ÉKWÉ INTERNATIONAL JOURNAL OF IGBO SCHOLARS FORUM FOR SOCIO-CULTURAL ADVANCEMENT (INC) Volume 19, No 4, OCTOBER, 2026 ISSN: 2476-8448

MERGE AND THE HUMAN CAPACITY FOR HIERARCHICAL THOUGHT: A MINIMALIST PERSPECTIVE ON THE EVOLUTION OF UNBOUNDED CREATIVITY

Dr. Cynthia Adaeze Onuegwunwoke,
cynthia.onuegwunwoke@alvanikoku.edu.ng

Dr. Godfrey Ngozi Ogbonna,
ogbonnagodfrey82@gmail.com

Dr. Amadi, Gloria Ukamaka,
gloria.amadi@alvanikoku.edu.ng

Dr. Eugene Chukwuemeka Onuoha,
Department of English Language and Literature, Alvan Ikoku Federal University of Education, Owerri, Imo State, Nigeria
eugene.onuoha@alvanikoku.edu.ng

And

Dr. Kevin Onyewuchi Ugorji
kelvinugorji70@gmail.com
Imo State Polytechnic, Awo-Omumma

Abstract

This conceptual study examines Merge as a minimal combinatorial operation and considers whether it can illuminate the evolution of human hierarchical thought and unbounded creativity. Within the Minimalist Program, Merge combines two syntactic objects into a larger object, and repeated application permits hierarchically structured

expressions without a fixed upper bound. The Strong Minimalist Thesis (SMT) proposes that the language faculty may be close to an optimal solution to conditions imposed by the conceptual-intentional and sensorimotor interfaces, with general computational principles doing much of the explanatory work. The paper evaluates this proposal against the bio-linguistic account of language, the "only us" debate, research on animal cognition, and hypotheses linking hierarchical syntax to pre-existing capacities for action planning and object manipulation. It argues that Merge is a powerful formal hypothesis but not, by itself, a complete evolutionary explanation: a recursive operation, the representations on which it operates, the neural mechanisms that implement it, and the systems that interpret or externalise its outputs must be distinguished. It concludes with testable interdisciplinary predictions and a research agenda for comparing linguistic hierarchy, action sequencing, and context-sensitive cognition.

Keywords: Merge, Minimalist Program, Hierarchical cognition, Language evolution, Third factor

Introduction

Human beings do not merely associate sounds with objects or respond to immediate stimuli. They construct representations in which one idea can be embedded within another, where relations among parts determine the interpretation of the whole, and where a finite repertoire supports indefinitely many novel thoughts. A person can think about a student who believes that a lecturer suspects that a text contains an ambiguity; can imagine a tool inside a box that is inside another box; or can formulate a counterfactual about what might have happened if a past decision had been different. These examples differ in content, but each depends on the ability to organise elements into structured relations. The

evolutionary question is not simply why humans communicate, but how a cognitive system capable of such open-ended hierarchical organisation emerged.

The Minimalist Program approaches this problem by asking how much of the complexity of human syntax can be derived from a small computational core together with interface conditions and general principles of efficient computation. Its central operation, Merge, combines two syntactic objects into a larger syntactic object. If α and β are combined, Merge forms {α, β}; crucially, the output can itself be merged with another object. The result is not just a longer sequence but a nested structure in which constituents stand in hierarchical relations (Chomsky, 1995, pp. 226–227; Chomsky, 2005, pp. 5–6). This simple description has far-reaching implications: repeated application makes it possible to generate an unbounded range of structures from finite resources.

Merge has therefore, become central to a particular account of what makes human language distinctive. Berwick and Chomsky (2016) argue that language is best understood primarily as an internal system for constructing thoughts, with speech and other modes of expression serving as externalisation systems. In this view, syntax is hierarchical rather than fundamentally linear; the ordering required for speech or sign is imposed at an interface rather than built into the core computational operation (pp. 89–108). Their "Why Only Us" framework raises a profound question: if the core operation is so simple, why is the resulting capacity so distinctive to humans?

This study addresses that question through a conceptual synthesis of minimalist syntax, evolutionary theory, comparative cognition, and philosophy of mind. It treats Merge not merely as a grammatical mechanism but as a candidate model for the minimal combinatorial capacity that

supports hierarchical thought. At the same time, it distinguishes the formal properties of Merge from claims about its biological origin. A theory can provide a parsimonious account of a cognitive computation without having demonstrated when, how, or through which evolutionary pathway that computation appeared.

Merge, hierarchy, and unbounded creativity

Merge as a minimal combinatorial operation: In the Minimalist Program, Merge is generally characterised as an operation that takes two syntactic objects and forms a new object containing them. In its simplest formulation, Merge(α, β) yields {α, β}. The elements may be lexical items or structures already built by previous applications of Merge. This recursion over its own outputs is the source of unbounded structural growth. For example, the words in "the child laughed" can be organised so that the noun combines with a determiner, the resulting nominal phrase combines with a verb, and the resulting predicate or clause can be incorporated into a larger structure. The exact derivation depends on the theoretical analysis, but the central point remains: constituent structure is built compositionally rather than assembled as a flat string.

The set notation highlights a key minimalist claim. Merge itself does not determine the linear order in which words are pronounced. A set such as {α, β} does not encode whether α precedes β. The distinction between hierarchical organisation and linear order is important because the same structural relations can be externalised differently across languages and modalities. Speech languages vary in word order; sign languages externalise linguistic structure through visual-manual resources; and internal representations need not be identical to the overt strings through which they are communicated. Berwick and Chomsky (2016) consequently

treat externalization as a component that interacts with, but is not identical to, the core operation that constructs hierarchical representations (pp. 89–108).

Hierarchical structure is not the same as sequence: A sequence contains an order: first A, then B, then C. A hierarchy additionally represents relations of containment or dominance. The distinction matters because the meaning of a sentence often depends on which constituents group together. In "old men and women," for example, different groupings can affect whether "old" applies to both men and women or only to men. Similarly, the relation between a clause and an embedded clause determines the interpretation of dependencies that cannot be explained by proximity alone. The human language system relies on hierarchical constituency and structural relations such as c-command, not only on adjacent words.

Hierarchical representations also make recursion possible. A clause can be embedded in a noun phrase ("the claim that the witness left"), and that noun phrase can appear within another clause, which may itself be embedded. There is no theoretical fixed upper bound on such embedding, even though performance is constrained by working memory, attention, processing time, and communicative circumstances. Thus "unbounded" is a statement about generative capacity, not a claim that human beings can actually process infinitely long expressions or that everyday speech routinely exploits deep embedding.

The idea of unbounded creativity is broader than syntactic recursion. People use familiar conceptual materials to invent new explanations, stories, plans, metaphors, and counterfactual scenarios. Merge offers a model for one important form of combinatorial openness, but it does not automatically explain the originality or value of every creative

act. Creativity also depends on the available concepts, knowledge, goals, affect, memory, cultural practices, and the ability to evaluate possible combinations.

Internal Merge and displacement: Minimalist theory distinguishes External Merge, which combines independently available objects, from Internal Merge, which remerges an element already contained within a larger structure. Internal Merge is commonly associated with displacement phenomena, such as the relationship between a wh-expression and the position where it receives an interpretation. In "Which book did Ada say that Chidi bought?", the phrase "which book" is interpreted in relation to the verb "bought" even though it appears at the left edge of the larger question. Minimalist accounts analyse such dependencies in terms of structural relations rather than treating them as arbitrary rearrangements.

The details of Internal Merge are theoretically contested, and not every aspect of movement is reducible to a single uncontroversial implementation. Nevertheless, the distinction illustrates the ambition of minimalism: a small number of operations should explain a wide range of structural phenomena. For evolutionary purposes, this creates a useful but difficult question. Is the human-specific innovation the operation of combination itself, a particular recursive use of it, the ability to remerge elements in structurally constrained ways, or the integration of these operations with conceptual and externalisation systems? A plausible evolutionary theory must specify which computational property it is attempting to explain.

The bio-linguistic programme and the Strong Minimalist Thesis

Language as a biological capacity: The bio-linguistic perspective treats language as part of human biological endowment rather than as a cultural invention that can be explained entirely by communication or imitation. In this framework, the language faculty is a natural object of inquiry, comparable in principle to other evolved cognitive capacities. This does not mean that particular languages are genetically encoded. Children acquire a specific language through exposure, interaction, and development, while the capacity to acquire and use language depends on biological, cognitive, and environmental conditions.

Hauser, Chomsky, and Fitch (2002) proposed a distinction between the faculty of language in the broad sense (FLB) and the faculty of language in the narrow sense (FLN). FLB includes sensory-motor systems, conceptual-intentional systems, and computational mechanisms implicated in language. They hypothesised that FLN might include recursion and might be the component most specifically human, while allowing that recursion-like computations could have evolved for reasons other than communication (pp. 1569–1579). The proposal helped reframe the research question: rather than looking only for animal communication that resembles human language, researchers should examine whether component capacities relevant to language occur in other cognitive domains.

The distinction remains influential but debated. “Recursion” can refer to different phenomena in syntax, cognition, mathematics, and behaviour, and evidence for one does not automatically establish another. A system that performs nested action sequences, for instance, may share abstract features with hierarchical syntax without representing linguistic constituents.

The Strong Minimalist Thesis: The Strong Minimalist Thesis is an ambitious proposal that the language faculty is close to an optimal solution to the conditions imposed by its interfaces and general principles of computation. Chomsky (2005) distinguishes three factors that shape the growth of language: genetic endowment, experience, and principles not specific to language or to the organism, often called the third factor (pp. 6–7). The third factor includes computational efficiency, properties of physical systems, and general constraints on learning and development. The Minimalist Program asks whether aspects of grammar traditionally attributed to specialised grammatical machinery can instead be explained by these more general principles.

The SMT does not claim that every feature of language is simple, or that experience and biology are unimportant. It asks whether the core computational system can be reduced to the simplest operation that satisfies the conditions imposed by conceptual-intentional and sensorimotor interfaces. If Merge supplies the needed combinatorial operation and general principles constrain its application, then a large inventory of language-specific mechanisms may be unnecessary. Chomsky's formulation of the Minimalist Program describes this as a search for the simplest possible computational operation that can yield the infinite array of hierarchically structured expressions required at the interfaces (Chomsky, 2000, pp. 89–91; Chomsky, 2005, pp. 1–22).

The evolutionary appeal is apparent. Natural selection does not need to construct a complex grammar rule by rule if a small computational change, in a brain already capable of conceptual representation, can generate a new class of structured thought. However, the inference from formal simplicity to evolutionary simplicity must be handled carefully. A mathematically simple operation may require complex neural implementation, developmental conditions,

representational primitives, and mechanisms for controlling its use. "Minimal" in a formal theory is not synonymous with "cost-free" in biology.

What the SMT explains, and what it does not: The SMT helps identify a target for explanation: the capacity to generate hierarchically structured expressions with no fixed upper bound. It also encourages explanatory restraint by asking whether a proposed grammatical principle follows from more general properties of computation. Yet the thesis is not a complete account of the origins of language. It does not, by itself, specify the genetic changes, developmental pathways, neural circuitry, or population history through which Merge became available. Nor does it independently explain the origins of lexical concepts, the diversity of spoken and signed languages, pragmatic inference, or the social conventions required for communication.

This distinction is important for the "only us" debate. Berwick and Chomsky (2016) argue that the core computational innovation may have appeared relatively suddenly and then been used for thought, with externalisation subsequently shaped by communication and cultural evolution. Other researchers favour more gradual scenarios in which language-specific capacities developed through recombination of pre-existing cognitive and communicative abilities. These alternatives should not be reduced to "one mutation versus no mutation": a sudden appearance of a cognitive phenotype can involve multiple biological and developmental processes, while gradual evolution can include thresholds and reorganisations.

The strongest contribution of minimalism is therefore methodological as well as substantive. It asks researchers to separate the properties of the computational system from the demands of communication and from the contingent histories

of particular languages. Its evolutionary claims become more compelling when they generate predictions that could be evaluated independently of the theoretical elegance of the formal model.

"Why Only Us?" competing evolutionary scenarios

The human-specificity problem: Language is unusual among biological traits because its central computational properties are not readily observable in the fossil record. Bones can reveal aspects of vocal anatomy, and archaeological artefacts can suggest planning, learning, and symbolic activity, but neither directly reveals the structure of internal syntactic computation. Researchers therefore infer language evolution from indirect evidence: comparative cognition, developmental patterns, genetic and neurological findings, archaeological records, and the formal properties of language itself. The inference is difficult because language has no exact nonhuman counterpart and because the target phenotype must be defined before its evolution can be explained (Bolhuis et al., 2014).

The "only us" question is best formulated precisely. It is not whether humans alone communicate, categorise objects, learn sequences, or solve complex problems. Many species display sophisticated abilities in these domains. The issue is whether any nonhuman species has a computational capacity that builds hierarchically organised, compositional representations in the specific sense required by human language. If the target is defined too broadly as "complex cognition," comparative evidence will appear abundant; if it is defined narrowly as Merge, direct evidence becomes much harder to obtain.

A relatively recent computational innovation: Berwick and Chomsky's account proposes that a simple computational innovation may have enabled the generation of hierarchical

structures, transforming human thought and making language possible (2016, pp. 1–52, 89–108). Bolhuis et al. (2014) likewise argue that the hierarchical structure of language is a key phenotype for evolutionary explanation and discuss the possibility of a relatively recent emergence. Such proposals can account for the apparent discontinuity between human language and other communication systems without requiring every component of language to have evolved at the same time.

The strength of this scenario is explanatory focus. If a single computational innovation changes what a mind can represent, then a qualitative cognitive transition may occur without a correspondingly dramatic change in every supporting system. A new combinatorial capacity could exploit older concepts, memory systems, motor capacities, and social learning. The weakness is evidential: the precise timing and biological mechanism of the hypothesised innovation remain uncertain. Claims about a particular date, a single mutation, or a sudden species-wide event should therefore be presented as hypotheses rather than established facts.

Gradualist and continuity-based accounts: A contrasting approach seeks evolutionary continuity between language and other cognitive capacities. Fujita and Fujita and Fujita (2022) argues that hierarchical syntax and the lexicon may have evolved from pre-existing abilities, proposing that Merge could have developed from motor action planning and that lexical categories may have predecessors in animal cognition. This approach rejects a strict saltational account and attempts to connect formal linguistic theory with comparative cognition and motor control. It is valuable because it specifies potential precursors rather than merely asserting that language must have evolved gradually.

Nevertheless, evolutionary continuity does not mean that the final human capacity is present in rudimentary form whenever an animal performs a sequence of actions. The same broad computational principle can be implemented differently across systems, and behavioural similarity can arise through different mechanisms. A chimpanzee's ability to organise tool-use actions, a bird's ability to learn a vocal sequence, and a human's ability to embed one proposition within another may all involve ordering, memory, and goal representation, yet they do not necessarily share the formal structure of Merge.

The risks of treating syntax as communication: Language is often explained as an adaptation for communication, but the minimalist approach emphasises that the core system may be primarily a system for thought, with communication as one important use. This shift matters because communicative efficiency does not straightforwardly predict the hierarchical properties of syntax. If the system is designed to satisfy conceptual and sensorimotor interfaces, some features may follow from computational constraints rather than from direct selection for message transmission. Berwick and Chomsky (2016) distinguish the internal computational system from the processes that externalise structures into speech or other modalities (pp. 89–108).

A thought-centred view should not be interpreted as denying the social importance of language. Human language is learned and used in communities; interlocutors coordinate attention, infer intentions, repair misunderstandings, and develop conventions. Rather, the claim is that the origin of the computational capacity and the evolution of its communicative uses are analytically separable. A comprehensive account must explain both why humans can generate structured representations and how those representations became integrated into socially transmitted systems.

Hierarchical object manipulation and action planning as possible precursors

Why action matters: One promising route to the evolution of hierarchy begins outside language. Everyday action often requires organizing subgoals into larger plans: obtaining a tool, using it to open a container, and retrieving an object; or placing one object inside another while preserving the relation between them. Such activities involve relations among actions and objects, anticipation of intermediate states, and the coordination of component operations toward a goal. If these capacities predate language, they may have provided resources that a later linguistic system could recruit.

Fujita and Fujita's (2022) proposal that Merge evolved from motor action planning offers a specific hypothesis of this kind. It connects the formal operation of syntax to a domain in which structured combinations are already behaviourally useful. The proposal is attractive because it offers a possible evolutionary bridge between nonlinguistic cognition and the capacity to build hierarchical linguistic representations. It also encourages researchers to investigate cognition beyond vocal communication, in line with Hauser et al.'s (2002) suggestion that relevant computational capacities might have evolved for reasons other than language (pp. 1569–1579).

The hypothesis should not be treated as a settled origin story. Motor plans are not necessarily syntactic objects, and nested action organisation may be supported by executive control, working memory, learned scripts, or representations of causal relations. To establish a relationship with Merge, researchers need to identify the representational property that is shared and show that it predicts more than general task complexity.

Object relations and nested representation: Objects can be organised in relations that are structurally richer than simple

adjacency. A cup may be inside a box, the box inside a cupboard, and the cupboard inside a room. This nested configuration is not equivalent to a linear list of "cup, box, cupboard, room": the containment relations determine what is inside what. Similarly, manipulating objects may require representing an action as a subgoal within a larger goal. Such cases make hierarchical cognition visible in a domain that does not depend on spoken language.

However, physical nesting and mental hierarchy must be distinguished. A set of objects can physically contain one another even if an agent represents the scene using associative or spatial mechanisms rather than a general-purpose combinatorial operation. Conversely, a language user can construct a hierarchy about abstract or imaginary entities that have no physical containment relation. The comparison is therefore suggestive rather than demonstrative. It raises the question of whether a common computational architecture supports both domains, not the conclusion that object manipulation is simply "Merge without words."

What comparative research should test: A rigorous comparative programme would need to distinguish at least four abilities: (a) ordering actions in a sequence; (b) representing hierarchical relations among subgoals; (c) recombining familiar components into novel structures; and (d) interpreting the structural relations in a compositional way. These abilities can dissociate. An animal may learn a long sequence without representing nested structure, or it may represent a hierarchy in a narrow domain without freely recombining it across contexts. Human language displays an unusually broad capacity to combine categories and propositions productively.

Research designs should therefore use matched tasks that vary hierarchical depth, linear order, familiarity, and contextual

support. For example, participants could learn object-action sequences in which the same elements occur in different nested arrangements. Tests could ask whether they generalise a structural pattern to novel objects, whether they distinguish structures with identical surface sequences but different groupings, and whether performance transfers between action and language tasks. Comparable animal studies would require species-appropriate methods and careful controls for training history, sensory demands, and motivation.

Even positive transfer would not prove that the same operation is involved. Transfer could reflect shared memory, attention, or abstract relational learning. Stronger evidence would require converging behavioural and computational models, developmental comparisons, and where possible neural evidence. The central prediction is not that all hierarchy must be linguistical, but that a genuinely shared combinatorial mechanism should produce systematic similarities across domains that cannot be explained by domain-general difficulty alone.

Developmental evidence and the problem of learning: Children acquire complex hierarchical language early, despite limited explicit instruction. This fact is consistent with the idea that humans possess specialised biological preparedness, but it does not determine exactly which aspects are innate or how the underlying computation emerged. Children also learn from structured environments, social interaction, and statistical regularities. A theory of Merge must account for the interaction between the learner's biological capacities and the evidence available in the environment.

Developmental studies can help distinguish a general capacity for combination from language-specific knowledge. If children's sensitivity to hierarchical relations in nonlinguistic tasks predicts their later mastery of syntactic dependencies,

that would support a possible shared foundation, although correlation would not establish causation. If linguistic hierarchy emerges independently of performance on matched nonlinguistic tasks, the result might suggest domain-specific constraints or differences in task demands. In either case, measures should avoid treating failure on a difficult experimental task as proof that a capacity is absent.

Context-sensitivity, conceptual interpretation, and the third factor

Why context matters: The conceptual proposal motivating this paper is that context-sensitivity may be central to human cognitive distinctiveness. Linguistic expressions are radically underdetermined without interpretation. The sentence "She is ready" leaves open who "she" refers to, what she is ready for, and which standards of readiness are relevant. "It is cold here" can be a report about temperature, an indirect request to close a window, or a complaint, depending on circumstances and intentions. Humans routinely integrate linguistic structure with shared knowledge, perceptual information, conversational history, beliefs, and inferred goals.

Merge cannot, by itself, account for these interpretations. It provides a formal mechanism for constructing structures; the interpretation of those structures depends on lexical meaning and the conceptual-intentional system, as well as pragmatic inference. The minimalist architecture makes room for this distinction by treating the interface with thought as central to the language faculty. Yet a full theory of context-sensitive interpretation must explain how representations are enriched, how intentions are inferred, and how interlocutors converge on relevant meanings.

Context-sensitivity is not unique to language in a broad sense. Animals adjust behaviour to environmental cues and social

conditions; human distinctiveness, if present, must be sought in the flexibility, scope, and compositionality of contextual integration. Humans can reason about absent entities, hypothetical worlds, nested beliefs, and norms that are not perceptually available.

The third factor as an explanatory opportunity: Chomsky's (2005) third factor includes principles that are neither part of the genetic endowment specific to language nor simply learned from experience. These include general properties of efficient computation and constraints that arise from the physical and cognitive systems in which language develops (pp. 6–7). The third factor offers an opportunity to explain why linguistic computation has certain properties without assuming that each property is explicitly encoded in Universal Grammar.

Context-sensitivity may intersect with the third factor, but the terms should not be conflated. Contextual information is often learned or acquired through experience, while general computational principles constrain how information can be accessed and combined. Context-sensitivity is therefore better treated as a candidate outcome of interactions among biological capacities, domain-general computation, memory, learning, and social cognition, not as a single third-factor principle in itself.

A minimalist evolutionary account could propose that Merge creates structured representations, while general computational constraints and contextual interpretation make those structures usable for thought. This would shift attention away from a richly specified inventory of language-specific rules and toward the interface between combinatorial structure and flexible cognition. But it would not establish that Universal Grammar is unnecessary in every form. The evidence would need to show which properties can be derived

from general principles and which require language-specific constraints.

Is context-sensitivity the truly unique evolutionary step? The suggestion that context-sensitivity, rather than a rich Universal Grammar, is the truly distinctive human step is provocative but requires refinement. Context-sensitivity is not one discrete capacity with a clear evolutionary timestamp. It includes reference resolution, pragmatic inference, perspective-taking, relevance assessment, common-ground tracking, and interpretation under uncertainty. These capacities may have different evolutionary histories and different neural or developmental bases. They may also be present in partial form in other species.

The more defensible claim is that human cognition combines hierarchical representation with unusually flexible contextual integration. Merge may enable the construction of complex structured thoughts; context-sensitive systems allow those thoughts to be related to intentions, beliefs, memories, social norms, and possible worlds. Neither component alone accounts for human creativity. A syntactic system without conceptual resources would have little to combine, while a rich contextual system without a productive combinatorial capacity might have difficulty representing arbitrarily nested propositions.

This synthesis also reframes the "only us" question. The human-specific property may not be a single isolated module, but a distinctive configuration of capacities: hierarchical combination, conceptual abstraction, context-sensitive inference, social learning, and flexible externalisation. Minimalism remains valuable because it asks how small the computational core can be. Comparative cognition remains valuable because it identifies the component abilities that may predate language. Pragmatics and cognitive science remain

necessary because they explain how structures become meaningful in situated thought and action.

Objections, limitations, and theoretical risks

The danger of equating formal simplicity with biological plausibility: A major limitation of Merge-based evolutionary explanations is the temptation to infer a simple evolutionary history from a simple formal description. The operation may be compact in a theory, but a functioning cognitive capacity requires neural implementation, developmental stability, suitable representational units, and integration with other systems. Evolution acts on organisms and populations, not on abstract operations considered in isolation. A convincing account must connect the formal proposal to biological mechanisms without assuming that the former automatically reveals the latter.

The same caution applies to claims about the timing of language emergence. Archaeological evidence may indicate changes in symbolic behaviour, but such evidence cannot directly date the emergence of a particular syntactic computation. Language may have existed before its clearest material traces, and cultural evidence may reflect multiple interacting capacities. The available record does not justify presenting a precise date for Merge as established fact.

The problem of defining recursion: Recursion is used in multiple ways. In formal grammar, it can refer to the repeated application of a rule or operation to its own output. In psychology, it may refer to nested representations or the ability to represent another agent's belief about a third agent's belief. In animal behaviour, it can be used loosely for repeated or nested action sequences. These uses are related only if a theory specifies the relevant representational and computational commonalities.

This is particularly important when evaluating claims that animals possess or lack recursion. A task may fail to reveal a capacity because it places heavy demands on memory, motivation, or training. Conversely, success may be achieved through strategies that do not involve the hypothesised computation. The appropriate conclusion is neither that human language is wholly incomparable with animal cognition nor that any nested behaviour constitutes Merge. Instead, research should specify the computation, define operational criteria, and test competing models.

Universal Grammar and explanatory under determination: The minimalist programme aims to reduce the role of language-specific innate structure, but the degree of reduction remains an empirical and theoretical question. The SMT is a guiding hypothesis, not a demonstrated theorem that all linguistic properties follow from Merge and general principles. Chomsky et al. (2023) showed that the technical analysis of Merge itself continues to develop including its interactions with interfaces and conditions on computation. If a theory needs additional constraints to account for particular phenomena, those constraints must be evaluated rather than dismissed as irrelevant to minimalism.

Likewise, the existence of context-sensitive interpretation does not disprove Universal Grammar. A system can possess biologically grounded constraints while relying heavily on experience and pragmatic inference. The useful question is comparative: which phenomena are explained more economically by general principles, which require language-specific assumptions, and what new predictions follow from each analysis? A conceptual preference for parsimony is not enough to settle the issue.

Humanistic risks: reducing thought to syntax: The phrase "what makes us human?" can invite overly simple answers. If

Merge is treated as the sole essence of humanity, the account risks reducing imagination, ethical reflection, art, social life, and self-understanding to a formal operation. Such a reduction is neither required by minimalism nor supported by current evidence. Merge may be a necessary component of a particular account of linguistic hierarchy without being sufficient for all forms of human cognition or creativity.

A humanistic perspective should instead recognise that the ability to construct structured thoughts matters because it expands the range of what humans can imagine, communicate, contest, and share. Language enables narratives about the past, possible futures, obligations, identities, and worlds that do not exist. Yet the meaning and value of these activities emerge through lived experience, culture, emotion, social relations, and historical circumstances. A theory of the computational basis of hierarchy can illuminate these capacities without exhausting their explanation.

What would count as evidence? A strong theory should generate discriminating predictions. Evidence for a shared hierarchical mechanism would be more compelling if the same structural constraints predicted performance across language and nonlinguistic tasks, if the relation emerged developmentally in a principled way, and if computational models showed that a common operation explained the data better than separate mechanisms. Evidence against a simple shared-mechanism account would include robust dissociations between linguistic hierarchy and carefully matched nonlinguistic hierarchy, or data showing that apparent structural transfer was fully explained by memory and familiarity.

Neural evidence should be interpreted with similar caution. Overlapping activation in two tasks does not prove that the same computation is performed, because many regions

support multiple functions. Different activation patterns do not necessarily prove separate computations, because a common abstract operation may be implemented through distinct networks in different contexts.

A research agenda for a Merge-centred account of human creativity

The conceptual framework developed here suggests five lines of research.

1. Linguists and cognitive scientists should distinguish the capacity to generate hierarchical structure from the capacity to interpret it in context. Experimental tasks can hold lexical material constant while varying constituency, or hold structure constant while varying the pragmatic setting. This would help isolate contributions from syntax, semantics, and pragmatic inference. The aim is not to deny interaction among components but to establish which factor explains which pattern.

2. Research on object manipulation and action planning should test structural generalisation rather than mere sequence learning. Participants could be trained on action plans with nested subgoals and then tested on novel objects and unfamiliar sequences. If they generalise the hierarchical relation across surface changes, the result would support a domain-general capacity for structured representation. If generalisation is limited to familiar action scripts, a learned-sequence explanation may be sufficient. Comparable comparative studies should use methods appropriate to the species and avoid assuming that superficially similar behaviour entails identical representations.

3. Developmental research should examine how hierarchical linguistic competence relates to nonlinguistic relational

reasoning over time. Longitudinal studies could test whether children's ability to represent nested object relations predicts their understanding of embedded clauses or structural dependencies after controlling for vocabulary, working memory, and general cognitive development. Such studies should not treat correlation as proof of a common mechanism, but they can identify developmental relations worth explaining.

4. Computational modelling should compare alternative architectures. One model could implement recursive combination with a small set of interface constraints; another could rely on domain-specific rules; a third could combine hierarchical structure with learned contextual inference. Models should be evaluated not only on their ability to reproduce familiar data but also on their predictions for novel structures, ambiguity resolution, and cross-domain transfer. The SMT gains empirical force when it makes predictions that could fail.

5. Evolutionary scenarios should be explicit about the level of explanation. A proposal about the emergence of a computational operation is not the same as a proposal about the origin of lexical concepts, the evolution of speech, or the cultural history of particular languages. Interdisciplinary work should specify which component is under investigation and what evidence would support or weaken the proposed pathway.

Conclusion

Merge provides a compact and influential account of how a finite cognitive system can generate hierarchically structured expressions without a fixed upper bound. Its importance lies not simply in the production of longer sentences, but in the possibility of constructing representations whose parts stand

in structured relations and whose interpretation depends on those relations. The Strong Minimalist Thesis extends this insight into an evolutionary hypothesis: the core of language may be computationally simple, with general principles and interface conditions accounting for much of its design.

The evidence does not warrant treating Merge as a complete explanation of human uniqueness or as a directly observed evolutionary event. Formal simplicity does not by itself establish biological simplicity; action hierarchy does not automatically amount to syntactic computation; and context-sensitive interpretation cannot be reduced to Merge. The most promising account is therefore integrative. A productive combinatorial operation may support hierarchical thought, while conceptual resources, memory, contextual inference, social learning, and externalisation make the resulting representations useful in human life.

References

Berwick, R. C., & Chomsky, N. (2016).*Why only us: Language and evolution.*MIT Press. https://doi.org/10.7551/mitpress/9780262034241.001.0001

Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved? PLoS Biology, 12(8), Article e1001934. https://doi.org/10.1371/journal.pbio.1001934

Chomsky, N. (1995). *The minimalist program*.MIT Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*.Cambridge University Press.

Chomsky, N. (2005). Three factors in language design. Linguistic Inquiry, 36(1), 1–22. https://doi.org/10.1162/0024389052993655

Chomsky, N., Seely, T. D., Berwick, R. C., Fong, S., Huybregts, M. A. C., Kitahara, H., McInnerney, A., &

Sugimoto, Y. (2023).*Merge and the Strong Minimalist Thesis.*Cambridge University Press. https://doi.org/10.1017/9781009343244

Fujita, H., & Fujita, K. (2022). Human language evolution: A view from theoretical linguistics on how syntax and the lexicon first came into being. Primates, 63(5), 403–415. https://doi.org/10.1007/s10329-021-00891-0

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

Studdert-Kennedy, M., & Terrace, H. (2017). In the beginning: A review of Robert C. Berwick and Noam Chomsky's Why Only Us: Language and Evolution. *Journal of Language Evolution*, 2(2), 114–125. https://doi.org/10.1093/jole/lzx005