

Regulatory Model of Language Origin: Imitation, Behavioral Forms, and Semantization

Alexey A. Nekludoff

AstraVerge Research

E-mail: an@astraverge.org

December 26, 2025

Abstract

This work proposes a regulatory ontology of language that departs from both representational and usage-based paradigms. Instead of treating language as a system of meanings, symbols, or socially stabilized rules, language is analyzed as a regulatory structure that stabilizes coordinated behavior under conditions of uncertainty.

Drawing on evidence from developmental psychology, gesture studies, neurolinguistics, and the emergence of non-verbal communicative systems, the analysis locates the origin of language in pre-symbolic mechanisms of imitation, bodily synchronization, and the stabilization of behavioral forms. Dictionaries and alphabets are interpreted not as primitive components of language, but as historically stabilized layers that externalize prior regulatory distinctions.

The proposed framework is articulated through a set of ontological commitments and their necessary consequences, which delineate principled limits of translation, information-theoretic models of communication, and purely text-based approaches to language acquisition. These limits are examined through conceptual analyses of translation failure, information theory, the search for extraterrestrial intelligence, and large language models.

The central claim is that what is observable as language constitutes only the externalized residue of historically viable regulation. Beyond this residue lie not semantic gaps or informational noise, but the ontological boundaries of regulation itself.

1 Introduction

Language is commonly approached either as a representational medium through which the world is described, or as a system of socially stabilized rules governing linguistic practice. Both perspectives have generated powerful analytical frameworks, yet both presuppose language in an

already constituted form: as a system of meanings, symbols, and norms available for reflection and use.

This work adopts a different point of departure. Rather than asking how language represents the world or how linguistic rules are applied, it asks how language becomes possible at all. The guiding hypothesis is that language originates not as a symbolic or propositional system, but as a regulatory mechanism that stabilizes coordinated behavior under conditions of uncertainty.

From this perspective, communication and meaning are not primary features of language, but late-stage effects of prior regulatory processes. Imitation, bodily synchronization, and the stabilization of behavioral forms constitute the minimal substrate from which dictionaries, alphabets, and linguistic structures eventually emerge.

The aim of the present work is therefore not to offer a comprehensive linguistic theory, but to articulate an ontological framework within which the emergence, limits, and failures of language can be understood as consequences of regulation. By doing so, the analysis delineates principled boundaries for translation, information-theoretic models, and contemporary claims about artificial language competence.

The argument proceeds by introducing a set of ontological commitments, deriving their necessary consequences, and examining the limits of observable language through a series of conceptual and applied cases.

2 Motivation and Problem Statement

Contemporary theories of language are characterized by pronounced methodological fragmentation. Representational approaches treat language as a medium for describing or encoding the world, whereas usage-based and pragmatic frameworks emphasize its role as a system of socially stabilized norms and practices. Despite their differences, both traditions share a common point of departure: they analyze language in its mature, culturally stabilized form and focus on already constituted semantic, grammatical, and discursive structures.

As a consequence, the processes through which language itself comes into being remain largely untheorized. In particular, dominant philosophical models tend to abstract away from pre-linguistic and extra-symbolic mechanisms of coordination that precede the emergence of explicit meaning and syntax.

At the same time, a substantial body of empirical evidence has accumulated in cognitive science, developmental psychology, and neurolinguistics, indicating the existence of such mechanisms. Imitation, bodily synchronization, and recurrent motor patterns of joint action are observed in early ontogenesis, in spontaneously emerging gestural systems, and in conditions of partial or complete impairment of speech. These phenomena point to the presence of regulatory structures that operate prior to semantic articulation and grammatical organization.

Despite their empirical accessibility, these regulatory mechanisms have not been integrated into a coherent philosophical account of language. What is missing is a conceptual framework

capable of relating pre-linguistic forms of coordination to later symbolic and discursive structures without reducing language either to representation or to social normativity.

The open problem addressed in this work is therefore the construction of a theoretical model that can:

- relate pre-linguistic mechanisms of imitation and synchronization to the emergence of dictionaries and languages;
- explain the transition from behavioral regulation to semantic and grammatical stabilization;
- characterize language in ontological terms, as a system of regulatory processes, rather than in purely epistemic terms, as a system of propositions or rules of use.

The present work proposes a solution to this problem in the form of a regulatory model of language. This model is grounded in an axiomatic description of uncertainty minimization and the stabilization of behavioral forms, and treats ontology not as a description of the world as such, but as a stabilized mode of extracting regulatory constraints from historical experience.

Methodological Remark

In this work, ontology is not understood as a description of the world “as it is in itself.” Rather, ontology is treated as a stabilized form of extracting regulatory constraints from historical experience. As such, it inevitably incorporates epistemic conditions of its own applicability, including criteria of distinction, operationalization, and stabilization.

This is not a limitation of ontology, but a structural consequence of its human origin: any ontology accessible to us is already mediated by the modes through which regulatory constraints are identified, tested, and retained.

3 The Dictionary as the Primary Regulatory Structure

Within the regulatory model proposed in this work, the *dictionary* is the primary structure from which language emerges. Here, the term *dictionary* does not denote a list of words, lexical entries, or symbolic definitions. Rather, it designates a system of regulatorily significant distinctions that enable coordination, stabilization of behavior, and the reduction of ontological uncertainty within and between localities.¹

The dictionary arises prior to language, symbols, and external representations. It precedes grammar, syntax, and semantics not merely in an empirical or historical sense, but in an ontological sense: without a dictionary there is nothing that could later be named, externalized, or stabilized as linguistic structure.

¹The term *locality* is used in an extended ontological sense developed in the author’s broader framework of coordination meta-structures; see [9].

3.1 Why the Dictionary Comes First

The dictionary comes first because survival, orientation, and coordination require distinctions before they require symbols. Any viable system must already discriminate between what is relevant and irrelevant, safe and dangerous, compatible and incompatible. Such discriminations are not linguistic; they are regulatory.

In this sense, the dictionary determines the minimal *capacity of distinction* required for viability. It specifies what must be distinguished in order to act coherently, which forms of behavior can be stabilized, and which expectations can be aligned across localities. Language does not create these distinctions; it later externalizes and stabilizes them.

3.2 The Ontological Content of a Dictionary

Ontologically, a dictionary consists of structured regulatory components, including:

- **Regulatory distinctions:** stable differentiations that constrain action and expectation;
- **Operations:** what can be done with these distinctions (activation, inhibition, reproduction, coordination);
- **Relations:** compatibility, conflict, precedence, and generative dependence between distinctions;
- **Forms of behavior:** stabilized patterns of action emerging from repeated coordination;
- **Mechanisms of alignment:** processes through which multiple localities converge on compatible behavior;
- **Proto-semantic correspondences:** early mappings between forms and expectations, prior to explicit semantic articulation.

At this level no symbols are required. The dictionary is fully operational without words, signs, or representations. What later appears as lexical content is a secondary external trace of a pre-symbolic regulatory structure.

3.3 Dictionary Versus Lexicon

A crucial distinction must be drawn between the dictionary, in the present ontological sense, and a *lexicon* (or vocabulary) understood as a symbolic inventory. A lexicon is an externalized artifact that presupposes the existence of a dictionary. The dictionary, by contrast, is a pre-symbolic regulatory structure.

While a dictionary may later be approximated by a list of words, such approximation obscures its ontological function. A *word* is not the origin of a distinction; it is a stabilized label

attached to a distinction that already operated regulatorily. Meaning does not precede the dictionary. Meaning emerges as a stabilized correspondence between regulatory form and expectation once externalization becomes possible.

3.4 Environmental Dependence of Dictionaries

Dictionaries are environment-specific. They arise in response to the demands of ecological, social, and cognitive contexts. This explains why different communities can develop radically different dictionaries even when their later linguistic structures appear comparable at a superficial level.

Some distinctions are widely recurrent because they correspond to broadly shared conditions of human existence. Others remain highly localized, reflecting specific modes of engagement with an environment. In this sense, the dictionary does not primarily encode *descriptions* of the world; it encodes conditions of viable interaction with it.

3.5 The Dictionary Does Not Define Linguistic Rules

A further consequence is that the dictionary does not determine grammatical structure or rules of sentence formation. It provides regulatory content for coordination, not formal constraints on expression. Grammatical organization emerges later as a secondary stabilization of externalized regulatory distinctions.

This separation helps explain both the diversity and the flexibility of linguistic systems: languages differ not because their dictionaries impose distinct grammatical logics, but because grammar is a subsequent layer of stabilization rather than the generative source of distinctions.

3.6 Dictionaries as Conditions of Access

Differences between dictionaries are often misconstrued as differences in information or expressive power. From a regulatory perspective, this framing is insufficient. A dictionary structures *access*: it determines which distinctions are available as regulatorily significant and which expectations can be aligned.

Consequently, dictionary differences can function as conditions of partial opacity between systems. Where dictionaries do not overlap, understanding does not fail because of missing information; it fails because there is no shared space of regulatory distinctions. This property becomes decisive for the subsequent analysis of translation limits and the ontological conditions of communication.

4 From Dictionary to Alphabet

The distinction between dictionary and alphabet marks a transition between two ontological layers of language. The dictionary establishes *regulatory content*: it specifies which distinctions are behaviorally and coordinatively significant within a given environment. The alphabet, by contrast, establishes *ontological infrastructure*: it specifies the conditions under which such distinctions can be externally fixed, reproduced, and stabilized beyond immediate interaction.

This transition is not a linear progression from content to form, nor a historical narrative of linguistic development. Rather, it reflects a structural shift from regulation enacted in behavior to regulation supported by external persistence. A dictionary can operate without an alphabet, remaining embedded in local and transient coordination. An alphabet becomes necessary only when regulatory distinctions must outlive their immediate enactment and circulate across time, space, or heterogeneous localities.

Understanding this transition is crucial for distinguishing failures of communication that arise from differences in regulatory content from those that arise from differences in externalization infrastructure. The following sections examine the ontological consequences of such mismatches.

5 The Alphabet as an Ontologically Fundamental Structure

In the regulatory model, the *alphabet* constitutes an ontologically fundamental structure that underlies the externalization and stabilization of language. Unlike the dictionary, which is primary with respect to regulatory function, the alphabet is primary with respect to ontological conditions of possibility. It defines the minimal structure through which distinctions can be fixed, reproduced, and made accessible beyond immediate behavioral coordination.

Crucially, the alphabet does not coincide with letters, phonemes, or writing systems. These are historically contingent realizations of a deeper ontological layer. The alphabet, in the present sense, designates a minimal system of elements, operations, and relations that makes external representation and stabilization of regulatory distinctions possible.

5.1 Temporal and Ontological Asymmetry

The alphabet exhibits a fundamental asymmetry between temporal emergence and ontological priority. Temporally, alphabets arise later than dictionaries: regulatory distinctions and forms of behavior precede any systematic external fixation. Ontologically, however, the alphabet is prior insofar as it provides the conditions under which a dictionary can be externalized, transmitted, and stabilized across time and localities.

This asymmetry explains why a dictionary can operate without an alphabet, while an alphabet without a dictionary is empty. Regulatory distinctions can guide behavior directly, but

without an alphabet they remain bound to local, transient coordination. The alphabet introduces a layer of ontological persistence, allowing distinctions to be fixed independently of their immediate enactment.

5.2 The Alphabet Is Not a System of Symbols

It is essential to emphasize that the alphabet is not a symbolic system in the narrow sense. Symbols presuppose stabilized correspondences between form and meaning, whereas the alphabet defines the preconditions under which such correspondences can arise. The alphabet does not encode meaning; it enables the possibility of encoding.

Accordingly, phonetic alphabets, logographic systems, ideographic scripts, and even non-linguistic inscription systems (such as diagrams or notational conventions) are to be understood as specific instantiations of an underlying ontological alphabet. They differ in modality and historical realization, but they rely on the same minimal structure of externalizable distinction.

5.3 Ontological Primitives of the Alphabet

At the ontological level, the alphabet can be decomposed into three interrelated classes of primitives.

Elements. Elements are minimal distinguishable units that can be externally fixed. They need not correspond to sounds or letters; rather, they instantiate the capacity for *distinguishability* itself.

- **Distinguishability:** the minimal ability to register “not-the-same”;
- **Compatibility:** the capacity to be combined or aligned with other elements;
- **Invariance:** what remains stable across transformations;
- **Uncertainty:** the divergence between a current configuration and a horizon of expectations.

Operations. Operations specify what can be done with elements once they are distinguishable and externally fixable.

- **Imitation:** copying of structure rather than signal;
- **Synchronization:** alignment of states across localities;
- **Stabilization:** extraction and retention of invariants;
- **Mapping:** the proto-semantic association between form and expectation.

Relations. Relations define the structural organization of elements and operations.

- **Equivalence of forms:** different configurations fulfilling the same regulatory function;
- **Compatibility of localities:** coordination without mutual disruption;
- **Directionality of alignment:** asymmetries of influence and adaptation;
- **Generation of new forms:** emergence of previously unavailable structures.

Together, these primitives define the alphabet as a minimal ontological infrastructure for externalization.

5.4 Alphabet and Externalization

The alphabet emerges as an ontological layer at the moment language becomes externalized. Externalization does not merely record pre-existing language; it transforms the conditions under which regulation operates. Once distinctions are externally fixed, they acquire persistence, transportability, and recombining beyond local coordination.

This process explains why alphabets often appear retrospectively: they are reconstructed once vocal or behavioral regulation is translated into stable external forms. Prior to externalization, the alphabet does not exist as an explicit system, yet it remains ontologically implicit as a necessary condition of any such translation.

5.5 Alphabet Versus Dictionary

The distinction between alphabet and dictionary is foundational. The dictionary answers the question *what must be distinguished in order to act and survive*. The alphabet answers the question *how distinction as such can be externally fixed and stabilized*.

The alphabet does not generate regulatory content; it provides the ontological means for its persistence. Conversely, the dictionary supplies regulatory distinctions but does not determine the structure of their externalization. Their relationship is therefore complementary rather than hierarchical.

This distinction becomes decisive in the analysis of communication breakdowns, translation limits, and the ontological conditions under which information-theoretic and computational models remain applicable.

6 Ontological Commitments of the Regulatory Model

The regulatory model of language proposed in this work rests on a set of ontological commitments. These commitments are not introduced as axioms in a formal sense, nor as self-evident

metaphysical truths. Rather, they articulate stable constraints that emerge from the analysis of language as a historically evolved system of regulation.

Taken together, they delimit the space within which the subsequent arguments, consequences, and limits of the model are formulated.

Commitment 1: Regulation Precedes Representation

Language is not ontologically grounded in representation. Its primary function is the regulation of coordinated behavior under conditions of uncertainty. Representational and descriptive uses of language presuppose prior regulatory stabilization and therefore cannot serve as the foundation of language itself.

Commitment 2: Distinction Precedes Symbolization

No symbol, word, or sign can function without a prior distinction that renders it regulatorily significant. Symbols do not generate distinctions; they externalize and stabilize distinctions that already operate within behavioral and coordinative processes. Accordingly, dictionaries precede alphabets functionally, even if alphabets later become ontologically fundamental for externalization.

Commitment 3: Imitation and Synchronization Are Primary Mechanisms

The emergence of stable linguistic structures presupposes mechanisms of imitation and synchronization. These mechanisms enable the reproduction and alignment of behavioral forms across localities and operate prior to semantic articulation or grammatical organization. Without such mechanisms, no shared regulatory structure can arise.

Commitment 4: Meaning Is Derivative of Stabilized Regulation

Meaning does not constitute an independent ontological layer. It emerges as a stabilized correspondence between regulatory forms of behavior and structures of expectation. Semantic content is therefore secondary with respect to the processes through which regulation is established and maintained.

Commitment 5: Rules Are Late Stabilizations

Rules of use, grammatical constraints, and normative structures are not primitive elements of language. They arise as late-stage stabilizations of already semantized regulatory forms. Consequently, normative descriptions of language capture its surface regularities, but not its ontological foundations.

Commitment 6: Ontology Retains Epistemic Conditions of Applicability

Ontology, as employed in this model, does not describe the world as such. It fixes stable forms of deriving regulatory constraints from historical experience. For this reason, ontological commitments inevitably retain epistemic conditions of their own applicability, including criteria of distinction, operationalization, and stabilization.

This inclusion of epistemic elements is not a conceptual flaw, but a structural consequence of the fact that any ontology accessible to us is articulated from within a historically situated regulatory perspective.

7 Necessary Consequences of the Regulatory Ontology

The ontological commitments outlined above entail a number of necessary consequences for how language can emerge, function, and be analyzed. These consequences are not presented as empirical generalizations or contingent theoretical claims. Rather, they follow from the internal structure of the regulatory model itself and delineate the limits within which alternative accounts remain viable.

Consequence 1: Logical Form Is Not Ontologically Primary

If regulation precedes representation and rules arise only as late stabilizations, then logical form cannot serve as the ontological foundation of language. Logical structure presupposes stable rules of combination and use, which themselves depend on prior regulatory and semantic stabilization. Accordingly, logical and formal models capture derived properties of language, not its generative conditions.

Consequence 2: Meaning Is Secondary to Form

Within the regulatory ontology, meaning cannot be primary. Stabilized forms of behavior must already exist before any correspondence between form and expectation can be established. Semantic content therefore emerges as a derivative layer, dependent on the persistence and reproducibility of regulatory forms. Any theory that treats meaning as ontologically primitive necessarily abstracts away from the processes that make meaning possible.

Consequence 3: Imitation Is a Necessary Condition for Language Emergence

Given the centrality of imitation and synchronization, no linguistic system can arise in their absence. Without mechanisms that allow behavioral forms to be reproduced and aligned across localities, neither shared dictionaries nor stable semantic structures can form. This consequence

places strict constraints on accounts that attempt to derive language from purely inferential or representational capacities.

Consequence 4: Rules of Use Cannot Ground Language

Usage-based descriptions correctly characterize how language functions once it is stabilized, but they cannot explain how language originates. Rules of use presuppose prior semantic and regulatory alignment and therefore cannot serve as an ontological starting point. Language games describe late-stage equilibria, not the processes that generate them.

Consequence 5: Language Is a Mechanism of Uncertainty Reduction

If the primary function of language is regulation, then its fundamental role is the reduction of ontological uncertainty through coordinated action. Communication, description, and information exchange are secondary effects of this process. Language must therefore be analyzed as a mechanism for stabilizing expectations and behavior, rather than as a system for transmitting representations of the world.

Consequence 6: Semantic Theories Detached from Behavior Are Incomplete

Any account of meaning that abstracts from the behavioral origins of regulatory forms necessarily remains incomplete. Semantic structures cannot be fully understood without reference to the forms of coordination from which they emerge. This consequence aligns philosophical analysis with empirical findings in developmental psychology and neurocognitive studies of language impairment.

Consequence 7: Thought Is Not Dependent on Linguistic Form

Because regulatory coordination can stabilize prior to semantic and grammatical articulation, cognitive structures need not depend on language in its externalized form. This consequence is consistent with evidence from aphasia and other conditions in which linguistic expression is impaired while complex reasoning and planning remain intact. Language enhances and extends cognition, but it does not constitute its ontological basis.

Consequence 8: Stable Imitative Communities Tend Toward Linguistic Emergence

Wherever sustained imitation and synchronization occur within a group, regulatory forms tend to stabilize and accumulate. Over time, this process gives rise to dictionaries and eventually to

externalized languages. The emergence of language in such contexts is therefore not accidental, but a structural consequence of sustained regulatory interaction.

8 Ontological Consequences of Alphabet Mismatch

Alphabet mismatch occurs when systems do not share the same ontological infrastructure for externalizing and stabilizing distinctions. This form of mismatch is often misconstrued as a problem of encoding, noise, or insufficient decoding procedures. Within the regulatory ontology of language, however, alphabet mismatch constitutes a deeper ontological limitation.

Because the alphabet defines the conditions under which distinctions can be fixed and reproduced externally, a mismatch at this level implies the absence of a common space of externalizable forms. In such cases, signals may be transmitted, but their fixation as stable, manipulable, and comparable structures is not guaranteed. What fails here is not interpretation, but externalization itself.

Alphabet mismatch has several ontological consequences:

- External forms produced by one system may not be isolable as elements by another;
- Operations of stabilization and recombination may not be mutually available;
- Equivalence between externally fixed forms cannot be established;
- Reversibility of encoding and decoding is not well-defined.

Under these conditions, communication cannot be restored by refining codes or increasing transmission fidelity. The problem is not that messages are hidden or distorted, but that the ontological conditions for defining a shared code are absent. What appears as a failure of decoding is, in fact, a failure of ontological alignment.

This analysis delineates a strict boundary for cryptographic and signal-based models of communication. Such models presuppose a shared alphabetic ontology and are valid only within its scope. When this presupposition fails, encryption ceases to be an invertible operation, and communication ceases to be a matter of information transfer.

Alphabet mismatch thus represents the first ontological limit of communication. It precedes semantic misunderstanding and informational loss, and it cannot be resolved by epistemic means alone. The next and more fundamental limit arises at the level of the dictionary itself.

9 Ontological Consequences of Dictionary Mismatch

Dictionary mismatch constitutes a more fundamental ontological limitation than alphabet mismatch. While the alphabet defines the infrastructure for externalizing distinctions, the dictionary defines the distinctions themselves as regulatorily significant. A mismatch at the level of the

dictionary therefore does not concern encoding or fixation, but the very space of possible regulation.

Epistemic models of communication typically assume that interacting systems already share a common dictionary and differ only in access to information, coding schemes, or interpretive resources. Within the regulatory ontology of language, this assumption is unwarranted. The existence of communication presupposes not merely signals or symbols, but a shared system of distinctions that render coordination possible.

When dictionaries do not overlap ontologically:

- no common set of regulatorily significant distinctions exists;
- forms of behavior cannot be stabilized as equivalent;
- expectations cannot be aligned or mutually constrained;
- uncertainty is minimized in incommensurable spaces.

Under such conditions, communication does not fail due to misunderstanding, noise, or missing information. It fails because there is no shared regulatory domain in which understanding could be defined. The absence of a common dictionary eliminates the very criteria by which correctness, error, or misinterpretation might be assessed.

Crucially, dictionary mismatch cannot be resolved through epistemic means. Increasing the volume of transmitted data, refining interpretive frameworks, or supplying additional contextual information does not generate shared regulatory distinctions. What is lacking is not access, but ontological compatibility.

This limitation has decisive implications for translation. Translation presupposes at least partial overlap between dictionaries, allowing regulatory distinctions to be aligned across systems. Where such overlap is absent, translation is not merely difficult or approximate; it is ontologically undefined. There is nothing to translate, because there is no common regulatory reference.

Dictionary mismatch therefore represents the deepest limit of communication. It precedes semantic divergence, informational loss, and symbolic incompatibility. Where dictionaries do not intersect, communication does not degrade—it does not arise at all.

The consequences of this limit extend beyond linguistic interaction. They apply equally to intercultural contact, artificial systems trained on stabilized textual corpora, and any attempt to infer regulatory structures from external traces alone. In all such cases, the absence of a shared dictionary marks an ontological boundary that no epistemic procedure can cross.

10 Limits of Translation

Within the regulatory ontology of language, translation is not an operation of transferring meanings, symbols, or messages between systems. Translation is a process of partial alignment be-

tween regulatory dictionaries, whereby distinctions, forms of behavior, and expectations become sufficiently compatible to support coordinated action.

This immediately imposes a strict ontological limit. Translation is possible if and only if there exists a non-empty intersection between dictionaries. Such an intersection need not be complete or symmetric, but it must be sufficient to establish shared regulatory reference points. Where this condition is not met, translation is not merely approximate or noisy; it is ontologically undefined.

Crucially, translation does not operate on externalized forms alone. Symbols, texts, or signals can be mapped only insofar as they are anchored in overlapping regulatory distinctions. In the absence of such overlap, any mapping remains arbitrary and fails to constitute translation in a strict sense.

From this perspective, many classical and contemporary models misconstrue translation as an epistemic problem: a matter of insufficient data, inadequate decoding, or imperfect alignment of representations. The regulatory model reveals a different structure. Translation fails not because too little is known, but because the conditions under which something could count as translatable are absent.

This explains why no amount of contextual enrichment, statistical correlation, or interpretive refinement can guarantee translation across radically different systems. These procedures presuppose a shared regulatory substrate that they cannot themselves generate.

Translation, therefore, marks a boundary rather than a bridge. It delineates the region in which communication is possible, not a technique for extending communication beyond its ontological limits.

11 Limits of Information Theory

The limits of translation identified above have direct implications for classical theories of information. Information theory, in its canonical form, provides a powerful and mathematically precise account of signal transmission under conditions of uncertainty. However, it operates under assumptions that remain ontologically prior to the regulatory structures discussed in this work.

In particular, information theory presupposes:

- a fixed alphabet of distinguishable elements;
- a predefined space of possible messages;
- stable criteria for equivalence and differentiation.

These presuppositions correspond to the existence of a shared alphabet and, more fundamentally, a shared dictionary. Information, in the Shannon sense, quantifies uncertainty over

an already constituted space of distinctions. It does not account for the emergence of that space itself.

When dictionaries do not overlap, the notion of information becomes ontologically indeterminate. There is no well-defined message space, no common alphabet of distinctions, and no shared criteria by which uncertainty could be measured or reduced. In such cases, failures of communication cannot be described as noise, loss, or distortion of information. The problem lies not in transmission, but in the absence of a shared regulatory foundation.

This observation does not diminish the validity of information theory within its proper domain. On the contrary, it clarifies that its domain of applicability is bounded by prior ontological conditions. Information theory describes how signals behave once regulatory distinctions are already aligned; it does not describe how such alignment comes into being.

Accordingly, attempts to extend information-theoretic models to domains lacking shared dictionaries—such as radically heterogeneous cultures, speculative non-human intelligences, or artificial systems trained solely on external traces—encounter a fundamental limit. Beyond this limit, the language of information ceases to be applicable, and ontological analysis becomes unavoidable.

The regulatory model thus situates information theory as a secondary, derivative framework: indispensable for analyzing communication within aligned systems, but silent on the conditions that make communication possible in the first place.

12 Implications for SETI: Noise, Signals, and Dictionary Dependence

Programs dedicated to the search for extraterrestrial intelligence (SETI) are traditionally grounded in an information-theoretic conception of communication. Within this framework, intelligence is expected to manifest itself through detectable signals exhibiting non-random structure, statistical regularities, or algorithmic compressibility. Signals that fail to meet these criteria are typically classified as noise and excluded from further consideration.

From the perspective of the regulatory ontology of language, this strategy rests on a strong and largely implicit assumption: that the sought-for intelligence shares, or at least approximates, a dictionary compatible with our own. Only under this assumption can signal structure be interpreted as information rather than noise.

However, as shown in the preceding sections, information-theoretic notions presuppose a shared dictionary and alphabet. Without such shared regulatory foundations, the distinction between signal and noise loses ontological determinacy. What appears as white noise relative to one dictionary may constitute a highly structured and regulatorily meaningful signal relative to another.

In this sense, SETI does not merely search for signals; it searches for signals that are al-

ready interpretable within a human-aligned dictionary of distinctions. This introduces a profound observational bias. The absence of recognizable structure cannot be taken as evidence of the absence of intelligence or coordination. It indicates only the absence of overlap between regulatory dictionaries.

Moreover, even the concept of *structure* is dictionary-dependent. Statistical regularity, compressibility, or repetition are meaningful indicators only within a predefined space of distinctions. A signal optimized for regulatory coordination within a radically different environment may deliberately avoid patterns that appear meaningful or efficient from a human perspective. What is filtered out as noise may therefore be precisely what carries regulatory significance in another ontological context.

This observation has an important consequence: SETI's reliance on deviations from randomness as indicators of intelligence implicitly conflates epistemic detectability with ontological existence. The regulatory model severs this conflation. Intelligence, understood as the capacity for coordinated regulation and uncertainty reduction, does not entail the production of signals that are informative under human dictionaries.

Accordingly, the failure to detect non-random signals does not justify conclusions about cosmic silence or the absence of non-human intelligence. It reveals the limits of an observational framework that treats communication as information transfer rather than as regulation grounded in shared distinctions.

In the extreme case, a fully functional regulatory system operating under a radically non-overlapping dictionary could produce external traces indistinguishable from white noise when observed through human-aligned informational filters. Such traces would not be encrypted messages awaiting better decoding, but manifestations of a regulatory order inaccessible to our current ontological commitments.

Thus, within the regulatory ontology of language, white noise cannot be categorically excluded as meaningless. It may represent the boundary at which our dictionary fails, rather than the boundary at which intelligence disappears.

13 Implications for Large Language Models

Large Language Models (LLMs) are often presented as evidence that language can be learned, approximated, or even mastered through exposure to large-scale textual data alone. Their apparent success in generating coherent, context-sensitive, and semantically rich outputs has led to the widespread assumption that language is, at its core, a statistical regularity over symbols.

From the perspective of the regulatory ontology of language, this interpretation involves a fundamental category error. LLMs operate entirely within the space of already externalized and stabilized linguistic traces. They presuppose the existence of a shared dictionary and alphabet and have no access to the regulatory processes through which these structures originally emerge.

Textual corpora encode the outcomes of regulation, not its generative mechanisms. They

contain stabilized correspondences between forms and expectations, but they do not contain the conditions under which distinctions become regulatorily significant. As a result, LLMs can reproduce patterns of usage, but they cannot establish or revise the underlying dictionary that renders those patterns meaningful.

This limitation can be stated precisely. LLMs are capable of:

- approximating distributions over externalized forms;
- reproducing statistically stable patterns of usage;
- generating contextually coherent continuations within a fixed dictionary.

They are not capable of:

- forming new regulatory distinctions;
- grounding distinctions in behavioral coordination;
- minimizing ontological uncertainty through regulation;
- generating a dictionary in the ontological sense defined in this work.

Importantly, this is not a claim about performance, intelligence, or utility. The success of LLMs in linguistic tasks is entirely compatible with the regulatory model. Indeed, it confirms it. LLMs succeed precisely because they operate within a densely stabilized dictionary produced by human linguistic practice. Their competence reflects the depth of prior regulation embedded in the data, not the emergence of regulation within the model.

Attempts to frame LLM training as “language acquisition” implicitly conflate exposure to externalized traces with participation in regulatory processes. Human language acquisition involves the formation of a dictionary through embodied interaction, imitation, synchronization, and uncertainty reduction. Text alone provides no access to these mechanisms. It supplies neither regulatory feedback nor criteria for relevance grounded in action.

This distinction becomes especially salient in discussions of semantic alignment and grounding. Alignment procedures adjust model behavior to human preferences within an existing dictionary; they do not generate the dictionary itself. Grounding, when understood ontologically rather than metaphorically, requires participation in regulatory coordination with an environment, not merely statistical coupling to symbols.

In this respect, LLMs represent the inverse case of SETI. Whereas SETI searches for signals without a guaranteed shared dictionary, LLMs operate within a dictionary so fully presupposed that its ontological origin becomes invisible. Both cases illustrate the same limit from opposite directions: language cannot be reduced to signal detection or pattern learning without losing its regulatory foundation.

Thus, within the regulatory ontology of language, LLMs are best understood not as language-forming systems, but as high-fidelity operators over the externalized residues of language. Their

achievements delineate the power of stabilization and reproduction, while simultaneously marking the boundary beyond which regulation, dictionary formation, and genuine linguistic emergence cannot be inferred from text alone.

14 Limits of Observable Language

The analyses presented in the preceding sections converge on a common conclusion: language, understood as a regulatory structure, is only partially accessible to observation. What can be observed—signals, texts, symbols, statistical regularities—are externalized residues of regulatory processes, not the processes themselves. The foundational mechanisms through which language emerges, stabilizes, and functions remain ontologically prior to their observable manifestations.

This distinction clarifies the limits encountered in diverse domains. In the case of translation, observable correspondences between symbols presuppose an underlying overlap of regulatory dictionaries. Where such overlap is absent, translation does not merely fail empirically; it becomes ontologically undefined. In information theory, observable measures of uncertainty operate over a predefined space of distinctions and therefore remain silent about the conditions under which that space is constituted.

The same limit appears in more contemporary contexts. SETI programs rely on observable deviations from randomness to infer intelligence, implicitly assuming a shared dictionary of distinctions. Large Language Models, by contrast, operate on massive collections of observable linguistic traces, presupposing a fully stabilized dictionary while remaining blind to its regulatory genesis. In both cases, what is observed is not language as regulation, but its externalized effects.

These examples illustrate a general principle: the observability of language is coextensive with its externalization. Regulatory coordination, dictionary formation, and the stabilization of distinctions occur at a level that cannot be directly inferred from signals, symbols, or statistical structure alone. Observation captures outcomes, not conditions of possibility.

This does not render empirical investigation irrelevant. Rather, it delineates its scope. Empirical methods can describe how language behaves once regulatory structures are in place, but they cannot, by themselves, recover the ontological foundations from which those structures arise. Attempts to do so inevitably conflate epistemic accessibility with ontological sufficiency.

Accordingly, the limits of observable language are not technological limits awaiting better instruments, nor statistical limits awaiting larger datasets. They are ontological limits, grounded in the distinction between regulation and representation. Language exceeds what can be observed because its primary function is not to produce observable artifacts, but to stabilize coordinated action under conditions of uncertainty.

Recognizing this limit does not close inquiry; it reorients it. It shifts the focus from extracting meaning from external traces to understanding the regulatory processes that make meaning

possible in the first place.

15 Conclusion

This work has argued that language is not primarily a system of representation or a collection of usage rules, but a regulatory structure that stabilizes coordinated behavior under conditions of uncertainty. By tracing language to pre-symbolic mechanisms of imitation and synchronization, the regulatory model delineates the ontological limits of translation, information, and observable linguistic form. What language makes visible are not meanings as such, but the stabilized residues of historically viable regulation. Beyond these residues lie not errors or noise, but the boundaries of our own regulatory and epistemic horizon.

References

- [1] Renee Baillargeon. “Object permanence in infants”. In: *Developmental Psychology* (1987).
- [2] Merlin Donald. *Origins of the Modern Mind*. 1991.
- [3] Susan Goldin-Meadow. *The Resilience of Language*. 2003.
- [4] Jack Goody. *The Domestication of the Savage Mind*. 1977.
- [5] Marc Hauser. *The Cognitive Foundations of Language*. 2002.
- [6] Gregory Hickok and David Poeppel. “Dorsal and ventral streams: A framework for language”. In: *Cognition* (2004).
- [7] Gregory Hickok and David Poeppel. “The cortical organization of speech processing”. In: *Nature Reviews Neuroscience* (2007).
- [8] Jean Mandler. *The Foundations of Mind*. 2004.
- [9] Alexey A. Nekludoff. “Philosophy of Discrete Being: Foundations and Structural Architecture”. In: (2025). DOI: [10.5281/zenodo.17690594](https://doi.org/10.5281/zenodo.17690594). URL: <https://doi.org/10.5281/zenodo.17690594>.
- [10] Walter Ong. *Orality and Literacy*. 1982.
- [11] Ann Senghas and Judy Kegl. “Nicaraguan Sign Language: A test case for the emergence of grammar”. In: *Annual Review of Anthropology* (1994).
- [12] Elizabeth Spelke. “Initial knowledge: Core domains in cognition”. In: *Cognition* (1994).
- [13] Michael Tomasello. *A Natural History of Human Thinking*. 2014.
- [14] Michael Tomasello. *Origins of Human Communication*. 2008.
- [15] Ludwig Wittgenstein. *Philosophical Investigations*. 1953.
- [16] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus*. 1922.