Language as a Generalized Computational-Semantic System

Sara Malik[1] and Naveed A. A.[1]

[1]*PakCrypt NPO*

Correspondence: smk@pakcrypt.org

Abstract

We propose a generalized concept of language that subsumes natural language, formal languages, mathematics, and systems exhibiting genuine nondeterminism. Within this framework, Chomsky's hierarchy is treated as a classical, deterministic subcase of a broader computational-semantic hierarchy that admits probabilistic and quantum extensions. We argue that a language, in the general sense, is any finitely specifiable symbol system paired with an interpretation function that grounds its expressions in a domain of reference—physical, abstract, or social. Natural language is the subset whose interpretation function is anchored in shared biological cognition and social practice. Mathematics is the subset whose interpretation function ranges over abstract structures, some of which have been empirically realized as physical descriptions. We show that this generalization dissolves several long-standing conflicts—between internalist and externalist semantics, between formal and natural language, and between deterministic and probabilistic computation—without requiring rejection of established results. We identify residual tensions with the symbol grounding problem, the hard problem of consciousness, and personal identity under information-theoretic reduction, and we argue that these tensions are features of the landscape rather than defects of the proposal. The framework is intended as a defensible philosophical hypothesis that organizes existing results, not as an empirical theory that overturns them.

Keywords: philosophy of language; Chomsky hierarchy; computation; semantics; quantum information; mathematical realism

1. Introduction

The question "what is language?" remains contested. Chomsky treats language as an internal computational capacity (Chomsky 1986, 1995). Wittgenstein treats it as a family of social practices (Wittgenstein 1953). Montague treats it as a formal system interpretable by model theory (Montague 1970). Each tradition captures genuine features. None subsumes the others.

We propose that these positions are not rival definitions of a single object but descriptions of adjacent regions within a larger space. That space is the space of generalized languages: symbol systems equipped with interpretation functions. Natural language, formal languages, mathematics, and probabilistic or quantum symbol systems are all instances. Our task is to give this generalization a precise statement, show that it is consistent with established results, and indicate where it yields new clarity.

The paper is organized as follows. Section 2 states the generalized definition. Section 3 places Chomsky's hierarchy within it. Section 4 extends the hierarchy to nondeterministic and quantum regimes. Section 5 treats mathematics as a language under the generalization. Section 6 addresses semantics and the grounding problem. Section 7 discusses computational limits and the Church–Turing thesis. Section 8 considers self-reference, reflexivity, and the open question of consciousness. Section 9 addresses information, identity, and residual conflicts. Section 10 concludes.

2. A Generalized Definition of Language

Definition. A language L is a triple (Σ, G, I), where Σ is a countable set of symbols, G is a finitely specifiable generative procedure that produces well-formed expressions over Σ, and I is an interpretation function that maps expressions to elements of some domain D.

Three points follow. First, the definition is neutral about the nature of D. D may be physical states, abstract structures, social commitments, or mental representations. Second, G need not be deterministic. It may assign probabilities to expressions, or—in the quantum case—amplitudes. Third, I need not be total. Some expressions may be undefined under I, either because D is incomplete or because the expression lies outside the grounded region of the symbol system.

Natural language arises when I is anchored in shared biological cognition and social practice (Tomasello 2008; Brandom 1994). Formal languages arise when I is stipulated. Mathematical languages arise when D is a domain of abstract structures. The generalization does not deny the distinctive features of any of these; it organizes them.

This definition absorbs several standing disputes. The Chomsky–Wittgenstein tension is recast as a difference over which component—G or I—is primary. Chomsky emphasizes G; Wittgenstein emphasizes I and its embedding in practice. Both are correct about their component.

3. The Chomsky Hierarchy as a Classical Subcase

Chomsky's hierarchy classifies formal grammars by the restrictions placed on their production rules, yielding regular, context-free, context-sensitive, and unrestricted languages, recognized respectively by finite automata, pushdown automata, linear bounded automata, and Turing machines. Natural languages are now understood to occupy the mildly context-sensitive region (Joshi 1985; Shieber 1985), which is polynomially parsable though more expressive than context-free.

Within our framework, Chomsky's hierarchy is the deterministic classical slice of a larger space. It fixes G as a rewrite system and leaves I implicit. This is appropriate for the mathematical study of syntax but leaves the semantic dimension unaddressed. The generalized definition restores I as a first-class component.

4. Nondeterministic and Quantum Extensions

The classical hierarchy can be extended by allowing G to be probabilistic. Probabilistic context-free grammars (Manning and Schütze 1999) are already standard in computational linguistics. A further extension allows G to be genuinely quantum: expressions are superpositions, and I maps them to amplitudes over D. Quantum finite automata and quantum pushdown systems have been studied formally (Moore and Crutchfield 2000); they recognize language classes incomparable to their classical counterparts.

We do not claim that natural language is quantum in any substantive physical sense. The evidence against quantum cognition at neural timescales is strong (Tegmark 2000). We claim only that the generalized notion of language admits such systems as legitimate members, and that the indeterminacy inherent in quantum measurement gives a precise formal analogue of the semantic indeterminacy familiar from Quine (1960). This is a structural parallel, not a physical hypothesis.

5. Mathematics as a Language

Under the generalized definition, mathematics qualifies as a language: it has symbols, generative rules, and an interpretation function ranging over abstract structures. This does not collapse the distinction between mathematics and natural language. The two differ in what D contains and in how I is fixed. Mathematical I is stipulated by axioms and definitions; natural linguistic I is shaped by biology and practice.

The user's intuition that applied mathematics is mathematics "with semantics in reality" can now be stated precisely. Applied mathematics is the subset of mathematical expressions whose interpretation function has been empirically realized by a mapping from abstract structures to physical states. Pure mathematics is not semantically empty; its I ranges over abstract D. Wigner's puzzle (Wigner 1960) concerning the effectiveness of mathematics in natural science becomes the question of why so many abstract D admit physical realizations. Our framework does not solve this puzzle, but it states it without confusion.

6. Semantics, Internalism, and the Grounding Problem

The apparent conflict between internalist and externalist semantics dissolves under the generalized definition. I has two components: a cognitive component that depends on the agent's internal representations (Frege's sense; conceptual role semantics), and a referential component that depends on the environment (Putnam 1975; Kripke 1980). Both are real, and both are necessary. The symbol grounding problem (Harnad 1990) identifies the point at which the cognitive component must make contact with non-symbolic sensorimotor content. We take this as a genuine and unresolved feature of the landscape.

A corollary: a sentence that is grammatically well-formed but semantically anomalous to a given interpreter is not thereby meaningless. It may be ungrounded in that interpreter's D while remaining well-formed under G. This matches the user's original intuition and preserves it without contradiction.

7. Computation, the Church–Turing Thesis, and Limits

A Turing machine corresponds to a decision procedure for a language—the set of inputs it accepts. A universal Turing machine simulates any Turing machine and therefore recognizes any recursively enumerable language. It does not recognize "any language" in the unrestricted sense: by a cardinality argument, the set of all languages over a non-trivial alphabet is uncountable, while the set of Turing machines is countable. Most languages, in the set-theoretic sense, lie outside Turing recognizability.

If the physical Church–Turing thesis holds (Deutsch 1985), then Turing computability bounds the computational resources available to any physically realized agent, and therefore bounds the languages any such agent can interface with. Natural language competence lies well within this bound. The generalized framework preserves this result and adds only that probabilistic and quantum extensions modify efficiency, not the class of decidable languages (Bernstein and Vazirani 1997).

8. Reflexivity, Self-Modification, and Consciousness

Natural languages are semantically closed: they contain their own truth predicates and can refer to and modify themselves. Formal languages, by Tarski's hierarchy (Tarski 1944), avoid this at the cost of expressive closure. Gödel's arithmetization (Gödel 1931) shows that even restricted formal systems admit genuine self-reference. Within the generalized framework, reflexivity is a property that some languages possess and others do not, and it is a defining feature of natural language.

We are cautious about the leap from reflexivity to consciousness. Hofstadter (2007) offers a suggestive account in which strange loops of self-reference constitute selfhood. Theories of consciousness such as Global Workspace Theory (Dehaene 2014) and Integrated Information Theory (Tononi 2008) do not require linguistic reflexivity. Proposals linking consciousness to quantum processes in neurons (Penrose and Hameroff 1996) face decisive objections from decoherence timescales (Tegmark 2000). We therefore do not claim that sufficiently rich reflexive languages are conscious. We claim only that reflexivity is a necessary ingredient of any system that models itself, and that this is a precondition—not a sufficient condition—for whatever cognitive phenomena underlie consciousness.

9. Information, Identity, and Residual Tensions

If an intelligent agent is individuated by the information encoded in its computational and linguistic state, then preserving that information preserves the agent. Physics increasingly supports the view that information is conserved under fundamental processes (Maldacena 1998; Penington 2020). But informational identity is type identity, not token identity. A perfect copy is the same pattern but not the same individual (Parfit 1984). The hard problem of consciousness (Chalmers 1995) introduces a further gap: functional duplication may not entail experiential duplication.

We accept these residual tensions. The generalized framework clarifies what is at stake without resolving them. An intelligent being is information in the sense that its cognitive and linguistic state is finitely specifiable in principle. It is not information in the sense that any instantiation of that specification is numerically the same being. These are distinct claims and should be kept distinct.

10. Conclusion

We have proposed a generalized concept of language as a triple (Σ, G, I), with Chomsky's hierarchy as its deterministic classical subcase. The generalization accommodates probabilistic and quantum extensions, treats mathematics as a legitimate member, and dissolves several standing conflicts between internalist and externalist semantics and between formal and natural language. It leaves open the symbol grounding problem, the hard problem of consciousness, and the metaphysics of personal identity. We take these residual tensions as genuine open questions rather than defects of the proposal. The framework's purpose is organizational: to state known results in a common vocabulary and to mark precisely where the remaining work lies.

References

Brandom, R. (1994). Making It Explicit. Harvard University Press.

Bernstein, E., and Vazirani, U. (1997). Quantum complexity theory. SIAM Journal on Computing, 26(5), 1411–1473.

Chalmers, D. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200–219.

Chomsky, N. (1986). Knowledge of Language. Praeger.

Chomsky, N. (1995). The Minimalist Program. MIT Press.

Dehaene, S. (2014). Consciousness and the Brain. Viking.

Deutsch, D. (1985). Quantum theory, the Church–Turing principle and the universal quantum computer. Proceedings of the Royal Society A, 400, 97–117.

Gödel, K. (1931). Über formal unentscheidbare Sätze. Monatshefte für Mathematik und Physik, 38, 173–198.

Harnad, S. (1990). The symbol grounding problem. Physica D, 42, 335–346.

Hofstadter, D. (2007). I Am a Strange Loop. Basic Books.

Joshi, A. (1985). Tree adjoining grammars. In Natural Language Parsing, Cambridge University Press.

Kripke, S. (1980). Naming and Necessity. Harvard University Press.

Maldacena, J. (1998). The large N limit of superconformal field theories. Advances in Theoretical and Mathematical Physics, 2, 231–252.

Manning, C., and Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.

Montague, R. (1970). Universal grammar. Theoria, 36, 373–398.

Moore, C., and Crutchfield, J. (2000). Quantum automata and quantum grammars. Theoretical Computer Science, 237, 275–306.

Parfit, D. (1984). Reasons and Persons. Oxford University Press.

Penington, G. (2020). Entanglement wedge reconstruction and the information paradox. JHEP, 09, 002.

Penrose, R., and Hameroff, S. (1996). Conscious events as orchestrated space-time selections. Journal of Consciousness Studies, 3, 36–53.

Putnam, H. (1975). The meaning of 'meaning'. Minnesota Studies in the Philosophy of Science, 7, 131–193.

Quine, W. V. O. (1960). Word and Object. MIT Press.

Shieber, S. (1985). Evidence against the context-freeness of natural language. Linguistics and Philosophy, 8, 333–343.

Tarski, A. (1944). The semantic conception of truth. Philosophy and Phenomenological Research, 4, 341–376.

Tegmark, M. (2000). Importance of quantum decoherence in brain processes. Physical Review E, 61, 4194–4206.

Tomasello, M. (2008). Origins of Human Communication. MIT Press.

Tononi, G. (2008). Consciousness as integrated information. Biological Bulletin, 215, 216–242.

Wigner, E. (1960). The unreasonable effectiveness of mathematics in the natural sciences. Comm. Pure Appl. Math., 13, 1–14.

Wittgenstein, L. (1953). Philosophical Investigations. Blackwell.