

From Competence to Usage: Distributional Semantics, Connectionism, and the Case for a Usage-Based Paradigm in Language Acquisition

Ezio Xie and Amu Chandra

Working paper, September 2026

Abstract

For much of the last half-century, the dominant framework for understanding human language has been the generativist paradigm associated with Chomsky, in which syntax is treated as an innate, modular faculty operating over recursive symbolic rules, largely independent of meaning, usage, and social context. This paper argues that two converging developments jointly motivate a shift away from this generativist paradigm towards an account in which linguistic knowledge emerges from statistical regularities in usage: the usage-based/constructionist tradition within cognitive linguistics (Langacker, Goldberg, Tomasello, Rosch), and the empirical success of distributional and neural approaches to language in computational linguistics (Firth's distributional hypothesis, word embeddings, and large language models). We trace the theoretical commitments of the generativist view, describe the usage-based and connectionist alternatives, and argue that the empirical behaviour of contemporary language models offers indirect but substantial support for the latter. We then consider the implications of this shift for the study of second-language acquisition (SLA), situating it in relation to Nick Ellis's frequency-driven usage-based account of SLA and Krashen's influential hypotheses, before addressing the limitations of the argument and the counterarguments still maintained within generative linguistics. The paper is a theoretical synthesis rather than a report of new empirical data, and its illustrative examples are offered as motivating cases rather than as proof.

1. Introduction

This paper addresses a question internal to linguistic theory: does recent evidence from cognitive linguistics and computational language modelling support a usage-based account of language over the classical generativist account, and if so, what follows for how we understand language acquisition? We take up this question in four steps. First, we set out the generativist paradigm and the theoretical commitments that follow from it. Second, we describe three consequences that this paradigm has for how language and language acquisition are conventionally understood, illustrating each with a motivating example. Third, we present the usage-based/constructionist alternative together with the connectionist (Parallel Distributed Processing) account of cognition that underwrites it, and we argue that the empirical success of distributional semantics and large language models constitutes independent, if indirect, evidence in its favour. Fourth, we examine how this alternative account relates to established work in applied linguistics, focusing on Nick Ellis's frequency-driven usage-based account of second-language acquisition and on Krashen's five hypotheses of second-language acquisition. We close by discussing the limitations of this synthesis and the responses available to defenders of the generativist position.

2. The Generativist Paradigm

The Chomskyan view of language can be characterised by three interrelated propositions. The first is that syntax, often referred to as universal grammar, constitutes an innate faculty for language, analogous to a bodily organ that grows and matures according to its own internal schedule, much as the eye or the pancreas does (Chomsky, 1975). This organ-like universal grammar is often termed linguistic competence, in contrast to the messier facts of actual language use, termed performance. The

competence/performance distinction has had a profound influence on how language has subsequently been studied and taught.

The second proposition is that syntax should be described by recursive symbol-manipulation rules (for example, $S \rightarrow NP + VP$, $VP \rightarrow V + PP$), such that in principle any well-formed sentence can be parsed into a clean hierarchical structure. The centrality of recursion to this picture has itself become contested within linguistics: Gibson (2024) argues that composition, rather than recursion or infinity per se, is the property that natural-language syntax actually requires, a claim that bears directly on how essential the rule system in this second proposition should be taken to be.

The third proposition follows from the second: sentences are understood as words composed according to these structural rules, with the rules themselves treated as the central object of linguistic inquiry and lexical semantics relegated to a peripheral concern. Some theorists have developed elaborate frameworks, for instance exo-skeletal approaches to argument structure (Borer, 2013), to explain how syntactic structure constrains word meaning. What has remained comparatively peripheral, however, is the grounding of lexical content itself: why a given word carves the semantic space in the way that it does. This is a narrower claim than it may at first appear, and it should be stated narrowly. Compositional and truth-conditional semantics is a highly developed enterprise conducted within and alongside the generative tradition (Montague, 1970; Heim & Kratzer, 1998), as is the study of how lexical semantic structure maps onto syntactic realisation (Levin & Rappaport Hovav, 2005). The point is not that meaning has been neglected, but that the distributional and experiential grounding of individual lexical items, the question a usage-based account takes as central, has been treated as largely external to the theory of grammar.

Decades of theoretical and experimental work have sought to characterise the hypothesised language faculty directly, but conclusive empirical evidence for an autonomous, dedicated syntactic organ has proven elusive. Real-time neurolinguistic studies, in particular, have not found the clean separation between an autonomous syntactic module and a distinct semantic one that this picture predicts: magnetoencephalographic (MEG) evidence indicates that syntactic and semantic composition are tightly interleaved during online sentence processing rather than proceeding through a separate syntax-first stage (Pylkkänen & McElree, 2006). Partly for this reason, a substantial body of work within linguistics, psychology, and cognitive science has moved away from strongly nativist, modular accounts of syntax (Elman, 2009; Goldberg, 2019; Piantadosi, 2024), even as generativist assumptions continue to shape how language is popularly understood and, we will argue, how it is often taught. It should be noted that this move is not unanimous even among researchers who take connectionist modelling seriously: Marcus (2001) argues, on largely independent grounds, that connectionist networks without built-in symbol-manipulation capacities cannot adequately capture the rule-like, systematic aspects of language and cognition, and that some innate combinatorial structure is likely still required. This is a hybrid, nativist-leaning position we return to in Section 8.

3. Three Consequences of the Generativist View for Language Acquisition

If syntax is treated as the central object of linguistic competence, and meaning and usage are treated as secondary, several consequences follow for how language acquisition is conventionally approached. We describe three such consequences below, each illustrated with a motivating example. These examples are offered as illustrations of the underlying theoretical point rather than as systematic evidence; we return to this limitation in Section 8.

A qualification is necessary before turning to these examples, since it bears on what the examples are able to show. A sophisticated defender of the generativist picture can grant every empirical pattern

described in Sections 3.1–3.3 (frequency entrenchment, collocational asymmetry, implicit sensitivity to adjective ordering) without conceding anything about competence itself, simply by locating all of it within performance. This is not a defensive manoeuvre invented to deflect the present argument; it is precisely what the competence/performance distinction was designed to do, and generativist accounts have always allowed performance to absorb effects of memory, processing cost, frequency, and pragmatics that leave an underlying idealised competence untouched (Chomsky, 1965). Read this way, the examples below are compatible with generativism rather than a refutation of it: they show that usage effects are real, systematic, and pedagogically consequential, not that they are incompatible with an innate, rule-based syntactic faculty. A generativist willing to keep re-describing such effects as performance can absorb all three examples without giving up the competence claim, and a purely competence/performance-based rebuttal of the theoretical point being made here is always available in principle.

What the examples are offered to show, more modestly, is that maintaining a sharp competence/performance split carries a rising explanatory cost as effects of this kind accumulate: a separate account is needed for entrenchment, for collocational asymmetry, and for implicit ordering preferences, each doing work that a unified usage-based account (Sections 4–6) does without positing any such split. The examples motivate that reframing and raise the price of the generativist's move; they do not, on their own, force it. We return to this distinction, and to the limits it places on the argument, in Section 8.

3.1 Overweighting of Grammaticality Relative to Usage Frequency

A view that privileges syntactic well-formedness over usage tends to treat any sentence that satisfies the grammar as equally worth teaching, regardless of how frequently that sentence, or sentences like it, actually occur in use. Consider, by way of illustration, a widely used commercial language-learning application that presented learners of Danish with the sentence “Jeg er ost,” glossed as “I am the cheese.” The sentence is unimpeachably grammatical, conforming straightforwardly to an NP + copula + NP structure, yet it is a vanishingly unlikely utterance in ordinary use. A cursory Google Books Ngram query indicates that while the string “I am cheese” is attested in English-language books, the string “I’m John,” instantiating the identical syntactic template, occurs several orders of magnitude more frequently, particularly in recent decades. If similar frequency asymmetries hold for the corresponding Danish sentences, then the two sentences are grammatically equivalent but usage-theoretically very different: entrenchment through repeated exposure to high-frequency utterances reduces the processing cost of producing and comprehending them in real time (Goldberg, 2019), a benefit that low-frequency but grammatical utterances do not confer. A pedagogy oriented solely around grammaticality has no principled basis for preferring one utterance over the other, since both satisfy the same rule; a pedagogy oriented around usage does.

3.2 Difficulty Accommodating Language Variation

A second consequence concerns language variation. Language varies continuously across time, region, register, and speaker, and this variation is not peripheral noise but a pervasive feature of language in use: words and constructions fall in and out of currency, dialects diverge, and individual speech is marked by false starts, filler words, and idiosyncratic phrasing. A grammar oriented towards an idealised, standard competence has difficulty accommodating this variation systematically, since much of it lies outside the “clean” structures that recursive rule systems are designed to generate. Materials built on this idealisation, whether traditional textbooks or many digital platforms, accordingly tend to present only standardised variants, in which sentences parse neatly into the trees the underlying gram-

mar predicts, leaving learners with little exposure to language “in the wild.” Such variation need not be auditory to pose a challenge: its written representation poses a comparable challenge. The rendering of Hagrid's speech in the Harry Potter novels, “Well, yeh might've bent a few rules, Harry, bu' yeh're all righ' really, aren' you?” (Rowling, 1997), is a case in point; a learner who has not previously encountered this kind of non-standard orthographic representation is likely to experience genuine processing difficulty when meeting it for the first time, despite its being unremarkable to a native reader.

3.3 Reliance on Translation and Explicit Rules to Convey Meaning

A third consequence follows from the comparatively peripheral status of lexical grounding under the generativist view: instructional materials influenced by this tradition tend to rely heavily on translation into the learner's first language, and on explicit rule statements, to convey meaning. First-language acquisition proceeds without translation, which demonstrates that translation is not a necessary condition for learning what words mean; yet adult foreign-language instruction relies on it almost universally, and adult learners characteristically struggle to grasp lexical nuances that native speakers acquire without difficulty. One explanation for the prevalence of translation is that it allows learners to draw on world knowledge unavailable to infants, which plausibly does accelerate early vocabulary learning; the concern raised here is not that translation is without value but that reliance on it can, at the margin, come at the expense of the translation-free route to word meaning that first-language acquisition demonstrates is available.

The point can be illustrated with a lexical example: the English verbs ostracise, expel, banish, and deport, despite their distinct collocational profiles, are commonly rendered by the same target-language equivalent (驱逐) in Mandarin translation. A competent English speaker readily judges that “John was ostracised from school” is anomalous where “John was expelled from school” is not. That is an intuition identical translations cannot themselves supply, and one that more fine-grained translation would only partially address, since native speakers do not appear to arrive at such judgements via translation at all. On a usage-based account, the relevant intuition instead derives from differential exposure: expel is frequently collocated with school-related contexts across a speaker's linguistic history, whereas ostracise is comparatively rare in that context and more often associated with historical or literary registers (e.g., ostracism in Ancient Greece). Reproducing this intuition in an adult learner, on this account, requires exposure to words together with the collocational contexts in which they recur, of a kind and quantity that is difficult to deliver economically through translation-based instruction alone.

A second illustration concerns explicit rule-based instruction. In English pedagogy for Chinese-speaking learners, adjective ordering is frequently taught as a discrete rule: a button that is red, wooden, and large should be described as “a big red wooden button” rather than alternative orderings. Learners often find the ordering difficult to internalise from the rule alone, and one widely used mnemonic renders the ordering as a memorable seven-character phrase mapping each character to a modifier category (determiner, opinion, shape/size, age, colour, origin, material) that the learner must then apply. Native English speakers, by contrast, order such adjectives correctly with little conscious effort and without recourse to any explicit rule; the ordering appears to be available to them implicitly, at low processing cost. On a usage-based account, this asymmetry follows from differential exposure to instances of the construction rather than from possession or non-possession of an explicit rule, suggesting that extensive exposure to naturally occurring stacked-adjective phrases, rather than rule drilling, is the more direct route to comparable fluency.

This example, unlike the two preceding it, has contested theoretical ownership, and should be handled with corresponding care. Cross-linguistic regularities in adjective ordering are among the phe-

nomena that the cartographic branch of generative syntax treats as evidence for a universal hierarchy of functional projections above the noun (Cinque, 1994; Scott, 2002), and on that reading the very effortlessness with which native speakers produce the canonical order counts as evidence for innate structure rather than against it. The usage-based reading offered here is therefore a competing interpretation of shared data, not a datum that only usage-based accounts can accommodate. What tells in its favour is that ordering preferences are in fact graded rather than categorical, varying across speakers, registers, and individual adjective pairs in ways that track corpus frequency; a fixed functional hierarchy accommodates the central tendency more readily than the variation around it. The pedagogical observation, in any case, stands on either reading: whatever the ultimate source of the native speaker's competence, the adult learner does not appear to acquire it efficiently from the rule.

4. The Usage-Based and Constructionist Alternative

An alternative to the generativist account has developed within cognitive and functional linguistics under the general heading of the usage-based approach, which treats linguistic structure not as prior to and independent of use but as emergent from it, shaped by the frequency and context in which forms occur.

4.1 Langacker's Cognitive Grammar

Ronald Langacker's *Cognitive Grammar* (2008) is among the foundational statements of this position. On Langacker's account, linguistic structures are not separable from their patterns of use but are shaped by the frequency and contexts in which speakers encounter and produce them; grammar itself is a structured inventory of symbolic units abstracted from usage events rather than a system of rules operating independently of meaning.

4.2 Goldberg's Construction Grammar

Adele Goldberg's *Construction Grammar* extends this position by arguing that units traditionally treated as atomic (words and idioms) are better understood as instances of a broader category of constructions: form-meaning pairings of varying size and schematicity, ranging from individual morphemes to fully general argument-structure patterns, that are stored in long-term memory together with information about the contexts, including multimodal information, in which they are typically encountered (Goldberg, 2019).

4.3 Tomasello's Social-Pragmatic Account

Michael Tomasello's work on the social-pragmatic origins of language provides a complementary developmental perspective, arguing that language is fundamentally a tool for joint attention and communicative interaction, and that its acquisition is inseparable from the social-cognitive capacities, such as intention-reading and joint attention, that make communicative interaction possible in the first place (Tomasello, 2003). This communicative view of language's primary function has recently found independent support from neuroscience: surveying evidence from the brain's language network, Fedorenko, Piantadosi, and Gibson (2024) argue that human language functions principally as a tool for communication rather than, as is sometimes assumed, a substrate for thought itself, and that it plausibly co-evolved with, rather than gave rise to, human reasoning capacities more broadly. This claim is itself contested: Watumull (2024) argues in direct reply that the evidence marshalled by Fedorenko and colleagues does not settle the relationship between language and thought, and that a purely communicative account understates the role language plays in structuring complex reasoning.

The point is accordingly best treated as a live debate that a usage-based, communication-oriented account of language can draw support from, not as a settled empirical fact.

4.4 Prototype Theory and “Leaky” Generalisations

Within the usage-based tradition, the generalisations that speakers extract from usage are not treated as exceptionless rules but as what have been termed “leaky” generalisations, subject to graded exceptions and contextual modulation. This view draws on Eleanor Rosch's Prototype Theory (Rosch, 1973), which holds that linguistic and conceptual categories are organised around graded, prototype-based membership rather than sharp, necessary-and-sufficient criteria. Taken together, the work of Langacker, Goldberg, Tomasello, and Rosch describes language not as a static rule system but as a dynamic system that emerges from, and remains continuously responsive to, patterns of usage, social interaction, and individual cognitive experience.

5. Parallel Distributed Processing and Connectionism

The usage-based position in linguistics finds a natural counterpart in the Parallel Distributed Processing (PDP) tradition within cognitive science, which offers an account of the computational substrate on which usage-based learning might operate (Rumelhart, McClelland, & the PDP Research Group, 1986).

5.1 Emergentism over Reductionism

A central commitment of the PDP tradition is emergentism: the view that complex cognitive phenomena arise from the interaction of many simple components, rather than being decomposable into discrete, independently characterisable modules. On this view, structured linguistic behaviour, potentially including phenomena of the kind captured by rules such as $S \rightarrow NP + VP$, can emerge from network dynamics without those rules being explicitly represented anywhere in the system. Spivey's (2007) dynamical-systems account of cognition develops a related picture, arguing that perception, cognition, and action are better understood as continuous trajectories through a high-dimensional state space than as sequences of discrete, modular mental events, a view that extends naturally to language processing as a continuous rather than a discretely rule-governed activity.

5.2 Connectionism

PDP is rooted in connectionism, the position that cognition arises from the interaction of many simple, densely interconnected processing units, loosely analogous to neurons. This architecture offers a natural computational analogue to the network of overlapping form-meaning associations posited by construction-grammar accounts of the lexicon and grammar, and suggests a mode of language processing that is graded and associative rather than governed by discrete symbolic rule application. The subsequent development of large-scale neural network architectures, synthesised by LeCun, Bengio, and Hinton (2015), demonstrates that this connectionist architecture scales to genuinely difficult pattern-recognition and sequence-modelling problems, including language, providing an existence proof that the PDP tradition's computational commitments are viable well beyond the smaller-scale models available when the tradition was first formulated.

5.3 Anti-Modularity

The PDP tradition is correspondingly sceptical of strong modularity, the view that the mind comprises largely autonomous, domain-specific processing modules, of which a dedicated syntactic module

would be one instance. On the connectionist account, cognitive capacities including language are more plausibly understood as interdependent products of shared underlying processing resources than as the output of an encapsulated language organ.

5.4 Ecological Validity

Finally, the PDP tradition places emphasis on ecological validity, the degree to which the conditions under which a cognitive capacity is studied resemble the conditions under which it is ordinarily exercised. Applied to language, this suggests that theories of linguistic competence should be answerable to patterns of language as it actually occurs in use, rather than solely to idealised or elicited judgments about well-formedness.

6. Distributional Semantics and the Empirical Turn

Independent of developments within theoretical and cognitive linguistics, the field of natural language processing has, largely for engineering reasons, converged on models of language that are strongly compatible with the usage-based and connectionist positions described above.

6.1 The Distributional Hypothesis and Word Embeddings

J. R. Firth's dictum that “you shall know a word by the company it keeps” (Firth, 1957), a distributional view of meaning that has never been fully embraced within mainstream generative linguistics, has become foundational within computational linguistics. Word2Vec (Mikolov et al., 2013) operationalises a connectionist approach to lexical meaning directly, representing words as vectors in a continuous geometric space learned from the statistical patterns of their co-occurrence with other words, such that semantic relationships between words are recoverable as geometric relationships between their vector representations. A frequently cited demonstration of this property is the approximate vector relationship “king – man + woman $\approx$ queen,” in which analogical relationships between word meanings emerge from purely distributional training without any explicit semantic annotation.

This particular demonstration warrants a caveat. Its apparent cleanness depends substantially on implementation choices: in standard evaluations the input vectors are excluded from the set of candidate answers, which materially inflates the result, and performance is considerably more modest under fairer evaluation protocols (Nissim, van Noord, & van der Goot, 2020). The broader claim does not rest on this case: that distributional training recovers substantial semantic and relational structure without explicit annotation is robustly attested across embedding methods and evaluation regimes, and it is that general finding, rather than any single showcase analogy, that bears on the argument made here. Subsequent developments (including FastText and GloVe as further word-embedding methods, and BERT and GPT as contextual and sentence-level embedding architectures) extend this approach to represent not only individual words but the sentential and discourse contexts in which they occur.

6.2 Implications for the Competence/Performance Distinction

The success of these distributional models bears directly on the theoretical debate outlined in Section 2. Piantadosi (2024) argues that the empirical performance of large language models poses a direct challenge to Chomsky's competence/performance dichotomy: such models achieve fluent, contextually appropriate language production and comprehension without any architecturally distinct module corresponding to an innate, rule-based competence, and without being given explicit grammatical rules during training. This outcome is more readily accommodated by a usage-based account, on which grammatical competence itself is not sharply distinguished from the statistical regularities of perform-

ance, than by an account that treats the two as categorically distinct. This reading is not uncontested, however; we take up direct challenges to it in Section 8.

6.3 Networks of Constructions

Goldberg (2024) addresses the relationship between constructionist theory and large language models (LLMs) directly, arguing that the constructionist framework, which rests on the recognition that constructions pair form with function at varying levels of specificity and abstraction and that linguistic knowledge is dynamic and grounded in usage, is corroborated rather than undermined by recent progress in transformer-based language models, alongside converging findings from corpus linguistics, laboratory phonology, neuroscience, and sociolinguistics. Subsequent work has begun testing this claim directly against model behaviour, examining whether the generalisations LLMs learn pattern more like graded, usage-sensitive constructions than like discrete symbolic rules (Weissweiler, Mahowald, & Goldberg, 2025). Related work within neural network research and cognitive science more broadly, from the original connectionist architectures described in Section 5, through Spivey's (2007) dynamical-systems account of cognition, to LeCun, Bengio, and Hinton's (2015) synthesis of large-scale deep learning, has converged from various directions on convergent conclusions regarding the capacity of interconnected, sub-symbolic systems to give rise to structured linguistic behaviour; within linguistics proper, Edward Gibson has argued on independent grounds against treating recursion as the indispensable core property of syntax that generativist theories have long assumed it to be (Gibson, 2024; see Section 2).

6.4 What Distributional Learning Suggests about Learnability

That contemporary language models acquire fluent linguistic behaviour primarily from statistical patterns in usage (without explicit translation, without curated grammatical instruction, and largely without built-in linguistic structure beyond a general-purpose learning architecture) is at minimum suggestive for theories of learnability more broadly. It does not, on its own, establish that human language acquisition proceeds by the same mechanism, since language models and human learners differ in the quantity, modality, and structure of their input as well as in their prior biases; the inference from model behaviour to human acquisition is therefore best treated as an existence proof that usage-based learning of this kind is possible, rather than as direct evidence that it is how humans do it. We return to this caveat in Section 8.

7. Relation to Established Second-Language Acquisition Theory

The usage-based and distributional account developed above can be related to established work in applied linguistics. We first situate it in relation to Nick Ellis's frequency-driven, usage-based account of second-language acquisition, which is the applied-linguistics tradition most directly continuous with the theoretical claims made in Sections 4–6, before turning more briefly to Stephen Krashen's influential, if also contested, set of five hypotheses (Krashen, 1982).

7.1 Ellis and Frequency-Driven Accounts of SLA

Within applied linguistics, the usage-based position developed in this paper finds its closest counterpart in the work of Nick C. Ellis, who has argued extensively that second-language acquisition is fundamentally a process of implicit statistical learning, in which the frequency and contingency of form-meaning mappings in the input a learner receives shape the strength and accessibility of the resulting linguistic representations (Ellis, 2002). On this account, phenomena as varied as phonotactic well-

formedness judgements, morphosyntactic accuracy, and the acquisition of formulaic sequences and verb-argument constructions are all sensitive to the frequency and distributional statistics of that input, in a manner directly continuous with the constructionist and connectionist accounts described in Sections 4 and 5 (Ellis & Wulff, 2015). This body of work supplies the applied-linguistics evidence base that the more purely theoretical arguments of Sections 2–6 require in order to bear on actual language pedagogy and acquisition research, and it stands as the closest ally, within second-language acquisition research specifically, of the broader usage-based paradigm shift this paper describes. The illustrative examples offered in Section 3 are best understood as informal gestures towards exactly the kind of frequency and collocation effects that this research programme studies systematically.

7.2 Krashen's Acquisition-Learning Distinction

Krashen distinguishes “acquisition,” the unconscious, incidental internalisation of language through exposure and interaction, from “learning,” the conscious, explicit study of grammatical rules. The usage-based account developed in Sections 4–6 gives a mechanistic gloss on the acquisition side of this distinction: acquisition, on this view, is the gradual entrenchment of form-meaning constructions through repeated exposure, a process structurally analogous to the training process by which distributional models acquire linguistic competence from usage data.

7.3 The Remaining Hypotheses, and a Caveat

Krashen's four remaining hypotheses can be treated more briefly, since their compatibility with the present account is suggestive rather than evidential. The Monitor Hypothesis, on which consciously learned knowledge functions only to edit output after the fact rather than to generate it, sits comfortably with the connectionist picture in Section 5, on which fluent real-time production is the product of graded associative processes, with explicit rule knowledge available for post hoc correction but not for real-time generation. The Input Hypothesis, on which acquisition is driven by comprehensible input slightly beyond a learner's current level (“ $i+1$ ”), is naturally compatible with an account predicting that entrenchment of new constructions proceeds through exposure calibrated to what a learner's existing network of constructions can partially, but not yet fully, process. The Natural Order Hypothesis, on which grammatical structures are acquired in a relatively fixed order independent of the difficulty of the explicit rule involved, receives a candidate explanation on the present account: acquisition order would be predicted to track the frequency and reliability with which constructions occur in usage rather than the formal complexity of the rule a generativist grammar would use to describe them, a prediction that is in principle testable against corpus-frequency data, and one that Ellis's (2002) research programme is in practice already designed to test. The Affective Filter Hypothesis, finally, is largely orthogonal to the usage-based/generativist debate addressed here, since it concerns the conditions under which input is processed rather than the representational format of linguistic knowledge.

These points of contact should be weighted accordingly, and the weight is light. Gregg (1984) offers an influential critique of the Monitor Model, arguing that its central theoretical terms (acquisition, learning, conscious, and subconscious) are not independently defined with enough precision to render the hypotheses empirically testable, and that the acquisition-learning distinction in particular risks being unfalsifiable as originally formulated. Demonstrating compatibility with a framework open to that objection cannot corroborate the usage-based account, and is not offered as doing so; the mappings above are illustrative contact with an influential applied-linguistics tradition, nothing more. The critique does not extend to the more specific, mechanistically grounded claims discussed in Section 7.1 regarding frequency and implicit statistical learning, which are separately testable against corpus and

experimental data, and it is there, rather than in the correspondences sketched here, that the applied-linguistics case for this paper's position actually rests.

8. Discussion and Limitations

The argument developed in this paper is a theoretical synthesis: it draws together an existing tradition within cognitive linguistics, an existing tradition within cognitive science, and recent empirical developments within computational linguistics, and argues that they are mutually reinforcing. It does not present new experimental data of its own, and several limitations should be acknowledged.

First, the illustrative examples used in Section 3 (the Danish sentence, the ostracise/expel contrast, and the adjective-ordering mnemonic) are offered to motivate the theoretical claims being made, not as systematic evidence for them, and, as flagged at the start of Section 3, they do not by themselves refute a generativist prepared to locate all three effects within performance. Substantiating these claims more rigorously would require corpus-based frequency analyses, controlled studies of collocational competence, and data-driven learning (DDL) research of the kind pioneered by Johns (1991) and subsequent work in corpus-informed pedagogy, none of which this paper attempts. Even such rigorous evidence would sharpen the cost-of-maintaining-the-split argument made in Section 3 rather than settle the underlying theoretical dispute outright, since the competence/performance distinction is, by design, not directly falsifiable by any finite set of performance data.

Second, the inference from the empirical success of large language models to conclusions about human language acquisition (Section 6.4) is indirect. Language models are trained on vastly more data than any human learner encounters, using learning algorithms (e.g., gradient-based backpropagation) with no established biological analogue, and architectures that differ substantially from human neural circuitry. Their success demonstrates that fluent linguistic behaviour can in principle emerge from statistical learning over usage without an innate rule-based grammar; it does not by itself demonstrate that human acquisition proceeds this way, and the analogy should be treated as suggestive rather than probative.

Third, the generativist position addressed in Section 2 is not static, and Piantadosi's (2024) argument in particular has drawn direct, substantive replies from defenders of generativist and nativist positions that this paper's thesis must be weighed against. That argument appeared in an edited volume rather than in a journal, the same volume from which Gibson (2024) is drawn. The replies were written against the widely circulated 2023 preprint rather than the published chapter, but their substance is unaffected. Katzir (2023) argues that large language models fail elementary adequacy tests for modelling human linguistic knowledge: unlike human learners, LLMs do not reliably acquire certain syntactic generalisations from limited exposure (for instance, constraints on parallel gaps in coordinate structures, which Katzir notes are reliably mastered by English-speaking children but not by contemporary language models); they fail to distinguish grammaticality from mere statistical likelihood in the way human speakers reliably do; and they conflate competence with performance in a way that is difficult to reconcile with well-documented dissociations, in human speakers, between grammatical knowledge and real-time processing limitations (e.g., the difficulty humans have with centre-embedded sentences despite recovering the intended structure given more time). Kodner, Payne, and Heinz (2023) raise a complementary objection: LLMs require several orders of magnitude more linguistic data than human children receive to reach comparable proficiency, which leaves the central puzzle of language acquisition, how children learn so much from so little input, largely untouched by LLM successes; they further argue that predictive success is not the same as scientific explanation, and that some independent theory of human language and cognition is still required to interpret what LLM behaviour does or does

not reveal about human learners, rather than treating LLM performance as self-interpreting evidence. Related positions remain live within the field more broadly, including poverty-of-stimulus arguments for grammatical constraints that could not plausibly be learned from the input available to children, and minimalist reformulations of generative grammar that substantially revise the framework Chomsky (1957, 1975) originally proposed. Notably, these are not solely post hoc objections to LLMs specifically: Marcus (2001) raised structurally similar concerns about pure connectionist networks more than two decades before the current generation of language models existed, arguing on largely independent grounds that systems lacking built-in symbol-manipulation capacities would fail to capture the rule-like, systematic aspects of language. That a comparable objection resurfaces in the LLM era suggests this is a durable disagreement about the architecture of language, not merely a response to any one model generation. A full treatment of this paper's thesis would need to engage these responses in greater depth than is possible here; our aim has been to establish that the usage-based and distributional evidence constitutes a serious challenge to the classical generativist picture, not that the matter is settled beyond dispute.

9. Conclusion

This paper has argued that the usage-based and constructionist tradition in cognitive linguistics, the connectionist tradition in cognitive science, and the empirical success of distributional semantics and large language models in computational linguistics converge on a common alternative to the classical generativist paradigm: an account in which linguistic knowledge emerges from, and remains continuous with, patterns of usage, rather than resting on an innate, modular, rule-based competence sharply distinguished from performance. We have argued that this shift has direct implications for how second-language acquisition is theorised, and we have related it to Nick Ellis's frequency-driven usage-based account of SLA and to Krashen's influential hypotheses in applied linguistics. The synthesis offered here is theoretical rather than empirical, and its illustrative examples and analogies to machine learning should be read as motivating rather than conclusive; we hope nonetheless that it sharpens the case for taking usage-based and distributional accounts of language seriously as an alternative to the generativist paradigm that has long dominated the field, and that it points towards specific, testable predictions (concerning acquisition order, collocational competence, and the role of input frequency) that could substantiate or challenge this account empirically in future work.

References

- Borer, H. (2013). *Structuring Sense, Volume III: Taking Form*. Oxford University Press.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1975). *Reflections on Language*. Pantheon Books.
- Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths Towards Universal Grammar: Studies in Honor of Richard S. Kayne* (pp. 85–110). Georgetown University Press.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd ed., pp. 75–93). Routledge.

- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547–582.
- Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2024). Language is primarily a tool for communication rather than thought. *Nature*, 630, 575–586.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In *Studies in Linguistic Analysis* (pp. 1–32). Blackwell.
- Gibson, E. (2024). Composition, not infinity: The irrelevance of recursion to theories of language. In E. Gibson & M. Poliak (Eds.), *From Fieldwork to Linguistic Theory: A Tribute to Dan Everett* (Ch. 1). Language Science Press.
- Goldberg, A. E. (2019). *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton University Press.
- Goldberg, A. E. (2024). Usage-based constructionist approaches and large language models. *Constructions and Frames*, 16(2), 220–254. <https://doi.org/10.1075/cf.23017.gol>
- Gregg, K. R. (1984). Krashen's Monitor and Occam's razor. *Applied Linguistics*, 5(2), 79–100.
- Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing*. *ELR Journal*, 4, 1–16.
- Katzir, R. (2023). Why large language models are poor theories of human linguistic cognition: A reply to Piantadosi. *Biolinguistics*, 17, e13153.
- Kodner, J., Payne, S., & Heinz, J. (2023). Why linguistics will thrive in the 21st century: A reply to Piantadosi (2023). *arXiv:2308.03228*.
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Levin, B., & Rappaport Hovav, M. (2005). *Argument Realization*. Cambridge University Press.
- Marcus, G. F. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. MIT Press.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
- Montague, R. (1970). Universal grammar. *Theoria*, 36(3), 373–398.
- Nissim, M., van Noord, R., & van der Goot, R. (2020). Fair is better than sensational: Man is to doctor as woman is to doctor. *Computational Linguistics*, 46(2), 487–497.
- Piantadosi, S. T. (2024). Modern language models refute Chomsky's approach to language. In E. Gibson & M. Poliak (Eds.), *From Fieldwork to Linguistic Theory: A Tribute to Dan Everett* (Ch. 15, pp. 353–414). Language Science Press.
- Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 539–579). Elsevier/Academic Press.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350.
- Rowling, J. K. (1997). *Harry Potter and the Philosopher's Stone*. Bloomsbury.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. MIT Press.
- Scott, G.-J. (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (Ed.), *Functional Structure in DP and IP: The Cartography of Syntactic Structures, Volume 1* (pp. 91–120). Oxford University Press.

- Spivey, M. (2007). *The Continuity of Mind*. Oxford University Press.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
- Watumull, J. (2024). Language is a “quite useless” tool: A rejoinder to Fedorenko, Piantadosi, and Gibson's “Language is primarily a tool for communication rather than thought.” *Biolinguistics*, 18, e15229.
- Weissweiler, L., Mahowald, K., & Goldberg, A. E. (2025). Linguistic generalizations are not rules: Impacts on evaluation of LMs. In *Proceedings of the Second International Workshop on Construction Grammars and NLP* (pp. 61–74). Association for Computational Linguistics.