaviour in artificial systems, engineering developments in AI models have always been inextricably linked to the flourishing of philosophical and epistemological frameworks (e.g., McCulloch & Pitts, 1943; Turing, 1950). Since the mid-twentieth century, the convergence of axiomatic reasoning theory, Boolean logic and intelligence theory has shifted the linguistic question from the analytic philosophy of language to philosophy of mind and AI (i.e., Putnam, 1960). These early, elegant descriptions, characterised by the research on formal systems, cognition, and intelligence, all shared a main feature, derived from Claude Shannon's theory: *information* is an entity distinct from the substrates that carry it, and therefore realisable on any medium (Fodor, 1975, Haugeland, 1989). Inspired by the optimism of the 1960s, Marvin Minsky, one of the pioneers of artificial intelligence, went so far as to say that: «within a generation […] few compartments of intellect will remain outside the machine's realm — the problems of creating artificial intelligence will be substantially solved» (Minsky, 1967). However, this elegant and influential theory eventually revealed its limits. The efforts of classical cognitivism to understand intelligence and language manipulation have only clarified what intelligence and language are not. Indeed, the '*strong' AI* programme failed to account for semantic competence intended as the 'mark' of intelligent behaviour (Searle, 1980; Dreyfus, 1992). The symbolic model explained syntax, but not meaning. In reaction, the embodied and connectionist theories of the 1980s and 1990s proposed that cognition is grounded in the body, perception, and action (e.g., Varela et al., 1991; Lakoff & Johnson, 1980; Barsalou, 1999; Gibbs, 2004, 2006, Desai, 2022). Meaning arises from the dynamic interaction between subject and environment, from sensorimotor

activations that constitute the core of representational content. Thus, the research on cognition and intelligent behaviour seemed to have moved away from machines and AI.

However, contemporary neural AI systems have introduced a new theoretical scenario. They produce meaningful content, and philosophical thought must account for it. Neural networks can produce *abstract representations* (Yang, et al., 2019; Butlin, 2021; Rogers et al., 2021; Johnston & Fusi, 2023), learn *autonomously* from new data (Bendale & Boult, 2016), and embed ontologies and structured knowledge within their *network* (Kumar & Sharma, 2023; DeCoste et al., 2024). Recent advances in AI have produced systems (i.e., LLMs) capable of writing texts, answering questions, summarising documents, and sustaining complex conversations (Naveed et al., 2023). At the same time, they lack several features that many traditional accounts of meaning have considered essential, such as embodiment, direct sensory experience, or human-like intentionality (Mitchell, 2021). This situation creates a conceptual dialectic. Contemporary AI systems exhibit behaviours that *appear semantically coherent*, despite lacking several features that many historical theories of meaning have considered central to semantic understanding.

This dialectic has produced polarised reactions. On the one hand, there is a tendency to describe these models in a strongly anthropomorphic manner as if they were *almost-human* minds. On the other hand, they are often described as purely statistical systems lacking genuine understanding, reduced to 'stochastic parrots' (Bender et al. 2021) or a mere 'jpeg of the web' (Chiang, 2023). However, the main issue, as will be argued in this article, is not empirical but theoretical. Both reactions presuppose conceptual frameworks that may no longer be adequate. Indeed, there is a lack of clear understanding of what kind of *meaning* can be attributed to artificial systems whose internal organisation differs radically from both traditional symbolic machines and embodied organisms.

This work proposes a *change of perspective*. The underlying hypothesis is that the history of artificial intelligence can be read as a history of different conceptions of meaning. Early symbolic approaches treated meaning as something external to the system, anchored to formal rules (Putnam, 1960; Fodor, 1975; Pylyshyn, 1980; Horst, 1990). Then, embodied approaches relocated meaning to the interaction between a bodily agent and its environment (Barsalou, 1999; Gallese & Lakoff, 2005). However, DNNs cannot be fully traced back to either of these two frameworks. Even though disembodied, they can learn statistical structures that reflect regularities in the world, construct vector spaces that can approximate geographical, temporal, quantitative or conceptual relationships, and can be sensitive to relationships that depend not only on linguistic co-occurrences, but also on a deeper alignment with real-world structures (Abdou et al., 2021; Williams, 2025). This suggests the possibility of a third description, in which semantic organisation is neither assigned by explicit rules nor directly grounded in bodily experience but depends on the relational structure of the system itself (Gärdenfors, 2000). In this paper, this claim is restricted to contemporary LLMs. The aim is not to establish a general ontology of meaning, nor to claim an equivalence between artificial and human cognition, but to examine whether neural architectures can support a *specifically structural form of semantic competence*[1].

This article argues that neural architectures require philosophical recognition of a *structural conception of meaning*. According to this perspective, meaning is not a property of isolated symbols nor the direct result of perceptual experience, but a *topological property* of systems. Neural networks construct high-dimensional spaces in which elements are defined by their mutual relationships, creating relational organisations capable of preserving stable regularities across domains. We will explore the possibility that LLMs exhibit forms of structurally organised semantic competence grounded in the relational organisation of their internal representations. This does not amount to human-like understanding (Mitchell, 2021), but reveals a new dimension of semantic emergence (Piantadosi, 2023; Cappelen & Dever, 2021, 2025; Grindrod, 2024). Understanding artificial semantics therefore could mean shifting the focus from single representations to the overall network organisation of relationships that makes generalisation, analogy and inference possible.

The expected contribution of this work is both *historical* and *philosophical*. On one hand, it proposes a conceptual reconstruction of the major paradigms of AI, highlighting the different conceptions of meaning that characterise them and how they have been linked to engineering developments. On the other, it develops the notion of *meaning* as a *structural property*, clarifying in what sense contemporary AI systems can exhibit forms of semantic organisation without possessing human-like understanding or consciousness. The aim is not to discuss the attribution of 'human understanding' to LLMs, but to articulate an intermediate level of semantic competence that is epistemologically relevant.

To do so, the article will be structured as follows:

1. The next chapter will focus the historical epistemology of AI. We will examine the symbolic and embodied paradigms, highlighting their strengths and limitations, and how the transformation of philosophical models shaped engineering developments.
2. Then, we will analyse the representational shift introduced by neural networks, with particular attention to distributed and relational forms of representation.
3. The last section will develop the central proposal of *meaning* as a *structural property* of networks and discuss its epistemological consequences. We will question if LLMs are showing *sparks of intelligence* (Grzankowski et al., 2024) and the evolution of neural models could reshape our very concept of meaning.

This work proposes a thesis with implications that go beyond philosophy of AI. If meaning can emerge from structural relationships within artificial systems, then traditional oppositions such as syntax/semantics and mind/body should be reconsidered. It is not simply a matter of rediscuss paradigms from a debate that has been raging since the middle of the last century. What we intend to do is to show how the intertwining of cognitive science, engineering implementation of LLMs, and philosophy of mind and AI, can lead to a new stage of the research on meaning. Our aim is to highlight how the emergence of abstract representations and advanced semantic competence in artificial systems can demonstrate that meaning emerges from sufficiently complex structures. It is not a side question. It is the gateway to a brand-new philosophy.

2. Where does meaning lie? An epistemological history of machines and semantics

The history of the development of intelligent machines is inextricably linked to the philosophical and cognitive theories that inspired or flourished alongside it. In this sense, theory and engineering have embraced and separated more than once. The first ideas about *what a thinking machine is* greatly anticipated its development. The entire theoretical basis for the development of GOFAI[2] is based on a rationalisation and mathematisation of language that dates to Thomas Hobbes (Leon, 2021), crosses the 19th century with Charles Babbage's first polynomial function calculator (Pasquinelli, 2023), and arrives at the analytical philosophy of language (e.g., Putnam, 1960). This chapter will therefore reconstruct these theoretical steps underlying the development of AI, to highlight the discontinuity with the new generation of DNNs in the following chapters.

2.1. Theories and first developments of thinking machines

Few analogies have shaped the early developments of artificial intelligence as deeply as the *mind/computer* one. Underlying this analogy was a simple and elegant framework (i.e., *classical cognitivism*) which proved extraordinarily effective in the development of the first generation of thinking machines. Classical cognitivism conceived both natural and artificial cognition as formal systems operating on symbols according to syntactic rules (Putnam, 1960; Minsky, 1967, 1969; Fodor, 1975; Pylyshyn, 1980; Horst, 1990). Linguistic expressions were analysed as structured symbol strings, and meaning was defined functionally, in terms of the role a symbol occupied within the computational system. The basis of this theory was therefore the idea that the rules of *syntax* determined linguistic handling, and therefore intelligent behaviour. Correct syntactic processing by a formal system would automatically bring out the semantic dimension of symbols, by virtue of their isomorphism with the real world. In other words, we can conceive that those strings of information 'express' something solely by virtue of the formal properties of the system (Haugeland, 1989). This approach was supported by the optimism of the mid-twentieth century, when it was widely believed that intelligence could be fully captured by rigid and syntactic symbol manipulation. It inspired an entire generation of developments and attempts to create artificial intentionality (Bobrow, 1964; Newell & Simon, 1964, 1995; Minsky, 1969[3]).

However, despite its formal elegance the symbolic paradigm progressively revealed a fundamental limitation. While it provided powerful models of inference and problem solving, it failed to account for the origin of meaning itself. Within such systems, symbols are manipulated purely on their formal properties, leaving their semantic interpretation external to the mechanism. A system may correctly process linguistic inputs and produce appropriate outputs while remaining entirely impermeable to their meaning (Searle, 1980; Dreyfus, 1992). By itself, formal symbol processing cannot explain semantic content. In this sense, the failure of *symbolic AI* was not primarily technical but conceptual. Intelligence and language processing cannot be reduced to the execution of syntactic rules over uninterpreted

symbols that leave meaning suspended in a *vacuum*. It is precisely this explanatory gap that motivated subsequent attempts to reconnect cognition with the world.

2.2. The revolution of embodied semantics

The *embodied turn* emerged as a direct response to this theorical impasse (Varela, Thompson & Rosch, 1991). If symbolic approach failed by operating on symbols detached from any form of experience, then grounding meaning in bodily interaction appeared as a promising alternative. Embodied approaches rejected the idea of cognition as abstract symbol manipulation and instead located meaning in the inextricable relation between an agent and its environment. Embodied semantics is based on somatosensory information stored by the cognitive system, which can be reactivated based on contextual stimuli (Barsalou, 1999, 2012; Solomon & Barsalou, 2004). It is possible to speak of intentional states in terms of perceptual symbols, recordings of neural activations that occur during sensory experience, operated by the sensorimotor regions of the brain (Gallese & Lakoff, 2005; Borghi, 2014). These symbols consist of mental simulations of perceptual state. When an experience occurs, they are recorded by the cognitive system through a multimodal representation stored in memory, which can be subsequently reactivated for operations of different complexity. Every process of understanding and processing, from metaphorical and analogical language to causal structures, is based on sensorimotor states.

The case of *metaphorical language* can be taken as an example of embodied semantics. In the context of conceptual metaphor theory (Lakoff & Johnson 1980; Gibbs, 2004, 2006; Desai, 2011, 2022), it can be argued that many of the metaphors we use to understand and describe the world are based on somatosensory information. There is a body of neural evidence supporting a cognitive perspective centred on the sensorimotor system, in which language and its processing can be translated into terms of embodied simulations within an action-perception dualism. For example, it has been shown that the premotor cortex and parietal lobe structures appear to be integrated at the neural level to construct concepts encoded from sensorimotor information (Gallese & Lakoff, 2005). Numerous neuroimaging studies have also examined the question of the foundation of abstract language, comparing activations in somatosensory and motor areas of the brain in relation to metaphorical and non-metaphorical expressions[4].

According to this framework, cognitive representations are not amodal tokens, but patterns of neural activation shaped by perception and action. Understanding language does not consist in manipulating arbitrary symbols, but in reactivating sensorimotor structures associated with sensorial experience. The turning point represented by embodied semantics has effectively demonstrated how the path of rigid, symbolic syntax of classical cognitivism (and the idea that it was possible to achieve true *artificial understanding*) has been exhausted. The anchoring of cognition to the body thus seemed to have put an end to the development of thinking machines. However, even this theory has an issue.

2.3. Limits of embodiment and the need for a new explanatory framework

Despite its explanatory power, embodiment does not provide a fully satisfactory account of linguistic handling. While embodied theories offer a solid explanation of the biological basis of semantics, they do not yet seem to fully account for higher-level processes such as abstract reasoning, symbolic manipulation, and context-sensitive inferential organisation. As a result, embodiment risks overextending the explanatory power of sensorimotor grounding, transforming what is a sufficient condition for some forms of meaning into a necessary condition for all. Because of this, some areas of research argue for the autonomy of syntactic competence (Moro, 2004; Moro et al., 2013) and the symbolic description of mental states.

In response to the limitations of both classical symbolic models and strongly embodied accounts, a range of pluralistic approaches to mental representation has been developed. These frameworks reject the idea that cognition relies on a single representational format, proposing instead that cognitive systems deploy multiple, functionally distinct resources. On this view, different types of representations such as sensory, linguistic, symbolic, or affective contribute in complementary ways to cognitive processing. Meaning would not reside in any single format but arises from the coordinated interaction between heterogeneous representational systems, while syntactic mechanisms retain a degree of functional autonomy in structuring and combining content (Moro, 2004; Dove, 2009). Neurocognitive research suggests that knowledge representation in the human brain is distributed across partially separate systems rather than grounded in a uniform sensory substrate[5] (Wang, 2020[6]; Bi, 2021[7]). More recently, also Shapiro (2019) argued that a purely embodied approach is not sufficient to explain the entire range of cognitive phenomena. However, while pluralistic accounts help clarify the limits of embodiment, they remain largely descriptive and do not yet explain how semantic structure can emerge in an artificial system. Considering this reconstruction, the reference to embodiment as the best explanation of cognitive and semantic processes is not without criticism, and research in this area is still ongoing. Taken together, these accounts indicate that semantic competence depends on the interaction of multiple representational formats, rather than on exclusive reliance on either sensorimotor grounding or abstract symbol manipulation.

This section of historical epistemology has highlighted the inextricable relationship between philosophy of mind, cognitive science and AI engineering. Technical developments have always followed theories of intelligence and cognition or intersected with them. This was the case with symbolic AI, whose relationship with classical cognitivism led to the theorisation and implementation of artificial intelligence based on syntactic rules and rigid algorithms. The end of that season of research into intelligent machines led to a detachment from the cognitive sciences, which were exploring the bodily domain of representation and semantics. However, today the embodied paradigm shows explanatory limitations. It does not account for high-level cognitive processes, and new hybrid theories have arrived on the scene. It is precisely in this context that AI has rejoined the debate.

3. Neural Networks and Sparks of Intelligence

The impasse outlined in the previous chapter did not directly lead to a new theory of meaning, but rather to a transformation in the way artificial systems were understood. The end of the symbolic paradigm of AI and the affirmation of embodiment seemed to have sterilised the claim to build thinking machines. However, the revolution represented by neural models (Hopfield, 1982, 1988; Krizhevsky et al., 2012; LeCun, Bengio & Hinton, 2015) and generative AI (Naveed et al., 2023) has reopened the question. Neural and connectionist approaches did not emerge as explicit philosophical responses to the limitations of symbolism or embodiment. They arose from different needs related to learning, adaptation, and the extraction of regularities from data (Kelleher, 2019). However, these architectures have introduced a profound change in the very concept of representation. It is precisely around this concept that our argument will develop. The following chapter will be divided in two sections. (1) We will firstly introduce the historical and epistemological main feature of the revolution of DNNs. (2) Then, we will go deeply into philosophical issues, focusing on the *structuralist turn* represented by LLMs.

3.1. Word embeddings and distributional spaces

Since the 1980s, AI research has begun to glimpse the possibility that a different model of intelligent machines could be achievable. Inspired by early pioneering work in the mid-century (McCulloch & Pitts, 1943; Rosenblatt, 1958; Hoff & Widrow, 1960), the revolution represented by DNNs marks a clean break with the symbolic conception of representation as a set of manipulable tokens. In connectionist and neural models, there is no privileged representational unit that corresponds to a word, concept or proposition. What the system represents is not embedded in any single element but emerges from the overall configuration of the network. Early work on parallel distributed processing (Rumelhart, Hinton & McClelland, 1986; Hopfield, 1982, 1988) had already highlighted that cognitive content in these systems is realised through patterns distributed across many units, rather than through symbolic structures. This insight was later developed in the philosophical debate through the idea of sub-symbolic computation (Lopez-Rubio, 2018), in which representations consist of relational and dynamic properties rather than discrete syntactic forms.

From this point of view, the contrast with classical cognitivism is not only methodological, but ontological. Symbolic systems presuppose representations whose identity and function can be specified independently from the overall state of the system. Instead, DNNs dissolve this independence. The representational content is inseparable from the overall organisation of the network. Learning does not consist in the acquisition of new symbols or rules, but in the reorganisation of the internal architecture to capture the regularities of the data. What is learned is not a set of explicit meanings, but a structure that constrains and enables future behaviour. Thus, the traditional question of *how symbols acquire meaning* gives way to a different problem:

- How can activity patterns support systematic behaviour without relying on predefined symbols and formal rules?

To answer this question, we need to consider word embeddings and distributional semantics. According to distributional theory (Firth, 1957), which states that the meaning of a word is determined by its relationship with other meanings and its use in linguistic contexts, semantic representation is progressively reduced to relational configurations within high-dimensional spaces. Distributional paradigms are based on the hypothesis that words that appear in similar contexts tend to have similar meanings (Harris, 1954; Sahlgren, 2008). Word *embeddings* transform contextual relationships into distances in vector space (Mikolov et al., 2013). The crucial point is not the quantitative nature of these representations, but rather their structural organisation. The proximity between vectors corresponds to semantic similarity[8], directions in space can capture systematic relationships, and groups of points delineate conceptual fields. Phenomena such as *analogy* are not programmed as symbolic rules but *emerge as geometric properties* of the representational space. In this sense, meaning is not stored as discrete content but manifests itself as a relative position within a network of relationships.

Embeddings thus show that it is possible to learn complex conceptual structures without resorting to either externally interpreted symbols or direct anchoring to bodily experience. What the system acquires is not a list of meanings, but a semantic topology, i.e. an internal organisation that embeds relationships, features and regularities. It is precisely this organisation, rather than the presence of symbolic units or perceptual correlates, that makes linguistically coherent behaviour possible in DNNs.

3.2. Artificial representations and 'models of the world'

The emergence of LLMs has reopened fundamental questions for both philosophy of mind and philosophy of language. These systems exhibit behaviours that, while not equivalent to human understanding, partially replicate it functionally. They produce coherent texts, trace relationships between entities in the world, infer implicit connections, and adapt to new contexts. In doing so, they seem to *construct distributed relational configurations* that reflect the structures of the domain they are trained on[9]. The philosophical problem then becomes inevitable:

• *What kind of semantics*, if any, can be attributed to such behaviour?

To answer this question, three sets of issues will be discussed here, related to (1) the epistemological status of artificial representations, (2) their relationship with embodied semantics, and (3) the transformation of the concept of semantics through the notion of network.

(*3.2.1.*) *Relational organisation and artificial representations.* Contemporary LLMs do not merely reproduce superficial regularities. Through distributed optimisation processes on high-dimensional spaces, they develop internal states that function as models of features of the world[10]. These models are not propositional maps or symbolic descriptions of reality, but relational organisations that constrain future processing and guide inference. The literature on *world models* has shown how DNNs can internalise latent data structures and organising information in a coherent and predictively effective manner (Ha & Schmidhuber, 2018; Bengio et al., 2013). In the case of LLMs, such internal configurations support forms of flexible inference, sensitivity to the hierarchical structure of language, and

compositional generalisation capabilities (Tenney et al., 2019; Rogers et al., 2021; Johnston & Fusi, 2023). Even in the absence of direct perceptual anchoring, these systems develop representations that reflect geographical, temporal, quantitative, or conceptual regularities present in the world, as shown by recent studies on the alignment between vector spaces and extra-linguistic structures (Abdou, 2021). Some hybrid approaches combine neural models with symbolic ontologies or structured knowledge embedded in the weights and connections of the network, producing systems capable of extremely accurate domain-specific reasoning (Kumar & Sharma, 2023; De Coste et al., 2024). However, even purely neural models show a surprising capacity for internal organisation. This raises a central question in the contemporary debate on representation: *are these internal states simple statistical correlations or do they constitute authentic representations in a robust sense?* (Grindrod, 2024). What is at stake is not the presence of intrinsic semantic content, but the *emergence of internally organised states* whose structure systematically tracks and preserves domain-relevant regularities. The question does not concern consciousness or phenomenal intentionality, but rather the ontological status of internal configurations. If they play a systematic role in guiding the behaviour of the system and preserving structural regularities, in what sense can they be considered representations?

(*3.2.2*) *Embodiment*, *grounding and relational organisation*. Embodied theories identify semantics with bodily interaction between organism and environment. According to this perspective, conceptual content is rooted in perceptual and motor systems, and even abstract concepts derive, at least indirectly, from sensorimotor experience (Gallese & Lakoff, 2005; Johnson, 2013). Within this framework, meaning is inseparable from the embodied dimension of experience. However, LLMs are disembodied systems, trained exclusively on textual data. Yet their internal representations capture regularities that seem to go beyond mere superficial linguistic co-occurrence. In several cases, these systems are sensitive to relationships that reflect structures in the world, even though they do not have direct access to the world itself. This suggests that alignment with real structures can occur through the internalisation of sufficiently rich statistical regularities. The philosophical debate on representation offers divergent positions here. For some authors, authentic representation requires a causal link with the environment (e.g. *Teleosemantics*; see Millikan, 1989; Papineau, 1996). Others argue that the functional or inferential role may be sufficient to establish representational status (e.g., Dennett, 1989). LLMs suggest another possibility[11]. Even though disembodied, LLMs can learn statistical structures that preserve stable relational regularities correlated with patterns present in their training domains, construct vector spaces capable of encoding geographical, temporal, quantitative, and conceptual relationships, and exhibit sensitivity to relational patterns that extend beyond simple linguistic co-occurrence[12] (Abdou, 2021; Williams, 2025). Do they therefore possess a form of meaning that is purely relational and emergent from their internal network dynamics? Does genuine meaning really require embodiment, as traditional accounts claim?

(*3.2.3.*) *The transformation of the concept of meaning through the network*. Classical semantic theories often presuppose a form of dualism, such as sign and referent, mind and world, representation and object. Neural architectures challenge this approach. Within a deep network, meaning is not assigned to

a symbol or located in a specific unit; it emerges as a distributed property of the overall configuration of the system. In this context, representation is inseparable from the topology of the network. What matters is not intrinsic content, but the position of a state within a structure of relationships. This transformation invites us to rethink the very concept of meaning. If semantic competence consists in the ability to preserve, coordinate, and generalise relational structures, then meaning could be conceived primarily as a structural property[13]. Such a hypothesis finds echoes in structuralist traditions in the philosophy of science and in contemporary approaches to representation that emphasise systemic organisation rather than intrinsic content (e.g., Gładziejewski, 2016). In this perspective, what biological and artificial systems have in common is not embodiment or phenomenal intentionality, but the ability to realise stable and coherent correlations.

These open issues converge into a single question. If deep neural networks can develop internally coherent relational organisations capable of supporting context-sensitive inference and generalisation, does this imply the emergence of a new form of semantics, capable of transcending the traditional opposition between mind and matter, between symbol and body? If this is the case, then meaning might not be a property added to representations, but a feature of certain kinds of organised systems. The next chapter will examine this possibility in philosophical terms.

4. Towards a structuralist theory of meaning?

Our previous arguments have shown that the contemporary debate on meaning has been dominated by a recurring alternative. On the one hand, the idea that semantics is rooted in perceptual-motor experience (Barsalou, 1999; Lakoff & Johnson, 2005; Borghi, 2014; Desai, 2022). On the other, the conception that it depends on the manipulation of autonomous formal structures (e.g., Moro, 2004, 2013). This opposition has played a decisive historical role, but the most recent neuroscientific and computational evidence suggests that it does not exhaust the space of theoretical possibilities (Striem-Amit et al., 2018; Wang, 2020; Bi, 2021).

However, the return of AI to the debate is redefining epistemic criteria and experimental evidence once again. Deep neural networks, and LLMs in particular, do not merely correlate linguistic sequences. Through distributed learning processes and optimisation within high-dimensional parametric spaces, they develop relational configurations that preserve statistically stable regularities across their training domains and can support forms of context-sensitive inference and abstraction (Ha & Schmidhuber, 2018; Rogers et al., 2021). These models are neither propositional descriptions nor explicit symbolic maps, but dynamically stabilised relational organisations that constrain future processing and guide inference (Johnston & Fusi, 2023). Artificial representations are not discrete entities with intrinsic content. They consist of overall states of the system, whose structure preserves relevant regularities of the domain on which the network was trained. In this sense, internal configurations may exhibit forms of structural alignment with conceptual, causal, or hierarchical relationships present in the data (Abdou, 2021; Gładziejewski, 2016). What is learned is not a list of meanings, but an internal topology that organises relevant features, enabling a form of systematic generalisation beyond the observed cases

(Bendale & Boult, 2016). This phenomenon cannot be adequately described in terms of pure syntax or bodily grounding. Neural networks show that it is possible to develop abstract, generalisable, and coherent representations even in the absence of direct sensorimotor experience (Butlin, 2021). Their competence does not derive from the external assignment of meanings, but from the *emergence of internal configurations that are sufficiently structured to support systematic inferences*. We claim that LLMs exhibit *structural semantic competence* (i.e., the capacity to form internal organisations that mirror the causal and semantic relations of their environments)[14]. This does not amount to human-like understanding (Mitchell, 2021), but it reveals a new dimension of semantic emergence[15] (Piantadosi, 2023; Cappelen & Dever, 2021, 2025).

The philosophical point is that such systems challenge the traditional localisation of meaning. If semantics is not contained in isolated symbols nor entirely rooted in bodily experience, then it could consist primarily in the form of internal organisation. In this perspective, meaning is not content added to a representation. It is a property of the relational configuration of the system. This position can be presented as a structuralist hypothesis of artificial semantics. According to this approach, a system can be considered semantically organised when:

1. Its internal *structure* preserves relevant domain relations.
2. It can *generalise* these relations beyond the observed cases.
3. It maintains *consistency* between different levels of representation.
4. Its *internal configuration* is not random, but imposes regular constraints on how the system infers, responds and generalises.

Structuralist Hypothesis of artificial semantics: A system exhibits structural semantic competence when its internal relational organisation supports stable, context-sensitive, and generalisable inferential coordination across domains.

In this context, LLMs take on a particular philosophical value. They show that semantic organisation can emerge from purely relational dynamics, without requiring a conscious subject or a direct biological anchor. This does not imply that such systems understand in the human sense of the term. Rather, it implies that structural organisation may be a sufficient condition for (a form of) semantics.

Understood not simply as technical architecture but as a self-organising relational topology, the network thus becomes the principle for a completely new philosophy. The philosophical question is therefore not whether machines have minds, but what structural conditions make the emergence of semantic organisation possible. In this sense, semantics would no longer be exhausted by the domain of intentionality, as if there is *'Meaning Without Anyone Meaning Anything'*.

Through this work, we did not want to prove a new theory, but rather to discuss a new moment in the relationship between AI, cognitive science and epistemology, which, in the face of the revolution represented by LLMs, is opening to structuralist approaches. This proposal will have to be proven

empirically, and it is not without issues. However, this area of research on intelligent systems is still young, and research is ongoing[16].

5. Concluding remarks

This work has attempted to reconstruct the history and prospects of meaning theory in relation to AI. The aim was twofold. On the one hand, we proposed an epistemologically aware history of AI, highlighting the intertwining of engineering developments and theories of mind. On the other hand, we wanted to explore the theoretical possibilities offered by LLMs. Indeed, their emergence introduces a new theoretical scenario. These systems are not limited to manipulating symbols, nor do they depend on direct sensorimotor anchoring. Following some influential works (e.g., Williams, 2025), we claimed that LLMs develop internal configurations capable of preserving structural relationships and supporting systematic inferences. In this sense, they reveal a conceptual possibility that previous alternatives tended to hide. *Forms of semantic competence may emerge from sufficiently integrated relational organisations capable of supporting stable inferential behaviour.* Indeed, they demonstrate that sufficiently complex internal configurations can support systematic inferences and build structured models of the reference domain, even in the absence of consciousness or a direct sensorimotor foundation. Our proposal is that intelligence may be an eminently structuralist property, and that the intertwining of theory of mind and developments in AI may lead to a confirmation of this thesis.

This article is a work of perspective. We have not brought new data but only raised evidence and shortcomings within the literature. New data and research efforts will be necessary to shape this framework.

Declarations

This research received no external funding. There is no founding source to be declared.

Author Contribution

FMC conceived, researched, and wrote the manuscript.

References

1. Abdou M, Kulmizev A, Hershcovich D, Frank S, Pavlick E, Søgaard A (2021) Can language models encode perceptual structure without grounding? a case study in color. arXiv preprint arXiv:2109.06129
2. Barsalou LW (1999) Perceptual symbol systems. Behav Brain Sci 22(4):577–660. https://doi.org/10.1017/S0140525X99002149

3. Barsalou LW (2012) The human conceptual system. *The Cambridge handbook of psycholinguistics*, *2012*, 239–258
4. Bendale A, Boult TE (2016) Towards open set deep networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1563–1572. https://arxiv.org/abs/1511.06233
5. Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021), March On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). https://doi.org/10.1145/3442188.344592
6. Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828. 10.1109/TPAMI.2013.50
7. Bi Y (2021) Dual coding of knowledge in the human brain. Trends Cogn Sci 25(10):883–895. 10.1016/j.tics.2021.07.006
8. Bobrow D (1964) Natural language input for a computer problem solving system
9. Borghi AM, Binkofski F (2014) Words as social tools: An embodied view on abstract concepts, vol 2. Springer, New York
10. Bottini R et al (2020) Brain regions involved in conceptual retrieval in sighted and blind people. J Cogn Neurosci 32:1009–1025. https://doi.org/10.1162/jocn_a_01538
11. Butlin P (2021) Sharing our concepts with machines. Erkenntnis 88(7):3079–3095. https://doi.org/10.1007/s10670-021-00491-w
12. Cappelen H, Dever J (2021) Making AI intelligible: Philosophical foundations. Oxford University Press
13. Cappelen H, Dever J (2025) Going whole hog: A philosophical defence of AI cognition. arXiv:2504.13988
14. Chiang T (2023) ChatGPT is a blurry JPEG of the web. *The New Yorker*, *9*, 2023. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web
15. De Coste M, Li Z, Khedri R (2024) A hybrid ontology-based semantic and machine learning model *pragmatics*, *36*(7), 1189–1210. https://doi.org/10.1016/j.pragma.2003.10.009
16. Dennett DC (1989) The intentional stance. MIT Press
17. Desai RH, Binder JR, Conant LL, Mano QR, Seidenberg MS (2011) The neural career of sensory-motor metaphors. J Cogn Neurosci 23(9):2376–2386. https://doi.org/10.1162/jocn.2010.21596
18. Desai RH (2022) Are metaphors embodied? The neural evidence. Psychol Res 86(8):2417–2433. https://doi.org/10.1007/s00426-021-01604-4
19. Dove G (2009) Beyond perceptual symbols: A call for representational pluralism. Cognition 110(3):412–431. https://doi.org/10.1016/j.cognition.2008.11.016
20. Dreyfus HL (1992) What computers still can't do: A critique of artificial reason. MIT Press
21. Firth J (1957) Papers in Linguistics, 1934–1951. Oxford University Press, London
22. Fodor JA (1975) The language of thought, vol 5. Harvard University Press

23. Gallese V, Lakoff G (2005) The brain's concepts: The role of the sensory-motor system in conceptual knowledge. Cognit Neuropsychol 22(3–4):455–479. https://doi.org/10.1080/02643290442000310
24. Gärdenfors P (2000) Conceptual spaces, vol 3. MIT Press, Cambridge, MA
25. Gibbs RW Jr, Lima PLC, Francozo E (2004) Metaphor is grounded in embodied experience. Journal of interdisciplinary J Philos 0:1–27. https://doi.org/10.1080/0020174X.2023.2296468
26. Gibbs RW Jr (2006) Metaphor interpretation as embodied simulation. Mind Lang 21(3):434–458. https://doi.org/10.1111/j.1468-0017.2006.00285.x
27. Grindrod J (2023) Distributional theories of meaning. In D. Bordonaba-Plou (Ed.), Experimental Philosophy of Language: Perspectives, Methods, and Prospects (pp. 75–99). Springer. https://doi.org/10.1007/978-3-031-28908-8_5
28. Grindrod J (2024) Large language models and linguistic intentionality. Synthese 204(2):71. https://doi.org/10.1007/s11229-024-04723-8
29. Gładziejewski P, Miłkowski M (2017) Structural representations: causally relevant and different from detectors. Biol Philos 32:337–355. https://doi.org/10.1007/s10539-017-9562-6
30. Grzankowski A (2024) Real sparks of artificial intelligence and the importance of inner interpretability. Inquiry: An for the prediction of spring breakup. Computer-Aided Civ Infrastruct Eng 39(2):264–280. https://doi.org/10.1111/mice.13074
31. Ha D, Schmidhuber J (2018) World Models. https://doi.org/10.5281/zenodo.1207631
32. Harris ZS (1954) Distributional structure. Word 10(2–3):146–162
33. Haugeland J (1989) Artificial intelligence: The very idea. MIT Press
34. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554–2558
35. Hopfield JJ (1988) Artificial neural networks. IEEE Circuits Devices Mag 4(5):3–10
36. Horst SW (1990) Symbols, computation and intentionality: A critique of the Computational Theory of Mind. University of Notre Dame
37. Johnson M (2013) The body in the mind: The bodily basis of meaning, imagination, and reason. University of Chicago Press
38. Johnson M, Lakoff G (1980) Metaphors we live by. University of Chicago Press
39. Johnston WJ, Fusi S (2023) Abstract representations emerge naturally in neural networks trained to perform multiple tasks. Nat Commun 14(1):1040. https://doi.org/10.1038/s41467-023-36583-0
40. Kelleher JD (2019) Deep learning. MIT Press
41. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
42. Kumar R, Sharma SC (2023) Hybrid optimization and ontology-based semantic model for efficient text-based information retrieval. J Supercomputing 79(2):2251–2280.

https://doi.org/10.1007/s11227-022-04708-9
43. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
44. Léon J (2021) Automating linguistics. Springer, Berlin, Germany
45. López-Rubio E (2018) Computational functionalism for the deep learning era. Mind Mach 28(4):667–688. https://doi.org/10.1007/s11023-018-9480-7
46. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133
47. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html
48. Millikan RG (1989) Biosemantics. J Philos 86(6):281–297. https://doi.org/10.2307/2027123
49. Minsky M (1967) Computation: Finite and infinite machines prentice hall. Inc., Engelwood Cliffs, NJ
50. Minsky ML (1969) Semantic information processing. The MIT Press
51. Mitchell M (2021) Why AI is harder than we think. arXiv preprint arXiv:2104.12871.
52. Moro A (2004) Autonomia della sintassi e tecniche di neuroimmagine. Lingue e linguaggio 3(1):133–130. 10.1418/13316
53. Moro A, Tettamanti M, Rotondi I, Perani D, Scotti G, Fazio F, Cappa SF (2013) Syntax without language: Neurobiological evidence for cross-domain syntactic computations. In *The equilibrium of human syntax* (pp. 317–341). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203083246-25/syntax-without-language-marco-tettamanti-irene-rotondi-daniela-perani-giuseppe-scotti-ferruccio-fazio-stefano-cappa-andrea-moro
54. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Barnes N, Mian AS (2023) A comprehensive overview of large language models. *ArXiv*. https://doi.org/10.48550/arXiv.2307.06435
55. Newell A, Simon HA (1964) Information processing in computer and man. Am Sci 52(3):281–300
56. Newell A, Simon HA (1995) GPS, a program that simulates human thought. In *Computation & Intelligence: collected readings* (pp. 415–428)
57. Papineau D (1998) Teleosemantics and indeterminacy. Australasian J Philos 76(1):1–14. https://doi.org/10.1080/00048409812348151
58. Pasquinelli M (2023) The eye of the master: A social history of artificial intelligence. Verso Books
59. Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. Nat Neurosci 32:15728–15736. https://doi.org/10.1523/JNEUROSCI.1953-12.2012
60. Piantadosi ST (2023) Modern language models refute Chomsky's approach to language. Manuscript. LingBuzz

61. Putnam H (1960) Minds and Machines in Sidney Hook, ed. In *Dimensions of Mind: A Symposium. New York University Press*
62. Pylyshyn ZW (1980) Computation and cognition: Issues in the foundations of cognitive science. Behav Brain Sci 3(1):111–132
63. Rogers A, Kovaleva O, Rumshisky A (2020) A primer in BERTology: What we know about how BERT works. Trans association Comput linguistics 8:842–866. https://doi.org/10.1162/tacl_a_00349
64. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386
65. Rumelhart DE, Hinton GE, McClelland JL (1986) A general framework for parallel distributed processing. Parallel Distrib processing: Explorations Microstruct cognition 1(45–76):26
66. Sahlgren M (2008) The distributional hypothesis. Italian J linguistics 20:33–53
67. Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424. https://doi.org/10.1017/S0140525X00005756
68. Shapiro L (2019) Embodied cognition. Routledge
69. Solomon KO, Barsalou LW (2004) Perceptual simulation in property verification. Mem Cognit 32(2):244–259. https://doi.org/10.3758/BF03196856
70. Striem-Amit E et al (2018) Neural representation of visual concepts in people born blind. Nat Commun 9:5250. https://doi.org/10.1038/s41467-018-07574-3
71. Sünderhauf N, Brock O, Scheirer W, Hadsell R, Fox D, Leitner J, Corke P (2018) The limits and potentials of deep learning for robotics. Int J Robot Res 37(4–5):405–420. https://doi.org/10.1177/0278364918770733
72. Tenney I, Das D, Pavlick E (2019) BERT rediscovers the classical NLP pipeline. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 4593–4601). Association for Computational Linguistics. https://doi.org/10.18653/v1/2019.acl-long.206
73. Turing AM (2021) Computing machinery and intelligence (1950). *Mind*, *59*(236), 33–60. 10.1093/oso/9780198250791.003.0017
74. Varela F, Thompson M, Rosch E (1991) The embodied mind. Cognitive science and human experience. The MIT Press
75. Wang X, Men W, Gao J, Caramazza A, Bi Y (2020) Two forms of knowledge representations in the human brain. Neuron 107(2):383–393. 10.1016/j.neuron.2020.04.010
76. Widrow B, Hoff ME (1988) Adaptive switching circuits. In *Neurocomputing: foundations of research* (pp. 123–134)
77. Williams I (2025) Can structural correspondences ground real world representational content in Large Language Models? arXiv preprint arXiv:2506.16370.
78. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J (2019) Task representations in neural networks trained to perform many cognitive tasks. Nat Neurosci 22:297–306. https://doi.org/10.1038/s41593-018-0310-2

Footnotes

1. This perspective can also be situated within a broader tradition of structural and distributional approaches to language, from early structural linguistics to Firth's context-based view of meaning, in which semantic content is understood through patterns of relation rather than through isolated symbols.
2. i.e., *good old-fashioned artificial intelligence*.
3. During the 1960s, seeking to answer the question of the possibility of artificial intelligence 'understanding' information, Marvin Minsky discussed several of the most promising programmes developed in that decade (Minsky, 1969). Between them, Daniel Bobrow's Student programme, which aims to solve algebraic problems formulated in words, is particularly relevant. According to Minsky, achieving this goal constitutes 'the demonstration of the power of meaning in solving linguistic problems'. The programme worked by breaking down sentences into smaller units, using the recurrence of certain keywords. If this breakdown did not lead to the resolution of the algebraic problem, the programme proceeded along another path, 'unpacking' the sentences according to other criteria. *Student* treated words as symbols and did by embedding linguistic-mathematical mapping rules (Bobrow, 1964).
4. In a recent study (Desai, 2022), a wide range of metaphorical expressions (MET; e.g. *The man grasped the idea*) are examined alongside abstract or non-sensory control phrases (ABS; e.g. *He had a bad day*). Literal action expressions (LIT; e.g., *The man grabbed the handle*) are also used in the research. The prediction is that sensorimotor areas should be activated by METs to a greater extent than ABSs. In the case of metaphors involving, for example, concepts of action, neuroimaging studies confirm the prediction in the superior motor areas associated with complex action planning and hand-object interaction in the left anterior inferior parietal lobe (Desai, 2011).
5. Neuroimaging research indicates that conceptual processing engages different neural regions depending on the type of content involved. While concepts closely tied to perception tend to recruit sensory-related areas, more abstract concepts preferentially activate associative and language-related networks. Several studies have highlighted a differentiated organisation within the anterior temporal lobe (ATL), suggesting that semantic representation is not supported by a single, uniform neural substrate. Evidence from fMRI investigations shows that distinct subregions of the ATL respond selectively to different conceptual properties. The dorsal portion appears to be more responsive to concepts that lack direct sensory features but retain clear object referents, whereas lateral regions are more strongly associated with highly abstract concepts that do not refer to concrete entities. By contrast, medial areas of the ATL show a tendency toward perceptually grounded concepts (Striem-Amit et al., 2018; Bottini, 2020; Wang et al., 2020). This pattern of activation suggests that semantic processing within the ATL is organised along at least two interacting dimensions: sensory accessibility and referential structure. Rather than encoding meaning in a homogeneous manner, the ATL appears to modulate its activity according to how concepts vary along these dimensions, supporting representations that range from concrete and perceptual to abstract yet internally structured. Converging evidence for this view is provided by

Peelen and Caramazza (2012), who demonstrated that ventral regions of the anterior temporal lobe encode not only perceptual attributes but also higher-level, abstract conceptual properties. Together, these findings support the idea that semantic representation in the brain cannot be reduced to sensory grounding alone but depends on internal organisational principles that operate across modalities, an organisation that reflects the representational flexibility observed in neural networks trained to integrate information across multiple tasks.

6. Wang et al. (2020) identify two broad representational pathways: one closely tied to perceptual systems and primarily involved in processing concrete concepts, and another of a non-sensory character, associated with linguistic and associative regions, which supports the representation of abstract content.
7. Bi (2021) claims that conceptual cognition should be described in terms of two parallel and complementary neural codes. According to this model, a sensory-based code rooted in perceptual–motor areas operate alongside an amodal, linguistically mediated code that enables understanding of concepts lacking a direct experiential correlate.
8. It should be noted, however, that contemporary systems such as LLMs do not use vector semantics but statistical inferences. Nevertheless, the revolution represented by this vector approach marked a decisive shift from symbolic AI to 'sub-symbolic' systems.
9. These structures should not be understood as fully grounded representations of reality, but as relational configurations capable of supporting context-sensitive inference, abstraction, and domain-relative generalisation.
10. i.e., activation patterns that can support the coordination of domain-relevant relational regularities across contexts.
11. The present proposal does not aim to replace causal or Teleosemantics/Biosemantics theories of reference, but to describe a weaker form of semantic organisation observable in contemporary neural systems. While Biosemantics approaches explain semantic content through biological function and causal history, the present framework focuses instead on the internal relational organisation that enables systematic inferential competence.
12. The question is therefore no longer whether embodiment grounds meaning, but whether structural alignment with worldly regularities can perform the same explanatory work.
13. In this sense, DNNs internal configurations may exhibit forms of structural isomorphism with aspects of the domains they are trained on, even in the absence of direct causal interaction.
14. However, not every structured or information-preserving system exhibits semantic competence. Simple systems such as thermostats, databases, or compression algorithms can preserve and process information without supporting context-sensitive inference, abstraction, or flexible relational coordination across domains. Structural semantic competence therefore depends not merely on the existence of organised information, but on the capacity of a system to sustain dynamically integrated inferential relations capable of generalisation across contexts.

15. In this framework, emergence should not be understood as a mysterious ontological transition, but as the functional stabilisation of relational organisations capable of supporting abstraction, inference, and cross-contextual generalisation
16. Although the present proposal is primarily philosophical, it may also suggest empirically testable distinctions between structurally organised semantic systems and systems based solely on information compression or statistical retrieval. If structural semantic competence depends on relational organisation rather than mere compression, systems preserving stable inferential relations across contexts should exhibit greater abstraction and cross-domain generalisation than systems based solely on statistical retrieval or information compression.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MANUSCRIPTINFORMATION.pdf