Psychological Realities of Embedded Recursion

Sadaf Khan

Department of Linguistics and Cognitive Science, Pomona College

LGCS191: Senior Thesis in Ling & Cog Sci

Dr. Robin Melnick

May 11, 2020

Submitted in partial satisfaction of the requirements for the degree
Bachelor of Arts in Cognitive Science at Pomona College.

Abstract

This paper studies how structural recursion is processed. Provided a number of constructions utilizing recursion with different levels of complexity (such as possession, coordination, tail-embedded and center-embedded relative clauses), does the ability to process increasing levels of structural recursion scale similarly across construction types? In order to examine the scalability of structural recursive processing, a self-paced reading task is employed with sentences spanning both aforementioned constructions and levels of embedding. A main effect of sentence length is observed, where longer sentences have quicker per-word reaction times than shorter sentences. There is also a main effect of type (construction x embedding level), where higher-embedded types are read more quickly per-word than lower-embedded types. An interaction effect between type and sentence length also exists, but does not explain why more complex, higher-embedded sentences take less time to read per word. When considered in combination with comprehension question accuracy degrading as embeddings increased across all constructions, it is likely that participants "give up" trying to parse longer, more embedded constructions, thus preventing meaningful comparison of the scalability of structural recursive processing.

Keywords: recursion, center-embedding, self-paced reading task, faculty of language

Introduction

The proposition of recursion as a central feature of language and/or linguistic processing has long been controversial in the field of linguistics and cognitive science. A great point of contention deals with experimentally validating such claims, a task which requires a series of assumptions in designing and interpreting such studies: do the experimenters follow generative ideas of syntax? Are the online and offline grammar systems interpreted as separate systems or one and the same? Was a particular center-embedded sentence difficult to comprehend due to the animacy of the verbs involved, wrap-up effects, or long-distance dependencies? These choices in interpretation may lead to later dismissal by researchers who are opposed to the conclusions, believing that a certain effect can or cannot be gleaned with the given parameters. The definition of recursion itself is ambiguous and adaptable to the researchers at hand, leading to further disagreement on the validity or significance of a finding.

To this effect, I hope to study how structural recursion is processed: provided a number of constructions utilizing recursion with different levels of complexity (such as possession, coordination, tail-embedded and center-embedded relative clauses), does the ability to process increasing levels of structural recursion scale similarly across construction types? This question was motivated by apparent gaps between the fields of psycholinguistics and theoretical linguistics, where theoreticians have often posited linguistic capabilities (i.e. the recursive *Merge*) without consideration of psycholinguistic findings that may directly contest their concepts. From a theoretical perspective, structural recursion itself may not be integral to the language capacity—it is rather procedural recursion that is the culprit, the manner in which humans build linguistic structures and process speech. If this is the case, linguistic structural recursion must be handled by the aforementioned recursive procedure, and in fact may be advantaged in comprehension. This is without consideration of limiting factors, such as working memory, the inclusion of which prevents online processing from being of particular help to theoretical claims about recursion.

Conversely, psycholinguistic studies have found that certain higher-level recursive structures such as center-embedded relative clauses have a rapid deterioration in acceptability judgements with increasing levels of recursion. The observed universality of this effect suggests that whatever procedural recursion may or may not exist, it is not particularly suited to structural recursion, nor can be easily observed or measured to make claims about.

In order to examine the scalability of structural recursive processing, a self-paced reading task is employed with sentences spanning both construction type (possession, coordination, tail-embedded and center-embedded relative clauses) and level of embedding (1, 2, and 3). A general psycholinguistic perspective suggests that there will be a main effect of embedding and construction across all types, reflected in increasing reading times for higher levels of structural recursion (tail-embedding, center-embedding) and for greater number of embedding. However, where lower-level recursions such as possession and coordination may produce a more linear increase in reading time, center-embedded recursion will have a cliff effect wherein all reading times increase drastically for the second and third level of embedding. Thus, a third interaction effect of construction and embedding would also be observable. It is unsure where tail-embedded recursion stands within this hierarchy.

In the other camp, a theoretical perspective also posits a main effect of construction and embedding level, with increased sentential reading times with higher levels of structural recursion, as well as greater number of embeddings. The “cliff effect”, if present, would be less pronounced in the higher levels of recursion, or explained by working memory constraints rather than an absence of underlying procedural recursion. There is also a possibility that the nature of the procedural recursion scales with individuals, such that those who are quick to comprehend with higher embeddings in lower-level recursion (possession, coordination) are also quicker to comprehend the same in higher-level recursion (tail-embedding, center-embedding). The difficulty here is to determine whether this is due to naturally higher reading rates, or some genuine underlying comprehension/structure-building process.

Background

It is critical to distinguish between procedural and structural recursion, both of which have been implicated in linguistic discussions. Procedural recursion involves a program calling itself within its execution; structural recursion comprises structures contained within structures of the same kind. The human language faculty—the suite of skills and features that allows for us to utilize a rule-based language where other species cannot—was for some time theorized to be predicated on the ability to recurse and/or create recursive structures.

2.1 Recursion as a design feature

Hauser et al. (2002) (this paper is also known as HCF) is a key paper in establishing recursion as a central design feature of language. Its foremost claim is that the faculty of language can be divided into two separate parts: the broad faculty of language (FLB), and the narrow faculty of language (FLN). The authors distinguish between the phenomenon of internal language ("I-language") as an idealized form of grammar and the performance of language as the application of I-language once filtered through the limited capacities of the mind and sensory-motor interface. A summary of their claims is as follows. It is the I-language that is integral to our understanding of how language exists in the mind. The FLB includes the FLN in conjunction with the sensory-motor interface and conceptual-intentional interface. The FLN is specifically the computational system that organizes language through internal representations (including syntax), while the FLB recruits the conceptual-intentional interface (to access meanings) and sensory-motor interface (to access phonology) to manifest the representation in spoken language. Those who subscribe to the FLN body of thought generally concur that it is recursion- that is, the ability to take a finite set of elements and reproduce them into an infinite set of outputs- that most minimally separates the FLN from the FLB, and humans' linguistic faculty from all other species.

There are three hypotheses the authors list as to how vital the FLB-FLN distinction is to understanding the human linguistic faculty. First is the hypothesis that FLB is analogous to animal communication, a body of thought non-linguists may conjecture. The second hypothesis is that the FLB is uniquely evolved in humans, and the FLN itself may not exist. That is, there are other cognitive and physical features in humans that are not originally specified to language that have evolved to handle language. The third hypothesis surmises only the FLN is uniquely adapted to the human capacity, whereas the FLB may be similarly evolved in non-human species. Additionally, the FLN is primarily composed of the recursion mechanism. This hypothesis is the one Hauser sides with and argue for the remainder of the paper, visiting the oft-cited "poverty of the stimulus" argument: that there is no learning mechanism that can take a finite input of positive evidence (the sum total of productions a child will hear in their life) and derive an infinite output (the sum total of potential productions that same child can produce after acquiring the language). Negative evidence (evidence of what is not grammatically permissible) is likewise not provided, further making the acquisition of language a theoretical paradox. Chomskyian bodies of thought have historically suggested that "universal grammar"—the idea that "linguistic dispositions" such as principles and parameters are innately present in the human mind—can explain how language is acquired. Non-human primates can do certain linguistic tasks, but cannot generalize to an infinity. Generalizing to an infinity is, in this conversation, an extension of the capacity for recursion: it is the conceived last stage in mastering and internalizing the successor function, which humans use to count. With this in mind, the theory of universal grammar is far from universally accepted by linguists and/or cognitive scientists alike. Other papers have found that it is quite possible to learn from the absence of evidence, contrary to the logic underlying the poverty of the stimulus argument (see Regier and Gahl (2004)).

When it comes to rule-based systems, the lowest-level hierarchy is a finite-state grammar (further explored in Jäger & Rogers (2012)) that cannot capture the complexity of human language as it can only handle local structures. Long-distance dependencies are rife in natural languages and thus minimally require phrase-structure grammar to explain their organization. To this end, HCF reviews a study by its same authors regarding the ability of tamarins to access either a finite-state or phrase-structure grammars, concepts that will be elaborated on in section 2.2. Fitch and Hauser (2004) (also known as FHC) examine whether tamarins (a New World monkey) can process different types of syntax. Two conditions are outlined, based on two grammars: a finite state grammar, which can only produce strings that followed the structure of $(AB)^n$ where A and B are each represented by one of eight different CVs, and a phrase-structure grammar, which had the capacity to form A^nB^n. Human speakers created a set of licit and illicit strings for both conditions that are played over speakers to tamarins following a familiarization-discrimination paradigm. The study finds that tamarins can distinguish between licit and illicit strings in the finite-state condition (i.e., ABAB as licit, with ABAA as illicit), but could not do the same for the phrase-structure grammar (AAABBB vs. AAAABBB). Therefore, they conclude that nonhuman primates cannot form long-distance dependencies and lack the recursive capacity which sets the human linguistic faculty apart.

There are several glaring issues with the design. Foremost, each string of As and Bs were uttered by different genders of humans—as utterances by the same one speaker were universally indistinguishable to tamarins—and gender-instantiated differences in pitch acted as a cue that A as opposed to B was being pronounced. This design does not pattern along natural language, where different speakers' utterances would never be interpreted as one utterance overall. The task was then to see if tamarins would notice in the finite-state condition that an AB unit wasn't repeatedly uttered (a relatively easy task), while the phrase-structure condition probed if tamarins would count the number of A units, compare them against another count of the B units, and then determine the two counts were different (far more cognitively challenging). It is unclear if this design really indexes a linguistic capacity or not. Existing literature on the counting capacity/ successor function acquisition in nonhuman primates predicts these results equally well. Concurrently, considering that societies that have mixed results on tests indexing the successor function (i.e. the Tsimane or Pirahã) also have lesser structural recursion or the absence of structural recursion in language, it is also possible that both capacities are innately conflated.

Jackendoff and Pinker (2005) provides a direct critique of HCF. They take issue with the dichotomy between the FLB and FLN that HCF propose, and point out that recursion is not entirely unique to the cognitive capacity, much less sufficient as a first development preceding other aspects of language. They point out: what is the benefit of developing computable productivity without a lexicon? Surely, more base computations such as simple concatenation would have been a major improvement to communication over entirely unordered and unrelated primate calls. Simultaneously, they point out the ambiguity of language used in HCF to describe the FLN as the computational mechanisms of “narrow syntax” and “mapping to the interfaces”, statements that they waver on and reframe in contradicting ways as the article proceeds.

While HCF suggest that recursion doesn’t naturally occur in fields other than cognition (with recursion in mathematics and computer programming being contingent on recursion in language), they fail to acknowledge recursion in visual processing with respect to gestalt principles. A triangle can be composed of many smaller triangles, with those triangles being composed of smaller triangles, and so on; such patterns occur in nature and have been picked out by our visual system as recursive. So, to reframe recursion as unique to the linguistic mind is not entirely accurate. The authors also take issue with Chomsky’s history of writing away much of what constitutes language when it is ill-fitting to his conception of grammar and computation. Idioms are not explained by a recursive operation, although they are rife in every language, and often maintained over long periods of time. Chomsky is happy to dismiss this as—at best—FLB, and at worst, the “periphery” of language. This purism—with regard to what is language and what is not—is problematic when one is attempting to define where language comes from, especially when one is keen to dismiss widely-attested phenomena that can reveal much about the linguistic faculty.

Perruchet and Rey (2005) attempt a replication study of FHC (2004). Chomsky’s grammatical hierarchy states that finite state grammars (FSGs) can concatenate, but phrase structure grammars (PSGs) can additionally embed strings within strings, leading to hierarchical structures and long-distance dependencies. They point out deficiencies in the design of FHC, where the phrase structure grammar (PSG) is represented by A^nB^n but makes does not explicitly code for center-embedding, as Chomsky’s hierarchy suggests is fundamental. Furthermore, by Chomsky’s own concession, certain types of output from a PSG can be similarly outputted by an

FSG, and thus differentiating the two grammars such that output can't be conflated is vital for an experiment of this nature. Perruchet and Rey suggest that humans in FHC did not learn a PSG for the A^nB^n output and a FSG for the $(AB)^n$ output, but rather kept in mind the number of female-to-male voice transitions in a given string, as the former condition only had one (AABB or AAABBB) and the latter had two or three (ABAB or ABABAB). If it cannot be unambiguously shown that humans used a PSG, it also cannot be ambiguously shown that tamarins *failed* to use a PSG. In their replication study, they found that humans were insensitive to noticing a center-embedded structure to the strings, rating violations according the acoustic patterns, regardless of the underlying grammatical structure. They also point out that tamarins are inclined to turn towards noise that are biologically significant, and the FHC acoustic data could be interpreted as biologically significant across conditions, leading to the lack of apparent discrimination between violations and consistent strings for tamarins. Finally, they argue that pursuing studies with nonhuman primates regarding their capability (or lack thereof) for human language by referencing the Chomskyian hierarchy is insufficient and ill-conceived. This is due to issues with perception and salience in nonhuman primates (i.e., how to overcome their biases for biologically significant sounds) but also the presumption that FSGs and PSGs are critical for human language and can be mastered by humans is in itself conceptually flawed.

Thus, a preliminary review of the subject reveals that the primacy of recursion in the language faculty is much debated. The recursive concept interacts directly with language hierarchy schemes, as theorists struggled to discern the place of natural language within formal language theory.

2.2 Formal Language Theory and boundedness

Formal language theory (FLT) is the study of formal languages, which are differentiated from one another purely on the basis of syntax above any other feature such as semantics. The concept and modern-day usage of FLT is preeminent in computer science and mathematics, where its application has yielded great developments in computation; however, its initial use was commenced by Noam Chomsky in an attempt to study the syntactic regularities of natural languages. FLT is less commonly utilized to understand natural languages today, due to limited application and relevance—natural languages are yet to be overtly categorized by its system.

Jäger & Rogers (2012) review FLT and the Chomskyian hierarchy, with languages organized as follows: computably enumerable languages, context-sensitive, context-free, and finite state languages. They build on this pre-existing order with mildly context-sensitive languages and subregular languages as phenomena meant to capture elements of natural languages that cannot be explained by the pre-existing hierarchy.

Of importance is that FLT is largely insensitive to meaning, frequency of use, and processing complexity—all factors that play critical parts in the processing of utterances in natural language. Context (as in, context-sensitive, or context-free) in this setting does not refer to pragmatic or semantic context, but rather the surrounding symbols in a grammar. Differences between rules can apply anywhere in the tree or terminal signals around it. Natural languages do not fit neatly into the hierarchy due to center-embedded constructions, defined as constructions "involving two dependent elements a and b that are not adjacent" and may contain another such construction between them. Because it is unclear whether there is an explicit upper bound to the number of center-embedded constructions considered grammatical in a given string, a natural language may be regular/finite state (embeddings are limited) or context-free/phrase structure (embeddings are unlimited). English allows center-embedding, while Swiss German allows cross-serial dependencies (which cannot be context free, because only nested embeddings may be "allowed" to proceed forevermore).

This finding was introduced by Bach et al. (1986) , which studied the counting mechanisms of the mind. Dutch and German clause-final verb clusters have different allowances: Dutch allows crossed dependencies (1), while German solely allows nested dependencies (2), as depicted below with the phrase "The men taught Hans to feed the horses.":

(1) De	mannen	hebben	Hans	de	paarden	leren	voeren.
The	*men*	*have*	*Hans*	*the*	*horses*	*teach*	*feed.*
(2) Die	Männer	haben	Hans	die	Pferde	füttern	lehren.
The	*men*	*have*	*Hans*	*the*	*horses*	*feed*	*teach.*

It was thought for some time that crossed dependencies would be easier to comprehend and produce than nested dependencies, which notoriously cause problems cross-linguistically. This study sought to evaluate the effect of such constructions on processing by recording Dutch and German participants' responses on two tasks: a comprehensibility rating and a test of correct analysis when confronted with crossed and nested dependencies, respectively. Sentences were constructed with 1, 2, or 3 levels of embedding, and specified to each language. Participants heard each sentence through headphones, and then answered a question about which NP corresponded to which VP, followed by a rating of how easy the sentence was to understand. Whereas 1 and 2 levels of embedding had about equal assessments of difficulty, level 3 was rated harder overall and also as especially harder for German speakers. With respect to mean correctness on the analysis questions, there were main effects of level and language as well, where Dutch speakers both rated higher-level clauses as easier and more often analyzed the construction correctly when compared to German counterparts. The authors suggest these findings indicate that crossed dependencies are easier for the mind to process than nested dependencies, as had been previously suspected, and in following a push-down stack could not serve as the main human language processor due to the fact it cannot construct crossed structures.

Nested dependencies are often looked to as manifestations of unbounded recursion, a feature of language many generative linguists cite as defining the language faculty itself. For it to be disadvantaged for processing in comparison to crossed dependencies points to a large flaw in the theory: how can the property that supposedly gave way to language be so poorly processed by speakers? With regards to the question between a push-down stack and alternate processors leading language processing, this paper was one of the first to open the debate for discussion.

Not only are highly embedded recursive structures poorly processed, but they can be wholly absent from languages, yielding further doubt on structural recursion as a site of expressing the human language faculty. Such languages are reviewed below.

2.3 Recursive-less languages

There are several languages that lack apparent structural recursion, the most prominently referenced being Pirahã. The Pirahã people are an indigenous hunter-gatherer group living in the Amazonian basin, numbering about 800 people. Most members of the group tend to be monolingual (for reasons related to resisting assimilation), although a proportion of some members utilize some words from Portuguese, mostly motivated by trade demands. The Pirahã language is part of the Mura family, of which Pirahã is the sole surviving language. It is largely SOV and notable for the fact that it has no cardinal or ordinal numbers, has a small phonemic inventory, and complex prosody and morphology (Futrell et al. (2016)).

Everett et al. (2005) reviews the Pirahã language, and note that the absence of overt counting phenomenologically could explain the absence of recursion and quantification linguistically. Furthermore, they suggest that the cultural setting of the Pirahã greatly affects the structure of their language, dipping their toe into Sapir-Whorf territory. The authors express doubt towards the notion of Universal Grammar, both because it is unclear as to what UG actually entails, as well as the fact that Pirahã presents a glaring exception to Chomsky's claim that all natural languages are built on the mental property of recursion. However, Chomskyian researchers rebut that Chomsky's claim states that while all language processing is contingent on this property, it may not physically express in the language itself (perhaps related to the notion of null or unpronounced components).

Gordon (2004) studies the number system of the Pirahã to better understand the language's lack of recursion. It is theorized that those that don't have a number system cannot have the recursive capability that allows for infinite language output, due to some relationship between recursion and number (a relationship that is often referenced, but not totally elaborated on). Some languages have number systems are not in base-10, but ultimately still indicate recursion as the successor function is continually called on (say, base-2, base-6, etc.). Pirahã has three main terms for numerosity that have been roughly translated as one (hói, with a falling tone), two (hoí, with a rising tone) and many (baagi or aibai). Through a variety of counting tasks, the author relates two major findings. One, although participants always used hoí (~two) for quantities greater than ones they had already numbered hói (~one), hói (~one) could reference quantities between 1-3. Hói also means small, in contrast to large in reference to mass or area size, perhaps explaining with the term that is purportedly "one" in meaning is actually closer to "small amount." Does this apparent lack of successor function mean that recursion is not present in Pirahã both linguistically and cognitively? How does recursion in numerical cognition differ from recursion in other types of processing (such as visual or auditory), and how do these two differ from linguistic recursion? The work opened the way for many successive studies on number in Amazonian indigene groups having to do with the successor function.

Pica et al. (2004) look at the number sense of Mundurukú speakers, another Amazonian people-group with a language that lacks terms for number beyond 1-5. Through a variety of counting tasks, they see that speakers used 1-5 consistently, but used approximate quantifiers (like "some", "many", "small quantity") less regularly after. Children did not have the term "5" yet and did not use it regularly, mirroring counting acquisition of children in the Western world as well. They summarize how this type of research is situated in the current literature: some researchers, like Chomsky, believe the number faculty derives from language, and is an abstraction of the discrete infinity that linguistic recursion opens the door for. Other researchers argue that all animals (including humans) have a number sense that is not always cardinal but definitely approximate (i.e. ANS), developed apart from language. Other researchers still postulate that the approximate number system (ANS) is universal but the cardinal number system is only acquired through explicit language and symbolism. Having been a student researcher in a cognitive science lab subscribed to the latter theory, it is difficult to investigate the role of mathematical number in linguistic recursive capacity. It seems as though studying languages that lack cardinal numbers after a certain point is a popular study, as it supports the lattermost theory of the human number sense and indirectly implies that human language does not create the ANS. How does the ANS complicate discussions of counting in grammars?

Futrell et al. (2016) produced a corpus of Pirahã based on prior interviews. They analyze the corpus to evaluate if the language possesses recursive features and also publicize the corpus for other researchers' future use. The authors review the debate on recursion in language, visiting HCF, Everett et al. (2005), and Nevins et al. (2009). Recursion is a potential linguistic universal (echoing sentiments of universal grammar as posited by Chomsky), although the manner in which this is suggested by HCF is dubbed as conceptually sloppy due to vagueness in both defining and range of terminology. Everett et al. (2005) combats this conclusion with evidence (or overt lack thereof) from Pirahã, as the language did not illustrate recursive structures at any level, purportedly caused by (or co-occurring with) the lack of regular number system in the language—that is, the cultural standard of lacking cardinality expresses in the language through the dearth of linguistic phenomena that are reliant on the concept of number, such as recursion and grammatical quantification (all, every, some). Nevins et al. (2009) refuted that Everett's claim represented a misunderstanding of HCF, which was speaking to recursion as the base, binary operation Merge (this is tenuous, as on my reading of HCF, there is no mention of Merge, but this proposition is within Chomsky's later body of work).

Merge applies whenever more than two words are strung together, and thus is dubbed recursive in the sense that the operation is called twice, the second time on its own output. This is a recursive operation, which is still different from a recursive structure. The former indicates an operation calling itself (such as a factorial, or the calculation of a Fibonacci number), while the latter is when a structure contains a structure of the same kind (such as stacked possessives like *Sasha's friend's mother*) Nevins et al. (2009) suggests HCF's folly was in presenting recursion as a structural phenomenon that is present in all languages as opposed to a operational phenomenon.

By this standard, Pirahã is recursive, as it contains sentences longer than two words. But Merge is so nebulously defined itself as to be always correct and nearly untestable. All natural languages have utterances with more than two words strung together, so all languages are recursive (operationally). But how do we study the nature of Merge as a cognitive operation? Is it utilized by the mind whenever more than two concepts are strung together as well? Is it implicated in spreading activation models? Can its efficacy be compared across individuals? Can we see children use it more productively with acquisition, and how is this mediated by processing limits? These questions will be revisited in the following section (2.4).

The authors of this paper pause to mention that this definition of recursion as operational is fine, but they move on to study recursive embedding (structural in nature) exclusively, referring to the ability of a linguistic constituent to contain a constituent of the same type. Recursive embedding matters because it is what theoretically allows for infinite output of a language with unbounded length. This paper's definition of structural recursion encompasses both direct and indirect recursion by Roeper's (2011) definition, reflected in the phenomena they evaluate: embedded possessives, reported speech/ sentential complements, relative clauses, and coordination (this paper is reviewed in more detail in section 2.4). If the corpus contains recursion, then it cannot by definition be a regular grammar, as outlined by Jäger & Rogers (2012). They mention that coordination does not itself provide evidence of a context-free language, which aligns with the fact that it is purportedly an example of direct recursion. With all of these phenomena, it is critical that they are not mistaken for juxtaposition or vice versa. English allows for both embedding and juxtaposition that appear strikingly similar (John went to jail because he drove drunk vs. John drove drunk. That's why he went to jail). Juxtaposition involves a semantic, but not syntactic, relation between productions, while embedding is a syntactic relation. Center-embedding is of particular interest to study because it necessarily makes the language at least context-free, if not higher on the Chomskyian grammar hierarchy (i.e. *Because [John drove drunk], he went to jail*). This is because an automaton producing an utterance has to keep an embedded clause in memory (in a "stack") as it builds up the rest of the utterance. If center-embedding is infinitely permitted, then automaton must have an infinite memory, and thusly is not a finite-state machine; therefore, the grammar cannot be regular. The authors mention that such utterances require a "pushdown automaton" that act on a context-free grammar.

The corpus review by the authors finds no unambiguous evidence that Pirahã is structurally recursive for all given phenomena, and thus suggest the grammar is regular. Does this imply that Pirahã has a limited number of utterances? While they concede that their corpus is still limited and potential sites for recursion are not definitively ruled out, either, but rather suggested to be something else due to more likely structures found overtly elsewhere in the data.

2.4 Merge

As it remains unclear whether natural languages permit unbounded center-embedded recursion, due in part to intervening factors such as human memory constraints, the theoretical approach to centering recursion has evolved from a focus on a structural phenomenon to base procedural act.

Powers (2002) explores the nature and validity of Chomsky's Merge, a concatenative mechanism that combines smaller units into larger units on a syntactic level. Chomsky's definition of Merge is that it combines solely two syntactic objects, and is a base requirement in the human computational system in order to produce language. Syntactic objects SOi and SOj thus become SOij, which operates as one unit thereon. Merge is a recursive process, because it calls itself on its output to create sentences, and Merge structures are not symmetrical. The author discusses how this predicts (and fails to predict) common productions in children's output across different European languages. Merge can operate at the word-level, sub-word level, and phrasal-level. At the word level, Merge somewhat predicts children's telegraphic speech in a way that X-bar syntax deals with in a more unwieldy and therefore unlikely manner. At the sub-word level, morphemes or other such segments—perhaps even phonemes—are merged to create phrases, which may explain some bilingual children's novel phrases that are two-word combinations of words from different languages (usually a functional word from one, and a lexical word from the other). At the phrasal level, different phrases can be combined, although the evidence provided for both this level and the paper holistically is sparse. The findings are dubious at best, as random examples serve to further the author's point that a child's two-word phrase must be Merge, because Merge creates an output of two units as one, and two-units-to-one is created by Merge, and so on, rendering the argument untestable and circular. The author concedes near the end that while Merge can describe many "novel constructions", it cannot explain all of children's output, which leaves the question of what those constructions are, and what operation could otherwise explain them.

Roeper (2011) elaborates on the nature of so-called recursion in language through naturalistic data and building on previous theory. Roeper first makes the distinction between direct and indirect recursion: direct recursion involves a category reproducing itself, resulting in a conjunctive structure. Coordination, for example, is direct recursion on Roeper's suggestion. Some noun phrase (NP) goes to a NP and some other NP, or some VP to a VP and another VP, and so on. The order of elements in direct recursion purportedly doesn't matter: *Mary and Doug climbed the tree* is indistinguishable semantically from *Doug and Mary climbed the tree*. However, Roeper's conclusion is questionable. A sentence such as *Sasha braided her hair and ran to class* is different from *Sasha ran to class and braided her hair*. Both sentences can be interpreted as having both actions take place concurrently, or have a sequential reading where Sasha braids her hair prior to running to class or after. These are semantically distinguishable. Direct recursion, however, does not interact with other common parts of grammar such as movement. Items cannot be extracted from a conjunction to be placed elsewhere. Conjunctions also don't inherently involve hierarchy or "dominance relations", perhaps playing a part in why movement is impermissible. Indirect recursion, conversely, involves an interpretive step, where one category produces another distinct category which can then reproduce the first category. This is illustrated through the stacking of possessive phrases (PossP): a determiner phrase (DP) may consist of a possessive as a determiner (a PossP) and its argument (an NP) such *Sasha's comb*. However, the PossP may consist of a DP and a *'s* to indicate possession, and that daughter DP may consist of another PossP and NP, resulting in a phrase such as *Sasha's dog's mother's comb*, which is different from *Sasha's mother's dog's comb* (thus exemplifying the necessity of interpretive steps). This latter type of recursion is tied to Chomsky's Strong Minimalist Thesis (SMT) which is that interpretation proceeds phase by phase or "computation of expressions must be restricted to a single cyclic/compositional process with phase." This description is still somewhat opaque to those not versed in Chomsky's body of theory.

Naturalistic evidence suggests that children first acquire direct recursion and deal with conjunctive readings prior to dominative ones. Suppose five colored balls are lined up in this order: red, green (X), blue, orange, and green again (Y). When 3-4 year olds are asked to point to "the second green ball" in this line-up, more than 50% pick X, suggesting a reading where the ball is conjunctively both second and green. The adjective phrase reproduces itself directly. Adults completing the same task largely choose Y, suggesting a reading where second modifies green ball in a dominative reading; the NP reproduces itself indirectly by reproducing itself after an adjective. However, the reading fails to mention if both children and adults select X when asked for “the green second ball”, which reads somewhat more conjunctively if for reasons related to the ordering of modifiers in English.

Both indirect and direct recursion are managed by Direct Merge and Indirect Merge, respectively. Direct Merge allows for categorical re-generation, and Indirect Merge invokes an identical category by first generating a non-identical one that comprises of the identical category. Merge broadly is defined as the “putative universal form of an operation that underlies syntactic hierarchical structures”, a recursive operation in itself that is recruited whenever more than two lexical items are next to each other (assumedly because phrases consisting of words up to two words can be concatenative and not recursive).

The paper goes on to describe the challenge for children of determining where recursion is permissible in their language, and the time it takes to acquire each phenomena is a function of “how much exposure is involved, the nature of the derivation, the intersection with morphology, and other factors.” This is very vague, and doesn’t make helpful predictions, or explain why center-embedding is so latently acquired.

In researching the terms from this work in other literature, I have found that not many other authors besides Roeper use the concepts of indirect and direct recursion, so it is difficult to give his theory too much weight. In the interest of understanding it better, I also reviewed Roeper (2014), an anthology covering the topic of recursion in cognition through a compilation of works from a range of authors. The first author, Chomsky, sets the problems and definitions for the remainder of the book.

Chomsky’s statements are generalizing and as such I quote swathes of it through this review for maximal clarity. He begins his work by stating that the most basic property of human language is that it provides “an unbounded array of hierarchically structured expressions that receive interpretations at two interfaces, sensory-motor for externalization and conceptual-intentional for mental processes.” Recursions is defined as “as enumeration of a set of discrete objects by a computable finitary procedure, one that can be programmed for an ordinary digital computer that has access to unlimited memory and time.”

What recursive procedures have this property? The general ethos with which the sciences conceive of theories is to begin with the simplest possible solution before elaborating or editing to encompass exceptional phenomena. Chomsky’s theory of universal grammar, i.e. that some aspect(s) of grammar are universally attested and pre-coded into the human linguistic faculty, has been carved down further over the years in the name of this pursuit of simplicity, from phrase structure grammar to X-bar theory to bare phrase structure, parallel simplification of transformational technology to Move’, and now finally to Minimalist Program. These developments have resulted in the following conclusions: Generative grammars Gi are recursive procedures that produce a set of hierarchically structured expressions at two interfaces: SM (externalized language) and CI (thought). Connecting this theory to the biolinguistics framework is a harder task. Chomsky states that it “seems well established... though much contested” that language “represents a domain-specific mental faculty, one that rests on structural organizing principles and constraints not shared in large part by other mental faculties, and in its processing and computation is automatic and mandatory.” This is already controversial, by his own admission. There is ample argument that language is not so neatly specified to one domain, and other reviews within this literature review discuss how the neuroscience of language incites activity in many disparate areas that are also recruited for non-linguistic tasks. Provided we concede this point, Chomsky also draws the distinction between I-language (I signifying internal, individual, and intentional) and every other form of language (including spoken language) being dubbed derivative. I-language comprises of the computational procedure P (or Gi, supposedly) that produces a total set of generated expressions at both the SM and CI interfaces. That which dictates the acquisition of language includes external data, a generic endowment that handles that data, and general principles of development that may be "extra-organic." The genetic endowment must then include guiding principles specific to language (UG), other cognitive processes (statistical analyses), and determinants of structure and organization in the brain generally. These seem to correspond to that which is the faculty of language narrow (FLN), the faculty of language broad (FLB), and the general organization of the mind. Every resultant surface language from P is that which is the optimal solution to the interface conditions (known as the Strong Minimalist Thesis, SMT). They emphasize that the Minimalist Program as defined is "theory-neutral" because it is not a theory itself but a research program; they take issue with the repeated misconstruction of its premise despite repeated attempts at clarification, as research programs can be misguided or poorly implemented but not refuted. Chomsky then covers some hypotheses of language acquisition:

- Continuity Hypothesis: Children attempt to put together their grammars by trying out all linguistic options possible (as dictated by UG) before setting parameters that match the language(s) they are exposed to.
- Maturation Hypothesis: Certain UG principles are only available as the child matures.

He goes into the finitary computational procedure *P*, which involves Merge: an operation that takes already constructed items and combines them into a new item, beginning with "atomic items." *P* has a working space, that is the lexicon and any items created by application of Merge. This theory supposedly corroborates with existing data on language evolution, which is that language didn't seem to exist over 50,000 years ago, and there has been "very little evolution of language" since our ancestors left Africa, about 50,000 years ago. The small gap of evolutionary time within which the modern phenomenon of language manifested is known as the "great leap forward", consisting of some mutation (or set of mutations) in the brain that allowed for the language of thought (LOT) and its externalization to a sensorimotor system that was pre-existing for many thousands of years. The purported mutation is likely Merge (according to Chomsky) but it can be other jumps, such as the externalization process itself.

The following chapter in the anthology, written by Hilda Koopman, covers how recursion is restricted in natural languages. This article is less relevant in content but rather the attitude and values it presents. There are, again, some bold assumptions: The abstract opens with "Since recursion is a fundamental property of human languages, it is puzzling that we regularly find cases where recursion is impossible or restricted." The author eventually argues that such restriction results from an "independently necessary property" each particular lexical items comes with, that "encodes sensitivity to phonological properties." Restrictions apply before Spell-Out—the externalization of Merge's final output into phonology—but after syntactic derivation (which conveniently makes it such that unbounded recursion can naturally occur in production, if not for processing demands and specific lexical-associated restrictions). She concludes, restrictions on recursion, although "accidental", can have their existence explained as "arising from the way that independently necessary properties interact in specific local syntactic environments." This justification seems to defy Chomsky's earlier principle that the data ought to be approached with the presumption of simplicity. It is possible I misunderstand her argument, but dismissing what does not fit the syntactic theory by nudging it to the realm of phonology or lexicon is suspect and a little outlandish.

Aravind K. Joshi penned a chapter regarding the indirect constraint of recursion in language. Joshi states that recursion in a grammar by design implies a potentially infinite set of output, using center-embedding as the oft-cited example. However, despite the potentially infinite depth of center-embedding, it is limited in practice due to human processing constraints. He states that it is unlikely that a limit on recursion depth is not directly constrained through grammar, but indirectly, for two reasons: 1) that such a constraint does not reflect the input necessarily and 2) that such a constraint makes it impossible to discuss a central aspect of language, which is the "potential infinity of output." The argument seems teleological. For Mildly Context Sensitive Grammars, he concludes that although recursion can be unbounded, there are gaps that arise after three or more levels of embedding which suggests constraints are indirect. This property is dubbed Indirectly Constrained Recursion (ICR) and it relates to the canonical and non-canonical forms of a particular string; gaps are part of the non-canonical form by the design of the grammar.

2.5 A gap between subfields

Christiansen and Chater (2017) lay out some of the current schisms and issues in generative linguistics. Many accounts of language acquisition formulated by the generative side rely on dubious assumption, i.e. children do not make mistakes as such but are rather trying out the grammars of other languages, or that the two-word stage comprises of many unpronounced or deleted elements at the underlying level. In the more technical domain, linguistics fails to provide relevant expertise to natural-language processing, illustrated by the following quote from Fred Jelineck, an IBM's engineer: "Every time we fire a linguist, the performance of our system goes up." There are no sectioned-off functions of the brain solely dedicated to language that neuroscientists have ever found, and purported language genes have yet to be isolated. In order to maintain validity of prior theory, generative linguistics has had to redefine and partition many of its terms, differentiating between competence vs. performance, language change vs. language evolution, core vs. periphery, and so on. These partitions impede natural dialogue between the fields, allowing them to further develop disparate notions that are uninformed by the other. Christiansen advocates for the construction-based framework over the traditional perspective for these reasons, arguing that it facilitates better dialogue (no pun intended) across fields and reframes the language acquisition problem as the child being a developing language-user who is "honing" their processing skills, instead of a mini-linguist attempting to parse out structure from nearly nothing but UG. The former framework does not necessitate a UG but only the ability to perceive probabilistic information from the input.

They elaborate on the Now-or-Never bottleneck, which describes the outcome of two conflicting facts about human cognition: that we have a quickly expiring auditory memory (around 3 seconds), but we produce speech dependent on what has just been said at a rapid rate (about 150 words per minute). So the mind must process incoming information nearly instantaneously in order to keep up and designate responses. The bottleneck may also explain why different kinds of modifiers tend to appear nearly directly adjacent to their objects and why sounds and words clump together to form larger phrases; language acquisition may be as simple as the task of delineating these local patterns. A construction-based account has no dedication to recursion, and no qualms with language without recursion at whatever level. It argues that languages only become more learnable with each iteration, as their properties are selected to the child's processing and learning biases. No claims are made within this article about properties of grammar that allow potentially infinite output, or the essentiality of such properties to defining language.

In continuing to highlight the divide between generative linguistics and psycholinguistics, Martorell (2018) summarizes the findings of many other papers compactly. The author starts by enumerating the psycholinguistic camps on the validity of generative linguistics. While some use the findings of generative linguistics to inform their experimental work in psycholinguistics, others advocate for reassessment of certain generative assumptions about processing, and still others argue that completely alternative frameworks for language should be adopted when discussing language. Even though real-time data is supposedly fairly consistent with generative predictions (a point I haven't previously come across, leading me to read the cited papers for this statement and reviewing them as well), generative linguists tend not to use psycholinguist experimental results to inform their theory-building. This has led to two different accounts with their own jargon of linguistic phenomena: formal grammatical phenomena is competence or computation, a domain entirely aside from real-time processing, which is performance or algorithm. Computation is a set of offline rules, algorithm is its online implementation. Algorithm therefore may involve computation, the latter of which is summarized as Chomsky's Merge. Merge was solely formulated as a phenomena of competence, with no real-time implications. It generates strings from last item to first, opposite to the structure of spoken language, and therefore has no predication or application to how real-time processing takes place, and whether this processing involves syntactic hierarchy. A simple revision to this theory—that Merge actually propagates first item to last—may be all that is needed to unite the computational-algorithmic split. Prior work has shown that processing can be both bottom-up and top-down, and that these two pathways serve different purposes. A top-down pathway predicts syntactic structure, while a bottom-up pathways takes the actual input and integrates it into said structure. A furthering of this hypothesis suggests that the top-down solely concerns itself with what is syntactically mandatory by focusing on verbs and its arguments, and bottom-up handles extraneous adjuncts.

Lewis and Phillips (2015) examine the difference between the one-system and two-system hypotheses of competence and performance. General consensus is that online data is the realm of the psycholinguist, and offline data that of the theorist. Offline data has no time restriction, is only registered after an entire linguistic unit is consumed, while online data is elicited during specific time windows, with limited units presented at a time. In the two-system hypotheses, online (language-processing) and offline (grammar) data are interpreted by two different systems. The grammar is a static set of rules, and language-processing is an applicable set of procedures. The motivations to argue for such a system is that it allows us to claim that the core of the human linguistic faculty is recursive ability in the grammar, without having to see it applied in language-processing. It supports the idea language cognition benefits the human mind in domains outside of solely communication, i.e. the access to recursion and infinity conceptually. It also legitimizes the observation that representing complex thought and optimizing efficient communication are different goals that need different tools. If there are two systems, how do they interact? We should expect different responses for the same material when processed offline versus online. If the system is singular, we can derive that the capacity for language was a single response to two different pressures: to have both an internal and external language. Under this system, offline and online responses may be the same (reflecting the singular system) or may be different (reflecting different snapshots in the processing pathway). There is some evidence that parsers do constrain online representations in the direction of offline representations, evidenced by ERP responses to grammatical anomalies, suggesting a one-system view, even if procedures for comprehension may be rougher than under ideal conditions. The authors argue that otherwise attested misalignments in online vs. offline responses can be explained by limitations of the rough heuristics utilized in online processing, and such misalignments are not robust or diverse enough to necessitate arguing two systems. These misalignments can be broadly categorized in the following: garden paths/revision failures, resource overload, consequences of memory access mechanisms, and internal stages of computation. Center-embedded relative clauses, a notoriously taxing representation are supposedly encompassed by the resource overload category. The category is defined by a failure to represent constructions that are technically licensed by the grammar. The authors argue that parsers can represent certain types of multiple-center-embeddings, although the Missing V effect (parsers judge three or more center-embedded clauses as basically grammatical even when one of the three+ verbs is missing) is less easily explained. There are some qualifiers: one, “basically grammatical” is a relative judgement that does not mean well-formed, substantiated by the fact that parsers cannot report a well-formed interpretation of such a missing-V center-embedded construction. Two, it is possible that parsers overlook a verb in the process in which they associate the last two NPS with the most deeply embedded verb, and the first NP with the outer verb, failing to notice the second NP is still in need of a verb by “forgetting.” That is to say, the misalignment is attributed to memory mechanisms, rather than competence and performance having different rules entirely. They conclude that it is more likely that a one-system hypothesis prevails over a two-system one in validity, and as such psycholinguists and theorists should use one another’s findings to inform their hypothesizing.

Petersson and Hagoort (2012) take to the biological realm and evaluate neurobiology in the purview of Chomskyian syntax. The brain can parse and generate grammars through connectivity and structured sequence processing, and artificial language learning (ALL) paradigms can reveal how these mechanisms are utilized. The parts of the brain in which language is processed extend beyond Broca and Wernicke's regions, including the left inferior frontal gyrus (LIFG) the superior-middle temporal cortex, the inferior parietal cortex, and basal ganglia (their justification for what technically counts as a linguistic processing instead of a generic processing is less clear). Production and comprehension are typically attributed to Broca's and Wernicke's region, but syntactic and semantic unification during is at least somewhat represented by activity in the LIFG for comprehension and in the superior-middle temporal cortex for production; unification is described as "real-time combinatorial operations." However, they specify that none of the parts of the brain that are recruited by language processing exclusively contend with language; the N400, P600, and LAN responses used in so many linguistic EEG studies are also incited by other stimuli, such as music, pictures, and gestures. Function does not map to structure in a one-to-one correspondence. This said, they try and examine syntax in an isolated context (which they agree is "somewhat artificial, because syntactic processing never occurs in isolation from the other linguistic levels").

Because the brain is finite in memory resources and organization, the authors suggest it must be a finite machine in the Chomskyian hierarchy. The speaker-hearer has both a finite memory to use for a grammar that can output infinite productions, and performance of language can be represented through a finite state automaton. This is in contrast to Jäger & Rogers's (2012) analysis of natural language, which created a new category of mildly context-sensitive to encompass the expansiveness of natural language. It is similarly untrue, as the authors imply, that finite state automatons necessarily create infinite output, or that a finite memory indicates a finite state automatons represent natural language processing; variations of other existing machines such as pushdown automata or Turing machines are also possible representations of the linguistic brain provided they are capped memory-wise. Bounded recursive processing, a central issue in discussing whether natural languages are finite state or context-free (and therefore the allowance of unbounded center-embedded recursion), is also discussed, but not with any conclusion: this is an outstanding issue in neurobiology research, as the processing of bounded patterns of non-adjacent dependencies is not well understood. Chomsky's theory circumvents the conclusion that the brain is a finite-state machine by distinguishing between performance and competence. The existing research suggests that bounded recursive processing is a latent capacity and unlikely to be the central tenant of the human language faculty.

2.6 Center-embedded clauses

As previously reviewed, some languages overtly lack grammatical structural recursion such as center-embedded clauses. However, even in English, higher-levels of structural recursion is rarely produced. In this context, lower-levels of structural recursion are coordination or possession, where smaller, identical constituents are structurally stacked next to one another. Higher-levels of structural recursion can involve tail- and center-embedding relative clauses, where heavier constituents (such as noun and verb phrases) are stacked either next to one another or nested.

Karlsson (2007) elaborates on the chasm between theoretical syntax and psycholinguistics by focusing on such center-embeddings. Whereas the former generally concludes that there are no numeric limitations to multiple center-embeddings (leading to the conclusion that syntax is unbounded recursion), the latter argues that the dearth of multiple center-embeddings in spoken language does suggest the presence of grammatical restrictions. There is still discourse about the delineation between acceptability as a grammatical judgement as opposed to processing difficulty.

Standard center-embeddings that do appear typically involve relative clauses (RCs) that modify the subject NP of the subordinate clause. Such RCs may act as noun-complements, thus allowing central embedding (CE). However, 2 center-embedded relative clauses (or 2-CE-RCs) do not appear in spoken language, much less CE-RCs exceeding that number; the author questions the validity of positing features associated with such grammatical phenomena as a central design feature of language re: recursion. Center-embedding is critically the distinguishing factor between natural language syntax being deemed fundamentally context-free (no upper bound on CE-RCs) and finite-state (upper bound exists).

Previous studies such as Miller and Isard (1964) find that acceptability judgements of a given sentence reduces with each additional CE, and Reich (1969) posits that the maximum acceptable number of CEs is likely 1 (this hypothesis is dubbed C1 max). This is countered by De Roeck et al (1982) as they find that 2 CEs can be processed with specific limitations (C2 max). Sampson (1996) argues that there is no demarcation between C2 and C3. Within the realm of language acquisition, Menyuk (1969) found that center-embedding is the last embedding position that children use, usually after age 7, (near the end of FLA). Diessel and Tomasello (2005) similarly found that 2-CE-RCs are basically absent from children's natural speech.

Karlsson examines the Brown corpus among others and finds 2+-CE-RC almost exclusively in written English, where if it is to appear, C2 is the norm, with C3 the absolute maximum, often appearing in far older texts and likely a development of Latin stylistics. More common than recursive center-embedding was tail-embedding (that is, recursive left or right-branching, such as genitives or PP chains). He concludes that tail-embedding is a concatenative process rather than recursive one.

Nickels et al. (2018) studies in what ways double-center-embedded relative clauses (hereon called 2CE-RC) can be altered to be more easily pronounced. Key details of 2CE-RCs include: 1) that they are difficult to comprehend, often deemed as ungrammatical/ deteriorating in grammar 2) that the use of pronouns as NPs make them somewhat easier to read, 3) they are subject to missing VP effects, wherein the second VP in a 2CE-RC ("VP2") can be omitted without comprehenders noticing a missing element. Different scholars attribute this difficulty of comprehension to different reasons:

- The parser cannot recursively call the same sub-routine (Miller & Isard 1964).
- A three-NP sequence with no relative pronouns is misparsed as coordination (Blumenthal 1966).
- Exponential increase in number of potential grammatical relationships (Fodor & Garrett 1967).
- The parser cannot assign both subject and object roles to NP2 (Bever 1970).
- The Sausage Machine parser can't correctly 'chunk' the word string (Frazier & Fodor 1978).
- 'Disappearing' syntactic nodes in complex tree structures (Frazier 1985).
- Syntactic prediction locality theory (SPLT, Gibson & Thomas 1999)

Such 2CE-RC constructions (i.e. *The rat that the cat that the dog chased killed ate the malt*) are both difficult to comprehend (with ample evidence that the "appropriate" syntactic structures are not created in a comprehenders' minds) and difficult to read aloud, usually taking on a "list intonation" as opposed to naturalistic prosody used for declarative sentences. The authors suggest that this is due to a mismatch at the interface of syntax and prosody. Through a variety of rating and pronunciation tasks, they find that shorter nested phrases and longer outer phrases reduces the list intonation and makes prosody more naturalistic, while the opposite was true when nested phrases were longer and outer phrases shorter. Sentences that participants read with more naturalistic prosody were also not subject to the missing VP effect, suggesting an accurate syntactic structure was represented due to prosodic alignment. The authors mention that

2CE-RCs are ideal for psycholinguistic research into the syntax-prosody interface, a relatively novel view among the other sources I researched, where 2CE-RCs were relegated to the domain of theoretical syntacticians. They make no mention of the role of center-embedded recursion in grammar or the linguistic faculty, but provide ample material sentences for future reference.

My previous research focused on this debate, and argued that within non-linguistic processing, it was evident that children generalized crossed dependency patterns earlier in life than nested patterns, and performed better overall with crossed dependency replication tasks than nested. In adults, this pattern was less evident, as they successfully generalized across both crossed and nested due to matured processing. However their performance in crossed dependency replication tasks was quicker (indicating more fluent processing), and across both tasks, their timing reflected the usage of a queue processor over a push-down stack. The non-linguistic domain is still another beast from linguistic specific processing, but a cursory look at processing in the human mind suggest we are not specifically customized to accommodate nested structures in any domain.

2.7 Summary

Due to the expanse of the literature reviewed, a summary of the concepts and their relevance to this study is in order.

In section 1, the history of recursion's prominence in debates about the language faculty was broadly visited. The topic seemingly polarizes many linguists, and the definition of recursion varies by researchers. Papers that seemingly provide evidence for recursion's centrality to the language faculty (FHC, HCF) are critiqued for insufficient methods (Perruchet and Rey (2005)) and unbased assumptions (Jackendoff and Pinker (2005)). Both procedural and structural recursion are tested in said studies. Linguistic structural recursion is demonstrated by a constituent being contained within a constituent of the same kind in a given phrase, whereas procedural recursion is the occurrence of a function or procedure operating recursively on an input in processing and producing language.

In section 2, formal language theory (FLT) is explored as a concept relating the property of recursion to languages. Jäger and Roger (2012) attempt to fit natural languages into the Chomskyian hierarchy, as it has no clear position in the hierarchy utilized by computer science and mathematics alike. Bach et al. (1986) explores how the mechanisms of formal language theory—such as stacks and queues—may or may not be utilized by the human mind when processing structural recursion. They find that nested dependencies are generally difficult for subjects to comprehend, regardless of if they are speakers of languages that permit nested dependencies or not; hence, it is unlikely that the mind uses a pushdown stack, which is better adapted to creating nested/recursive structures than crossed/non-recursive structures.

In section 3, languages that lack structural recursion such as Pirahã and Mundurukú are observed. Everett et al. (2005) and Pica et al. (2004) suggest that the absence of counting and approximate number sense (ANS) among these people-groups is the reason for the lack of observed quantification and structural recursion in their respective languages. This observed gap is corroborated by Futrell et al. (2016). Unfortunately, the literature's discussion of the successor function, mathematical recursion, cognitive recursion, and linguistic recursion are often conflated and difficult to untangle from one another.

In section 4, *Merge* is explored as the procedural recursion that many generative linguists now use as the reference point for recursion's centrality to the language faculty. Powers (2002) outlines *Merge* as a function that builds sentential structure by calling itself on two items, and then again on the output of those two items to create a three-item structure, and so on for longer structures. It is recursive primarily because it calls itself on its output. It is difficult to test the validity of *Merge* as a theory because it is a general combinatorial operation that may operate on non-linguistic levels as well. Roeper (2011) and Roeper (2014) discuss direct and indirect recursion, implying that the latter constitutes a more complex recursion as it requires are intermediate step. Chapters within Roeper (2014) also conjecture that structural center-embedded recursion is indirectly constrained by natural language's grammar as opposed to constrained by working memory or other such cognitive factors.

In section 5, the division in field development between generative/theoretical linguists and psycholinguists is sketched. Christiansen and Chater (2017) speak to the lack of coordination between the two subfields and argue a construction-based account for language acquisition is best suited to uniting the two discourses. Martorell (2018) attempts to rectify psycholinguistic issues with Chomsky's *Merge* by suggesting it propagates first item to last instead of the other way around. The author argues that both bottom-up and top-down processing likely play a part in how this *Merge* predictively operates. Lewis and Phillips (2015) endeavors to parse whether the human language system is singular or double when it comes to processing online and offline data, and find that a single system is likely at play. Petersson and Hagoort (2012) surmise that that, due to the lack of uniquely and solely linguistic activation in the brain when it comes to language processing, recursion is likely a latent linguistic property that has little to do with the language faculty.

In section 6, structural recursion is investigated through a notorious construction: 2-center-embedded-relative-clauses (2-CE-RCs). Karlsson (2007) conducts a corpus study to find that 2-CE-RCs mostly occur in written English, if at all. Nickels et al. (2018) review the unique properties of 2-CE-RCs (that they are difficult to comprehend; that the use of pronouns as NPs make them somewhat easier to read; that the second VP can be omitted without comprehenders noticing a missing element) and why researchers believe misparsings of them frequently occur.

In essence, linguistic recursion spans many dichotomies: structural/procedural, psycholinguistic/generative, direct/indirect, nested/crossed, regular/phrase-structure, and so on. This study aims to test how processing of structural recursion scales with increased embeddings and construction. Construction types include direct recursion (coordination) and indirect recursion (possession, center-embedding, and tail-embedding). Because tail and center-embedding are more difficult to process than possession (as suggested by Karlsson (2007) and Nickels et al. (2018), coordination and possession are both considered lower-levels of recursion even though they span different categories by Roeper's (2011) definition. Due to the muddiness of *Merge* as a measurable phenomenon, this study also looks to see if any processing of recursion varies across participants; that is, just as working memory varies by individual, so may the ability to process recursion. This would be indicated by those who handle greater levels of embedding more quickly in one construction type similarly performing more quickly across other construction types. This could suggest that in the case that *Merge* exists, it is a procedure that fluctuates in efficacy and nature across the population. However, most psycholinguistic studies have shown that comprehension of increased embeddings in center-embedded relative clauses rapidly deteriorate, and thus will not scale across to other construction types: it is perfectly possible to comprehend three levels of coordination embedding (*Mary and John and Kara went to the park*) more quickly than others and still have an equally degraded processing of a 3-CE-RC (*A friend that a mother that a girl saw knows loves*) as any other. In this case, a "cliff" effect will be observed for center-embedded relative clause structures, where comprehension quickly deteriorates after one level of embedding across most participants.

Methods

Materials

72 sentences were created by the author with direction from Bach et al. (1986), with 6 sentences designed for each sentence construction and embedding: coordination, possession, center-embedded relative clauses, and tail-embedded relative clauses (with three levels of embedding each). A sample is shown below.

	a (1 embedding)	b (2 embeddings)	c (3 embeddings)
Possession	Tina's friend caused problems.	Tina's friend's dog caused problems.	Tina's friend's dog's collar caused problems.
Coordination	Tina and Marie went to the play.	Tina and Marie and Ron went to the play.	Tina and Marie and Ron and Tony went to the play.
Relative Clauses (center-embedded)	The cat hissed.	The cat that the dog chased hissed.	The cat that the dog that Marie walked chased hissed.
Relative Clauses (tail-embedded)	The dog barked.	Marie walked the dog that barked.	Marie walked the dog that barked that chased the cat that hissed.

Names and subject nouns were not reused across sentences. Three practice sentences were also included to acclimate participants to the task.

Each sentence was paired with a respective question probing the content of the sentence. Each question had three potential answers, one of which was correct; the non-target answers were either alike to the correct answer in a semantic sense (algebra vs. geometry vs. trigonometry) for lower levels of embedding or referenced another part of the sentence for higher levels of embedding. This was done as lower levels of embedding did not possess obvious alternate answers, whereas higher levels of embedding naturally had a greater number of constituents to draw from in creating foils/non-target answers.

All sentences and questions were displayed through a digital screen, in the face Time New Roman and in the color black against a white background. Answers were displayed similarly except for their blue color on the screen. The full set of sentences and their paired questions are listed in Appendix A.

Participants

55 native-English speaking participants (*M* = 35) between the ages of 21 and 61 years old (an average of 33.62 years) were recruited through Amazon MTurk, a service which offers anonymized surveyors for human intelligence tasks (HIT).

Procedure

A standard self-paced reading task was administered through IbexFarm, a free server for hosting psycholinguistic studies online. Participants were directed to the website on which the study took place through an MTurk assignment outgoing link. The first page of the study acquired demographic information (MTurk Worker ID, sex, age) and consent, as well as explaining the nature of the task. Participants were instructed to click a space bar to proceed through a sentence, with one word appearing at a time, and to answer comprehension questions by clicking or typing the appropriate number associated with the intended answer. Three practice sentences and one practice question were presented following the instructions in order to familiarize the participant with the task, after which target sentences began to appear.

Sentences appeared one page at a time. After the participant had finished reading the sentence, a comprehension question with three potential answers were shown on a new page. the target sentences were presented. The order of sentence types was randomized, with each participant receiving a different ordering of sentences. All participants read all of the sentences, following a within-participants design. The three potential answers per question were also randomized in their relative position. Participants were allotted 30 minutes to complete the task.

At the conclusion of the study, a unique survey code was generated for participants to input in MTurk and receive a compensation of $1.50. The task, in total, took an average of 18 minutes 26 seconds. The total span of data was collected within four hours of posting the MTurk task.

Results

Data was scrubbed and reformatted through Excel and R. Due to nature of MTurk encouraging rapid completion of many tasks in order to receive compensation, there was a wide range of collected data that reflected non-cooperative participants. Data points that represented per-word reaction times that fell outside two standard deviations of the data were excluded to reduce this noise.

An aggregation of total reaction times per type was computed, as well as total reaction times divided by total number of words in each type. This value ("Normalized RT") represents the averaged total time spent on a given word within this type.

Table 1. Reaction Times per Type

Type	**Total RT (ms)**	**Total # of Words**	**Normalized RT (ms)**
1- Coordination	719012	38	18921.37
2- Coordination	864854	46	18801.17
3- Coordination	1085757	62	17512.21
1- Possession	382061	18	21225.61
2- Possession	613321	30	20444.03
3- Possession	875038	44	19887.23
1- Center-Embedded	285549	14	20396.36
2- Center-Embedded	750717	41	18310.17
3- Center-Embedded	1045613	58	18027.81
1- Tail-Embedded	247760	12	20646.67
2- Tail-Embedded	774070	41	18879.76
3- Tail-Embedded	1170165	66	17729.77

As is expected, the total reaction time and number of words increase as embeddings do. Less expected, however, is that the average total time spent on a type per participant within a given construction style (coordination, possession, etc.) universally *decreases* with embedding (Table 1). This is counter to the general literature on processing, where more complex constructions take longer to process. This raises the question as to if participants somehow found constructions with lesser embeddings harder to process.

Table 2. Reaction Times per 1-Embedding

Type	**Total RT (ms)**	**Total # of Words**	**Normalized RT (ms)**
1- Coordination	719012	38	18921.37
1- Possession	382061	18	21225.61
1- Center-Embedded	285549	14	20396.36
1- Tail-Embedded	247760	12	20646.67

When comparing across constructions but within the same embedding level, total reaction times are longest for coordination across all embeddings. In the first level of embedding, possession takes the second longest, center-embedding the third, and tail-embedding the fourth (Table 2). Center-embedding and tail-embedding are closely matched at this level due to identical structures. Normalized reaction times show that coordination in fact took the least amount of total time per word, with possession ranking the highest instead. Center- and tail-embedded types are even closer matched after the normalization.

Table 3. Reaction Times per 2-Embedding

Type	**Total RT (ms)**	**Total # of Words**	**Normalized RT (ms)**
2- Coordination	864854	46	18801.17
2- Possession	613321	30	20444.03
2- Center-Embedded	750717	41	18310.17
2- Tail-Embedded	774070	41	18879.76

By the second embedding level, total reaction times are second longest for tail-embedding, third for center-embedding, and fourth for possession. When looking at normalized reaction times, possession again ranks the highest, with tail-embedding, coordination, and center-embedding following respectively—although the three latent types are highly matched at this level, with coordination and tail-embedding set apart by only about 80 ms. This is less expected as well, as by this point, higher levels of recursion like center- and tail-embedding should begin to take more time than the lower level constructions. Instead, possession seems to be taking the most time, suggesting embedded or "stacked" possession is harder to comprehend than anticipated.

Table 4. Reaction Times per 3-Embedding

Type	**Total RT (ms)**	**Total # of Words**	**Normalized RT (ms)**
3- Coordination	1085757	62	17512.21
3- Possession	875038	44	19887.23
3- Center-Embedded	1045613	58	18027.81
3- Tail-Embedded	1170165	66	17729.77

By the third level of embedding, total reaction times are highest for the tail-embedding, with coordination, center-embedding, and possession in following. When looking at normalized reaction times, possession took the longest per participant, with center-embedding, tail-embedding, and then coordination following. This is somewhat more in line with expected results, as the embeddings outrank coordination, but due to possession's high rank in combination with the fact that normalized reaction times per construction went down with embedding overall, it is still quite disparate from what was hypothesized by both psycholinguists and generativists alike. That is, with increased embeddings, reaction times should be going up regardless; it was the way in which reaction times increased that differed, with generativist thought suggesting reaction times scaling similarly across types, while psycholinguistic research suggesting a cliff effect for the center-embedding where reaction times rapidly increase due to lack of parsing.

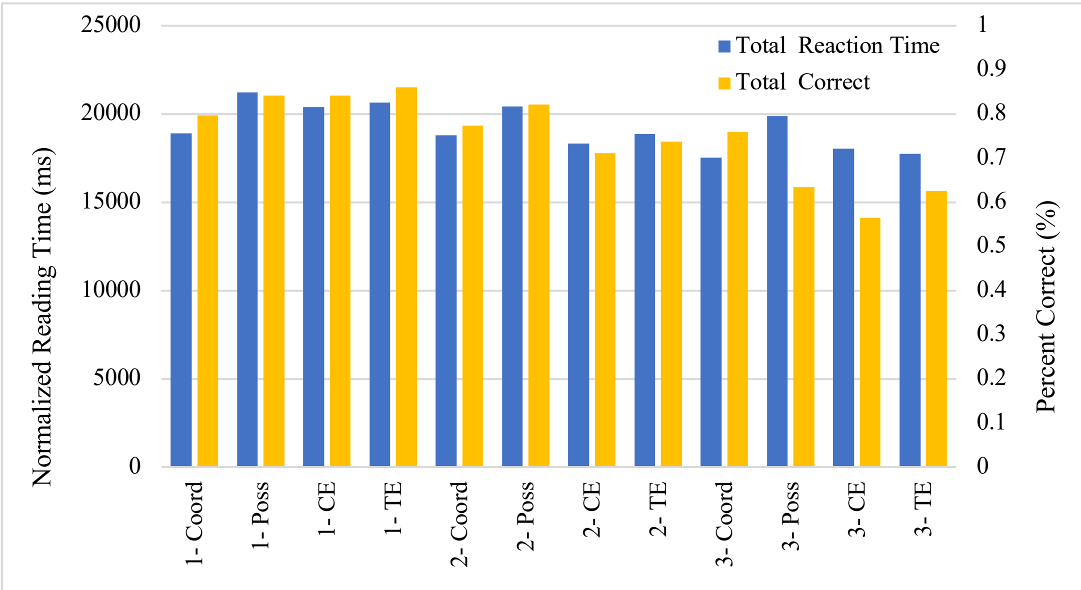
In order to affirm whether sentences were indeed being parsed, the average rate of accuracy per type question was also aggregated.

Table 5. Accuracy per Type

Type	**Correct**
1- Coordination	79.63%
2- Coordination	77.37%
3- Coordination	75.92%
1- Possession	77.73%
2- Possession	82.22%
3- Possession	63.39%
1- Center-Embedded	84.11%
2- Center-Embedded	71.11%
3- Center-Embedded	56.47%
1- Tail-Embedded	86.00%
2- Tail-Embedded	73.77%
3- Tail-Embedded	62.61%

Across nearly all types, accuracy went down with higher levels of embedding; the only exception was, possession, which had a greater accuracy for two levels of embedding than one level, but a lesser accuracy than both for three levels of embedding (Table 5).

Figure 1. Accuracy and Reading Times by Type



This data oppositionally suggests that participants were indeed having a harder time processing greater levels of embedding, despite lesser average reading times with said levels of embedding (Figure 1). Even more conspicuously, coordination's accuracy went down about 1.85% per additional embedding, whereas accuracy went down per additional embedding about 13.82% for center-embedding and 11.70% for tail-embedding, indicating that these higher types of recursions were more impacted by greater embeddings than the lower types of recursion. Chance was at 33% per question, so for three levels of embedding, possession and tail-embedding were answered at about 33% over chance, while center-embedding was answered correctly at 23% over chance. This is preliminary evidence that the ability to process recursive structures did not scale across types, in line with the psycholinguistic predictions.

Data points that were answered incorrectly for the associated question were filtered out to focus on items that were ostensibly correctly parsed. Average reaction times per item within a type were aggregated, and them normalized to represent average reaction times per any word within a type.

Table 6. Average Reaction Times per Item-Type

Type	**Avg. RT (ms)**	**Avg. # of Words**	**RT per word (ms)**
1- Coordination	406.14	6.33	65.02
2- Coordination	404.35	7.67	52.94
3- Coordination	377.86	10.33	36.97
1- Possession	454.23	3.00	151.41
2- Possession	448.45	3.17	143.51
3- Possession	435.35	7.33	60.23
1- Center-Embedded	453.30	2.33	203.35
2- Center-Embedded	391.78	6.83	57.62
3- Center-Embedded	374.76	9.67	38.84
1- Tail-Embedded	463.75	2.00	231.88
2- Tail-Embedded	404.37	6.83	60.52
3- Tail-Embedded	373.40	11.00	34.04

Even with incorrectly answered items removed, the general trend persists: both reaction times per item and reaction times per word went down with increased embeddings within constructions. The differences in reading time per word are fairly dramatic, with the reading times for the third level of embedding in coordination, center-embedding, and tail-embedding falling to 34-38ms long; these values are small enough that it brings into question whether participants were actually reading and processing items, or were effectively giving up on parsing the item. This data, in combination with the rates of accuracy (Table 5) highly suggest that greater levels of embedding for higher types of recursion were misparsed and/or unparsed.

Linear model regressions were performed to see how the per-word normalized reading rate was predicted by type (that is, construction x embedding level) and/or number of words in the item.

Table 7. Linear Regressions

	Coordination		Possession		Center-Embedding		Tail-Embedding	
	Estimate (ms)	P-val	Estimate (ms)	P-val	Estimate (ms)	P-val	Estimate (ms)	P-val
Baseline	**129.631**	<.0001	**185.67**	<.0001	**425.796**	<.0001	**238.236**	<.0001
2-Embedding	**-23.058**	0.1612	**-10.044**	0.6317	**-271.037**	0.0002	**-122.453**	0.0013
3-Embedding	**-44.047**	0.0027	**-41.694**	0.0246	**-356.56**	<.0001	**-169.208**	0.0194
Number of Words per Item	**-10.201**	<.0001	**-11.42**	0.0084	**-95.333**	<.0001	**-3.181**	0.6581
2-Embedding x # of Words	**3.206**	0.1538	**-5.55**	0.2853	**81.118**	<.0001	**-4.906**	0.5495
3-Embedding x # of Words	**5.497**	0.0024	**NA**	NA	**92.189**	<.0001	**NA**	NA
	RMSE: 2.055 Adjusted R^2: 0.976		RMSE: 6.715 Adjusted R^2: 0.973		RMSE: 6.643 Adjusted R^2: 0.993		RMSE: 9.931 Adjusted R^2: 0.988	

The baseline row represents the estimated, average time to read a word provided that sentence length is controlled (that is, the number of words is 0) and there is only one level of embedding. 2- and 3- Embedding rows represent how this estimated average time changes with two and three embeddings respectively for a given type. Number of words per item indicates how each additional word also modifies this baseline reading time. Finally, the last two rows show how type and embedding level interact; the recorded value indicates by how much the "number of words per item" effect on the baseline is augmented.

There is a large span among the estimated baseline values across types (Table 7). This model predicts that center-embedding words take the most time to read, with tail-embedding, possession, and coordination in following with all other factors controlled for. This is somewhat predicted by the existing psycholinguistic literature, where center-embedding is a canonically difficult construction to parse, with tail-embedding being a little easier. Possession still takes some more time in the baseline than coordination, as hinted at by Roeper; coordination is still only direct recursion, whereas possession is the more complex indirect recursion. However, these baseline values are modulated by the effect of embedding level on the estimated average per word reading time; contrary to expected, the impact is not only negative (that is, greater embeddings universally predict faster per word reading times), but is more negative for higher levels or recursion (center-embedding, tail-embedding) than for the lower levels.

Additionally, across all types, there is a main effect of sentence length (e.g. number of words per item), where each subsequent word decreases the average reading time per word. This is not necessarily expected, although it is possible that longer sentences offer more information that can increase predictability of following words and thus quicken reading time. This effect is confounded by embedding level, as higher levels of embedding inherently involve greater sentence length. The positive values in the interaction rows for coordination and center-embedding suggest interaction effects that effectively lessen the impact of the sentence length main effect of lessening per-word reading times. However, these values do not fully explain away the main effect of embedding level, where greater embedding levels predict lesser reading times per word. Moreover, the interaction values for possession and tail-embedding values are either negative, or not available, and thus are opposite to the interaction effects for coordination and center-embedding.

Discussion

Recursive structures have been differently interpreted by various bodies of linguists, such as psycholinguists and generative linguists. Although there is no clean or definitive line dividing views between the subfields—in fact, within subfields, there is still a diversity of thought and argument—typically, the flawed processing of recursive structures is attributed to working memory issues by generative linguists, and to a broader lack of both working memory and internal processing by psycholinguists. Generative linguists often posit that an individual's "I-language" and/or "Merge" method can not only handle, but is specifically attuned to parsing recursive structures due to the primacy of this property to the narrow faculty of language (FLN); psycholinguists conversely argue that the observable issues with processing recursive structures across a range of tasks make it highly unlikely that recursion is a central property of humans' FLN, if this so exists.

In order to attempt again at measuring the application of a recursive processing method, a self-paced reading task was employed using different kinds of recursive structures informed by previous literature (coordination, possession, center-embedding, tail-embedding) with increasing levels of embedding (1, 2, and 3). Psycholinguistic perspectives suggested that while the ability to process embedding may scale within a given construction for coordination and possession, manifesting as increased reading times per greater embedding level, certain constructions like center-embedding were canonically near-impossible to parse and thus would not scale with embedding level and instead "fall off" by the third embedding and have no evidence of comprehension. Generative perspectives suggested scaling could occur across all constructions, with those who could scale better than cohort members at a third embedding for lower level recursions (coordination, possession) could similarly perform better than cohort members for higher level recursions (center- and tail-embedding). This view also puts forth that scaling within constructions may not be similar across constructions—thus, center-embedding may still cause very real issues due to aforementioned acknowledge of working memory constraints—but some modicum of parsing should take place.

A precursory analysis of the collected data for this study has not lined up clearly with either subfield's general outlook on structural recursion. As sentence length increased across constructions, average per word reading times went down. As embeddings increased, so again did average per word reading times go down. Because sentence length was confounded with embedding level (higher embedding levels necessitate longer sentences), some of the lessening reading time effect could be attributed to sentence length, but not enough that it would explain why participants seemingly processed higher embeddings more quickly. Both a generative and a psycholinguistic approach would anticipate greater reading times with higher embeddings; it was only the manner in which reading times would increase that differed among the two perspectives. A review of performance on associated comprehension questions indicated that quicker reading times did not predict whether correct parsing had occurred. While question accuracy for coordination degraded only slightly with each embedding, possession, tail-embedding, and center-embedding degraded drastically with each embedding. Center-embedding had the smallest accuracy score for three-embeddings among the constructions, supporting existing psycholinguistic literature of the difficulties center-embedded relative clauses present.

There are several explanations for these preliminary findings. The main effect of sentence length is explicable through literature where longer sentences offer more information for the reader to predict following words from, thus hastening reading time (Smith & Levy (2013)). The secondary main effect of embedding level in light of the unrealistically rapid per-word reading times and decreased question accuracy suggests that participants gave up trying to read and/or parse sentences with greater levels of embedding, and instead quickly clicked through such sentences. Unfortunately, this prevents the observation of how participants would read types when attempting to build a parsing, and thus prevent comparing results against the two hypotheses for data scaling established prior to data collection.

Further analysis of the gathered data would do well to look at read times per critical regions, such as the second verb for 3-Center-Embedding, or the verb and any words following for 2-Possession. If pauses exist at these regions despite the currently observed “giving up” effect, there could still be evidence of how the ability to process structural recursion scales overall.

There were several other issues with the design of this experiment that could be remedied for future directions. Center-embedded relative clauses are notoriously hard to parse when the relative pronoun is made into a null or unpronounced constituent. That is, *the cat that the dog chased hissed* is easier to read that *the cat the dog chased hissed*. The second sentence causes garden-path like effects on reading, whereas the first, less so (although parsing issues still exist). In this design, these relative pronouns were left in due to the belief parsing difficulty would still be observed with their inclusion; in fact, the concern was that participants would be encouraged to “give up” parsing without such relative pronouns, as the underlying structure of the item becomes even more opaque without them. Seeing as “giving up” has likely occurred even with the inclusion of these relative pronouns, it could be beneficial for the materials to exclude them and therefore be able to directly compare performance on such items to attested performance on 2-CE-RCs and such from the reviewed literature.

Further experimentation must ameliorate the “giving up” effect. MTurk is a convenient service for aggregating many participants remotely, as was required both by the quarantine circumstances of the time at which this task was being conducted and by the nature of the scaling effect likely being of small magnitude—hence needing a larger sample size to increase statistical power. However, because MTurk workers are encouraged to finish tasks as quickly as possible for best financial returns, a “giving up” effect is more likely to happen in order to hasten the duration of the task. Longer sentences were visibly longer on the screen as indicated by a greater number of dashes, so participants were also cued into when they would need to click through a sentence more quickly in order to maintain a quick pace of completion. The self-paced reading task was essentially in competition with other tasks on the MTurk worker hub, thus reducing the level of effort involved in completing it. To reduce this effect, this task could be done in-person; however, this is highly inconvenient for the desired number of participants in order to glean whatever type of scaling effect exists. Another solution would be to pay workers more for the same task, such that it no longer feels as though money is being lost when spending “too much time” on a present task.

This study has failed to advance the understandings of how structural recursion is processed and how this processing scales. However, answering the question at hand remains compelling and highly relevant to understanding something so fundamental to the entire field of linguistics and cognitive science as the human faculty of language. Therefore, a revised institution of this study is recommended to continue chipping away at how the human mind operates with and for language.

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, *1*(4), 249-262.

Christiansen, M. H., & Chater, N. (2017). Towards an integrated science of language. *Natural Human Behavior*, *1*(0163), 10-1038.

Everett, D., Berlin, B., Gonalves, M., Kay, P., Levinson, S., Pawley, A. & Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current anthropology*, *46*(4), 621-646.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377-380.

Futrell, R., Stearns, L., Everett, D. L., Piantadosi, S. T., & Gibson, E. (2016). A corpus investigation of syntactic embedding in Pirahã. *PloS one*, *11*(3), e0145289.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. Science, 306(5695), 496-499.

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.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211-225.

Jäger, G., & Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1956-1970.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*(2), 365-392.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, *44*(1), 27-46.

Martorell J. (2018). Merging Generative Linguistics and Psycholinguistics. *Frontiers in psychology*, *9*, 2283. doi:10.3389/fpsyg.2018.02283

Nickels, S., Schott, E., & Fodor, J. D. (2018). Center-embedded sentences: What's Pronounceable is Comprehensible. In *On Concepts, Modules, and Language: Cognitive Science at its Core*. New York, NY: Oxford University Press.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates?. *Psychonomic Bulletin & Review*, *12*(2), 307-313.

Petersson, K. M., & Hagoort, P. (2012). The neurobiology of syntax: beyond string sets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1971-1983.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499-503.

Powers, S. M. (2002). Merge As a Basic Mechanism of Language: Evidence from Language Acquisition. *Basic Functions of Language, Reading and Reading Disability*, 105-117.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition*, *93*(2), 147-155.

Roeper, T. W. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics*, *5*(1-2), 057-086.

Roeper, T., & Speas, M. (Eds.). (2014). *Recursion: Complexity in cognition* (Vol. 43). Amherst, MA: Springer.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69-90.

Appendix A

	a (1 embedding)	b (2 embeddings)	c (3 embeddings)
Possession	1. Tina’s friend ran. q. Who ran? a. Tina’s friend. 2. Lisa’s flower grew. q. What did Lisa’s flower do? a. grow 3. Gertrude’s lamp glowed. q. What was glowing? a. Gertrude’s lamp 4. Amanda’s mirror cracked. q. What did Amanda’s mirror do? a. cracked. 5. Miranda’s chair rocked. q. Whose chair rocked? a. Miranda’s. 6. George’s computer dings. q. What does George’s computer do? a. dings	1. Marie’s friend’s dog barks. q. Who barked? a. Marie’s friend’s dog. 2. Tom’s plant’s leaf withered away. q. What did Tom’s plant’s leaf do? a. wither away 3. Henry’s house’s window shattered. q. Whose house’s window shattered? a. Henry’s. 4. Frank’s car’s gas tank ran low. q. What happened to the gas tank? a. it ran low 5. Robert’s shop’s sales went up. q. Whose shop’s sales went up? a. Robert’s 6. Glen’s teacher’s cousin rides a motorcycle. q. What does the cousin ride? a. a motorcycle	1. Jenna’s friend’s boyfriend’s mother wrote a book. q. Who wrote a book? a. Jenna’s friend’s boyfriend’s mother. 2. Darren’s backyard’s pool’s bottom is deep. q. What is deep? a. pool’s bottom 3. Melissa’s artwork’s lines’ ink bled through the canvas. q. Whose artwork’s lines’ ink bled through? a. Melissa 4. Claire’s high school’s principal’s parking space is occupied. q. What is the principal’s parking space? a. occupied. 5. Samuel’s project’s opening dates’ hours are 5pm-9pm. q. What are the opening hours? a. 5pm-9pm 6. Ron’s dad’s commute’s length spans a few hours.

			q. What spans a few hours? a. commute's length
Coordination	1. Phoebe and Evan went to the play. q. Who went to the play? a. Phoebe and Evan 2. Mona and Andrew ran an event together. q. What did Mona and Andrew run? a. an event 3. Fernando and Sophie designed a fashion line. q. Who did Fernando design a fashion line with? a. Sophie 4. Mary and Philip taught geometry. q. What do Mary and Philip teach? a. geometry 5. Dean and Amy stargazed for fun. q. What do Dean and Amy for fun? a. stargaze. 6. Brenda and Adam baked a cake. q. Who did Adam bake a cake with? a. Brenda.	1. Nick and Anna and Haley participated often. q. Who participated often? a. Nick and Anna and Haley 2. Emily and Tyler and Harold roasted a chicken. q. What did they roast? a. a chicken 3. Therese and Nicole and Michael shared a room. q. What did they share? a. a room. 4. Blake and Samantha and Jade played video games. q. Who played with Blake and Samantha? a. Jade. 5. Shane and Beth and Judy made pottery. q. What did they make? a. pottery. 6. Ernie and Poe and Gary sketched a scene. q. Who did Ernie and Poe sketch a scene with? a. Gary.	1. Andy and Eamon and Norah and Harold packed boxes. q. Who did Andy and Eamon pack boxes with? a. Norah and Harold. 2. Erin and Daniel and Lucas and Sarah trained for finals. q. What did they train for? a. Finals. 3. Scott and Jordan and Perry and Finn developed an application. q. What did they develop? a. an application. 4. Rina and Kim and Ali and Steve hiked up a mountain. q. Who hiked with Kim and Ali and Steve? a. Rina. 5. Jerry and Sally and Richard and Hannah arranged a party. q. Who arranged a party with Jerry and Sally? a. Richard and Hannah. 6. Gemma and Julian and Eric and Kyle

			turned on all the lights. q. What did they turn on? a. all the lights
Relative Clauses (center-embedded)	1. The cat hissed. q. What hissed? a. the cat 2. Zach practiced tennis. q. What did Zach practice? a. tennis 3. Jonas sneezed. q. What did Jonas do? a. sneeze 4. Josh kneeled. q. Who kneeled? a. Josh. 5. Margot crawled. q. What did Margot do? a. crawl. 6. Helen panicked. q. What did Helen do? a. panic.	1. The rat that the dog chased squeaked. q. What squeaked? a. the rat 2. The bell that the mayor hung rang. q. What was hung? a. the bell. 3. The banana that Lara ate went bad. q. What did Lara eat? a. banana 4. The headphones that Tara wore played. q. What played? a. the headphones 5. The face mask that Jack tried smelled. q. What did Jack try? a. The face mask 6. The mustard that Callie ate oozed out. q. Who ate the mustard? a. Callie	1. The trainer that the elephant that Shay hired trumpeted laughed. q. Who hired the trainer? a. Sheila 2. The plant that Quinn that Sheila bought watered wilted. q. Who watered the plant? a. Quinn 3. The bird that Noah that Ava spotted drew tweeted. q. Who spotted the bird? a. Ava 4. The phone that the technician that Katherine bought repaired rang. q. Who bought the phone? a. Katherine 5. The light switch that Charlie that Gerald painted switched broke. q. What broke? a. The light switch. 6. The vase that Johnny that Victor

			sculpted purchased held flowers. q. Who sculpted the vase? a.Victor
Relative Clauses (tail-embedded)	1. Melanie shrieked. q. Who shrieked? a. Melanie 2. Joanna exercised. q. What did Joanna do? a. exercise 3. Milo snored. q. What did Milo do? a. snore 4. Zara typed. q. Who typed? a. Zara 5. Chloe stretched. q. What did Chloe do? a. stretch 6. Max cooked. q. What did Max do? a. cook	1. Shirley drank the soda that bubbled. q. What did the soda do? a. bubble 2. Julia fixed the fridge that broke. q. What did Julia fix? a. The fridge 3. Joseph picked the oranges that ripened. q. What ripened? a. the oranges 4. Maya wore the shoes that glowed in the dark. q. What glowed in the dark? a. The shoes 5. Juliet painted the canvas that reflected light. q. What did Juliet paint? a. The canvas. 6. Veronica sliced the carrot that helped eyesight. q. What did Veronica slice? a. The carrot.	1. Angela grew the flowers that helped the bees make honey. q. What helped the bees? a. the flowers. 2. Eve created an application that stored instructions that guide tech repair. q. What did Eve create? a. an application 3. Ariana directed the movie that used visuals that caused epilepsy. q. What caused epilepsy? a. visuals 4. Brandon called the mailman that delivered the mail that shocked him. q. What shocked Brandon? a. the mail. 5. Miles cooked the meal that used ingredients that upset his stomach. q. What did the meal use? a. ingredients that upset his stomach

			6. Shawn surfed the wave that washed out beach-goers that enjoyed the day. q. Who enjoyed? a. beach-goers