UNIVERSITÀ DEGLI STUDI DI VERONA

DEPARTMENT OF FOREIGN LANGUAGES AND LITERATURES

GRADUATE SCHOOL OF HUMANITIES

DOCTORAL PROGRAM IN FOREIGN LITERATURES, LANGUAGES AND LINGUISTICS

XXXVI Cycle / 2020

Language and Perception

Investigating Linear and Hierarchical Implicit Statistical Learning across the Visual, Auditory, and Tactile Sensory Domains

S.S.D. L-LIN/01

Coordinator:	Prof. Dr. Stefan Rabanus (University of Verona)
Tutor:	Prof. Dr. Denis Delfitto (University of Verona)
Tutor:	Prof. Dr. Maria Vender (University of Verona)

Doctoral Student: Arianna Compostella

Language and Perception: Investigating Linear and Hierarchical Implicit Statistical Learning across the Visual, Auditory, and Tactile Sensory Domains - Arianna Compostella
Doctoral Thesis
Verona, 2024

Table of Contents

Acknowledgments

First and foremost, I am deeply grateful to my supervisors, Denis Delfitto and Maria Vender, for introducing me to this research topic and for their continuous support and encouragement over the years. Maria and Denis were the first to believe in this project. Even before my Ph.D., their fascinating lectures during my master's ignited my curiosity and passion for research. Thank you, Denis, for being the first to discuss this research topic with me, for always challenging me, and for your invaluable advice and guidance. Your mentorship has helped me grow both as a person and as a researcher. Maria, I am especially thankful for your constant feedback, unwavering support, and characteristic kindness and professionalism. Without your support, this journey would not have been the same. I am grateful to both Denis and Maria for the many interesting conversations we had and for welcoming me into their research group.

I also extend my gratitude to the thesis reviewers, Diego Krivochen and Jennifer Culbertson, for their invaluable feedback that shaped this thesis. Diego, your expertise in formal systems and insightful advice regarding Lindenmayer systems, Fibonacci grammar, and formal language theory have been incredibly helpful throughout my journey. Thank you for inspiring me with your research in this field and for your constant availability and passion. Jennifer, your comments greatly improved the structure of my thesis, revealing aspects that I had not considered. Without your help, this thesis would not have been the same. Thank you for your valuable comments and feedback, which reflect your expertise and intellectual vibrancy. Thank you also for your availability, kindness, and warmth, and for welcoming me into the research group at the Centre for Language Evolution in Edinburgh for two months during my Ph.D. My time in Edinburgh was incredibly stimulating. I also thank Simon Kirby for welcoming me into the group, for his invaluable feedback on the project, and for sharing his expertise. A heartfelt thanks to the Ph.D. students and postdocs in the research group in Edinburgh for making me feel at home, for the wonderful moments we shared, and for teaching me so much through their fascinating research projects.

Special thanks to Chiara Melloni for her feedback on the first chapter of my thesis and for her availability and kindness. I also thank Massimiliano Canzi for his feedback on the statistical analysis and Elger Abrahamse for sharing his experience in creating a tactile protocol during the early months of my Ph.D. I am also grateful to everyone who provided valuable feedback on different parts of my project during conferences. A special thank you to Doug Saddy, Samuel Schmid, Julie Franck, Vitor Zimmerer, and Beth Phillips for their valuable discussions on the Fibonacci grammar.

I am grateful to all the Ph.D. students and postdocs in Verona for the fun and joyful moments we shared, especially during summer and winter schools. In particular, thanks to Andrea Brugnoli and Greta Viale for sharing this journey with me. I cherish the wonderful friendship that developed with Greta, with whom I shared many adventures and important moments of this path. Marta Tagliani, thank you for the conversations, laughter, and shared moments in the lab and on the train, as well as for your feedback and for teaching me various tricks for statistical analysis with R. Thank you for collaborating with me on side projects and for making our collaboration stimulating. Maura Panozzo Chiomento, thanks for your companionship during conferences and your help in testing participants in the tactile experiment. Thank you for your kindness and the calmness you bring.

I also thank all the participants of the reading groups in both Edinburgh and Verona. These initiatives expanded my research horizons and allowed me to delve into various interesting research topics, some of which are explored in this thesis. Without our in-depth conversations, I would not have discovered certain papers and topics.

I also thank all the participants who took part in the experiments.

Lastly, I am profoundly grateful to my family for their unwavering support and belief in me. Thank you for instilling in me a love for knowledge and a desire for constant improvement. Finally, I want to extend my deepest gratitude to Roberto, who has been my unwavering support throughout this journey. Your love and encouragement have lightened the challenging moments and made the joyful ones even more meaningful.

Abstract

This thesis explored how humans process and form recursive hierarchical structures arising from temporally ordered sequences of stimuli, across the visual, auditory, and tactile sensory domains. As we will explain throughout this thesis, we posit that the ability to form recursive hierarchical abstract representations from temporally ordered stimuli is a cognitive ability involved in human syntax processing and acquisition. Language unfolds in a linear fashion. Words follow one another, creating sentences that, on the surface, appear as linear sequences of sounds or symbols. However, a purely sequential arrangement of words alone falls short in encompassing the complexities of human language syntax. It is evident that the syntax of human languages has a fundamental hierarchical dimension, where constituents are organized in a way that is intricately linked to their linear order. Among the various syntactic phenomena that depend on this hierarchical organization, recursion is one of the most fascinating and controversial in the study of language. Recursion in human syntax, understood as the characteristic of embedding constituents within constituents of the same kind, has long been considered a fundamental and distinctive feature of human language. Therefore, the cognitive ability to deal with recursion has been viewed as crucial for language capacity, possibly representing a uniquely human faculty at the core of language ability. However, this topic is highly controversial. Despite the importance attributed to recursion in linguistics, several questions remain open. What is the role of recursion in human language? Is the ability to handle recursion specifically tied to the human language faculty? What is the mechanism underlying the cognitive ability to form recursive abstract representations in language, considering both the linear and hierarchical nature of syntax? To analyze this topic, this thesis will delve into three critical issues at the core of theoretical and experimental linguistic debates. The first issue addresses the debated role of recursion in human language syntax. The second issue examines the contributions of recursive hierarchical abstract representation and statistical learning to the acquisition and processing of human syntax. The third issue, intimately connected to the second, examines the existence of domain-specific representational and learning constraints, alongside the influence of domain-general learning abilities on this process. Our research had

two main objectives: Firstly, we aimed to determine whether sequential statistical learning and the formation of recursive hierarchical abstract representation operate independently as distinct levels of language analysis or if they work together synergistically as complementary learning mechanisms. If they complement each other, we sought to understand the cognitive processes involved in transitioning from linear to recursive hierarchical dimensions. Secondly, we investigated whether the ability to form recursive hierarchical abstract structures from sequential stimuli is a language-specific ability or a domain-general ability, shared across different modalities and whether there are domain-specific differences in this ability between sensory domains. To address these inquiries, we employed the Artificial Grammar Learning paradigm, conducting three Serial Reaction Time tasks. Three distinct groups of adult participants were presented with a sequence of stimuli featuring the rules of a non-canonical binary grammar belonging to the Lindenmayer systems: The Fibonacci grammar (Fib). The choice to use this grammar was driven by its exceptional suitability for thoroughly investigating this research topic in all its various facets. On one hand, it allows for the investigation of the application of recursive algorithms for predicting points in the string, while simultaneously examining the relationship between sequential statistical learning and the creation of recursive hierarchical representations. On the other hand, this paradigm permits the examination and direct comparison of these cognitive abilities across different sensory modalities. In the three tasks, the symbols of Fib were encoded onto auditory tones, vibrotactile impulses, or colorful visual shapes. Through analysis of reaction times and accuracy data in response to perceived stimuli, we explored whether participants implicitly learned the regularities of Fib across all three sensory domains and potentially domain-specific learning differences. Our findings suggested a close linkage between the ability to form recursive hierarchical representations and the capacity to grasp low-level transitional regularities. With this regard, we introduced a cognitive parsing algorithm hypothesizing the cognitive mechanisms involved in transitioning from sequence to hierarchy. Furthermore, we observed that the cognitive ability to process and learn these structures, which underpin human language, is a domain-general ability present across diverse sensory domains. However, we also identified domain-specific differences, with auditory and tactile modalities exhibiting a distinct advantage over the visual domain. In summary, our results indicated that sequential statistical learning and recursive hierarchical abstract representation synergize as complementary modes of learning, rather than operating as distinct levels of language analysis. Moreover, our findings suggest that the capability to from recursive hierarchical abstract structures arising from temporally ordered stimuli is not a language-specific ability but rather a domain-general capacity present across different sensory modalities, potentially interacting with language in specific ways.

General overview

In this thesis, we embark on an exploration of the cognitive mechanisms underlying the processing and formation of recursive hierarchical abstract representations arising from temporally ordered stimuli, across the visual, auditory, and tactile sensory domains. These structures, as we will see, represent a particular structural phenomenon of human language. Despite the sequential nature of human language, it is widely acknowledged that the relationships between single words in a sentence are not based on rigid linear positions (Tettamanti et al., 2009). Phenomena such as long-distance dependencies, recursive sentence structure, movement, and sentence transformation constitute a hallmark of human natural language (Fitch, Friederici, 2012). Among the many structural (i.e., hierarchical) phenomena present in language, our focus in this thesis will be on recursion, intended as the characteristic of human language to potentially have constituents embedded into constituents of the same kind (Pinker, Jackendoff, 2005, p.203). This phenomenon, in language, can occur multiple times, potentially giving rise to complex structures with multiple levels. Our journey is motivated by two fundamental inquiries: firstly, to elucidate the cognitive processes involved in transitioning from linear sequences to recursive hierarchical structures, and secondly, to ascertain whether this ability is domain-general or domain-specific across sensory modalities. The decision to investigate this research topic arises from the recognition of two pivotal issues that lie at the heart of both theoretical debates and empirical inquiries in linguistics. The first revolves around the significance of abstract hierarchical representation and statistical learning in the intricate processes of human language acquisition and comprehension. The second issue, intricately intertwined with the first, delves into the existence of domain-specific constraints on representation and learning within language, juxtaposed with the broader influence of domain-general cognitive abilities in this multifaceted endeavor. The thesis unfolds across six chapters, each contributing to a comprehensive understanding of these phenomena.

Our exploration into the mechanisms at the heart of language acquisition and processing begins with Chapter 1, where we delve into the longstanding debate between nativist and usage-based approaches on the topic of language acquisition.

In this chapter we explore the arguments supporting each viewpoint, with a particular focus on the topic of the acquisition of syntax. Importantly, we review psycholinguistic studies which provide evidence for the crucial role of both statistical learning and abstract structural representation in syntax processing and acquisition. In addition, we will also provide a brief overview of studies conducted using neural networks that have tested the potential to achieve high-level linguistic competence based on statistical learning mechanisms. The use of statistical learning mechanisms and the development of abstract structural representations are key points emphasized by usage-based and nativist theories, respectively. Rather than viewing these theories as opposites, we will challenge the notion that they are completely incompatible and demonstrate how they can partially complement each other. Therefore, we argue for a significant shift in perspective, urging modern theories of language acquisition to acknowledge and embrace the critical role of statistical learning operating within hierarchical boundaries and constraints in human cognition during the process of language learning. This perspective will guide us through our thesis.

In Chapter 2, our primary focus will be on recursion, the specific type of structural phenomenon found in human language that will constitute the central topic of our study. Recursion refers to the capacity to create complex, multi-level hierarchical structures, where a part of the structure can reflect the same organizational pattern as the whole. A significant challenge in the current (psycho)linguistic literature is the lack of a universally agreed-upon and precise definition of recursion. Thus, our objective will be to provide a clear and detailed definition of this concept. We will examine the interplay between sequentiality and hierarchy in human language, underscoring the necessity of considering both aspects when studying syntactic recursion. Additionally, we will explore the shift from linear to hierarchical structures, focusing on the cognitive processes that underpin this ability. In the latter part of the chapter, we will discuss methodologies for experimentally investigating the formation of recursive hierarchical representations from sequential inputs. This will include an introduction to implicit statistical learning, followed by an in-depth analysis of the Artificial Grammar Learning (AGL) paradigm, which is highly effective for studying implicit statistical learning. Furthermore, we will delve into Formal Language Theory (FLT) and the Chomsky hierarchy of grammars, frequently employed in AGL studies within psycholinguistic research to explore the computational capabilities underpinning human language. We will specifically review studies that have explored recursion using formal languages from the Chomsky hierarchy. Finally, we will address key issues in the study of recursion in psycholinguistics, laying the groundwork for our extensive investigation into how humans form recursive structures from temporally ordered sequences.

In Chapter 3, we will review research on sequential implicit statistical learning and the ability to form recursive hierarchical structures across different sensory modalities. This includes examining the debate over whether implicit statistical learning is domain-specific or domain-general. We will also consider the effects of domain-specific spatiotemporal structures and qualitative differences between sensory modalities.

Chapter 4 will introduce the Fibonacci grammar (Fib), a grammar belonging to the Lindenmayer systems (L-systems). Fib's unique features, such as self-similarity and aperiodicity, make it an ideal tool for studying the formation of recursive hierarchical representations from sequential stimuli. As we will discuss, to effectively study recursion with Fib, however, it is essential to develop experimental designs that address the challenges of investigating cognitive abilities related to recursive processes. We will propose a recursive parsing algorithm for processing Fibonacci strings and argue that it may be the only mechanism compatible with human cognitive resources for predicting points in Fibonacci sequences during a Serial Reaction Time task. The chapter will conclude with a summary of the main findings from studies which investigated the learnability of the Fibonacci grammar so far.

In Chapter 5, we present the design and results of our experimental study, which consists of three Serial Reaction Time tasks in which we expose participants to string generated by the Fibonacci grammar, encoded onto different types of sensory stimuli. With this experimental study, we investigate the capacity form recursive hierarchical structures stemming from sequentially presented input across the visual, tactile, and auditory sensory domains. Our aim is twofold: (i) to ascertain whether this capacity is domain-general and to illuminate any potential differences in learning specific to each modality; (ii) to elucidate the computational mechanisms involved in acquiring and processing these structures, with a particular focus on the interplay between sequential statistical learning and the formation of recursive hierarchical abstract representations.

Finally, in Chapter 6, we expand the scope of our findings by examining their theoretical implications. We conclude by contextualizing our findings within the framework of language acquisition, illustrating how our results advance our understanding of the fundamental processes involved in language processing and acquisition. Moreover, we discuss limitations and propose avenues for future research.

1. Introduction

Language is one of the most fascinating and, at the same time, controversial issues in the study of cognition and mind. Language is a universal human trait found in all known human societies. Crucially, despite there being great variability among human languages, several principles and commonalities are consistent across them (Chomsky, 1959; Pinker, 1994). In addition to the presence of linguistic universals, it is surprising to note the simplicity and similarity with which humans learn language. Children's ability to attain complete mastery of language is simply astonishing, especially if considering the limited set of sentences to which they are exposed in a limited amount of time. How do they succeed in such a complex task, without effort, with no specific intent to learn, and without explicit teaching? Language is the hallmark of the human species, and its acquisition represents a natural developmental process. All children are up to the task, and curiously they manage to acquire it following the same steps, at a nearly similar pace, independently of the type of language to which they are exposed[1] (Guasti, 2002). Children end up having an abstract representation of their language's properties. All speakers intuitively know whether a sentence is well-formed or ill-formed just by listening to it, without conscious effort. Crucially, the linguistic input children receive is dramatically impoverished as compared to the full competence they develop. They come across a finite number of sentences, but they end up with the ability to produce a potentially infinite number of sentences (Guasti, 2002)[2]. Over the years, numerous theories have emerged to explain the enigma of language acquisition. Among these, the nativist theory and the usage-based theory stand out as the most prominent. Simplifying, the nativist theory posits that language is innate, domain-specific, and richly structured (cf. Chomsky, 1957), while the usage-

[1] We refer here to typically developing children. The natural process of language acquisition can be compromised in case of language impairments, hearing impairments, neurological damages, or intellectual disabilities.

[2] Infinity refers to the size or cardinality of a set. It has not been definitively established that languages are infinite, whether countably or uncountably so. (Langendoen and Postal, 1984).

based theory emphasizes that language acquisition occurs through domain-general statistical mechanisms applied to the linear sequence of children's utterances (cf. Tomasello, 2003).

In this chapter, we will delve into the intriguing debate between the nativist and the usage-based approaches in the realm of language acquisition. We will explore both the arguments in favor of the former and those rooted in the latter, embarking on a journey through the most significant and recent studies in both strands of research, with a special focus on the acquisition of syntax. Along this path, clear evidence emerges, emphasizing the crucial role of both structural constraints and external experience in the process of language acquisition. In this exploration, we will observe that numerous studies have put forth proof of the existence of structure-dependent abstract representation, along with the ability to utilize statistical learning mechanisms to acquire linguistic phenomena at the syntactic level, two key aspects that nativist and usage-based theories have respectively emphasized. This chapter aims not to present merely a dichotomy, but to serve as a bridge. Often, nativist and usage-based perspectives have been regarded as separate and conflicting theories. Despite this, we pave the way for a reconciliatory view between the two positions. We will challenge the notion that these theories are inherently incompatible and showcase the presence of some points of convergence between them. Indeed, in this chapter, we assert the paramount importance of experiential statistical learning in language acquisition and underscore the undeniable existence of hierarchical structures and constraints within this process. We effectively call for a paradigm shift, urging contemporary language theories to embrace the vital role of statistical learning operating within hierarchical boundaries and constraints. This perspective will serve as the guiding framework for the present thesis. Indeed, our research will delve into the intricate relationship between these two phenomena, specifically the acquisition and processing of sequential statistical information, and the formation of recursive hierarchical structures, in temporally ordered sequences.

1.1 Statistical learning or UG?

1.1.1 Universal Grammar, domain-specific phenomena, and structure-dependent abstract representation

The main point of the Poverty of the Stimulus (POS) argument is as follows: Because the knowledge required to develop linguistic ability far exceeds that provided by linguistic inputs in the environment, it is hard to believe that children could gain this knowledge by pure exposure to the linguistic stimuli in the environment. The POS argument still represents an important source of support for the nativist theory (Berwick et al. 2011). According to this hypothesis, language capacity is both richly structured and innate. Humans possess a set of domain-specific, hard-wired rules and constraint mechanisms that form the foundation of language acquisition. This would explain why humans manage to acquire language in parallel fashion in the absence of a rich input. The genetic mechanisms posited to enable humans to acquire language is known as Universal Grammar (UG) (Chomsky 1975; 1986). "UG defines the range of possible variation, and in so doing it characterizes the notion of possible human language." (Guasti, 2002, p.18). Linguists among the generativist framework have often stressed the relevance of structure-dependence in supporting the nativist theory. "Generative grammar proposes that the form for expressing rules is innately constrained, and one putative constraint is structure-dependence […]" (Crain & Nakayama, 1987, p.522).

Compelling evidence for the presence of an abstract structure in language arises from studies on the creation of new creoles in sign languages and investigations into home-sign languages (Goldin-Meadow, 2005). Nicaraguan Sign Language (NSL) represents a tangible proof. In year 2004 there were around 800 deaf signers of NSL, from 4 to 45 years old (Senghas et al., 2004). NSL is quite a recent language. Starting from the 1980s, a community of deaf Nicaraguans have created this new sign language which initially showed the properties of a pidgin. Indeed, individuals of this group had not previously encountered a fully developed language. This was due to the fact that until the 1970s, deaf individuals in Nicaragua, both children and adults, were predominantly isolated from each other, leading to limited interactions. Prevailing societal attitudes resulted in the majority of deaf individuals remaining at home, while the educational institutions and clinics that did exist catered to only a small number of children. Consequently, a distinct sign language did not evolve within this context. This absence of a sign language is supported by the fact that the present-day adults who are over 45 years old lack linguistic abilities in this regard (Senghas et al., 2004). In that situation, deaf persons developed different home-sign languages, which allowed them to communicate with their family members. Only starting from the 1980s, deaf people had the opportunity to spend more time together, being more integrated in the society, being progressively enrolled in schools for special education, and they started to meet and socialize also outside school hours (Senghas et al., 2004). From that moment, based on their own home-sign language, children began to develop a new sign language to communicate with each other. Scholars have observed that the grammar of NSL changed over time. Specifically, the changes firstly appear among preadolescent; in a second moment they spread to younger children. Interestingly, however, they did not affect adult-signers' language (Senghas, 2003). As a consequence, younger signers are more fluent as compared to older ones, who retained the original, less developed form of the language. This fact offers an interesting case of study for the exploration of the inception of the universal hallmarks of language, permitting to observe if these properties can emerge naturally during the process of language learning, even in the case of people who have never been exposed to them (Senghas et al., 2004). Senghas et al. (2004) conducted an experiment in which they observed and compared signs and gestures used to describe complex motion events[3] among three groups of 30 NSL signers. The signers were divided into three cohorts, based on the year that they were first exposed to NSL: 10 before 1984, 10 between 1984 and 1993, 10 after 1993. The choice of observing gestures referring to complex motion events is not casual: "the description of motion offers a promising domain for detecting the introduction of segmented, linear, and hierarchical organization of information into a communication system. […] Signing that dissects motion events into separate manner and path elements, and assembles them into a sequence, would exhibit the segmentation and linearization typical of developed languages and unlike the

[3] E.g. climbing up a wall; rolling down a hill;

experience of motion itself." (Senghas et al., 2004, p.1780). Interestingly, scholars found that the older signer used a language that was closer to a gestural model, whereas the second and the third groups of younger signers preferentially used a more language-like communication system, using segmented and sequenced constructions. Hence, over the years, NSL underwent a transformation, from being a set of gestures through which complex expressions were holistically conveyed, becoming then discrete and combinatorial in nature, characteristics which constitute the hallmark of human language. Hence, with this experiment, scholars have proven the emergence of combinatorial patterning and discreteness in the new NSL. This confirms that even in situations where there is no presence of discrete elements and hierarchical combinations in the language context, human learning capabilities can generate these structural aspects anew (Senghas et al., 2004).

In conclusion, generativist linguists claim that given the impoverishment of the stimulus as opposed to the rich linguistic abilities that children show to have and given the presence of universally shared traits in languages, there should be an endogenous, domain-specific, biologically grounded explanation for language emergence.

1.1.2 Statistical learning, domain-general phenomena, and linear order

Alongside the nativist theory, we find very different positions on the topic. Indeed, the issue concerning how language is processed and acquired has been taken into consideration both by psychologists and linguists, sometimes from rather different perspectives. "Traditionally, linguists have emphasized the role of innate knowledge in language, with the influence of the child's environment playing a relatively minor role. In contrast, psychologists studying language development have to explain how the interaction of innate knowledge and the child's environment account for the developmental progression of language ability" (Redington, Chater, 1998, p.129). Specifically, in the last two decades of the 1990s, we assisted in a series of influential publications which seemed to offer new, favorable opportunities in the exploration of language acquisition, renewing a serious interest in the investigation of the learning possibilities from linguistic input, and opening new horizons for the range of opportunities that were taken into consideration by innate positions. Specifically, usage-based approaches started to seem a sustainable alternative, or at least, possible valid complementation to the UG theory. "Noam Chomsky, the founder of generative linguistics, has argued for 40 years that language is unlearnable; he and his followers have generalized this belief to other cognitive domains, denying the existence of learning as a meaningful scientific construct […]" (Bates, Elman, 1996, p.1849). From various fields of study, in fact, scientific results began to arrive that would constitute an important alternative to the hitherto prevailing thesis that language ability is innate and domain specific. Among them, the Connectionist approach, often referred to also as Neural Networks or Parallel Distributed Processing (PDP) (Guasti, 2002; Ramsey, Stich, 1990), and the Statistical Learning approach to language acquisition (SL). Connectionism and SL have been developed in different research areas. However, both approaches are based on the idea that distributional information in the linguistic input, such as co-occurrence of elements, constitute a powerful cue that can be exploited during the learning process. For this reason, they have often gathered under the term *Distributional learning mechanisms* (Redington, Chater, 1998). Connectionism aimed at developing computational models that process and learn language by exploiting statistical cues in the linguistic input they are fed with. SL aimed at investigating humans' computational abilities to grasp and rely on the statistical information that is intrinsically present in language, during the process of language processing and acquisition. Both these fields of study brought important evidence of the presence of very powerful statistical mechanisms that can be used to process and learn linguistic material. These statistical mechanisms would appear to be domain-general, being exploitable in different domains to acquire complex information from external input. Contrary to the nativist perspective, which posits that language ability is largely innate and domain-specific, the Connectionist and Statistical Learning approaches challenge this notion by emphasizing the role of environmental input and statistical cues in language acquisition. While nativists emphasize the primacy of innate knowledge, these alternative theories suggest that learning from linguistic input, enriched with statistical information, plays a crucial role in language development. Rather than relying solely on pre-existing linguistic structures encoded in the mind, these approaches highlight the ability of individuals to extract patterns and regularities from their linguistic environment through exposure and experience. This stands in stark contrast to Chomsky's assertion that language acquisition cannot be explained by learning mechanisms, underscoring a fundamental disagreement between nativist and usage-based perspectives.

Hence, summarizing, influential studies coming from different fields provided new evidence for: (i) The possibility to develop neural networks which showed evidence for the potentiality to learn from the input: Several studies on neural networks have provided evidence for the possibility to extract regular patterns by simple exposure to impoverished output; (ii) the presence of a rich source of statistical information available in the linguistic input; (ii) children's ability to grasp these regularities; (Christiansen, Allen, Seidenberg, 1998). The results obtained from Connectionism and SL contributed toward stealing the limelight from the nativist approach, which constituted one of the most accredited theories in language acquisition in those years, whereas usage-based approaches gained increasingly more credit.

In the next section, we will see more in detail two of these important milestones that marked the history of the usage-based line of research: The development of Rumelhart and McClelland's Neural Network for Past Tense (1986) and the publication of the article Statistical Learning by 8-Month-Old Infants by Saffran, Aslin, Newport in 1996.

1.1.2.1 Connectionist models. 1986: Rumelhart and McClelland's neural network for the English past tense

In 1986, the two psychologists David Rumelhart and James McClelland announced to have created a neural network model for the past tense, which represented "a turning point in linguistics", quoting the title of a review in the Times Literary Supplement (Sampson, 1987). The reviewer expressed his astonishment saying that the implications of Rumelhart and McClelland's study were "awesome". He stated: "To continue teaching [linguistics] in the orthodox style would be like keeping alchemy alive" (Pinker, 1999, p.104). Rumelhart and McClelland's model represents one of the milestones of a new research program in cognitive science: Connectionism, often referred to as Parallel Distributed Processing (PDP). Connectionist models are inspired by the neuronal brain architecture: these models consist of a relatively high number of very simple units that are connected in a net, dimly resembling the architecture of neurons in the brain (Pinker, 1999; Ramsey, Stich, 1990). Networks usually include layers of units (or nodes): one input layer, an output layer, and, between them, one or more hidden layers. The units are linked by weighted connections that transfer activation signals between them so that one unit can inhibit or excite another one. The activation signal consists of a function of the sending unit's activation level and the connection weight. In feed-forward networks, the process goes only in one direction, whereas in more complex networks, as in the case of recurrent networks, the communication between nodes can be bi-directional and there might also be feedback loops (Ramsey & Stich, 1990). Rumelhart and McClelland's neural network represented a revolution as compared to the cognitive models available at that time among generativists. Indeed, this model was not anymore based on the idea of combinatorial rules and symbols manipulation but rather it was based on laws of resemblance (or association) and contiguity (Pinker, 1999). "[…] Pre-connectionist model builders have presupposed computational architectures that perform operations best described as 'symbol manipulations'. In such systems, information is generally stored in distinct locations separate from the structures performing computational operations. Information processing in such devices consists of the manipulation of discrete tokens or symbols, which are relocated, copied, and shuffled about, typically in accordance with rules or commands which are themselves encoded in a manner readily discernible by the system" (Ramsey, Stich, 1990, p.189). Rumelhart and McClelland developed a model for English past tense which involved a network capable of generating the past tense of a verb (output) starting from the sound of its stem (input). The model is based on statistical principles, specifically on distributional statistics (Redington, Chater, 1998). The mechanism of the model exploits the fact that in English past-tense forms are constructed incrementally from mini-regularities that are common across verbs (cf. Pinker, 1999). At the input layer, verb stems are represented by 460 neuron-like units, each capable of being activated or deactivated. These units encode various phonetic features present in English verbs, such as specific vowel and consonant combinations. Rather than having dedicated units for individual verbs, the activation of specific units corresponds to the sounds present in a given verb. Similarly, the output layer mirrors the structure of the input layer and contains the past-tense forms. Connections between the input and output layers are synapse-like and adjustable in strength, allowing for the encoding of associations between input verb stems and their corresponding past-tense forms (Pinker, 1999). "In effect each connection is a probabilistic microrule that states something like, *If the stem contains a stop consonant followed by a high vowel, the past-tense form is likely to contain a nasal consonant at the end*" (Pinker, 1999, p. 196). In the beginning, in a not yet trained network, the connections between nodes have the value of zero. This means that regardless of the input, the output would be absent. In the learning phase of the model, the connections between units undergo modification. This process occurs as the model is trained with a specific set of verbs paired with their correct past-tense forms, presented multiple times. As the model is exposed to positive evidence—instances where the input verb corresponds accurately with its past tense—the weights of connections between units are adjusted. This adjustment of connection weights enables the model to learn and encode associations between verbs and their respective past-tense forms more effectively. Rumelhart and McClelland trained their network using a dataset comprising 420 verbs, each presented 200 times. The model successfully computed the correct sound patterns for the majority of these verbs. Following this, the network was tested with 86 previously unseen verbs. For approximately three-quarters of the new regular verbs, the model accurately generated the past-tense forms with '-ed'. Interestingly, when faced with new irregular verbs, it exhibited reasonable overgeneralization errors, producing forms like 'digged' and 'catched'. Additionally, the model displayed tendencies akin to language acquisition in children, making errors such as 'gived' for verbs it had previously produced correctly. It also demonstrated the tendency to analogize new irregular verbs to existing families of similar-sounding irregular verbs (Pinker, 1999). Hence, the essence of Rumelhart and McClelland's model lies in a clever approach: instead of directly linking one word to another, it connects the phonological characteristics of a word to those of another word. This strategy facilitates automatic generalization through similarity. Importantly, these associations are overlaid across the various words in the training dataset (Pinker, 1999).

In conclusion, connectionist models[4] represented a completely new and powerful alternative to Chomsky's innatism. Indeed, these empirical models seemed to provide the early evidence for the fact that learning can take place even in the absence of pre-existing innate dispositions or knowledge (Ramsey & Stich, 1990).

1.1.2.2 Statistical Learning to extract statistical cues from linguistic input

Besides Rumelhart and McClelland's neural network, during the 80s and 90s, several studies on neural networks have provided evidence for the possibility to extract regular patterns by simple exposure to impoverished output. Moreover, as said above, there have been interesting discoveries concerning the rich source of statistical information available in the linguistic input. However, although there had been consistent evidence for the fact that neural networks can learn by simply exploiting statistical regularities available in the linguistic input, various scholars remained deeply skeptical about the hypothesis that humans might learn in the same way, being able to deal with the statistical regularities available in the speech stream they hear. One reason was that children's memory and attention seemed too limited to support this kind of learning (Bates, Elman, 1996). It is only by taking into consideration the state of the art at that time that we can fully appreciate the centrality of the discovery made by Saffran, Aslin and Newport in 1996. Indeed, these scholars provided a completely revolutionary kind of evidence: They demonstrated that infants as young as eight months old could segment nonwords

[4] Besides Rumelhart and McClelland's model, other connectionist models have been developed in those years to simulate language processing. Among the others, PARSNIP, created by Hanson and Kegl in 1987; St. John and McClelland (1988); Fanty (1985); Cottrell (1985); Waltz and Pollack (1985); Selman and Hirst (1985); Charniak and Santos (1986); Elman (1990); Plunkett & Marchman, 1991; Seidenberg & McClelland, 1989; Zemel, 1993.

consisting of three-syllable sequences after being exposed to a continuous string of nonsense syllables for just 2 minutes. Remarkably, they achieved this segmentation solely by relying on statistical cues present within the string. "[…] The nature of this learning is surprising: a purely inductive, statistically driven process, based on only 2 min of incidental input, with no reward or punishment other than the pleasure of listening to a disembodied human voice. […] it contradicts the widespread belief that humans cannot and do not use generalized statistical procedures to acquire language" (Bates & Elman, 1996, p.1849). The relevance of the discovery is ensured by the fact that, in this study, Saffran and colleagues managed to investigate pure statistical learning by avoiding the presence of other possible acoustic cues in the string, which could have helped children in the segmentation process: the only cues in the artificial speech stream were transitional probabilities between syllables. To build their pseudowords, indeed, Saffran and colleagues exploited one property of natural language: "Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another within a word, whereas transitional probabilities spanning a word boundary will be relatively low" (Saffran, Aslin and Newport, 1996, p.1927). For example, in English, given the two words pretty#cold, the transitional probability between *pre* and *ty* is higher than that between *ty* and *co*. To test whether babies might detect word boundaries by exploiting the sequential statistical information in a concatenated speech stream, Saffran and colleagues designed an experiment in which they exposed children to two types of auditory stimuli. Babies were tested through the familiarization-preference procedure (Jusczyk and Aslin, 1995). In the first part of the experiment, children were exposed to a 2-minutes continuous speech stream of three-syllable nonsense words, which were randomly repeated. No prosodic cues were provided. The speech stream sample comprises the orthographic sequence *bidakupadotigolabubidaku…* Word boundaries were discernible solely through transitional probabilities between syllable pairs, which were consistently higher within words (1.0 in all instances, such as *bida*) compared to between words (0.33 in all instances, such as *kupa*) (Saffran, Aslin and Newport, 1996, p.1927). In a second moment, children were presented with two types of stimuli: items that were presented during the familiarization phase and items that were highly similar to those previously presented but that have never been encountered before. Every baby was exposed to multiple repetitions of a particular three-syllable sequence in each testing session. Among these sequences, two were considered "words" from the artificial language used during the familiarization phase, while the other two were three-syllable "nonwords" containing the same syllables as heard during exposure but arranged differently from their presentation as words (Saffran, Aslin and Newport, 1996). Children controlled the duration of the trials by fixating their gaze on a blinking light. If infants have successfully gathered the essential information from the familiarization materials, they might exhibit varying periods of focused attention (listening) in the two types of tests trials (Saffran, Aslin and Newport, 1996). Results showed that children succeeded in discriminating between the familiarization syllable order and the novel one, as confirmed by longer listening times for the latter. This indicates that the children exhibited a preference or heightened attention towards the novel sequence, suggesting their ability to detect and differentiate unfamiliar patterns or sequences from familiar ones (“novelty preference" or "dishabituation effect”, cf. Saffran, Aslin and Newport, 1996, p.1927). Since the fact that children succeeded in detecting serial order information is not enough to prove that they detected word boundaries, a second experiment was carried out. The second experiment aimed at verifying if babies could distinguish recurrent syllable sequences from strings of syllables that spanned word boundaries, hence sequences of syllables occurring less frequently. To test if infants could differentiate between syllable pairs within words and those spanning word boundaries, an analogy was drawn to English, where 'pretty#baby' represents an internal syllable pair ('pretty') versus an external pair ('ty#ba') (Saffran, Aslin and Newport, 1996, p.1927). As in the previous experiment, the first part consisted of a familiarization phase in which babies were exposed to three-syllable nonsense words in a continuous speech stream. However, the test phase consisted, this time, of two words and two “part-words”. The part-words were formed by combining the last syllable of one word with the initial two syllables of another. Consequently, these constructs comprised three-syllable sequences previously encountered by infants during familiarization but did not correspond to actual words within the dataset. Determining these part-words as new or unfamiliar relied on the infants' ability to have learned the words robustly enough so that sequences spanning word boundaries appeared relatively unfamiliar (Saffran, Aslin and Newport, 1996). Surprisingly, children listened longer to part-words, suggesting that they correctly succeeded in discriminating between word and part-word items. This astonishing result represented a turning point in linguistics, and during the following year, further interesting results provided by other studies strengthened the statistical learning hypothesis, giving it increasingly more credit among scholars.

1.1.2.3 Different types of statistical information: Conditional statistics and distributional statistics

During the 1990s, the early years of statistical learning research, scholars primarily focused on investigating transitional probabilities between syllables, as exemplified by the study conducted by Saffran, Aslin, Newport, 1996). Statistical learning, indeed, emerged from the exploration of speech segmentation abilities in infants (Thiessen, 2017). These studies concentrated on transitional probabilities between syllables, which refer to the probability of co-occurrence between two syllables. Transitional probability indicates the likelihood that an element Y occurs given the presence of an element X. In other words, it represents the probability that X and Y occur together, as illustrated in Figure 1.

$$\mathbf{Y|X} = \frac{\textbf{Frequency of XY}}{\textbf{Frequency of X}}$$

Figure 1. Transitional probability between X and Y. (Saffran, Aslin and Newport, 1996, p.1928, note 12).

Regarding syllable boundaries, the transitional probability from one syllable to the next is higher when the two syllables occur within a word, whereas it is lower when the two contiguous syllables span a word boundary (Saffran, Aslin, Newport, 1996). For example, as discussed in the previous section, given the sound sequence many#children, the transitional probability from 'ma' to 'ny' is greater than the transitional probability from 'ny' to 'chi'.

Subsequent studies have extended statistical learning to other linguistic research areas, demonstrating the extensive range of opportunities within the field. Moreover, scholars have begun investigating the effect of statistical cues on different types of stimuli, such as visual or auditory ones. Statistical learning studies have also been conducted with adults and even with animals. Additionally, other statistical learning mechanisms have started to be explored (Thiessen, 2017). Some scholars have proposed a distinction between different statistical learning mechanisms that learners might employ when dealing with various structures (Thiessen, 2017). Thiessen and colleagues (2013) elucidate that individuals are attuned to a broader spectrum of statistical information that extends beyond conditional relations determined by transitional probabilities. Much of the research in statistical learning has focused on sensitivity to conditional relationships, mainly in the context of word segmentation abilities. However, individuals across various age groups and species also display sensitivity to other types of statistical patterns that cannot be fully explained by conditional relationships alone (Thiessen, Kronstein, Hufnagle, 2013). As these authors highlight, statistical learning can be categorized into three subcategories: conditional statistics, distributional statistics, and cue-based statistical learning. The term statistical learning has been utilized to describe various instances where learners acquire statistical information within the input structure. While these instances fall under the umbrella of statistical learning, a comprehensive explanatory mechanism that adequately addresses the complexity and diversity of all cases has been lacking for many years. Following Thiessen et al. (2013), we will delve more deeply into what conditional, distributional, and cue-based statistics entail.

Conditional statistics reflects the predictive relationship between two elements X and Y. Transitional probability, the most investigated phenomenon within conditional statistics, quantifies the likelihood of event Y occurring given that event X has occurred. High transitional probabilities indicate that X frequently predicts

Y, while low transitional probabilities indicate that X rarely predicts Y. For instance, if event X occurs 100 times and the X-Y sequence happens 40 times, the transitional probability of Y following X is 40%. Unlike simple co-occurrence frequencies, conditional statistics provide a more accurate measure of the strength of the relationship between two events. This is because high co-occurrence frequencies can result from both events being common independently, rather than being predictive of each other. For example, the phrase "the dog" might occur often since both "the" and "dog" are frequent words. However, because "the" can precede many different words, the transitional probability between "the" and "dog" remains low (Thiessen et al., 2013). After Saffran, Aslin and Newport's study (1996), other scholars have replicated the result that babies can discriminate between elements occurring with high transitional probabilities and those less coherent items (Aslin et al., 1998; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). Moreover, as Thiessen and colleagues (2013) elucidate, studies have demonstrated that humans can detect transitional probabilities also between non-adjacent elements, even if the task has proven to be more difficult as compared to the detection of transitional probability between adjacent elements (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004).

Distributional statistics capture the central tendencies and characteristic features of a set of elements by considering both their frequency and variability. These statistics are termed "distributional" because they reflect how learners are attuned to the frequency and variability of examples in the input. A classic example of distributional statistical learning comes from the experiments by Maye et al. (2002), which investigated the impact of phonemic exemplar distributions on infants' discrimination abilities. The study found that infants' ability to differentiate between phonetic categories, such as /d/ and unaspirated /t/, could be influenced by the frequency distribution of these sounds. When infants were exposed to a bimodal distribution, where clear examples of /d/ and unaspirated /t/ were frequent, they were more likely to discriminate between these categories. Conversely, when exposed to a unimodal distribution, where an intermediate sound between /d/ and unaspirated /t/ was most common, infants showed less discrimination between the categories, despite the exemplars being equally present in both training scenarios.

This finding has been replicated for other phonetic distinctions (e.g., Maye et al., 2008). This sensitivity to the frequency of phonemic examples helps explain how infants adapt to the phonemic structure of their native language during their first year (Werker & Tees, 1984). Typically, sounds that are central to a phonemic category occur more frequently than those near the category boundaries (Werker et al., 2007). Another important aspect of distributional statistics is variability. Learners exposed to high-variability distributions are more likely to accept a broader range of examples as belonging to a category, but they also show less certainty when judging stimuli near the category boundary (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008). Conversely, low variability in the input results in sharper category boundaries (Thiessen et al. 2013). Another distributional cue that humans can detect and exploit to form categorical distinctions is the context in which an element occurs. If two similar elements systematically appear in different contexts, humans have a proclivity to represent them as belonging to two different categories (Honey & Hall, 1989; James, 1890; cf. Thiessen, Kronstein, Hufnagle, 2013). As Thiessen and colleagues (2013) discuss, besides phonetic discrimination, distributional statistics has been investigated also in word learning and in the discovery of syntactic structure (Reber & Lewis, 1977). In addition, as for conditional statistics, distributional statistics has proven to be effective in other processes besides language learning, as in auditory perception and object categorization (Rakison, 2004; Younger & Cohen, 1986). Moreover, this ability has been detected also in animals (Lotto et al., 1997).

Cue-based statistics involve the ability of learners to identify perceptual features in the input that signal the presence of certain properties that are not directly observable. This type of statistical learning helps infants discern which perceivable attributes in their environment correspond to attributes that cannot be directly perceived, and how they prioritize certain cues over others. A prominent example of cue-based learning in the realm of statistical learning is the identification of acoustic cues to word boundaries, such as pauses, phonotactics, and lexical stress. For instance, in English, words typically start with a stressed syllable (Cutler & Carter, 1987). By the age of 8–9 months, infants learning English begin to use stress patterns as indicators of word beginnings (Johnson & Jusczyk, 2001). Learners not only detect these cues but also generalize their knowledge to new contexts. Once an infant understands that stress can predict the onset of words, this knowledge is applied broadly. This generalization aspect of cue-based statistical learning is significant because it influences subsequent learning processes, a phenomenon often described as "learning how to learn" (Harlow, 1949; Thiessen & Saffran, 2003; Yerkes, 1943). After learning which perceptual features cue underlying structures, these features shape how learners identify and understand new structures in the future (e.g., Curtin, Mintz, & Christiansen, 2005). Importantly, individuals are adept at recognizing these consistent patterns in cues when the cues operate with a degree of probability rather than absolute certainty (Gratton, Coles, & Donchin, 1992; Thiessen & Saffran, 2007). Interestingly, this learning process is not limited to linguistic domains and extends to various contexts, such as visual and auditory tasks (Thiessen et al., 2013).

While numerous studies have documented how learners are sensitive to various types of statistical relations in their input, these studies often isolate one specific phenomenon—be it conditional, distributional, or cue-based statistical learning. Thiessen et al. (2013) argue that a more holistic approach is necessary to fully understand statistical learning. They propose a memory-based framework that combines two crucial processes: extraction and integration. According to Thiessen et al. (2013), extraction involves identifying statistically coherent clusters of perceptual features, such as word forms, and storing them as discrete representations in memory. Integration, on the other hand, entails comparing these clusters to discern commonalities and the central tendency of the input. The authors argue that neither process alone suffices to explain the full range of statistical learning phenomena. Instead, a combined approach enables a comprehensive understanding of how learners process and utilize statistical information. Extraction, while effective at identifying conditional statistical relations, falls short in explaining distributional and cue-based statistical learning. For instance, clustering models can segment sequences into chunks but do not account for the similarities among these chunks, which is essential for category learning. Similarly, integration models that focus on distributional statistics fail to segment input into meaningful chunks, thus missing conditional relationships critical for language learning. Thiessen et al. (2013) propose that combining extraction and integration processes addresses these shortcomings. Extraction provides a lexicon of candidate words, while integration allows learners to detect central tendencies and generalize across exemplars. This synergy not only enhances sensitivity to conditional relations but also improves the recognition of distributional and cue-based regularities. For example, models of long-term memory that integrate information across multiple exemplars can explain how learners generalize from specific instances to broader categories. This is vital for understanding how infants and adults alike identify phonological regularities and apply this knowledge to word segmentation and other tasks. By integrating extraction and integration, this unified framework accounts for a broader spectrum of statistical learning phenomena. It explains how learners detect conditional relations and leverage distributional information, thus providing a more complete picture of statistical learning processes. This approach suggests that learning is not merely about storing chunks of information but also about understanding the relationships and regularities within that information. Thiessen et al.'s approach also opens new avenues for modeling statistical learning beyond traditional boundaries. By recognizing that learners use both extraction and integration, researchers can better simulate complex learning tasks, including syntactic learning and other higher-order processes. This comprehensive model highlights the importance of considering multiple statistical learning processes in tandem, rather than in isolation, to capture the richness of human learning capabilities. In summary, Thiessen et al. (2013) advocate for a memory-based framework that combines extraction and integration to provide a holistic understanding of statistical learning. This approach not only addresses the limitations of focusing on a single type of statistical learning but also offers a unified explanation for how learners process a wide range of statistical information.

1.2 Evidence for structure-dependent abstract representation and statistical learning at the syntactic level

Numerous studies have showcased the capacity to utilize statistical information in acquiring various linguistic phenomena. Nonetheless, limited attention has been given to the potential acquisition of syntactic-level information, such as phrase-structure grammar, through the application of statistical mechanisms. This point likely represents the main dividing line between nativists and proponents of the usage-based theory. Nativists argue that there is not much to be acquired regarding syntax. The operations Merge and Agree, according to this view, are provided by Universal Grammar (UG) (Chomsky, 1995), while the rest is attributed to external factors, i.e., the "third factor" (Chomsky, 2005). The concept of *third factor* refers to influences external to Universal Grammar, such as principles of cognitive economy or constraints arising from the structure of the human brain. However, there is no clear consensus on what the "third factor" includes. According to Johansson (2013), it may encompass principles of data processing, economy of derivation, interface conditions, general cognitive capacities, architectural and computational constraints, developmental constraints and canalization in embryology, physical and mathematical laws. These factors would contribute to the final shape of syntax without being an intrinsic part of innate grammar or specific linguistic experiences. On the other hand, proponents of the usage-based theory argue that exposure to linguistic material might bootstrap the development of phrase-structure grammar, rejecting the necessity of an innate universal grammar. They acknowledge that grammar development can be influenced by various cues present in linguistic input, such as semantic, morphological, pragmatic, or phonological cues. Importantly, however, distributional information between words or word classes is also considered to play a significant role in grammar development (Redington, Chater, 1998). Despite these insights, the possibility to acquire phrase-structure grammar remained for several years one of the most contentious and controversial issues in language acquisition (Redington, Chater, 1998; Kidd, 2012). This is undoubtedly partly attributable to the scarcity of studies investigating this phenomenon, both in the field of psycholinguistics and in connectionist studies. Indeed, initial psycholinguistic studies in Statistical Learning (SL) primarily explored speech segmentation abilities in infants. As discussed earlier, a notable work in SL by Saffran, Aslin, and Newport (1996) revealed the remarkable ability of 8-month-old children to detect transitional probabilities between syllables. Further studies have extended their focus beyond phonology to investigate phenomena like morphology and vocabulary acquisition. Additionally, apart from the initial studies involving infants, subsequent research has also explored the abilities of adults. However, the great majority of statistical learning studies have focused on the investigation of early-stage language acquisition phenomena (i.e., speech segmentation and lexicon formation). On the other hand, very few works have analyzed the role of statistical learning in the acquisition of syntax, even if, many scholars proposed that statistical regularities may play a role at this level as well.

The same observation applies to investigations using neural networks. During the first years of research in the field of connectionism, apart from the early works by scholars in computational linguistics and machine learning aimed at solving various practical issues, particularly those involving linguistic corpora, very few studies focused on the acquisition of syntax. Only some years later, scholars within the field of cognitive science began to explore this issue by designing computational models based on distributional information to hypothesize how the child may acquire syntactic categories (Redington et al. 1998). Indeed, cognitive scientists started to see in neural networks an interesting tool for the investigation of language acquisition. That could have happened, certainly thanks to, on one side, the rapid development of computer technology in the previous years, which provided cognitive scientists with new powerful computational techniques and technological tools. On the other side, thanks to the important discoveries in neuroscience concerning brain structure and functioning. Indeed, with Parallel Distributed Processing, scientists aimed at creating computers that were directly inspired by the brain structure (Christiansen, Chater, 2001). Since the very beginning, connectionist approaches to the study of language acquisition have been sharply criticized and remained for several years highly controversial (Christiansen, Chater, 2001). Despite that, connectionist approaches have provided several interesting results in the investigation of different subfields of language acquisition, from low-level acoustic discrimination to words segmentation, and even morphological processing, as we have seen with the well-known case of Rumelhart and McClelland's neural network for past tense. By contrast, two areas have been less investigated: sentence processing and speech production, constituting for years a considerable challenge for the connectionist approach (Christiansen, Chater, 2001). Despite the long-standing line of connectionist research, some issues have not been fully addressed for many years. This fact is a direct reflection of the complexity of human language. "[…] Progress in this area has been much slower than in most other learning tasks, undoubtedly due to the inherent complexity of natural language" (Langley, 1987, p.5).

On the other hand, there has always been almost unanimous consensus on the fact that, at the syntactic level, all languages function in a structure-dependent way. "[…] a structure-dependent operation is one which is based on the abstract structural organizations of word sequences. By contrast, structure-independent operations apply to sequences of words themselves, and include operations like NEXT and CLOSEST which are contingent on linear order" (Crain & Nakayama, 1987, p.522). Children never utter structure-independent sentences: phrase construction and movements do always respect structure-dependent rules. This is the case, for example, of auxiliary fronting in English polar interrogatives (Boeckx & Hornstein, 2004; Crain & Nakayama, 1987; Crain & Pietroski, 2001; Legate & Yang, 2002). Nevertheless, despite the significance of structure-dependence theories in language processing and acquisition, critical psycholinguistic experiments testing the role of structure-dependence in language were notably lacking until at least 1987. Indeed, the first study that thoroughly investigated this issue was carried out by Crain & Nakayama, in 1987, testing auxiliary fronting phenomena.

Hence, it is intriguing to ascertain whether a bias toward structure indeed exists during the acquisition of syntactic-level phenomena. Additionally, it is crucial to comprehend whether statistical learning also plays a role in acquiring the fundamental mechanisms that underlie syntactic structure. In recent times, noteworthy psycholinguistic studies have presented significant evidence supporting the existence of structure-dependent constraints in the acquisition of syntactic-level phenomena (Coopmans et al., 2022; Culbertson & Adger, 2014; Lidz et al., 2003; Martin et al., 2019; 2020), while also furnishing compelling evidence of the capability to process statistical information at the syntactic level (Gerken et al., 2005; Gomez, 2002; Kidd, 2012; Saffran, 2001; Saffran, Wilson, 2003; Thompson and Newport, 2007). Moreover, significant and intriguing results concerning the possibility of acquiring syntax have emerged from computational language models, which have made tremendous strides, especially in recent years. In the following sections, we will provide an overview of the most important results obtained in recent years from these two different perspectives: the psycholinguistic and computational branches.

1.2.1 Evidence for structure-dependent abstract representation

Crain & Nakayama (1987) were the first to experimentally test auxiliary fronting phenomena by carrying out three experiments in which they investigated movement transformation, specifically the inversion between subject and AUX in sentences with relative clauses.[5] In the first one, they elicited the production of yes/no questions from children (age: 3 to 5)[6]. The second one aimed at analyzing the type of errors children made in the first experiment. In the last one, they compared the acquisition of interrogatives based on a structurally-based account with the acquisition based on semantic generalization. The result of their experiment strongly supported the hypothesis according to which children learn and process a language by inferring structure-dependent rules, regardless of the computational complexity of sentences. In other words, the rules that children hypothesize do not

[5] Example of sentences with relative clause (i), the relative structure-dependent (ii) and structure-independent (iii) interrogative forms:

(i) The man who is tall is in the other room.

(ii) Is the man who is tall __ in the other room?

*(iii) *Is the man who __ tall is in the other room?*

(Crain & Nakayama, 1987 p.525)

[6] Example of sentences from Crain & Nakayama's elicitation task: *"Ask Jabba if the boy who is watching Mickey Mouse is happy".*

apply to sentences as linear strings of words. Children's computational hypotheses are based on abstract structures, which can be detected by recognizing the internal structure of sentences in which words are seen as terminal symbols (Crain & Nakayama, 1987).

Further confirmation that language processing occurs in a hierarchical manner has been proven by several recent psycholinguistics studies (Coopmans et al., 2022; Culbertson & Adger, 2014; Lidz et al., 2003; Martin et al., 2019; 2020). In addition to the case of AUX inversion, another interesting phenomenon to test the presence of structure-dependent rules is that related to word order. The arrangement of words within sentences could either be learned through linear statistical surface mechanisms or reflect the hierarchical structure in which words are organized. If the latter hypothesis is true, the linear word order would be the result of the linear one-dimensional transposition of an n-dimensional structure that encodes the relationships these words have with each other.

A particularly interesting phenomenon is related to the ordering of nominal modifiers. Greenberg's Universal 20 (Greenberg, 1963), as restated in Cinque (2005) is as follows:

> *In prenominal position the order of demonstrative, numeral, and adjective (or any subset thereof) conforms to the order Dem > Num > A; in postnominal position the order of the same elements (or any subset thereof) conforms either to the order Dem > Num > A or to the order A > Num > Dem.* (Cinque, 2005)

Universal 20 has been attested through various typological studies and supported by controlled samples of 576 languages (Dryer, 2018). However, little had been said about the reasons underlying this phenomenon. Indeed, this phenomenon could be dictated by historical, cultural causes, or instead reflect a cognitive constraint (Culberton, Adger, 2014).

Culbertson & Adger (2014), by carrying out two experimental studies, tested an intriguing hypothesis, according to which Dem-Num-Adj-N and N-Adj-Num-Dem are most common orders because they maintain an isomorphism between scope and surface order. On the other hand, orders like N-Dem-Num-Adj and Adj-Num-Dem-N are non-isomorphic, which may explain their rarity or absence in human languages (Culbertson, Adger, 2014). In experiment 1, they presented participants with examples from an artificial language by randomly assigning them to four conditions: each condition consist of a specific type of postnominal noun-modifier (i.e., N-Adj; N-Dem; N-Num). Importantly, there were no phrases with more than one modifier. During the testing phase, they had to infer the relative ordering of noun phrases with more than one modifier, by choosing from binary option. In this way, they tested the two hypotheses: if participants infer the order relying on semantic scope relations, they should choose N-Adj-Dem, N-Num-Dem, and N-Adj-Num. Indeed, these orders are isomorphic to the semantic scope. On the opposite, if they chose the order based on surface transitional probabilities, English participants should prefer N-Dem-Adj, N-Dem-Num, and N-Num-Adj, given that the order between the modifiers reflect the most frequent one in English. Results showed that participants choose the isomorphic order over the one that was more similar to English surface linear order, in all three conditions. Interestingly, however, the effect was more marked in the Dem-Adj case than Dem-Num or Num-Adj. The authors suggest that this result might reflect the structural distance between modifiers: the semantic scope relation manifests itself more evidently when the structural distance between modifiers is greater. In experiment 2, the authors extended the investigation to the whole phrase. Indeed, according to the structure-dependent hypothesis of processing, participants should infer a scope-isomorphic order in the noun phrases with all three modifiers as well. They trained hence participants with all three types of modifiers. Again, results showed a preference for scope-isomorphic order. In conclusion, Culbertson and Adger's experiments showed that learners consistently preferred the order that maintained the semantic scope relation between modifiers over the one which was linearly more similar to English (the linear order of modifiers was identical to English, but in a postnominal position).

Martin et al. (2020) wanted to go deeper and shed more light on the result found in Culbertson & Adger (2014). Indeed, these results could potentially have been caused by a metalinguistic strategy used by the participants that may have led them to choose sequences that complied with the semantic scope without, however, having access to an abstract representation of the constituent structure. In fact, the participants may have arrived at the result by using a strategy of flipping the order of the modifiers from that of their L1 (Martin et al., 2020). Then, through three experimental studies they aimed at verifying whether the preference for orders respecting the hierarchy was still found when participants do not use metalinguistic strategy of flipping. In Experiment 1, to reduce the possibility that participants might visually flip the word order of their L1 by applying it to the artificial language they are learning, they used new stimuli that had no correspondences in the participants' L1. In addition, all words and phrases were presented both auditorily and orthographically. After conducting a noun training phase and a modifier training phase to ensure that the participants had learned the meanings of the words in the artificial language, the participants were taught phrases with a noun and a single modifier. Each of the participants then learned two types of modifiers, depending on the condition they were randomly assigned to (either Adj and Dem, or Num and Dem, or Adj and Num). In the testing phase, participants had to guess the order of the modifiers when both were present. In this phase, a picture appeared at each trial with two descriptions under it. Participants were asked to click on the corresponding description. Both descriptions always included the correct lexical items in post-nominal position. However, only one of them had homomorphic order (e.g., N-Adj-Dem vs. N-Dem-Adj). The results showed a preference for homomorphic order only in the Dem-Adj condition (e.g., N-Adj-Dem order preferred over N-Dem-Adj). However, no preference is found in the Dem-Num condition nor in Num-Adj (i.e., no preference between N-Adj-Num and N-Num-Adj orders, nor between N-Num-Dem and N-Dem-Num orders). Among the possible causes for this result could have played a role (i) the type of the task: having to choose instead of producing the phrases, participants may have been less likely to fully acquire the lexical items and create a mental representation of the meanings. This would imply a weaker influence of the underlying structure or greater focus on linear order; (ii) having given participants both available options in the testing phase may have weakened what would otherwise have constituted a stronger bias. In Experiment 2, therefore, instead of using the forced-choice task in which participants must choose between two orthographically presented options, they chose to use a production task where participants were required to produce the phrases verbally through a microphone. In addition, the stimuli, which were created using the same artificial language as in Experiment 1, are presented here only orally. The procedure is the same as in Experiment 1 but readjusted to oral production. Thus, in the testing phase, images representing nouns combined with two modifiers are presented and participants are asked to verbally produce the corresponding syntagma. The results are interesting: in line with what was found in Experiment 1, a preference toward homomorphic ordering in the Dem-Adj condition is reconfirmed. However, a preference toward homomorphic order is also found here in the Dem-Num and Num-Adj conditions. The hypothesis that the production task would have reinforced the preference toward homomorphism is thus confirmed: the fact that the participants had to orally produce the words would have led them to acquire the lexical meanings at a higher level and thus to have a better representation of the meaning or syntactic categories of the items. Having to verbally produce the phrases, moreover, might have encouraged the mental formation or activation of hierarchical representation, which in turn would have led to a stronger preference toward homomorphic structures. In Experiment 3, participants are asked to verbally produce phrases even during the training phase, unlike in Experiments 1 and 2. The results are in line with those of study 2: indeed, there is a preference for homomorphism in all three conditions. Wrapping up, the difference between the results of study 1, on the one hand, and those of studies 2 and 3, on the other hand, would seem to be dictated by task type. Moreover, through post-hoc analysis it is reconfirmed that the preference for homomorphism is more pronounced for the Dem-Adj combination than for the other combinations. This result is again interpreted as evidence that Dem and Adj are structurally more distant from each other than Dem-Num or Num-Adj. This hypothesis is based on the assumption that the entire hierarchical structure of the nominal phrase is present or activated in the mind of the speaker/listener even when one of the categories is not expressed. This would be in line with what is assumed by the theory proposed by Cinque (2005). However, other explanations could underlie the phenomenon, which would not imply the abstract representation of the whole phrase structure every time a nominal sentence is produced: for example, the differences in the strength of associations found could be a reflection of the semantic or conceptual structure

underlying the syntactic hierarchical representation. In other words, a reflection of the way objects relate to their properties, numerosity, etc. in the world, not just in linguistic expressions, as proposed by Culbertson, Schouwstra & Kirby (2020).
Coopmans et al. (2022) carried out an interesting experimental work on word order in noun phrases as well. They conducted two behavioral experiments to test whether participants interpreted noun phrases based on their internal abstract hierarchical structure rather than linear order. Then, they trained a neural network on the same task they used in the behavioral experiment to check if the network would have shown the same or a different behavior of participants. Experiment 1 aimed to investigate how participants interpret noun phrases containing ordinals, color adjectives, and nouns referring to the shape of a target object. Two conditions were used: convergent and divergent. In the convergent condition, the hierarchical (nonintersective) and linear (intersective) interpretations of the noun phrase led to the same answer. For example, "the second blue ball" could refer to both the second among blue balls (hierarchical) and the ball that is blue and in the second position (linear). In the divergent condition, the hierarchical and linear interpretations of the noun phrase led to different answers. For instance, "the second blue ball" might be blue (linear interpretation) but not the second among blue balls (hierarchical interpretation). Results showed that participants predominantly gave hierarchical answers in response to divergent trials. However, this result has to be taken carefully: indeed, an alternative interpretation that did not rely on constituent structure could have caused the result. This interpretation involved considering "second blue" as a complex adjective applied to the noun "ball" (e.g., a ball that is second among blue items). This alternative approach (left-branching) yielded the same target as the hierarchical interpretation, but it did not necessarily require hierarchy, as opposed to the right-branching interpretation. In the right-branching 'hierarchical' structure depicted in Figure 2A, there exists a connection between the element 'second' and a constituent, indicating modification of the constituent. However, such a constituency-based relationship is unnecessary for representing the meaning of the left-branching structure illustrated in Figure 2B (Coopmans et al., 2022).

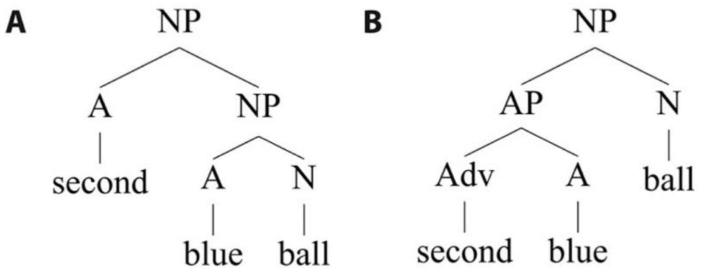


Figure 2. Representations of the phrase 'second blue ball' can be depicted in both right-branching (A) and left-branching (B) structures. Figure taken from Coopmans et al., 2022, p. 425.

Hence, to distinguish between the two interpretations, a second experiment was conducted. Experiment 2 was similar to Experiment 1, but with the addition of blue and green triangles to the array of items. This addition introduced two shapes, making the noun (ball or triangle) crucial for identifying the target. Each trial now contained two potential targets for phrases like "second blue ball." Both right-branching and left-branching interpretations were always available in each trial but never converged on the same item. Results showed that participants were significantly more likely to use right-branching interpretations than left-branching interpretations. Hence, taken together the two experimental studies strongly support the significance of hierarchical constituent structure for semantic interpretation in the context of the given noun phrases. In Experiment 3 they trained and tested a long short-term memory (LSTM) model on a computational version of the behavioral experimental task involving noun phrases. Crucially, the model demonstrated the ability to give hierarchical answers when trained on unambiguously hierarchical datasets. However, when the training data contained both unambiguously hierarchical and ambiguous trials, the model strongly favored the linear interpretation, even though the hierarchical interpretation was a better fit for the overall data. Moreover, the model did not systematically generalize to novel items not seen during training, unlike human participants who showed a bias towards interpreting language hierarchically despite ambiguous input data. Summing up, Coopmans et al. (2022) found that the model's behavior differed significantly from that of humans, who seem to have a bias to interpret language in accordance with its underlying hierarchical structure, even when the input is ambiguous. They concluded by suggesting that the model would require different inductive biases to achieve human-like generalization.

1.2.2 Psycholinguistic evidence for the ability to track statistical regularities at the syntactical level

In this section, we will examine several studies that have presented evidence regarding the capacity to process statistical information at the syntactic level. As we delve into the section, it becomes apparent that recent scholarly interest has turned towards investigating this matter, encompassing evaluations of sensitivity to both frequency and transitional data within the realm of syntax. Specifically, the studies have considered certain phenomena as indicative of statistical markers for syntax acquisition. These include the frequency of syntactic structure (Kidd, 2012), transitional probabilities between words to segment phrases (Thompson and Newport, 2007), dependencies to discern phrase boundaries (Saffran, 2001) utilization of low-level statistical output (word segmentation) as input information for computing higher-level phenomena (syntax) (Saffran, Wilson, 2003), distributional cues for constructing syntactic categories (Gerken et al., 2005) and transitional probabilities between adjacent and nonadjacent dependencies (Gomez, 2002).

Saffran (2001) investigated if participants, adults, and children, would have succeeded in exploiting predictive relationships as a cue to have access to the hierarchical phrase structure of an artificial language. With this study, Saffran (2001) aimed at shedding light on the extraordinary human ability to access the hierarchical organization of language given linearly ordered sentences as an input. How do humans manage to move from linear order to hierarchical structure? Which cues might help them in this process? It is possible that dependencies could act as a statistical hint for identifying phrase boundaries. Dependencies have not been thoroughly investigated as indicators for recognizing phrasal units that might not be explicitly indicated in the input. When learners encounter dependencies in the input, they might naturally combine related elements into phrases, even if there are no other clear indicators typically associated with phrase boundaries (Saffran, 2001). Resting upon these observations, Saffran (2001) tested this hypothesis in two artificial grammar learning experiments, the first one with adults, and the second one with children (from 6 to 9 years old). The first experiment aimed to investigate whether learners could discern phrase structure in language input solely based on predictive dependencies between form classes, without other cues. Participants were exposed to sentences from an artificial language, where phrase structure rules were governed by the distribution of words across categories, devoid of any semantic, prosodic or referential cues. Hence, the artificial language used contained consistent predictive dependencies as the primary cue to phrasal units. One difference from natural language, however, is that the predictive dependencies crucial for uncovering phrases were conveyed simply through adjacent form categories. This simplicity contrasts with what is usually seen in natural languages. Participants were divided into two groups. The first group underwent intentional learning (i.e., intentional condition), hence they were explicitly instructed about the grammatical rules; the second group underwent incidental learning (i.e., incidental condition), meaning their exposure to the artificial language occurred passively while they performed an interfering task. This setup provided an opportunity to contrast the results of implicit and intentional learning, aiming to determine if the mechanisms supporting this statistical learning mechanism can function similarly to the exposure-driven learning seen in natural language development. Participants listened to a tape of multiple legal sentences for four times (30-minutes session). This training procedure was repeated two times, in separated sessions. Subsequently, they underwent three distinct forced-choice tests. One forced-choice test (i.e., Rule Test) was administered right after each of the two listening sessions to evaluate the participants' understanding of the generalizations over form classes that produced the input. Participants were presented with two new sentences: one adhering to the language's rules and the other violating them. Participants were instructed to designate which of the two sentences was more acceptable based on the language they were previously exposed to. The Rule Test offered insights into how well participants grasped the language's structure. However, understanding phrase structure was not necessary for them to perform well. For instance, participants could have scored above chance without needing to understand the phrase structure, but simply recognizing that a word belonging to a specific category never followed a word belonging to another specific different category. Subsequently, another forced choice task (i.e., Fragment Test) was administered to directly evaluate learners' representation of input in terms of phrase groupings. Each trial presented two sentence fragments: one forming a phrase and the other spanning a phrase boundary. The hypothesis posited that successful grouping of input strings into phrases would render phrase fragments more natural than non-phrase fragments. A control group was included in the study, comprising participants who underwent the three forced choice tests without prior exposure to the language. This measure aimed to ascertain that any performance above chance among the experimental participants stemmed from learning, rather than potential biases in the test materials. Saffran (2001) carried out a second experiment to evaluate the existence of these learning mechanisms among children, using the same materials and procedure as that in the first experiment with adults. Results of the first experiment confirmed the ability among adult learners to identify phrasal units even when lacking explicit cues beyond predictive dependencies. Their performance on the rule tests suggests a comprehensive acquisition of information, not only regarding the occurrence of individual categories but also concerning the more complex conditional rules that dictate relationships between these categories. Moreover, the predictive relationships established between form classes within the experiment played a pivotal role in facilitating the statistical learning process associated with phrasal groupings. In addition, it is noteworthy that the experimental group exhibited superior performance compared to the control group in the Fragment Test. Remarkably, participants achieved these results within the incidental paradigm, indicating that learning occurred without them consciously intending to learn. This suggests that they were able to acquire the language structure even though their primary focus was not on learning it directly. In summary, the results of the first experiment provided strong evidence for the fact that adults managed to learn the hierarchical structure of the artificial language by relying on predictive dependencies, as these structures were the only one cue that had been inserted in the artificial language and that could have been exploited as a statistical cue by participants to detect phrasal units. Results from the second experiment indicated that children managed to acquire some rudimentary aspects of the structure of the language as well, but the result was weaker as compared to adults' performances. Overall, adults outperformed children on all the tests. In general, the results of Saffran's study represent an important piece of evidence for the fact that linguistic input contains statistical information that represents a goldmine for learners. Indeed, the types of statistical cues that are present in language have been revealed to be strictly linked to the kind of statistical abilities that are in the faculty of humans' computational possibilities. Saffran (2001) clarified that these findings uphold the idea that human learning mechanisms can derive hierarchical structures from statistical connections between form classes, through implicit learning processes. The extent to which these mechanisms are specifically adapted for language acquisition, or if they arise from broader cognitive and perceptual properties, poses an important question for further empirical investigation (Saffran, 2001).

Saffran and Wilson (2003) carried out the first experiment in which they tested in a single experiment the abilities to track statistical information at different linguistic levels. Indeed, as Saffran and Wilson (2003) explained, up to that moment, experiments on statistical learning have focused on analyzing one single phenomenon at a time, investigating, for example, speech segmentation abilities, word category formation, lexical acquisition, or syntactic learning. No experiments, up to that moment, had investigated two or more phenomena at the same time, using a single paradigm. However, as Saffran and Wilson pointed out, in a natural language environment, learners manage to acquire and process different levels of the linguistic structure under the same circumstances. Hence, the question arose: do learners use the output of lower-level statistics as an input to access higher-level linguistics aspects? They tested this hypothesis in 12-month-old infants by exposing them to artificial utterances, whose words were ordered based on a finite-state grammar and were presented as a continuous stream. The results were impressive: children, after only a few minutes of exposure, first managed to segment words from the continuous stream, and subsequently, they also successfully discovered the ordering relationships between words. In other words, they managed to track, in a first moment, the transitional probabilities at syllables level and then, distributional statistics between words. From this result, the authors concluded that the learning processes being examined are both highly effective and meticulously crafted to address the challenges encountered by individuals learning human language (Saffran, Wilson, 2003).

Gomez (2002) investigated children's and adults' abilities to track nonadjacent dependencies. The ability to deal with nonadjacent dependencies is at the core of the faculty of language. Indeed, several linguistic phenomena, especially at the syntactic level, are expressed through relationships between linguistic elements that are not based on sequential proximity but on hierarchical dependencies. These phenomena manifest themselves through long-distance dependencies. Nonadjacent dependencies seem to be particularly challenging to learn and process, as they require tracking connections between non-contiguous elements, bypassing intervening nonrelevant linguistic material. Example of long-distance dependencies in natural language are provided by auxiliary-inflectional morphemes relationships (e.g., Laura is walking; Marc has arrived) and number agreements (e.g., the flowers in the garden are yellow; the girls in the picture are Sara and Mary) among the other (Gomez, 2002). The author exposed adults and infants to sets of artificial sentences composed of three non-words each, generated by two miniature artificial languages. The two languages contained the same adjacent dependencies but different dependencies between the first and the third elements. In other words, both languages shared identical adjacent dependencies, requiring learners to discern between them solely through acquiring nonadjacent dependencies, which related to the connection between the first and third elements. Each language generated sequences like "vot-kicey-rud or "pel-wadim-jic". Despite starting and ending with the same words and featuring identical pairwise adjacent transitions, the languages diverged in the dependencies between their first and third elements. Consequently, learners could solely discern between the languages by grasping the nonadjacent dependencies. To further explore this, Gomez (2002) manipulated the context variability by systematically expanding the pool from which the middle element was selected. This experimental setup allowed to examine two conflicting hypotheses: If learners incorporate lower-order dependencies within higher-order ones, exposure to smaller sets should enhance the learning of nonadjacent dependencies due to the higher transitional probabilities between adjacent elements in smaller sets compared to larger ones. Conversely, if increasing variability prompts learners to concentrate on other sources of consistent structure, sensitivity to nonadjacent dependencies should increase with set size. In the adult experiment, participants underwent training sessions during which they listened to auditory sequences generated by one of two artificial languages. Throughout this training phase, learners were directed to focus on the sequences they heard. Before the testing phase, participants were informed that the sequences heard during training followed particular rules regarding word order. During the subsequent testing phase, participants were presented with new strings. Half of these strings adhered to the word order rules encountered during training, while the others did not. Participants were directed to press particular keys on a keyboard indicating whether they believed a string followed the word order rules from their training or not. In the experiment involving infants, the stimuli mirrored those used in the adult experiment, with the exception that infants were trained on two nonadjacent dependencies instead of three. Additionally, the head-turn preference procedure was employed. The duration an infant spent oriented toward the test stimulus was measured. A notable contrast in listening time between trained and untrained strings would suggest that infants developed some sensitivity to the nonadjacent dependencies outlined by their training language. Interestingly, Gomez (2002) found that neither adults nor infants showed evidence of integrating lower-order dependencies into higher ones. Similarly, there was no gradual improvement in discrimination as the set size increased. Instead, discrimination spiked notably only with the largest set size. The inability to discriminate among smaller set sizes, coupled with the sudden surge in discrimination with the largest set size, suggests that both infants and adults initially processed adjacent dependencies by default. They shifted their attention to nonadjacent dependencies only when the former proved unreliable enough. These findings are significant as they demonstrate that even very young learners can adapt to the informational demands of their learning environment, displaying a considerable level of adaptability in their learning process. Hence, this important result suggests that the statistical information that is present in the linguistic structure strongly influences learners' behaviors. As Gomez (2002) elucidate, these findings highlight the fact that specific circumstances could influence learners to prioritize nonadjacent rather than adjacent dependencies, shedding light on how learning might be influenced by statistical patterns dynamically. Learners are inclined to identify consistent patterns within the stimulus array. When transitional probabilities are high, adjacent elements are perceived as stable. However, in scenarios where high variability disrupts adjacent dependencies, learners will look for alternative sources of predictability. These results align with the perspective that humans are proactive and adaptable learners, eager to exploit any regularities they encounter (Gomez, 2002).

Gerken et al. (2005) investigated children's ability to form syntactic categories by relying on distributional cues in the input. As the authors clarify, the ability to understand and form an unlimited number of syntactically correct sentences is founded upon the ability to form discrete syntactic categories. With this experiment, Gerken and colleagues aimed to verify whether humans could accomplish this task by simply relying on distributional cues in the input. In natural language, several cues, such as morphological markers indicating the word's position in the sentence and phonological properties, could provide humans with useful statistical information, which can be exploited for correct syntactic categorization (Gerken et al., 2005). The idea that distributional cues alone are sufficient to form syntactic categories is called the distributionally-based category formation hypothesis. On the other hand, some scholars strongly believe that distributional cues alone are not sufficient, and that semantic information is necessary for correct categorization (i.e., semantic bootstrapping hypothesis). Gerken and colleagues explored the distributionally-based category formation hypothesis. They tested American infants (1.5 years old) in three experiments, in which they were exposed to a series of artificial words with gender morphological markers present in Russian. In the three studies, using the head-turn preference procedure, they investigated infants' ability to discriminate grammatical from ungrammatical test trials, with the final goal of testing children's ability to form syntactic categories relying on morphological cues of gender. This paradigm was chosen for several reasons. Firstly, Russian gender is a rich and complex system that children actually acquire, making it intriguing to see if infants could grasp some of its complexities in a short lab session. Secondly, they aimed to introduce infants to a linguistic category absent in English. Lastly, previous studies have shown that American adults can learn this paradigm when exposed to stimuli with partially correlated cues, providing a benchmark for comparison with infants' learning behavior. They developed a Russian gender paradigm featuring masculine and feminine lexical stems, each paired with two distinct case endings. Certain words in both genders were termed 'double marked' as both their stem endings and case inflections conveyed gender information. In contrast, other words were 'singly marked,' indicating that they were only distinguished by the case inflection. Experiment 1 involved two groups of infants: one group was exposed to the Russian gender paradigm before the test (i.e., familiarization phase), with the expectation that they would distinguish grammatical from ungrammatical test trials. The second group received no prior exposure and served as a control. During the test phase, both groups encountered grammatical and ungrammatical words, with the latter created by applying incorrect case endings while maintaining the same format as the grammatical trials. Results were in line with their hypothesis and indicated that only familiarized infants were able to distinguish grammatical from ungrammatical Russian words. Experiments 2 and 3 followed the same procedure and were carried out to account for potential phonological artifacts that might have influenced the results (Experiment 2) and to shed more light on the difference between the double marked condition and the single marked condition (Experiment 3). Experiment 2 addressed two potential phonological issues identified in Experiment 1. Despite this, it still replicated the effect of infants being able to distinguish between grammatical and ungrammatical items, albeit to a lesser extent compared to what was found in Experiment 1. Experiment 3 involved testing two groups of infants. The first group underwent the same familiarization and testing procedure as the infants in Experiment 2, potentially enabling a replication of those findings. The second group underwent the same testing process but was exposed to stimuli lacking double marking of gender categories. Consequently, the only gender indication stemmed from the two feminine and two masculine case inflections. Based on prior findings indicating that adults rely on double marked stimuli to infer word category, Gerken and colleagues hypothesized that only infants exposed to the double marking condition would successfully differentiate between grammatical and ungrammatical test items. The outcomes from Experiment 3's double marked condition strongly mirror the effect observed in Experiment 2. Conversely, infants in the single marked condition were unable to differentiate the same test stimuli that were successfully distinguished by infants in the double marked group. The complete absence of any hint of a grammaticality effect in the single marked condition underscores the idea that infants, similar to adults in prior studies, show a greater ability to grasp the categorical structure of a paradigm when multiple cues to that structure are available within a subset of the items they are exposed to during familiarization. In summary, Gerken et al. (2005) demonstrated that 1.5-year-old children can discriminate novel grammatical words from non-grammatical ones when exposed to a partial Russian gender paradigm for about two minutes. Importantly, however, children only showed the ability to discriminate grammatical words when some of the familiarization stimuli contained two category cues. This suggests that children may use a combination of distributional cues to learn the structure of syntactic categories. Their results are in line with previous findings exploring this ability in adults, which, when exposed to paradigms containing two category cues, exhibited similar behavior. Importantly, as the authors pointed out, the cues tested reflect genuine features found in the Russian language. Consequently, it suggests that infants may be able to utilize similar cues in real-world language acquisition settings. Overall, Gerken and colleagues' results challenge the Semantic Bootstrapping hypothesis, arguing that referential information is not necessary to form syntactic categories, while, on the other hand, distributional information is enough.

Thompson and Newport (2007) conducted an experiment with a miniature artificial language to test whether transitional probabilities between words play a role in the segmentation of phrases. Their goal was to verify whether this process would ultimately lead to the acquisition of grammar. As elucidated by the authors, up to that moment, transitional probabilities had already been found to play a role in the acquisition of phonology (Maye, Werker, & Gerken, 2002), in the segmentation of words (Newport & Aslin, 2000, 2004; Saffran, Newport, & Aslin, 1996), and even in the formation of word categories (Hunt, 2002; Mintz, Newport, & Bever, 2002). Thompson and Newport (2007) extended these investigations by exploring whether transitional probabilities would be an exploitable cue even at the level of syntax acquisition. The hypothesis was that the same small range of statistical effects might be used by humans to deal with different aspects of language at different levels, from speech segmentation to syntax acquisition. To investigate the role of transitional statistics, they took into consideration all those phenomena that in natural language create peaks and dips within phrases and at their boundaries. They carried out 4 experiments with adults. In the first experiment, they investigated the role of optional phrases in determining transitional probabilities between phrases. Participants showed evidence of sensitivity towards this distributional information, and they succeed in learning the phrases and word order of the artificial language. In the following experiments, they added other phenomena that are naturally present in natural language and that would provide participants with additional distributional information: they tested whether participants would have been sensitive to the transitional probabilities generated by the presence of moved phrases, word classes of different sizes and repeated phrases. Moreover, in experiment 4, they pushed even further by increasing the complexity of the language used in experiments 1 and 2, and the result was that participants performed even better in this condition, demonstrating their ability to deal with complex statistical information at the level of syntax. Taken together, the results of these experiments show that humans can exploit transitional probabilities between words to form phrases. Importantly, this operation seems to lay at the basis of the discovery and learnability of the input structure. Hence, Thompson and Newport’s results provided brand-new evidence for the fact that the same kind of transitional probabilities that are at work during low-level language acquisition can be exploited to acquire higher-order levels of language. Furthermore, these results confirm the hypothesis that the rich complexity of natural language structure might facilitate learners during the non-trivial task of language learning by providing them with important, exploitable information. The authors concluded that further experiments should verify whether the statistical abilities deployed for the acquisition of syntax in artificial languages among adults are also those at work in the process of natural language acquisition by infants. Moreover, they underlined the fact that, despite having demonstrated that distributional information constitutes an important cue for the acquisition of syntax, other kinds of computations might be necessary to handle the full richness of natural language structures, suggesting that future research should address this issue (Thompson, Newport, 2007).

Kidd (2012), noticing the discrepancy between the number of theories according to which statistical learning would play a role in the acquisition of syntax and the number of studies that provided empirical evidence in favor of that, carried out a syntactic priming experiment to investigate children's abilities to acquire grammar. The aim of the study was to examine the correlation between implicit statistical learning and the acquisition of syntax in children. The study investigated individual differences in various cognitive abilities, involving 100 children aged 4 to 6. The children completed tests of implicit statistical learning, explicit declarative learning, and standardized tests of verbal and non-verbal abilities testing vocabulary, nonverbal IQ, and declarative memory. Additionally, they participated in a syntactic priming task that provided a dynamic index of their ability to detect and respond to changes in the frequency of linguistic input. The final goal was to ascertain whether there were correlations between these abilities, specifically between implicit and explicit learning abilities and syntactic priming. Building on earlier research, Kidd (2012) predicted a direct connection between implicit learning skills and syntactic priming. Specifically, the author expected to find a direct relationship between implicit statistical learning and the long-term effects of priming. Conversely, no association between explicit learning abilities and syntactic priming were anticipated. Implicit statistical learning abilities were investigated through a Serial Reaction Time (SRT) task in which the ability to learn repeating 10-item sequence patterns was tested. In this task, children were presented with a single visual stimulus that moved between four spatial locations on a computer screen. The task required participants to press buttons corresponding to the location of the visual stimulus, and their reaction times were measured. The sequence was presented over four blocks, with the final block presenting the visual stimulus in a random order. The primary dependent variable of interest was the participants' reaction times, with decreases in reaction times in non-random blocks indicating implicit learning. The syntactic priming task aimed to assess the priming effect on the production of full be passive sentences in English. As the author explain, this construction was selected because it has proven highly effective in language acquisition research and is relatively rare in spoken language, typically mastered only after formal instruction. This fact thus enhanced the likelihood of detecting a priming effect since children typically do not spontaneously produce passive constructions. It also raised the probability that any increase in the use of passives could be linked to learning. The task used different pictures depicting transitive scenes that could be described with either an active or a passive construction. The task consisted of three blocks: baseline, test, and posttest. In the baseline block, children were instructed to describe a picture without any input or guidance from the experimenter. In the subsequent test block, children underwent priming. They were informed that the experimenter would describe one picture, and they would then describe the next. Their task was to echo the experimenter's prime sentence, which consistently featured a full "be" passive structure with a "by" phrase. In the posttest phase, children described further pictures without priming, hence testing for potential long-term priming effects. The results aligned with Kidd's hypothesis: Success on the implicit statistical learning task correlated with the persistence of the syntactic priming effect during the posttest phase, where no further primes were given. Conversely, explicit learning did not forecast priming. Overall, Kidd's findings indicate a direct link between children's performance on an independent statistical learning test and the sustained retention of primed syntactic structures over the long term. As the author suggest, these findings provide empirical evidence showcasing the direct relationship between implicit statistical learning and children's acquisition of syntax.

1.2.3 Tracking statistical regularities at the syntactical level in computational language models

Scholars that investigated the learnability of syntax through neural networks pursued the ambitious goal to demonstrate that the ability to correctly process and form syntactic structures can be gained by networks through learning, hence by pure exposure to language, without the need of prespecified symbolic rules (Chritiansen, Chater, 2001). However, among them, we find two rather different approaches to the investigation. In the less ambitious one, scholars have created models designed to learn syntax based on input sentences in which each syntactic element has been previously tagged with the information concerning its syntactic role (e.g. dogs = plural, noun).[7] Typically, the purpose of these network models is to identify the grammar (or a portion of it) that matches the example structures. This implies that the structural elements of language are not learned through observation, but are instead inherent (Chritiansen, Chater, 2001). The second class of model has a more ambitious aim: it provides evidence for the learnability of syntax by simply feeding the network with rough sentences, without any other information. In the 1990s, these models started to provide the first empirical interesting results accounting for sentence processing phenomena. Hence, several scholars in those years began to think that these models could hold significant potential for shedding light on crucial issues concerning the psychology of language, as they could reveal the possibility of language learning without pre-existing linguistic knowledge (Chritiansen, Chater, 2001). Here, we will focus on this second class, reviewing some of the most interesting models that have been developed over the years.

Elman (1991) conducted one of the earliest and most influential studies aiming ambitiously to test whether a network could learn grammar solely through exposure to input sentences, without any prior information about the grammar. (Christiansen, Chater, 2001). In 1991, Elman trained a Simple Recurrent Network (SRN) intending to explore whether it would have succeeded in predicting words in simple sentences. The project of training an SRN to investigate sentence processing had already been undertaken by Elman one year before (Elman, 1990). Simple Recurrent Networks have revealed to be a revolutionary tool that, in the following years, have found interesting application, not only in sentence processing but also in different fields such as speech and handwriting recognition, music composition, robot control, machine translation, financial forecasting, etc. In general, these networks are particularly efficient in those tasks which require prediction. This is not surprising, because they have been designed with the specific goal of discovering and processing patterns among sequential data. Sentences are

[7] Among this class of models, we find:
PARSNIP (Hanson & Kegl, 1987); VITAL (Howells, 1988); Pollack, 1988, 1990; Chalmers, 1990; Niklasson & van Gelder, 1994; Sopena, 1991; Stolcke, 1991.

sequences of temporally-ordered words, time is indissolubly tied to language, and it is starting precisely from these considerations that the Simple Recurrent Network has been conceived (Elman, 1990). Elman noticed, indeed, that one of the major problems with previous Parallel Distributed Processing (PDP) was the difficulty to represent time, an intrinsic issue dictated by the very nature of neural networks that are, as their name recall, parallel processors, as opposed to the serial nature of human language[8]. How the time factor has been included in the SRN network, and its consequence on the specific way in which the network works, represent a major difference from previous PDP networks. "In parallel distributed processing models, the processing of sequential inputs has been accomplished in several ways. The most common solution is to attempt to parallelize time by giving it a spatial representation. However, there are problems with this approach, and it is ultimately not a good solution. A better approach would be to represent time implicitly rather than explicitly. That is, we represent time by the effect it has on processing and not as an additional dimension of the input" (Elman, 1990, p. 180). This is precisely what Elman has achieved with SRN. The Simple Recurrent Network appeared to be a revolutionary tool. However, this approach was not completely new. The SRN designed by Elman is an implementation of the recurrent network described by Jordan (1986). The fundamental feature of this type of network is its possession of recurrent connections, which allow hidden nodes to access their preceding outputs. This capability enables these outputs to influence subsequent behaviors. Therefore, due to recurrent connections, the network exhibits memory (Elman, 1990). In other words, the processing system is dynamic, and its operations are influenced by temporal factors. In this manner, Elman successfully represented time through its effect on processing: the system's dynamic properties respond to temporal sequences, made feasible by the network's memory facilitated through recurrent connections (Elman, 1990). To delve deeper into the network's operation, we refer the reader to Elman (1990). Elman (1990) tested the SRN in various tasks, such as the learnability of the Exclusive-Or function (XOR) and more complex sequential

[8] Elman does not negate the importance of hierarchical structure in human language; however, he argues that theoreticians have often uniquely focused on this aspect, with the result of neglecting the importance of serial order in language.

patterns where the duration of patterns was variable, thus challenging the networks with more intricate tasks. He fed the network structured sequences of letters and later exposed it to simple utterances composed of two or three-word sentences to assess its ability to predict words in a sequence. These sentences were generated using a sentence generator provided with 13 categories of nouns and verbs[9], 29 lexical items, and 15 sentence templates[10], resulting in 10000 sentence frames of two or three words each. The frames were filled randomly with appropriate lexical items, and then the SRN was trained. Elman (1990) aimed to determine whether the SRN could predict subsequent words in a sentence. He did not expect the network to predict the exact lexical item but rather one belonging to the correct category. As Elman (1990) stated, "Successors cannot be predicted with absolute certainty; there is a built-in error which is inevitable. Nevertheless, although the prediction cannot be error-free, it is also true that word order is not random. For any given sequence of words, there are a limited number of possible successors. Under these circumstances, the network should learn the expected frequency of occurrence of each of the possible successor words; it should then activate the output nodes proportional to these expected frequencies" (Elman, 1990, p. 197). Surprisingly, the results showed that the SRN identified major categories of words and developed internal representations of input vectors that reflected information about their sequential order. While the network could not predict exact word sequences, it recognized that certain classes of inputs (such as verbs) typically followed others (such as nouns) within the corpus (Elman, 1990). Thus, the SRN can be viewed as a dynamic system capable of predicting word categories and subcategories. The error rate for predicting the correct lexical word was high but deemed acceptable given the nondeterministic nature of the task. Elman's SRN successfully built internal representations of input vectors that encoded important information about sequential order. Crucially, Elman suggested that these categorical representations

[9] E.g. VERB-TRANSITIVE {chase, see, …}; NOUN-HUMAN {woman, girl, man, …}; VERB-PERCEPTION {hear, smell, …}.

[10] E.g. WORD 1=NOUN-HUMAN + WORD 2=VERB-PERCEPTION + WORD 3=NOUN-FOOD.

WORD 1=NOUN-ANIMAL + WORD 2=VERB-TRANSITIVE + WORD 3=NOUN-ANIMAL.

were hierarchical in nature and emerged simply from the temporal distribution of sequential information. Importantly, this hierarchical structure emerged without predefined architectural constraints regarding the form or space of the hierarchy within the network. Elman's Simple Recurrent Network (1990) marked a significant milestone in connectionist studies, leading to further exploration of language learning using SRNs. Researchers have applied SRNs extensively to investigate syntactic structures, including the learnability of finite-state grammars (Giles et al., 1992; Giles & Omlin, 1993; Servan-Schreiber et al., 1991). Moreover, SRNs have been tested on more complex grammatical structures, extending beyond the capabilities of finite-state devices, in simulations of formal language theory and language-like grammars (Christiansen & Chater, 1999).

Elman (1991) delved deeper into the induction of grammatical structure in the learning process, examining whether SRNs could develop internal representations encoding hierarchical relationships between constituents. While previous findings showed SRNs could predict word categories based on lexical representation, critical issues regarding grammatical structure remained unresolved (Elman, 1991). To address these, Elman trained SRNs on sentences containing multiply-embedded relative clauses, testing their ability to discover and represent complex hierarchical and recursive structures. Importantly, the network received no explicit information about lexical categories, grammatical roles, or number, forcing it to autonomously discover these features. Interestingly, the results provided early evidence for the fact that the network succeeded in the task by developing distributional representations based on the hierarchical structure of constituents and their grammatical relations (Elman, 1991). Elman explained the centrality of his results with these words: "The important result of the current work is to suggest that the sensitivity to context which is characteristic of many connectionist models, and which is built-in to the architecture of the networks used here, does not preclude the ability to capture generalizations which are at a high level of abstraction. Nor is this a paradox. Sensitivity to context is precisely the mechanism which underlies the ability to abstract and generalize". (Elman, 1991, p. 116). However, as Elman pointed out, these results were not conclusive. Despite providing new, interesting evidence, the work was preliminary, and some major issues remained. First, the investigation did not take into account semantics aspects, focusing only on syntactic distributional properties. Second, it remained to be probed if the ability of the SRN to learn in a relatively simple syntactic environment was extendable to more complex sentence structures. Third, in a standard learning situation, human learners are exposed to more complex, rich, and diversified linguistic stimuli as compared to the limited linguistic phenomena that were taken into consideration in this experiment (Elman, 1991). Hence, despite the important development they had undergone in those years, connectionist models were in 1991 still far from providing an ultimate tool for the study of human cognition. In addition to the point just discussed, another pending issue was the one concerning the difference in computational power between connectionist models and human cognitive resources. "What is not currently known is effect of limited resources on computational power. Since human cognition is carried out in a system with relatively fixed and limited resources, this question is of paramount interest. These limitations provide critical constraints on the nature of the functions which can be mapped; it is an important empirical question whether these constraints explain the specific form of human cognition" (Elman, 1991, p. 115). Specifically, this last issue was subsequently taken into consideration by Elman (1993).

Elman (1993) began with the observation that language acquisition occurs early in life, when cognitive resources are not fully developed and within a limited time window. He considered the hypothesis that these developmental constraints on cognitive resources are not a hindrance to language learnability but rather a prerequisite. This hypothesis emerged from an experiment where he trained connectionist networks with complex sentences to predict subsequent words. The experiment aimed to explore the networks' capacity to learn and represent embedded structures. Surprisingly, Elman discovered that when the networks were provided with fully developed, "adult-like" resources, they failed in language learning. In contrast, when the networks had limited working memory capacity, they successfully learned language and managed complex syntactic structures. "There are circumstances in which these models work best (and in some cases, only work at all) when they are forced to start small and to undergo a developmental change which resembles the increase in working memory which also occurs over time in children" (Elman, 1993, p. 72). In addition to this, both children and neural networks seem to be more sensitive and open to learning during the early stages of the learning process. Hence, summarizing, Elman (1993) found early evidence for the fact that, in humans as in neural networks, there might be interesting parallelism between the development of computational resources and the learnability of complex tasks, such as language. Hence, Elman (1993) concluded highlighting the importance of taking into consideration developmental phenomena when investigating issues related to language learning.

Elman's work had a profound influence on connectionist research in subsequent years. His results provided crucial experimental confirmation of the potential of neural networks as tools for understanding the acquisition of syntax in humans. Early findings with neural networks highlighted intriguing parallels in learning processes between networks and humans, prompting scholars to pursue further investigations in this area. Additional support for the similarities between neural networks and the human brain in language learning was demonstrated by several researchers. Weckerly & Elman (1992) observed similar behaviors between SRNs and humans in processing center-embedded sentences. Building on Elman's research, Christiansen (1994) explored cross-dependencies and extended the understanding of SRN capabilities. Notably, he found that trained SRNs could successfully learn complex constructions, revealing similarities in complexity between SRNs and human cognition. Christiansen and Chater (1999) investigated the learnability of complex recursive structures with SRNs, discovering qualitative parallels in processing center-embedding, right-branching recursion, and cross-dependencies. Specifically, both SRNs and humans exhibited similar computational limitations when handling recursive depth. "Connectionist networks can learn to handle recursion with a comparable level of performance to the human language processor" (Chistiansen & Chater, 1999, p.199). This was the first time that a detailed comparative study had been carried out with different types of recursive structures, by comparing the neural networks' results with a statistical benchmark based on n-grams (Chistiansen & Chater, 1999). Reali and Christiansen (2005) investigated the learnability of yes-no questions. They carried out experiments with bi-grams and tri-grams statistical models. The aim was to check if the models would have learned to correctly front auxiliaries in polar interrogatives. After training 10 SRNs with exposure to positive evidence (Bernstein corpus), they tested the SRNs by making them discriminate grammatical versus ungrammatical interrogatives. Interestingly, they found that the SRNs succeeded in the task by preferring grammatical forms over ungrammatical ones.

However, while these earlier connectionist models achieved some success, it is important to recognize that they did not provide a definitive solution to the challenge of language acquisition. In the 1990s and early 2000s, connectionist approaches in syntactic studies were still in their infancy, and researchers acknowledged both the potential and limitations of this line of research. The use of Simple Recurrent Networks (SRNs) initially showed promise but faced criticism for their narrow scope and limited ability to represent input comprehensively. Christiansen and Chater (2001) acknowledged that these models often operated with simplified grammar fragments and small vocabularies, prompting concerns about their scalability to real-world language complexities. Critics argued that because connectionist models typically learn from finite and repetitive datasets, they struggled to demonstrate the capacity to develop full linguistic competence from the often sparse and incomplete input data encountered in natural language acquisition (Bickerton, 1996; Guasti, 2002; Berwick et al., 2011). One major critique from generativists was that connectionist models had not adequately addressed the Poverty of the Stimulus (POS) argument. Berwick et al. (2011) pointed out that results from Christiansen and Reali's (2005) study on auxiliary fronting could potentially be explained by basic statistical facts, such as simple bigram distributional statistics, rather than true syntactic comprehension. They questioned whether SRNs could effectively handle more complex interrogative structures, including those with embedded relatives, which were not explored in the 2005 study. Another argument put forward by Berwick et al. (2011), not specifically targeting connectionist models but applicable to usage-based approaches in general, was that these approaches did not adequately address the structure-dependence of linguistic rules. Usage-based studies often failed to demonstrate that their models adhered strictly to structure-dependent rules. Even when these models successfully handled complex hierarchical structures, they might still operate in a structure-independent manner. According to scholars in the generative stream, "[…] structure dependence of rules is an abstract property of certain grammars, and it is orthogonal to the property of generating expressions that exhibit hierarchical constituency" (Berwick et al., 2011, p. 1229). Thus, according to these scholars, usage-based approaches to language acquisition have not offered a valid answer to the poverty of stimulus argument. Despite the early successes of connectionist models, significant gaps remained. Connectionist researchers like Christiansen and Chater (1999) acknowledged the limitations of their small-scale simulations and stressed the need for more comprehensive evidence. The critical challenge was generalizing beyond the provided input. Critics such as Bickerton (1996) and Guasti (2002) emphasized that, unlike human learners who can generalize and develop robust linguistic competence from degenerate input, connectionist models will learn a degenerate language if the input they receive is degenerate. In summary, while connectionist approaches made promising strides in those years, they did not fully resolve the complexities of language acquisition. The need to demonstrate the ability to handle richer, more varied input and to provide evidence of mastering structure-dependent phenomena continued to drive further research and debate in the field in the following years.

In recent years, advancements in neural networks for natural language processing have significantly pushed the state of the art, resulting in unprecedented achievements in human language learning. Key advancements have been made by models such as Long Short-Term Memory networks (LSTMs) (Hochreiter & Schmidhuber, 1997), Gated Recurrent Units (GRUs) (Cho et al., 2014), and Transformers (Vaswani et al., 2017). These models are classified as deep neural networks due to their incorporation of multiple layers within their architectures. LSTMs and GRUs have emerged as solutions to longstanding issues with traditional RNNs, particularly their struggle to learn effectively from sequences with long-range dependencies. GRUs introduce gating mechanisms within the network architecture, enhancing their ability to manage long-term dependencies with fewer parameters. These gates regulate how information in the hidden state is updated across words, enabling precise tracking of dependencies spanning extensive sequences. LSTMs now dominate as the primary variant of RNNs in natural language processing (NLP) (Linzen, Baroni, 2020). In contrast, transformers represent a significant departure from traditional RNNs, LSTMs, and GRUs. Built on self-attention mechanisms, transformers eliminate the need for recurrence. They feature deep architectures incorporating multiple layers of self-attention and feed-forward networks, enabling efficient parallel processing of sequences. This design has underpinned breakthroughs such as BERT and GPT. BERT, introduced by Google AI in 2018, leverages transformers to generate context-aware word embeddings by considering both left and right contexts simultaneously, achieving state-of-the-art performance across various NLP tasks. Similarly, OpenAI's GPT, unveiled in 2018, utilizes transformer-based architectures to generate coherent and contextually appropriate text, highlighting the versatility and power of transformers in language generation tasks. The effectiveness of both gating mechanisms in GRUs and attention mechanisms in transformers is not predetermined but determined by weights learned during training. This adaptability, coupled with advancements in specialized hardware, enables transformers to be efficiently trained on vast corpora, enhancing their capabilities in handling complex natural language tasks (Linzen, Baroni, 2020). Certainly, these models were not created with the aim of testing linguistic hypotheses but for practical purposes. Indeed, they were not designed based on specific theories to test predictions derived from those theories. Nonetheless, this does not undermine their value as scientific models, as their origins do not affect their scientific validity (Cichy & Kaiser, 2019). For this reason, several cognitive scientists have started using DNNs as models of human behavior and brain responses. Various syntactic phenomena have been tested using DNNs. Linzen and Baroni (2020) provide an interesting overview of these studies, and we encourage readers to refer to their paper for further insights. Here, we only mention a few of the phenomena that have been tested. Linzen et al. (2016) trained an LSTM on English sentence prefixes to predict the number of the upcoming verb, achieving over 99% accuracy on new prefixes. However, this high accuracy does not necessarily show the network's ability to identify the subject's head noun, as the subject is often the noun closest to the verb. Even with sentences containing up to four attractors, the network maintained 82% accuracy, indicating it was not frequently misled by attractors. Bernardy and Lappin (2017) demonstrated that other DNN architectures, such as GRUs and convolutional networks, can also successfully perform the number prediction task. This suggests that Linzen et al.'s (2016) findings are not specific to LSTMs. Gulordava et al. (2018) showed that LSTMs can learn long-distance agreement by predicting the next word in a corpus without focusing specifically on subject-verb agreement. They tested this by comparing the probabilities assigned to singular and plural verb forms after a sentence prefix. LSTMs trained this way showed high accuracy in agreement prediction across four languages (English, Hebrew, Italian, and Russian) and various dependency types. Additionally, LSTMs performed well on grammatically correct but semantically implausible sentences (e.g., colorless green ideas), indicating they can compute agreement without relying on lexical or semantic cues. This suggests that word prediction training alone can teach networks about long-distance agreement and syntactic categories. Wilcox et al. (2018) examined how well LSTM language models detect English filler-gap dependencies. Additionally, McCoy et al. (2020) investigated the case of auxiliary fronting in English question formation (Linzen, Baroni, 2020).

Overall, DNNs demonstrate remarkable efficiency in learning and producing language. Crucially, their language production capability resembles that of humans, with low error rates even when handling complex syntactic structures. They achieve this without relying on innate linguistic constraints such as abstract symbols and rules, traditionally deemed necessary by generative linguists for language acquisition. However, this does not mean these models are entirely unconstrained. As Linzen and Baroni (2020) argue, their biases stem from initial weights and the structure of their architectures, though these differ from constraints proposed by generative linguists. Of course, this does not rule out the possibility of innate biases like a universal grammar or cognitive biases of different nature in the human brain. Nonetheless, these models provide significant evidence that such innate mechanisms are not essential for developing complete linguistic competence. Regarding the influence that the constraints of models may have on how they process linguistic material, Matusevych and Culbertson (2022) conducted an intriguing study. These authors investigated whether different types of RNNs exhibit a "homomorphism bias," which refers to a transparent correspondence between the underlying hierarchical structure of the noun phrase and the linear sequence of its elements. This bias is reflected in the tendency to preserve hierarchical relationships between words even when the linear order changes, a phenomenon supported by several studies (Culbertson & Adger, 2014; Martin et al., 2019, 2020; Culbertson et al., 2020). Matusevych and Culbertson (2022) tested three computational models on their capacity to demonstrate this bias and mimic human-like preferences for noun phrase word order. Importantly, these models differed in their architecture: one was a linear RNN (LSTM), and the other two were hierarchical RNNs (an ON-LSTM and an RNNG). They pre-trained these models on English language data and then exposed them to an artificial language with a different word order for noun phrases. The models were subsequently tested on their ability to predict the order of modifiers in the artificial language. The findings revealed a clear distinction between the linear and hierarchical RNNs. Only the hierarchical models, ON-LSTM and RNNG, exhibited the homomorphism bias, successfully replicating the human-like preference for maintaining hierarchical relationships within noun phrases, although they displayed different behaviors for which the authors could not provide detailed explanations. In contrast, the linear LSTM failed to demonstrate this bias. This suggests two important points: first, that the human preference for specific word orders, even in unfamiliar languages, might stem from the inherent hierarchical nature of language processing; and second, the study highlights the importance of hierarchical representations in language modeling, providing evidence that hierarchical RNNs are better equipped to capture these complexities compared to linear models.

Despite the great success that recent deep DNNs have been achieving in recent years, two fundamental issues arise when considering these models as cognitive tools for studying biological language. The first concerns the disparity in training on vast datasets compared to the more limited exposure during human learning. Moreover, they lack a real-world environment that supports language learning, unlike the interactive environment in which children learn (Linzen, Baroni, 2020; Piantadosi, 2023). Another challenge stems from the complexity of neural networks, as the exact mechanism leading to their extraordinary results remains incompletely understood. While it is possible to precisely measure the activation of each unit after processing individual words (Linzen, Baroni, 2020), there is considerable interest in understanding how these activations arise from the network's internal dynamics described by vectors containing hundreds or thousands of real numbers manipulated through arithmetic operations guided by millions of parameters. Specific techniques are needed to make these complex vectors understandable to humans. Although this is a challenging task, some researchers have explored various methods to address it (Linzen, Baroni, 2020). For example, Manning et al. (2020) question whether these models can learn the hierarchical structure of language, despite lacking explicit representations of syntax in their input. By using 'structural probe' methods, they have developed a similarity metric for the internal word representations in DNNs that reflects their syntactic distance in a sentence analysis. Manning et al. (2020) suggest that a linear transformation of these learned embeddings effectively captures parsing tree distances, allowing for an approximate reconstruction of tree structures commonly used by linguists. However, as noted by Linzen and Baroni (2020), it is crucial to recognize that the ability to extract information from a network's representation does not necessarily mean the network actively uses such information to influence its behavior. This suggests that the production of hierarchical representations by the network does not necessarily imply the active use of hierarchical strategies. It may simply reflect the application of distributive statistics over a large dataset containing such structures. Speculatively, this could be a potential difference between the biological brain and artificial neural networks. As we have emphasized, recent psycholinguistic studies indicate that the human brain possesses not only statistical abilities but also a cognitive bias favoring structural preferences, as demonstrated by significant results from psycholinguistic studies presented in Section *1.2.1*. On the other hand, given that human language is produced by the human brain, it would not be surprising if its form reflects the cognitive biases of the system from which it originated. To shed more light on these open issues, Piantadosi (2023) suggests the importance of introducing additional architectural constraints and principles inspired by the human brain into artificial neural networks. This could include constraints that improve optimization or reflect the cognitive limitations of human learners. These questions are central to cognitive sciences and drive the current 'The

BabyLM Challenge' (Warstadt et al., 2023), which aims to develop models capable of learning with data quantities that mirror human development. Finally, while some scholars criticize the utility of these models in the context of studying biological language (Chomsky, 2023), it is important to consider that continued study of these models could open numerous research opportunities. Although they do not provide immediate answers, they are essential for deepening our understanding of human language. Future studies could address these issues from different perspectives and refine investigative techniques, thereby contributing to improving our understanding of human language and formulating compelling theories about how structure and statistics interact, as also suggested by Piantadosi (2023).

1.3 Reconciling the different positions: Statistical learning meets hierarchical structure-dependent constraints

As observed in the preceding sections, psycholinguistic studies have provided evidence for the existence of structure-dependent abstract representations and the ability to use statistical learning mechanisms to understand phenomena at the syntactic level. Additionally, we have seen how recent developments in neural networks have demonstrated human-like linguistic capabilities, leveraging statistical computations on large datasets, notably without innate biases proposed by generativists such as Universal Grammar (UG). The existence of structure-dependent phenomena within language and the potential for learning through statistical mechanisms represent focal points that nativist and usage-based theories have respectively emphasized. The former by advancing the idea that language ability is innate, domain-specific, and richly structured. The second by showing that language is learnable through a set of potentially domain-general statistical mechanisms computable on the linear sequence of utterances to which children are exposed. Frequently, nativist and usage-based approaches have been perceived as two distinct and incompatible theories. However, the reality differs. Indeed, within the two theories we find aspects that are in fact compatible and not mutually exclusive. First, within the two theories there is not necessarily a relation of implication between different postulates. To illustrate, the richness of language's structure does not necessarily indicate its domain-specific nature. Moreover, the postulates of one theory are not necessarily irreconcilable with those of the other theory. For example, the presence of sequential statistical learning capability does not rule out the presence of hierarchical structural constraints. Several scholars support the prospect of reconciling the postulates formulated by these distinct theories. Scholars under the usage-based approach, both from Connectionism and SL, have often emphasized that distributional learning is completely compatible with UG theories of language acquisition. "Distributional analysis, as a way of learning aspects of language, is wholly compatible with generative grammar, cognitivism, and modern epistemology. Even if children are innately equipped with a universal grammar, there are still many aspects of the particular of their native language that must be picked up by experience. Distributional learning mechanisms provide a potentially useful means which might contribute to this process" (Redington, Chater, 1998, p.135). In fact, usage-based approaches to language acquisition do not represent a return to the Tabula Rasa Theory (Bates, Elman, 1996). As Saffran, Aslin and Newport (1996) highlighted, commenting on the result they obtained, the discovery that eight-month-old infants, exposed for only 2 minutes to a continuous string of nonsense syllables were able to segment nonwords by relying solely on statistical cues available in the string, does not have to be interpreted as a proof for the fact that all that children can linguistically achieve is due to general, unconstrained learning abilities. Saffran, Aslin and Newport (1996) do reject the nativist hypothesis according to which, language cannot be learnt; however, they do not exclude the possibility that the statistical learning mechanism at work during language acquisition might be innately biased, arguing that "[…] some aspects of early development may turn out to be best characterized as resulting from innately biased statistical learning mechanisms rather than innate knowledge" (Saffran, Aslin and Newport, 1996, p.1928). Several scholars adhering to the usage-based approach emphasize the notion of innately driven learning, in essence, a constrained learning mechanism. "Even if we assume that a brain (real or artificial) contains no innate knowledge at all, we have to make crucial assumptions about the structure of the learning device, its rate, and style of learning, and the kinds of input that it prefers to receive” (Bates, Elman, 1996 p. 1848- 1849). Ramsey and Stich (1990), taking into consideration Nativism and Connectionism, pointed out that the incompatibility between them has been much exaggerated (Ramsey, Stitch, 1990, p.20). “The task of the language acquisition mechanism is an inductive learning task. And […] any successful inductive learning strategy must be strongly biased” (Ramsey, Stitch, 1990, p.197). Therefore, it is undeniable that these distinct approaches have frequently exhibited notable divergences in their investigative perspectives, methods, and overarching research objectives. However, as numerous scholars have pointed out, the presence of innate structural biases and statistical learning abilities are not mutually exclusive. In this regard, we report Readington and Chater’s assertion: “The claim that some interesting aspects of language can be learnt does not imply the claim that all aspects of language can be learnt from scratch. Indeed, as discussed previously, empiricist and nativist positions alike are compatible with distributional learning mechanisms” (Readington and Chater, 1998, p.135). In this vein, Yang proposed a reconciliation between the two approaches, highlighting potential advantages that could arise from their convergence: “Language acquisition can then be viewed as a form of ‘innately guided learning’, where UG instructs the learner ‘what cues it should attend to’. Both endowment and learning are important to language acquisition – and both linguists and psychologists can take comfort in this synthesis” (Yang, 2004 p.455).

1.4 Conclusion

In conclusion, recent advancements in psycholinguistic research have provided compelling evidence supporting the effectiveness of experiential learning in language acquisition. These studies highlight the specific types of statistical information that learners can track at different levels of linguistic analysis, from the phonological level to, importantly, the syntactic level. The crucial role of statistical learning has been highlighted by research with deep neural networks as well, which, as we have seen in this chapter, has led to surprising results in recent years. These results demonstrate the potential to create machines without “innate language faculties” that can learn linguistic abilities purely through exposure to linguistic data, by tracking the statistical regularities in the data, with linguistic abilities that are almost entirely in line with those of humans.

At the same time, recent psycholinguistics studies have also demonstrated that language is richly structured and constrained. Indeed, they provided compelling evidence of the presence of structural abstract representation during the acquisition and processing of syntactic phenomena. These findings highlight that language learners, when dealing with syntax, rely not only on surface-level statistics but also on the underlying structure of language. Nowadays it seems clear that adequate theories of language processing and acquisition investigating syntactic mechanisms, must acknowledge the crucial role of statistical learning, taking into account the evidence that this occurs within hierarchical boundaries and constraints (Coopmans et al., 2022).

In this chapter, we have seen that recent studies provide compelling evidence that, when dealing with noun phrases, we develop abstract representations of their constituting elements based on hierarchical, rather than purely sequential, relations. Hierarchical structures are present at various levels in human language; besides the syntactic level, they are also found at the phonemic and morphological levels. However, it is important to emphasize that not all aspects of language are hierarchical. Some language processes are strictly local in a computational sense and involve only linear mechanisms, rather than hierarchical ones (Culicover, 2013). One of the most extensively studied and debated phenomena in the syntactic domain is the presence of hierarchical structures. Within syntax, there are different types of hierarchical structures, ranging from "simple" hierarchical phenomena to the complex realm of recursive hierarchical phenomena. Recursive embedding is a notable feature of human syntax, where a sentence can be embedded within another sentence, and a portion of a structure can exhibit the same organization as the entire structure itself. This capability allows for the creation of multi-level complex structures in which constituents are embedded in constituents of the same category, a remarkable feature of human syntax, as we will see in Chapter 2. However, not all hierarchical structures in syntax are recursive; recursion is just one manifestation of structure-dependence phenomena (i.e., hierarchical relationships among constituents) in language. The focus of this thesis is to shed light on the mechanisms underlying the formation of recursive hierarchical abstract representations. Specifically, given the sequential nature of language, we will explore how recursive hierarchical structures emerge from temporally ordered sequences of stimuli. Numerous studies have focused on the concept of recursion in human language, but many issues remain unresolved. Despite recursion being defined as the fundamental property and hallmark of human language, universally present across all human languages (Hauser, Chomsky, & Fitch, 2002; Moro, 2016) for several years, in the literature, there has been a lack of a precise and shared definition of "recursion" in the context of language. Additionally, there has been a scarcity of suitable tools for empirically investigating recursive cognitive abilities.

In this thesis, following the approach outlined in this chapter, we will investigate the emergence of recursive hierarchical abstract representations from temporally ordered stimuli. As we will explain throughout this thesis, we posit that this ability plays a role in human syntax processing and acquisition. Specifically, we will investigate this ability in different sensory domains (i.e., visual, auditory, and tactile), shedding light on the potential domain-specific and domain-general components of this cognitive function. Additionally, we will focus on the statistical mechanisms and structural abstract representations involved in the process, thereby illuminating the complex relationship between linear order and hierarchy. This investigation will provide insights into how these two seemingly distinct aspects of language are intricately connected. Furthermore, it will help us better understand the ability to form recursive abstract representations from sequential stimuli. Traditionally, recursion has been considered a unique linguistic capacity, innate to the human linguistic system and not shared with other cognitive systems or species (Hauser et al., 2002). However, recent studies are challenging this view, revealing a more complex picture than previously assumed. The present thesis aims to further explore this phenomenon and investigate whether recursion is not solely a linguistic feature. It has been demonstrated that human language learning exhibits cognitive biases in terms of boundaries and constraints within which the acquisition process unfolds. Regarding the source of these cognitive biases, multiple hypotheses exist; they could pertain to either domain-general cognitive processes or be unique to language (Culbertson, Smolensky, & Legendre, 2012). Our research seeks to shed light on the domain-specific and domain-general components of this ability, by exploring this capacity across various sensory domains.

In the upcoming chapter, we will focus on *recursion*, the specific type of structural phenomenon found in human language that will constitute the central topic of our work. Recursion refers to the capacity to create multi-layered hierarchical representations, in which part of a structure mirrors the organizational pattern of the whole. In this context, we will address a critical issue prevalent in the existing (psycho)linguistic literature: the absence of a universally accepted and precise definition of the term recursion. Hence, we will endeavor to offer a rigorous and explicit definition of the term. We will then analyze in detail the relationship between sequentiality and hierarchy in human language, emphasizing the importance of considering both these dimensions in the study of recursion in syntax. Moreover, we will focus on the issue concerning the transition from the linear to the hierarchical dimension, with a special emphasis on the cognitive mechanisms that underpin this ability. In the second part of the chapter, we will delve into the methodologies available for experimentally studying the ability to form recursive hierarchical abstract representations from sequentially fading input. We will start by presenting the research traditions which investigated the abilities to implicitly learn structured information from the environment. As we will see, the topic has captured the attention of several scholars over the years, and, interestingly, it has undergone a renewal in interest in more recent years. We begin this investigation by retracing the main stages in this study tradition and presenting the state of the art in the field. Importantly, we will show a recent contribution to the realignment of two different major lines of research that, over the years, investigated this ability despite remaining largely separated: *Implicit Learning* and *Statistical Learning*. Then, we will embark on an in-depth exploration of the Artificial Grammar Learning (AGL) paradigm, which is exceptionally well-suited for the investigation of implicit statistical learning. Additionally, we will introduce Formal Language Theory (FLT) and the grammars belonging to the Chomsky hierarchy. FLT, initiated by Noam Chomsky in the 1950s, aims to systematically study the computational basis of human language by describing the mathematical and computational properties of various language classes. The Chomsky hierarchy proposes four nested levels of grammars ordered by complexity, each corresponding to a specific automaton that generates the strings of their respective languages. Automata, abstract representations of computational systems, can recognize or reject strings based on their computational power. Grammars belonging to the Chomsky hierarchy have been extensively used in AGL studies in psycholinguistic research to shed light on the computational abilities at the core of the human language faculty. Specifically, we will examine studies that have sought to investigate recursion through formal languages belonging to the Chomsky hierarchy. In the last part of the chapter, we will address some pivotal issues that characterize the study of recursion in psycholinguistic research. Firstly, we will note that numerous studies have struggled with a mismatch between their chosen methods and their study objectives. In other words, in several cases, the ability to process recursive hierarchical structures has been investigated using unsuitable tools, resulting in experiments that were unable to provide precise answers about the phenomenon, leading to misinterpretations. Secondly, we will argue that studying human language abilities using languages from the Chomsky hierarchy can be problematic if important factors related to the generative power of these systems are not taken into account, as we will discuss in the upcoming chapter. Recent psycholinguistic studies show that the complexity defined by the Chomsky hierarchy does not directly correspond to cognitive complexity. These two aspects are interconnected in ways that are not straightforward. Therefore, the topic is more nuanced and complex than it might initially appear and requires careful consideration. Finally, we will address another issue found in several studies: the confusion between the algorithmic properties and the representational abilities of recursion. We will emphasize the distinction between them, advocating for a focus on distinctive behavioral signatures as indicators of cognitive processes related to recursion. Specifically, we will explain that definitions of recursion primarily focused on algorithmic properties may not offer the most relevant framework for empirical research. To gain insights into how human cognition represents recursive structures in behavioral experimental tasks, it is crucial to focus on the distinctive signatures of recursion. Hence, we will explore the identification of behavioral indicators that hold significant promise in facilitating the study of recursion. This chapter will serve as the foundational basis for our comprehensive examination of the mechanisms by which humans form recursive abstract structures from temporally ordered sequences, the central focus of this thesis.

2. From Sequence to Hierarchy: Exploring the Emergence of Recursive Hierarchical Representation Arising From Temporally Ordered Stimuli

2.1. What is recursion?

Human language possesses a remarkable feature that sets it apart from all other forms of expression: discrete infinity. This concept, which is at the heart of the versatility and power of language, refers to the ability of a linguistic system to generate an unlimited number of expressions from a finite set of discrete elements. In other words, it enables us to create an endless array of sentences and meanings from a relatively small inventory of words, phrases, and rules. The characterization of language as a potentially limitless array of expressions achievable with limited resources has been recognized for a long time (van der Hulst, 2010). Already Descartes (Descartes, 2003 [1637]), suggested a significant distinction between humans and animals lies in the ability to organize speech in various ways, a skill that every human possesses. Similarly, Wilhelm von Humboldt (1999 [1836]) emphasized human language's capacity to achieve endless diversity using limited resources (Sauerland & Trotzke, 2011). However, none of these works have specifically mentioned the term recursion to account for this property of human language. In the 1950s, in his early work, Chomsky introduced the term *recursion* to account for this fundamental feature of human language that allows us to convey a virtually unlimited range of ideas, from the simplest statements to the most intricate narratives (Chomsky, 1956). "[…] the whole point of introducing recursion into linguistics was to account for the fact that speakers/hearers show a continuous novelty in linguistic behaviour — a novelty that does not appear to be capped in any meaningful respect. Further, since speakers/hearers cannot possibly store all the possible sentences they understand or utter, the cognitive state accounting for this linguistic behaviour must be underlain by a finite mechanical procedure — an algorithm." (Lobina, 2011, p.155). In Chomsky's work, recursion is taken as the key to the generative power of language, enabling us to create new and meaningful expressions. Recursion, in essence, is the necessary ingredient that transforms language from a finite set of words and rules into a medium of boundless creativity, setting human language apart from other forms of communication. Recursion hence was a crucial element in Chomsky's phrase structure based approach to language (Sauerland & Trotzke, 2011). In *Syntactic Structures* (Chomsky, 1957), recursive devices are described as to be essential to linguistic theory. Crucially, however, despite the importance that Chomsky attributed to recursion, he never offered a clear and precise definition of what recursion in human language is (van der Hulst, 2010; Tomalin, 2011). Moreover, over the years, Chomsky attributed different meanings to the term recursion, with a lack of consistency in the use of terminology. As Chomsky's Generative Grammar theory progressed, also the definition of this property of human language changed (Coolidge et al., 2011; Lobina, 2011). Chomsky initially attributed the recursive property to the transformational system of the grammar (i.e., a component that transformed certain phrase markers into different phrase markers while maintaining the underlying structure) but later assigned it to the base component (i.e., rewriting rules that generated strings with linked phrase markers). Subsequently, most rewriting rules were eliminated from its syntactic theory. However, despite these changes, it is crucial to recognize that recursion, as a general property of generative systems, remained a central concept in Chomsky's theory. This holds true whether we are talking about production systems with rewriting rules or the more recent concept of Merge (Lobina, 2011). A renewal of interest, which generated hype around the research on recursion, coincided with the publication of the influential paper *The faculty of language: What is it, who has it, and how did it evolve?* by Hauser, Chomsky, and Fitch (Hauser et al., 2002). "[The term recursion] seems to have gained a disproportionate amount of attention ever since Hauser et al. (2002) hypothesized (for that is what it was) that this property may be the central and unique feature of the faculty of language." (Lobina, 2011, p.151). Hauser and colleagues (2002) formulated a new hypothesis on the faculty of language. Specifically, they proposed a distinction between the faculty of language in the broad sense (FLB) and the narrow sense (FLN). The former comprises various components, including sensory-motor and conceptual-intentional systems, alongside computational mechanisms for recursion, which enable the creation of an unlimited range of expressions from a limited set of elements. Importantly, they suggested that FLN only comprehend recursion and is the unique human component of the human language faculty. Hauser et al.'s provocative article received several criticisms. However, what interests us is that even they did not offer a precise definition of what recursion is, merely echoing the definitions and concepts that linked it to the property of discrete infinity, much like other linguists did before them. "[…] a core property of FLN is recursion […] FLN takes a finite set of elements and yields a potentially infinite array of discrete expressions. This capacity of FLN yields discrete infinity (a property that also characterizes the natural numbers). Each of these discrete expressions is then passed to the sensory-motor and conceptual-intentional systems, which process and elaborate this information in the use of language. Each expression is, in this sense, a pairing of sound and meaning. […] The core property of discrete infinity is intuitively familiar to every language user. Sentences are built up of discrete units: There are 6-word sentences and 7-word sentences, but no 6.5-word sentences. There is no longest sentence (any candidate sentence can be trumped by, for example, embedding it in "Mary thinks that . . ."), and there is no nonarbitrary upper bound to sentence length. In these respects, language is directly analogous to the natural numbers […]" (Hauser et al., 2002, p.1571). In this text, therefore, the authors did not shed light on the definition of recursion but rather speculated about the origins of FLN, the relationship there is between FLN and animal communication systems, and other human cognitive domains. In 2005, Pinker and Jackendoff offered a comprehensive critique of Hauser et al.'s arguments and introduced their own clear definition of recursion, since, as they claimed, Hauser and colleagues had not provided a precise one. Importantly, in their definition of recursion, Pinker and Jackendoff highlighted the crucial aspect of *embeddedness*. They described recursion as "a procedure that calls itself, or to a constituent that contains a constituent of the same kind." (Pinker, Jackendoff, 2005, p.203). Pinker and Jackendoff's definition possibly altered Hauser et al.'s original notion of recursion, which emphasized the link with discrete infinity and its role in generating an infinite range of thoughts (Coolidge et al., 2011). Importantly, their definition focused on two crucial aspects: *embeddedness* and *constituent of the same kind.* Besides Pinker and Jackendoff, other scholars have defined recursion based on these two concepts. Kirby (2002, p.1) defines recursion as "...a property of language with finite lexica and rule-sets in which some constituent of an expression can contain a constituent of the same category." As we will see later in this chapter, these two fundamental concepts have been considered more recently by other scholars in the effort to clarify, precisely, and ultimately define the concept of recursion in human language.

After Hauser et al.'s work (2002), a great number of publications have extensively focused on the concept of recursion within human language. These publications have presented numerous and divergent definitions, ranging from the vague to the highly intricate, leading to a state of terminological confusion and further propagating conflicting interpretations. The notion of recursion has long been a subject of contention in the field, and the diversity of definitions in use has complicated the interpretation of empirical results. In the realm of cognitive sciences, recursion remains a topic of extensive debate and disagreement, as Martins (2012) aptly noted, "Recursion is one of the most controversially discussed terms in the cognitive sciences. [...] although recursion has been hypothesized to be a necessary capacity for the evolution of language, the multiplicity of definitions being used has undermined the broader interpretation of empirical results." (Martins, 2012, p.2055). As we have seen in this chapter, the state of confusion surrounding recursion's definition has deep historical roots. Watumull et al. (2014) underscored this confusion, describing how "the concept of recursion as articulated in the context of linguistic analysis has been perennially confused." (Watumull et al., 2014, p.1). As Tomalin (2011) observed, "there were profound ambiguities surrounding the notion of recursion in the 1950s, and this was partly due to the fact that influential texts such as Syntactic Structures neglected to define what exactly constituted a recursive device. As a result, uncertainties concerning the role of recursion in linguistic theory have prevailed until the present day." (Tomalin, 2011, p. 297). The ambiguity surrounding recursion is becoming a significant issue in the literature. The term has seldom been explicitly defined, leading to a situation where significant misunderstandings are at risk of spreading throughout the literature.

(Fitch, 2010). In general, we observe several problems related to the definitions of recursion found in linguistic literature. As Parker (2006) explains, some definitions equate recursion with discrete infinity (Adger, 2003; Carnie, 2002; Lobeck, 2000). Others confuse it with the concept of phrase structure rules (Christiansen, 1994; Horrocks, 1987; Lobeck, 2000; Pinker, 2003) or with the concept of iteration (Radford, 1997). Firstly, although the term recursion was introduced to explain the concept of discrete infinity in language, the two concepts should not be conflated and confused. Indeed, recursion is not equivalent to discrete infinity, but it can be understood as one of the various mechanisms available that instantiate this property of language (Parker, 2006). Moreover, we should be cautious when discussing infinity in relation to language. Indeed, *infinity* refers to the size or cardinality of a set. It has not been definitively established that languages are infinite, whether countably or uncountably so (Langendoen and Postal, 1984). Additionally, the concepts of recursion and iteration are quite different, as we will see in the upcoming section. However, it is important to note that in the literature there are also clear and correct definitions of recursion, which are based on the concept of embedding into elements of the same category (Kirby, 2002; Martins, 2012; Parker, 2006; Pinker and Jackendoff, 2005; Trask, 1993), and which emphasize the difference between recursion and iteration (Hurford, 2003; Martins, 2012; Parker, 2006; Pinker and Jackendoff, 2005) and between recursion and hierarchical embedding (Martins, 2012; Parker, 2006). We refer the reader to Parker (2006) for a detailed overview. We will briefly explain these concepts in the upcoming section.

It should be noted, however, that the concept of recursion is not exclusively related to linguistics but is also present in computer science. Recursion is indeed eminently a formal notion (cf. Post, 1943; 1944). In computer science, more formal definitions are observed compared to those in linguistics. However, even here, the definitions lack a single common thread (Parker, 2006) and there are definitions that again confuse the concept of recursion with iteration, as observed in Loeper et al. (1996, p. 153): "[r]ecursion and iteration are two equivalent ways in programming for repeatedly performing a specific task." Moreover, it is interesting to note that in the computer science literature, some scholars refer to recursion in terms of a structural property of an object, while others in terms of a procedure or algorithm. In other words, some refer to what has been defined as *structural* recursion, while others refer to *procedural* recursion (Parker, 2006). On one hand, structural recursion refers to the object's structure and the possibility of an object to be defined in terms of itself. On the other hand, procedural recursion refers to the procedure, the algorithm, or the function that calls itself (Parker, 2006). Within the definitions of procedural recursion, the one by Liu & Stoller (1999) is particularly interesting as it emphasizes that a recursive algorithm requires a push-down stack to function. "Recursion refers to computations where the execution of a function or procedure calls itself and proceeds in a stack fashion" (Liu & Stoller, 1999, p. 73). Recursion, defined in these terms, are thus markedly different from iteration, which is generally carried out using loops (Liu & Stoller, 1999; Parker, 2006). Iteration, differently from recursion, does not need a stack. Iteration utilizes a loop structure to execute a set of instructions a certain number of times. Conversely, recursion employs a function to repeatedly call itself until it reaches a base case. In computer programming, a stack is a data structure that controls how data is accessed. Data can be added to the stack with a 'push' operation or removed with a 'pop' operation. The stack operates on a last-in, first-out (LIFO) principle. This means the most recently added element is the first to be removed. For recursion to work properly, it is essential to keep track of the current position in the procedure. When a recursive function calls itself, it needs to remember where to continue once the recursion is complete. A stack facilitates this by allowing data to be stored incrementally and then retrieved in reverse order, ensuring that each return point is correctly followed (Parker, 2006). In the same vein, Loudon (1999) describes recursion as having two distinct phases: winding and unwinding. During the winding phase, each call to the recursive function triggers another call to itself, continuing the recursion. This phase ends when a call meets a specified termination condition. Afterward, the process shifts to the unwinding phase, where the function instances are resolved in the reverse sequence of their calls (Parker, 2006). This concept, although originating from computer science, is also very interesting for understanding a specific type of recursion found in language, namely, *nested recursion*. We will revisit this concept in the upcoming section, when we will examine the different types of recursion present in language, providing linguistic examples for each.

It is crucial to note that any result achieved through a recursive algorithm can also be attained using an iterative algorithm (Parker, 2006). Both approaches can solve the same problem, but their efficiency and appropriateness may vary depending on specific contexts and constraints. For example, a classic instance of a recursive algorithm used in computer science is calculating the factorial of a non-negative integer number (Parker, 2006). The factorial of a non-negative integer *n*, denoted as n!, is the product of all positive integers less than or equal to *n*. In (1) are some examples:

(1) $0! = 1$ (by definition);
$1! = 1$;
$2!=2\times1=2$;
$3!=3\times2\times1=6$
$4!=4\times3\times2\times1=24$; and so on…

The algorithm in (2) computes the factorial of a non-negative integer *n* using recursion (Example adapted from Parker, 2006, pag. 180).

(2)
```
FUNCTION Factorial (num):
    IF num = 0 THEN:
        return 1
    ELSE:
        return num * Factorial(num - 1)
```

Crucially, as shown in (3), an iterative algorithm can reach the same result (Parker, 2006). (Adapted from Parker, 2006, p. 184).

```
(3)     FUNCTION Factorial(num):
        result = 1
        FOR temp = num; temp > 0; temp--
                result = result * temp
         return result
```

Another frequently cited example in computer science of a recursive algorithm is the one used to solve the Towers of Hanoi problem, which can be stated as follows: You have three pegs (A, B, and C). Initially, all disks of different sizes are stacked on peg A in decreasing size (largest at the bottom, smallest at the top). The goal is to move all the disks to peg B using peg C as an auxiliary (spare) peg. You can only move one disk at a time. A larger disk cannot be placed on top of a smaller disk. The recursive algorithm in (4) is very suitable to solve the Towers of Hanoi problem. (Adapted from Parker, 2006, p. 180).

```
(4)     FUNCTION MoveTower(disc, source, dest, spare):
          IF disc == 1 THEN:
            move disc from source to dest
          ELSE:
            MoveTower(disc - 1, source, spare, dest)
            move disc from source to dest
            MoveTower(disc - 1, spare, dest, source)
          ENDIF
```

How does the algorithm work? When “disc == 1”, it means there is exactly one disk to move. In this case, simply move the disk from the *source* peg to the *dest* peg. If there are more than one disk (“disc > 1”), the function does the following: Move the top *disc - 1* disks from the *source* peg to the *spare* peg, using the *dest* peg as an auxiliary. Then, move the bottom disk from the *source* peg to the *dest* peg. Finally, move the *disc - 1* disks from the *spare* peg to the *dest* peg, using the *source* peg as an auxiliary.

In summary, the difference between a recursive and an iterative algorithm lies primarily in their approach to problem-solving. A recursive algorithm solves a problem by calling itself with a smaller instance of the same problem, whereas an iterative algorithm uses loops to repeat a set of instructions until a condition is met. While any problem that can be solved using a recursive algorithm can also be solved using an iterative algorithm, there are notable differences in terms of clarity, conciseness, and memory usage. Recursive algorithms are often more straightforward and easier to understand because they closely mirror the problem's structure. However, they can be less efficient in terms of memory usage, as each recursive call consumes stack space. Depending on the type of problem, one approach may be more effective than the other. For example, problems with a naturally recursive structure can be more elegantly solved using recursion. In contrast, problems that require simple repetition are better suited for iteration.

In the upcoming section, we will explore the concept of recursion from a cognitive science perspective. Specifically, we will clarify the differences between various types of recursion found in language, such as *tail recursion* and *nested recursion*, as well as different types of iteration, such as *iteration with embedding* and *iteration without embedding*.

2.1.1. Defining recursion in cognitive science: distinction between iteration without embedding, iteration with embedding, tail recursion and nested recursion

Given the ongoing challenges in defining and understanding recursion in language, it is clear that a more precise and consistent conceptualization of this fundamental linguistic concept is necessary. We will provide a clear definition of recursion, distinguishing it from iteration and hierarchical embedding. Additionally, we will present examples of linguistic instances that fall under these categories and examine the different types of recursion found in language, specifically tail recursion and nested recursion. Following Martins (2012) definition, ***iteration*** in its most basic form (i.e. without embedding), involves a sequence of discrete iterative steps, with

each step being independent from all the others. Individual steps can be concatenated indefinitely, giving rise to a set of sequences with potentially no upper limit, but they lack the capacity to capture dependencies between them or create nested, hierarchical structures. Hence, the fundamental constraint of iteration without embedding lies in its inability to encode dependency relationships. Without the capacity for embedding, it cannot create new hierarchical structures where certain constituents are reliant on others (Martins, 2012).
If we think about language, an example of iterative structure can be (5):

(5) Mary ate [$_{NP1}$ an apple $_{NP1}$] and [$_{NP2}$ a yoghurt $_{NP2}$] and [$_{NP3}$ an egg $_{NP3}$].

The NPs in (5) are organized in flat structure: the NPs are essentially independent of one another. To confirm that the structure is iterative, and that each NP is independent, we could invert the order of the NPs without changing the overall meaning as in (6). (Parker, 2006).

(6) Mary ate [$_{NP1}$ an egg $_{NP1}$] and [$_{NP2}$ an apple $_{NP2}$] and [$_{NP3}$ a yoghurt $_{NP3}$].

As Parker (2006) explains, (5) is semantically equivalent to (6), where equivalence is measured in terms of truth conditions. In other words, propositions a-d in (7) are true in both instances.

(7) a. ate (Mary, apple)
b. ate (Mary, egg)
c. ate (Mary, yoghurt)
d. ate (Mary, egg) & ate (Mary, apple) & ate (Mary, yoghurt).

Although iteration is an algorithm used in human language, it is clear that an iterative algorithm of this kind cannot fully accommodate all the sentences generated in human language; indeed, it inherently falls short of capturing the full spectrum of linguistic expressions. The presence of a two-dimensional space becomes evident when we consider language. Language unfolds in a linear fashion. Words follow one another, one after the next, creating sentences that are, on the surface, linear sequences of sounds or symbols. This linearity is the vehicle of communication, allowing us to convey information step by step, from the beginning to the end of a sentence. Importantly, however, a rigidly linear arrangement of words alone falls short in encompassing the complexities of human language rules. The mere fact that a word placed far from others within the same sequence can exert an influence on them implies the necessity of introducing an additional dimension to accurately represent syntactic relationships (Moro, 2016). As a matter of fact, it is widely acknowledged that, in natural language grammar, we never find rules based on rigid linear positions (Tettamanti et al., 2009). Phenomena such as long-distance dependencies, hierarchical organization of phrase structure, movement and sentence transformation constitute a hallmark of human natural language (Fitch, Friederici, 2012). A purely sequential arrangement of words alone falls short in encapsulating the intricacies of human language's underlying rules.

Differently from iteration (without embedding), ***iteration with embedding*** (i.e. *hierarchical embedding*) can generate hierarchical structures by generating dependencies and relationships among constituents. This algorithm is what accounts for what linguists refer to as "phrase structure", that is the organization of sentence structure based on constituent hierarchies (Chametzky, 2000; cf. Parker, 2006). This well-established principle of language states that it consists of units that combine to form larger units. Lexical items combine to form phrases, and these phrases further combine to create larger phrases or sentences. Phrases are considered constituents, acting as unified units in specific ways (Parker, 2006). For instance, while we can transform sentence (8) into (9), we cannot transform it into (10); (example adapted from Parker, 2006, p. 209):

(8) Amy spoke to the waitress.
(9) It was the waitress that Amy spoke to.
(10) *It was waitress that Amy spoke to the.

The reason behind this limitation is that only complete constituents can be moved; fragments of constituents cannot be displaced independently (Parker, 2006). Phrase structure describes how elements in a sentence are organized. A significant implication of this organizational principle is that sentences are not merely linear sequences of words or phrases; instead, they are arranged hierarchically, as we can see in (11) (Parker, 2006).

(11) $[_{1}$ Lora $[_{2}$ saw $[_{3}$ the girl $[_{4}$ from Italy $_{4}]$ $_{3}]$ $_{2}]$ $_{1}]$

Unlike what we have seen in (6), in this case, we cannot freely rearrange the constituents in the sentence without affecting its semantic meaning. Indeed, (11) is not semantically equivalent to (12).

(12) Lora from Italy saw the girl.

As Martins (2012) interestingly pointed out, each hierarchical level can potentially contain a single or a multi-constituent. In this second case, the algorithm generates structures with long-distance dependencies. Within iterative processes featuring embedding, constituents can be nested within the same level to an infinite extent.

Crucially, however, in *iteration with embedding*, each hierarchical level is represented individually. That is, the creation of entirely new hierarchical levels is contingent on predefined rules either explicitly incorporated into the algorithm or acquired through the input data. This limitation is where recursion emerges as a solution (Martins, 2012).

Recursive embedding offers the possibility to construct hierarchical structures, and, crucially, it allows for the formation of new hierarchical levels, without the need for additional rules. Rules involve the embedding of one constituent from a specified set into another constituent belonging to the same set. Within the same set, all elements share common attributes, based on their membership in that set, even though they may differ in other characteristics. Recursive embedding is evidently an algorithm at works in human syntax. Indeed, within the realm of language, a sentence can be embedded within another sentence, and a portion of a structure can exhibit the same organization as the entire structure itself. Crucially, however, it is evident that not all structures within human syntax are recursive; some are purely hierarchical (embedded). Furthermore, not everything in language is structural (Culicover. 2013). Nonetheless, the presence of recursive hierarchical structures undoubtedly constitutes a feature of human syntax. Importantly, recursion can take various forms in human language. There are at least two types of recursion: *tail recursion* and *nested (i.e. embedded) recursion* (Parker, 2006). Following Parker (2006), we define in these terms the two types of recursion.

Tail recursion is a type of recursion where a phrase or sentence embeds within another of the same type, and this specifically happens at the beginning or end of a sentence. This phenomenon is frequently observed in natural language constructions. To further clarify, tail recursion can be categorized into two types: left-branching and right-branching. Left-branching recursion occurs when embedding is at the left end of a phrase or sentence, as exemplified in (13). On the other hand, right-branching recursion involves embedding at the right end, as shown in (14). The examples are adapted from Parker (2006).

(13) Johns's friend's brother's dog's bone

(14) The boy that kissed the woman that Mark met in the restaurant that Josh recommended.

In (13), the entire noun phrase (NP) illustrates tail recursion where each nested NP is embedded successively towards the left edge of the sentence. Conversely, in (14), the embedding occurs towards the right edge of the sentence.

Nested recursion, on the other hand, refers to embedding within a phrase or sentence where material exists on both sides of the embedding, rather than at the edges. An example of nested recursion is demonstrated in sentences like (15), which is taken from Parker (2006, p.174).

(15) The mouse the cat the dog chased bit ran.

In (15), each subject noun phrase is paired with a verb located elsewhere in the sentence. Specifically, the first subject NP is connected to the final verb, the second subject NP to the penultimate verb, and the third subject NP to the initial verb. This arrangement ensures that each sentence embeds within another phrase at its center. For example, "the dog chased" is embedded within "the cat bit," which in turn is embedded within "the mouse ran." This nesting is surrounded by additional context on both sides, distinguishing it from tail recursion. One key distinction between tail recursion and nested recursion lies in their handling of dependencies. Nested recursion introduces long-distance dependencies, where elements or positions within a phrase or sentence can be separated by significant distances. Long-distance dependency refers to a relationship where one element or position within a phrase or sentence is influenced by another, even if they are separated by intervening material. Nested recursion exemplifies such dependencies because embeddings occur centrally within a sentence or phrase. The phrase beginning before the embedded segment continues beyond it, indicating a dependency between the start and end of the phrase, mediated by intervening content. In contrast, tail recursion lacks such long-distance dependencies as the embedded phrase appears at the boundary, not affecting the rest of the sentence or phrase in the same way (Parker, 2006).

Some important aspects we would like to emphasize at the end of this section, before moving on to the next one;

(i) recursion is not so tied to the concept of discrete infinity. Certainly, a recursive embedding algorithm can potentially generate set of sentences with no fixed upper bound or sentences with no fixed upper bound to their length, starting from and using a finite set of elements. However, it is important to note that other algorithms, such as iteration (with or without embedding), can also do the same. For example, we could create a sentence with no upper bound limit to its length by applying an iterative algorithm on a finite set of nouns: “I saw a cat and a dog and a mouse and a horse." "I saw a cat and a dog and a mouse and a horse and a fish." I saw a cat and a dog and a mouse and a horse and a fish and a chair," ... Thus, discrete infinity does not imply recursion.

(ii) not all the structural phenomena that involve hierarchy are recursive. Thus, hierarchy does not imply implies recursiveness. Recursion, intended as recursive embedding, is only one of the many structural phenomena characterizing human syntax.

(iii) long-distance dependencies do not necessarily indicate recursion. Only nested recursion produces sentences with long-distance dependencies, whereas tail recursion does not. Additionally, non-recursive algorithms, such as iterative embedding, can also create long-distance dependencies. Thus, the presence of long-distance dependencies alone does not imply recursion (Martins, 2012, Parker, 2006).

(iv) as Parker (2006) pointed out, while computer science literature acknowledges that recursive algorithms, like the factorial algorithm, can be implemented iteratively, in natural language this does not seems to be the case. "[…] in natural language, semantics forces tail recursion and iteration to be understood differently as it indicates a difference in structure which is not visible simply by considering the strings. The strict ordering requirement in tail recursion that is lacking in iteration can thus be used as a tool to differentiate these in natural language." (Parker, 2006, p. 188). "An iterative description of the structure of the sentences is not true to the complex meaning they reflect. That is, in human language, semantics thus gives us the extra information required to identify the correct structure, where the string alone does not tell us if we are looking at iteration or tail recursion." (Parker, 2006, p. 227). As Parker (2006) correctly notes, one could argue that a linguistic system lacking semantics would not require recursion. In other words, if there were no meaning to convey, then iteration would be adequate for the syntax of the communication system.

2.1.2. Examining the Uniqueness of Recursion in Human Language

In this section, we will discuss whether we can accept the hypothesis by Hauser, Chomsky, and Fitch (2002) according to which recursion is the unique property of human language. According to their perspective, recursion is an inherent aspect of human language. It is the only trait that is exclusively characteristic of human language. Recursion is unique to the language faculty and is found universally across all human languages. Additionally, according to them, recursion is a distinctive feature of the human mind (van der Hulst, 2010). To check Hauser et al. (2002) hypothesis, we could pursue different lines of investigations. Among the others: (a) Is recursion unique to the human language faculty and absent in other cognitive domains? (b) Is recursion absent in animal cognition? (c) Is it possible for language to exist without recursion? (d) Is recursion the sole unique feature of human language? "If recursion truly is unique to language, there are three places we should not find it: (i) human non-linguistic cognition, (ii) non-human non - communicative cognition; (iii) non-human communicative cognition." (Kinsella, 2010, p.179). Of course, the aim of this thesis is not to pursue all these different lines of research. For an interesting overview of these various issues, we refer the reader to van der Hulst (2010). However, it is noteworthy that, although recursion has been defined as a fundamental and unique property of human language (Hauser, Fitch, Chomsky, 2002), several authors suggest that the situation is more complex and nuanced than these authors propose (van der Hulst, 2010). Firstly, recursion is not as abundantly present in human language. "If recursion is a defining feature of human language, as has been claimed, we would expect to find evidence of it in everyday talk, the primary form of language." (van der Hulst, 2010, preface, xxxiii). In reality, the situation appears to be quite different. Karlsson (2010) asserts that in written language, complex nested syntactic recursion beyond three levels is absent. This applies to sentences, noun phrases, and prepositional phrases. Additionally, even two-level nesting is exceedingly uncommon in written text. In spoken language, nested recursion deeper than one level is virtually nonexistent. Verhagen (2010) also indicates that the significance of recursion in language is often overemphasized. Moreover, regarding question (c), Everett (1986, 2005) has proposed that a language without recursion may exist. The Pirahã language, for instance, does not utilize CP embeddings or recursive possessors. Nonetheless, Pirahã can convey these concepts through other methods (van der Hulst, 2010, Parker, 2006; Kinsella, 2010). Concerning (d), Kinsella 2010 has provided an interesting analysis. According to her, recursion alone cannot be the defining feature that makes language unique. It is merely one aspect among many. Indeed, a list of features unique to language, and notably independent of recursion, can be identified easily. Among the others, structure-dependence; also the duality of patterning is a core feature of human language: meaningful linguistic units can be decomposed into smaller, meaningless units which can be recombined to form different meaningful units (Hockett, 1960). This characteristic is absent in other communication systems and human cognitive processes and operates independently of recursion. Additionally, human language employs numerous syntactic devices—such as case, agreement, pronouns, articles, quantifiers, auxiliaries, tense—that are unique to language and do not rely on recursion. The lexicon itself exists independently from the realm of recursion and is not found in other cognitive domains or non-human communication systems (Kinsella, 2010). Regarding (b), the situation is confused. We refer to Parker (2006) for a comprehensive overview on the issue. As Parker (2006) suggests, evidence for recursion in the non-communicative cognitive processes of other species is minimal and remains speculative. Hypotheses about recursion in areas such as kinship and dominance relations, and even more tenuously, in complex action sequences, are still largely unproven. Additionally, non-human communication systems present significant challenges when evaluated for recursive properties. Some interesting results regarding the hypothesis according to which recursion might be present in non-human cognitive abilities have been achieved in a recent work by Ferrigno et al. (2020). These authors investigated recursive abilities using a cross-population design in which they conducted a nonlinguistic sequence generation task to determine if participants could learn to produce center-embedded structures and generalize to novel stimuli. The study included children, U.S. adults, adults from a Bolivian indigenous group, and three monkeys. All the three groups of humans naturally induced recursive structures from ambiguous training data. Monkeys, managed to do that, however, they required additional exposure to achieve similar results. The findings indicate that recursive hierarchical strategies are inherent in human cognition, evident early in development and across different cultures, though the capacity is not exclusive to humans. Nonhuman animals can represent and generate new sequences with recursive, hierarchical, and center-embedded structures. While abstract hierarchical structuring was not the initial strategy for monkeys, two out of three monkeys eventually learned to generalize and generate novel center-embedded sequences with more exposure. Regarding (a), it is interesting to explore whether recursion is a cognitive ability unique to the human language faculty or if it is also present in other cognitive domains. In this context, Hauser et al. (2002) propose that recursion may not have originally developed for linguistic purposes. Instead, its early evolution might have been a response to other challenges faced by our ancestors, such as navigation. Parker (2006) and Kinsella (2010) highlight that recursion appears in several areas of cognition beyond language, including number, navigation, games, vision, social cognition, and music. For a thorough discussion of these domains, refer to Parker (2006). As Parker (2006) explains, numerical reasoning may be closely tied to language from an evolutionary standpoint (Hurford, 1987; Chomsky, 1988), making it challenging to distinguish between numerical and linguistic capacities. Indeed, it has been suggested that numerical ability might have evolved as a by-product of language (Pinker & Jackendoff, 2005; Chomsky, 1988) or that our numerical knowledge could stem from our linguistic skills (Hurford, 1987; Bloom, 1994). In the context of navigation, wayfinding can be viewed as a recursive process, similar to the divide-and-conquer algorithms used in computer science for searching and sorting (Parker, 2006). However, Parker (2006) also notes that humans might employ alternative strategies, such as iterative methods, for path creation tasks. To determine if recursion is essential for these tasks, further research is needed to explore the strategies people actually use (Parker, 2006). Thus, while recursion might be beneficial in navigation, it remains uncertain whether it is indispensable. Wayfinding represents a domain where recursive operations could be advantageous, yet recursion in navigation might be part of a broader cognitive technique rather than a fundamental characteristic of navigation itself (Parker, 2006; Pinker & Jackendoff, 2005). As explained in Parker (2006), Lerdahl and Jackendoff (1983) discuss recursion in music in terms of its grouping structure. They suggest that listeners perceive music as progressively larger units, where smaller units are contained within larger ones. However, this form of hierarchy (i.e. embedding) does not involve self-embedding, which is a key criterion for recursion discussed in this chapter. As Parker (2006) explains, we can state with certainty that music is structured hierarchically, but identifying whether musical phrases are recursively embedded within each other is challenging when only considering strings without reference to the underlying structure. Determining if a musical piece contains nested phrases of the same type is problematic since musical phrases lack semantic content. Hence, it is difficult to ascertain whether musical phrases repeated in a musical piece is the result of tail recursion or simply iteration (Parker, 2006). Crucially, however, a very interesting case of recursion present in music is presented in Hofstadter (1980) when talking about key change modulation. In simple musical pieces, it is proposed that listeners process key changes using a shallow memory stack, storing and resolving tonic keys as the music modulates. A piece of music might start in one key and then modulate to another key midway. During this modulation, the listener needs to remember the original tonic key throughout the section played in the new key. When the music eventually returns to the original key, the listener retrieves the stored tonic key from memory, by popping it from the stack. This process is analogous to the nesting of linguistic phrases, in the sense that a musical key is nested within another (Parker, 2006). In more complex compositions, such as Bach's Little Harmonic Labyrinth, key modulations are frequent and rapid, often leaving listeners disoriented regarding the tonic key (Parker, 2006). Hofstadter (1980) suggests that this indicates a limit to the level of recursive embedding humans can process in music, similar to the limits of processing deeply nested linguistic structures. Crucially, hence, unlike navigation, the complex key modulations in Bach's music appear to necessitate recursive processing, where simple iteration might not suffice (Parker, 2006). Regarding vision, Pinker and Jackendoff (2005) suggest that the way we visually group objects and decompose them into parts might offer evidence for recursion beyond language. Object recognition involves comparing an object to a stored mental representation. When we encounter a scene, our visual system assigns meaning by recognizing each object within it. This is achieved by breaking down the objects into smaller parts through a bottom-up process, which continues until the parts cannot be further divided. We then mentally reconstruct these parts, assigning meaning to each one. Once all objects are recognized in this manner, the entire scene is understood. This procedure is undoubtedly recursive and specifically a case of nested recursion (Parker, 2006).

Summing up, as Parker (2006) explains, in some cases, recursive interpretations of cognitive processes are not the only possible interpretations. In some domains outside language, recursion could be at work but is not necessary, suggesting that recursive strategies might have been a later addition, potentially evolving from capacities initially developed for language. Examples of optional recursion include navigation and games (Parker, 2006). Therefore, in these cases, the presence of optional recursion outside language would not negate the central role or distinct relevance of linguistic recursion proposed by Hauser et al. (2002). However, as Parker (2006) explains, there are clear instances of necessary non-linguistic recursion in human cognition, such as in music (Bach's embedded key changes), visual perception, social cognition, and theory of mind. For further details on social cognition and theory of mind, refer to Parker (2006).

In the next section, we will delve into a crucial aspect that characterizes syntactic recursion in human language: the indissoluble relationship between linear order and structural (i.e., hierarchical) dimensions. Indeed, a key characteristic of recursion in human syntax is that recursive hierarchical structures emerge from temporally ordered, fading sequences of stimuli. Crucially, as we will see, this feature links the recursion found in language to that of key change modulation in music: Both the two instances of recursive processes arise from sequentially, temporally fading sequences of elements. On the contrary, this is substantially different from the type of recursion observed in vision, which, as we will explore in detail in Chapter 3, arises from *static* (i.e., spatial) elements rather than *sequential* (i.e., temporal) ones.

2.2. Linear order and hierarchical structure in human language

Space and time are the framework within which the mind is constrained to construct its experience of reality.

Immanuel Kant.

In this section, we will delve into the issue of linear order and hierarchy in syntax and how these two facets intertwine in the production and comprehension of human language. Syntax is a complex system that encompasses both sequential and hierarchical aspects within its structure. On one hand, a pure linear sequence of words, on its own, falls short in adequately capturing the intricacies of human language[11]. Language, as a tool for expression and communication, is significantly more complex than a simple sequence of words arranged linearly. A purely linear order of words fails to capture the complexities of the principles underlying human language. The influence of a word on others, even when it appears far apart in the sequence, underscores the necessity of adding another dimension to accurately represent the intricate network of syntactic relationships. (Moro, 2016). Indeed, the general consensus is that rigid, linear positions do not constitute the foundation for syntactic rules. The attributes of human language extend beyond mere word order, encompassing phenomena like long-distance dependencies. In essence, it is widely accepted that a solely sequential arrangement of words falls short in capturing the nuanced and intricate underlying principles that govern the syntax of human language. On the other hand, however, the relevance of the linear order of human language cannot be denied either. Sequentiality, understood as the temporal unfolding of language, is a fundamental dimension of human language. "[…] when looking at the structure of a sentence in a human language we may be struck by the strange analogy with snowflakes: minimal components, combined with simple rules that are recursively applied, give birth to geometric patterns of great complexity. The major difference with respect to snowflakes is that sentences must undergo a process of linearization that flattens out the hierarchical bidimensional structure into a linear one […]" (Moro, 2016, p. 28). The role of linear order in language has been a topic of diverse perspectives among scholars. Notably, some researchers argue that linear order should not be regarded as an intrinsic element of the language's core (Berwick and Chomsky, 2017; Chomsky, 2020). Instead, they claim that it functions as a distinct process, bridging the internal language system with sensory-motor systems. In this view, linear order is distinct from the fundamental essence of language. In other words, linear order does not significantly influence the conceptual-intentional level or contribute directly to semantic interpretation. Consequently, properties like linear order and other aspects tied to externalization are best understood as external to the internal language system, often denoted as I-

[11] Nonetheless, it is crucial to highlight that not all facets of language are hierarchical. Certain linguistic processes are strictly local from a computational perspective, relying solely on linear mechanisms instead of hierarchical structures (Culicover, 2013).

language. Chomsky (2020) further elucidates that while auditory perception adheres to linear order, the cognitive process responsible for constructing internal linguistic structures largely disregards this linear sequence. This prompts an essential question: Why does linear order persist in language at all? According to Chomsky (2020), the answer lies in the imperative dictated by the sensory-motor system, which has roots that date back millions of years before the emergence of language as we know it today. In summary, the presence of linear order in language is, in essence, a concession to the demands of the sensory-motor system. In contrast, the core of language, guided by the fundamental computational operation known as Merge, inherently produces linguistic structures without a strict requirement for linear order. Merge represents a theoretical process wherein a minimal binary tree is constructed by combining two conceptual entities, X and Y, to yield a novel object Z = {X, Y} (Chomsky, 2013). The resulting element Z, constituted of X and Y, can be the object of subsequent merge operations, thereby giving rise to more elaborate tree structures. Importantly, the result of the merge process is hypothesized to be an unordered pair (Chomsky, 2013)[12]. Consequently, this perspective on language posits that the internal representation of syntax abstracts away from the temporal aspect of word sequences. Thus, according to this view, at the deep syntactic level, there exists no temporal or ordinal information, but only structure (Dehaene et al., 2015). This arrangement enables language to exhibit structure dependence while permitting flexibility in linear sequencing.

Hence, summing up, in the realm of linguistic theory, there has been a long-standing assumption that the relationship between hierarchy and linear order in language is highly flexible. According to this view, the two elements can be freely associated with one another, allowing for considerable variation in the way they interact. Importantly, however, not all scholars agree in attributing the sequential dimension of language a role of secondary importance, relegating it to the sole product of the interface effect with the sense-motor system. One of the most important works that challenged this view is that of Kayne (Kayne, 1994; 2022).

[12] It is important to note, however, the presence of other theories where order is considered as part of the definition of Merge, such as Stabler's formalization of minimalist grammars where linguistic expressions are defined as *binary ordered trees* (Stabler, 1997).

Kayne, indeed, questioned the prevailing notion that hierarchy and linear order in language can exist in a flexible, interchangeable manner. On the contrary, he asserted that there is a rigid connection between hierarchical structure and linear order, emphasizing the essential role of linear order in syntax and underscoring its influence on the human language faculty. As a matter of fact, the strong correspondence between phrase structure and linear order of terminals is most evident in the fundamental principle that heads must always precede their associated complement positions, whereas adjunctions must consistently occur to the left and never to the right. This principle applies not only to adjunctions within phrases but also to adjunctions to heads. The strict rules governing linear order extend to specifiers as well, which Kayne contends are a form of adjunction. Therefore, specifier positions must invariably appear to the left of their associated head, never to the right. Kayne (1994) introduced the *Linear Correspondence Axiom* (LCA), challenging the conventional view that X-bar theory is a primitive component of Universal Grammar (UG). As he explained, the source of all the major properties that have been attributed to X-bar theory is in fact the Linear Correspondence Axiom. In other words, X-bar theory derives from the LCA, and what is primitive in UG is the LCA. Indeed, according to him, X-bar theory, at its core, embodies a collection of properties that exhibit antisymmetry. Crucially, this inherent antisymmetry is, essentially, derived from the foundational antisymmetry present in the linear ordering of terminal symbols, which has three defining properties: it is transitive, total, and antisymmetric. According to Kayne's LCA, linear precedence of terminal symbols is strongly linked to asymmetric C-command. In other words, A precedes B iff A asymmetrically c-commands B. Hence, in a nutshell, Kayne claims that linear/temporal order is part of the core syntax, as opposed to Chomsky's view. Concerning the source of this fact, Kayne (2022) highlights the importance of Merge, which, crucially, would work differently from what has been assumed in previous works (cf. Chomsky, 2013). "[…] Merge should always be taken to form the ordered pair <X,Y>, rather than the set {X,Y}". (Kayne, 2022, p. 10). It is precisely through the concept of ordered pair in the algorithm of Merge that Kayne manages to explain how this mechanism plays a fundamental role in integrating linear, temporal order into core syntax.

“Antisymmetric linear/temporal order is part of core syntax. Temporal order is partly (though not fully) integrated into core syntax via Merge itself. When two elements X and Y are merged, a relative linear/temporal order is assigned to them.” (Kayne, 2022, p. 1). Hence, the distinction between forming ordered pairs <X, Y> versus sets {X, Y} is a critical aspect in the discussion of linear order within core syntax. Importantly, however, as he further explains, the process of forming ordered pairs via Merge creates a *partial* linear ordering within core syntax, signifying relationships between elements but not within their internal substructures. Specifically, it signifies that X is linearly ordered before Y. However, it does not explicitly convey information about the linear order of subconstituents within X or Y (Kayne, 2022).

In summary, the debate surrounding the role of linear order in core syntax is complex and multifaceted. While some linguistic theories, such as Kayne's notion of antisymmetry, emphasize the inherent connection between hierarchical structure and linear order, others, like Chomsky's recent work, propose that linear order becomes relevant only during externalization. In this thesis, we will embrace the idea that linear order is an inner property of language. Specifically, we believe that, in order to understand and fully appreciate the uniqueness and intricacies of human language, it is fundamental to consider language as a whole, integrally in its manifestation, not focusing only on one part of its manifestation. Over the years, scholars have focused and given different weight to specific aspects of language, often taking into consideration, and focusing on a single aspect to explain the peculiarity of language. Conversely, we believe that the true peculiarity of human language cannot be reflected in a single feature but is the result of an interweaving of elements. As we explained, recursion constitutes a specific type of structural-dependence phenomenon in human language. Rather than focusing on the recursive hierarchical structure itself, we believe that the unfolding and inexorable intertwining of recursive hierarchical structures with the linear, sequential, and temporal dimensions represent a distinctive feature of human language. Indeed, human language, being a cognitive phenomenon, would not exist apart from the temporal dimension of the reality in which we are immersed. The importance of integrating the temporal dimension has been taken into account by various scholars dealing with language, in different fields of study, from neurophysiology, cognitive science, to natural language processing. Elman (1990) addressed the issue of integrating the temporal dimension into the representation of human language. His aim was to create connectionist models capable of adequately processing human language. Elman supported the claim that temporal integration is a fundamental and intricate aspect of language and cognition. "Time underlies many interesting human behaviors. […] Time is clearly important in cognition. It is inextricably bound up with many behaviors (such as language) which express themselves as temporal sequences. Indeed, it is difficult to know how one might deal with such basic problems as goal-directed behavior, planning, or causation without some way of representing time." (Elman, 1990, p.179). Despite the importance of the time dimension in language, Elman suggests that linguistic theorists may have overlooked the importance of representing and processing temporal aspects in utterances. They might have assumed that all the information in an utterance is instantly available in a syntactic tree. However, findings in natural language parsing indicate that solving this problem is not as straightforward as presumed. "[…] what is one of the most elementary facts about much of human activity -that it has temporal extend- is sometimes ignored and is often problematic." (Elman, 1990, p. 180). "There are many human behaviors which unfold over time. It would be folly to try to understand those behaviors without taking into account their temporal nature." (Elman, 1990, p.207). In the same vein, already 30 years before Elman's work, albeit with a different objective, the eminent US neuropsychologist Karl Lashley, pointed out the paramount role that temporality plays in language, such as in other cognitive domains, like music and complex motor actions, highlighting the primary importance of the problem of temporal integration. "I have chosen to discuss the problem of temporal integration here, not with the expectation of offering a satisfactory physiological theory to account for it, but because it seems to me to be both the most important and also the most neglected problem of cerebral physiology." (Lashley, 1951, p. 114). Lashley proposed that the human ability to deal with actions that unfold in a sequential manner, be it language, music, or motor skills, poses significant challenges to our comprehension of brain function. Crucially, he pointed out a crucial flaw in existing models that relied on a simplistic stimulus-response chain. While these models could capture the sequential nature of routine actions, they overlooked the essential role of sustained goals and subgoals in more intricate tasks. In complex actions, overarching goals must persist while subgoals are initiated and completed. Such action sequences deviate from a chaining model where each completed action triggers the next one, and any omission disrupts the entire sequence. Lashley argued that these observations highlight the necessity for models of complex action sequences. Thus, he emphasized the notion that cognitive functions, such as language and music, represent complex structured phenomena intricately intertwined with temporality, a concept widely accepted in contemporary discussions. (Lashley, 1951; Fitch, Martins, 2014). Indeed, his insights have been confirmed and expanded upon by subsequent studies. Interestingly, Fitch and Martins (2014) examined a broad range of contemporary data related to the processing of music, language, and other complex sequential actions, discovering that these findings align closely with a revised neuroanatomical interpretation of Lashley's hypotheses, thus confirming that hierarchy in language and music is constructed upon a foundational sequential action system. Specifically, Fitch and Martins (2014), revisiting Lashley's ideas, shed light on the hypothesis that the similarities in language, music, and complex actions are not coincidental but rooted in the hierarchical nature inherited from basic features of motor planning. In other words, the idea is that hierarchical structuring of temporal sequences is a fundamental ability supporting human music and language, originating from deeper evolutionary roots in action processing. Unlike Lashley, their review utilizes brain imaging and lesion data, particularly examining the role of Broca’s area in processing hierarchical structures in the temporal dimension. Key considerations included whether sequential hierarchy processing is a specific capability, distinct from other types of hierarchical processing, such as visuospatial static hierarchical processing. Moreover, they were interested in verifying which neural mechanisms support this cognitive ability. Importantly, their evaluation incorporated extensive behavioral, neuroanatomical, and brain-imaging data, which were not available during Lashley's era. Fitch and Martins started their investigation by first offering a terminological clarification of the terms *hierarchy*, *set* and *sequence*. “[…] a set is an unordered collection of distinct, unique objects;

while a sequence is a collection of objects, perhaps including duplicates, ordered by some rule. Although the set {a, b, c} is identical to the set {b,c,a},the sequences [abc] and [bca] are distinct and different. Furthermore, because sequences but not sets can contain duplicate items, sequences are not, strictly speaking, a type of set." (Fitch and Martins, 2014, p. 88). "Hierarchy denotes a set or sequence of elements connected in the form of a rooted tree (a connected acyclic graph, in which one element is singled out as the root element). Hierarchies thus possess the following key properties: (1) all elements are combined into one structure (connectedness); (2) one element (the root) is superior to all others; and (3) no element is superior to itself (that is, there are no cycles, direct or indirect)." (Fitch and Martins, 2014, p. 89). This terminological clarification is indeed crucial for investigating temporally ordered hierarchies, where order is significant, in contrast to static hierarchies, where order is typically irrelevant. Indeed, as they explained, the commonality identified across language, music, and action is the imperative to establish the correct temporal ordering on subelements. "[…] temporal hierarchies incorporate an additional ordering component, where at least some elements at any given level represent a sequence rather than a set." (Fitch and Martins, 2014, p. 89). Fitch and Martins (2014) explored whether there is evidence for abstract and modality independent hierarchical structuring in the brain. By searching the literature, they aimed to (i) find evidence supporting the neural distinction in the processing and representation of static hierarchical structures on one hand, and temporal structures on the other. Furthermore, (ii) they sought compelling evidence indicating common neural substrates involved in processing diverse types of hierarchical sequences, encompassing language, music, and action. Firstly, they observed that recent studies have delved into the brain regions involved in processing hierarchical sets in various domains (Kravitz et al., 2011; Kumaran et al., 2012). In the social domain, the hippocampus seems to encode dominance relationships, while also being active in encoding hierarchical ranks in nonsocial domains (Kumaran et al., 2012). As a matter of fact, in the visuospatial domain, correct integration of landmarks recruits the parahippocampus (Aminoff et al. 2007), along with the medial temporal lobe (MTL) and the retrosplenial cortex (Kravitz et al., 2011). Crucially, The MTL system is proposed to encode high-order hierarchical associations in motor and linguistic domains as well (Meyer et al., 2005; Opitz & Friederici, 2003; 2007; Schendan et al., 2003). In contrast, temporal hierarchical processing, whether in music, language, or other tasks, consistently activates the posterior prefrontal cortex, particularly Broca's area (BA 44/45), in the left emisphere (Amunts et al., 2010). Increased Broca's activation correlates with higher working memory demands, such as processing long-distance dependencies. Also in musical syntax, neuroimaging studies consistently reveal activation in BA 44/45, in both hemispheres, but in some cases the activation tends to lean toward the right side (Brown et al., 2006; Fadiga et al., 2009; Koelsch et al., 2000; Maess et al., 2001; Patel et al., 2008; Sammler et al., 2011). Further confirmation of the function of the Inferior frontal Gyrus (IFG), which includes Broca’s area, in music processing and memory have been reported by several studies (Herholz et al., 2012; Janata, Parsons, 2013; Koelsch, 2013; Patel, 2013). Moreover, as far as motor action is concerned, some studies presented supporting evidence indicating that Broca's area has a specific function in the planning of hierarchically structured action (Dehaene et al., 1997; Koechlin, Jubault, 2006). Interestingly, Fitch and Martins (2014) reported that, regardless of the input domain, whether auditory or visual patterns, even nonlinguistic visual symbols (Bahlmann et al., 2009) evoke Broca's activation when sequentially presented. Hence, they concluded that, (i) despite overlaps in the medial temporal lobe, the neural mechanisms for processing hierarchical sequences do not completely coincide with those for hierarchical sets (Fitch and Martins, 2014); (ii) neuroimaging findings substantiate Lashley's hypothesis concerning a shared foundation for music, language, and certain action planning processes. Specifically, Broca's area is identified as a pivotal component in this interconnected system. This last evidence aligns with what Koelsch (2012) has referred to as the *syntactic equivalence hypothesis* (Fitch and Martins, 2014). In conclusion, building upon these pieces of evidence, Fitch and Martins (2014) put forth the following hypothesis: The cortical resources within the Inferior Frontal Gyrus (IFG), encompassing at least BA 44 and BA 45, act as a storage buffer that can be scanned by other cortical and subcortical circuits involved in sequential behavior. This buffer is essential for executing supra-regular hierarchical sequence processing, and the processing load intensifies with the depth and complexity of the hierarchy under consideration. Regarding the origin of this exceptional human ability to process hierarchical sequences, Fitch and Martins (2014) proposed that it might have roots in a pre-existing form of action syntax, predating the development of human music and language. As they explained, Lashley's idea regarding structured phenomena intricately intertwined with temporality, viewed through a modern lens, is associated with widespread brain circuits, primarily located in prefrontal regions, notably Broca's region. These prefrontal areas play a foundational role in the hierarchical planning and sequencing of actions, a function likely shared with other primates, particularly chimpanzees, which are known for their intelligent use of tools (Fitch, Martins, 2014). However, as they explained, while this capability likely originated in primates, its scope would have expanded significantly during human evolution to encompass both perception and the production of various hierarchical sequences. According to their hypothesis, the substantial enhancements to this Broca-centered action sequencing capacity would have been instrumental in the emergence of both music and language.

In this section, we have explored the debate surrounding the relationship between hierarchy and linear order in language. As we have seen, different linguistic theories present contrasting views on this issue. In particular, we focused on scholars who have considered the importance of the sequential, linear, and temporal dimension in language, in addition to the structural hierarchical dimension. We started with the work of Kayne, who challenged the commonly held belief that hierarchy and linear order in language are two separated levels. Instead, he argued for a rigid connection between hierarchical structure and linear order, asserting the crucial role of linear order in syntax and highlighting its profound impact on the human language faculty. Specifically, Kayne posited that linear/temporal order is an integral component of the core syntax, in contrast to Chomsky's perspective, who proposed that linear order becomes relevant only during externalization. Afterwards, we briefly reviewed the ideas of scholars who, in different fields of study, have advocated for the importance of integrating and taking into consideration the sequential, thus temporal dimension when dealing with the study of language. Specifically, we focused on Lashley’s work, who proposed that cognitive functions, like language, music, and certain types of motor actions, are complex structured phenomena in which the temporal dimension plays a major role. After that, we delved into the reexamination of Lashley's ideas by Fitch and Martins (2014). Interestingly, they validated Lashley's insights, by reviewing and examining contemporary data on music, language, and action processing. They proposed that the similarities in these domains stem from a shared ability to deal with sequential hierarchical structures, rooted in motor planning. Importantly, their study, reviewing brain imaging and lesion data, highlighted Broca's area as a key player in processing sequential hierarchies. Moreover, they concluded that neural mechanisms for processing sequential hierarchies differ from those for static hierarchies. Specifically, the Inferior Frontal Gyrus, particularly BA 44 and BA 45, was suggested as a crucial cortical resource for executing hierarchical sequence processing, with the processing load increasing with hierarchy complexity.

2.2.1. From sequence to hierarchy: Cognitive mechanisms at play

At the core of language lies the essential capacity to process and hierarchical structures arising from sequentially ordered stimuli. As demonstrated in the preceding section, this cognitive skill is not only pivotal in the realm of language but also extends its influence on other cognitive domains, such as music and certain types of complex motor actions. Building upon the compelling evidence that underscores the significance of both hierarchical structure and temporal sequential order in language, as well as in other cognitive faculties, we want to shed light on how cognition derives hierarchical patterns from sequentially presented input.

Dehaene and colleagues (2015) introduced a taxonomy that classifies diverse forms of cognitive internal representations that can arise when processing a sequence of temporally distributed stimuli. As they explained, a sequence of stimuli, can be processed and stored in different ways, at different levels of detail. Specifically, they put forth a classification system outlining five distinct cerebral mechanisms for coding sequences, with increasing degree of abstraction: (i) transitions and timing knowledge; (ii) chunking; (iii) ordinal knowledge; (iv) algebraic patterns; and (v) nested tree structures. For each mechanism, they examined existing experimental paradigms and outlined their behavioral and neural markers.

(i) *Transitions and timing knowledge* refers to the knowledge of the shifts from one item to the next, encompassing the identification and approximate timing of the subsequent item in relation to the previous item; in other words, it refers to the capacity to depict the temporal gaps between sequential elements and utilize these temporal representations in basic computations. Several studies have reported that sensory circuits can internalize the temporal patterns of regular sequences, generating an endogenous response even in the absence of sensory input, solely in anticipation of an anticipated event. An additional notable trait of temporal sequence encoding is its automatic nature. Hence, *transition and timing knowledge* is an initial, basic stage of sequence representation, in which sequences are stored by recording the transitions between items and their approximate timing. At this stage, the processing mechanism operates at an item-specific, superficial level.

(ii) *Chunking* involves grouping multiple consecutive items into a unified entity, which can then be manipulated as a cohesive unit at the subsequent hierarchical level; upon the recurrence of a sequence of elements, these elements can be represented as a cohesive entity referred to as a "chunk," consolidating them into a singular unit for storage. Hence, a "chunk" can be defined as a set of adjacent items that consistently reoccur together, allowing for subsequent manipulation of this element as a singular entity.

(iii) *Ordinal knowledge* entails the knowledge of the elements' order in the sequence, distinguishing the first item from the second and so forth, irrespective of their temporal arrangement; hence, this mechanism abstracts from precise timing details and focuses solely on the relative temporal order of elements, discerning which of them comes first, second, or third. As Dehaene et al. (2015) explain, having distinct mechanisms for timing and ordinal knowledge proves beneficial. In situations where event timing is fixed and predictable, timing mechanisms are essential. However, in scenarios where it is possible to predict that something will happen or how many events will occur without knowing when, the utility of ordinal knowledge becomes apparent.

(iv) *Algebraic patterns* involve the mental representation of more abstract properties which capture the relationships between successive stimuli. Through this mechanism, an input sequence is internally coded by a corresponding sequence of abstract relationships, concepts, or categories. Hence, this mechanism is characterized by the ability to abstract from the specific identity and timing of sequence items and grasp the pattern underlying them, assigning items to abstract categories. The categorization of stimuli can be based on concepts such as sameness and difference. For example, the word "papavero" can be represented through the abstract pattern schemas AABC that match the sequential regularities within syllables, consisting of the repetition of a syllable followed by two different ones.

(v) *Nested tree structures* involve the embedding (or nesting) of groups of items (i.e., chunks) within each other, forming a hierarchical structure of any depth. This mechanism allows for an underlying mental representation forming tree structure, in which sequences of items (i.e., chunks) are linked together. Crucially, this process might involve the recursive utilization of the same elements at different levels; hence, recursive hierarchical structures are a specific type of nested structures.

As Dehaene and colleagues (2015) explain, cognitive phenomena such as language, music, motor action, and mathematics cannot be accounted for by flat mechanisms such as (i) - (iii) but require the formation of algebraic patterns (iv) and nested tree structures (v). Indeed, as they observe, already in the 1970s, Restle and colleagues put forward the fact that, in these complex sequential cognitive phenomena, basic sequences are grouped and represented in ways which transcend a simple, flat associative chain, hypothesizing the necessity of the representation of abstract tree structure (Restle, 1970; Restle and Brown, 1970). An important point worth discussing is related to the mechanisms involved in the formation of nested tree structure. As we have seen, Dehaene et al. (2015) emphasize that the formation of hierarchical structures might potentially involve the recursive utilization of the same elements across different hierarchical levels. "[…] at this level, characteristic of human languages, a sequence can be ''parsed'' according to abstract grammatical rules into a set of groupings, possibly embedded within each other, forming a nested structure of arbitrary depth, and possibly involving the recursive use of the same elements at multiple levels." (Dehaene et al., 2015, p.2). The application of recursive mechanisms in the formation of nested hierarchical structures is an ability at play in the human language faculty, as discussed in this chapter. However, it is important to highlight the fact that not every hierarchical phenomenon is recursive. In fact, the utilization of a recursive procedure is possible but not required for the formation of hierarchical structures. Indeed, as Fitch and Martins (2014) rightly emphasize, while every recursive tree is hierarchical, not all hierarchies are recursive. “For many types of hierarchy […], such as motor actions embedded within plans, it is unclear what self-embedding would even mean.” (Fitch and Martins, 2014, p.98). Crucially, however, language, like music in specific circumstances, such as the processing of key change modulation (Hofstadter, 1980; cf. Section *2.1.2.*) entails the formation of a particular type of hierarchical structure arising from sequentially ordered stimuli, namely *recursive* hierarchical structures. Indeed, both syntactic structures and key change modulation in music can be not only hierarchical but also recursive, thus allowing for self-embedding (Chomsky, 1995; Hofstadter, 1980; Karlsson, 2010; Mithun, 2010; Roeper, 2009). More specifically, considering that in both language and music, sequential order is intricately bound to hierarchical structure, as we have discussed in this chapter, we believe it is more accurate to refer to these structures as recursive structures arising from temporally ordered sequences of stimuli. This is indeed the terminology we adopt throughout the present investigation.

Taking these facts into consideration, the present thesis aims to delve deeper into the intricate relationship between sequence and hierarchy. Our specific goal is to elucidate the cognitive mechanisms underlying the transition from linear order to recursive hierarchical structure during the processing of sequences of stimuli featuring recursive hierarchies. How does cognition derive recursive hierarchical patterns from sequentially presented input? This investigation may provide insights into how our mental capabilities and constraints, operating within the constant framework of space and time, shape the way we acquire and process recursive hierarchical structures from a temporally ordered fading sequence. This process, serving as a cognitive mechanism underlying the human language faculty, could be a window that helps us understand more about the mechanisms and structures of language. Language, indeed, being a byproduct of the human mind, is inherently shaped and constrained by our cognitive boundaries. In Chapter 5, we will explore this cognitive ability in various sensory domains. This investigation could illuminate the question we discussed in Section *2.1.2.*, specifically regarding the possibility of finding this ability outside language. It might suggest whether this ability is strictly linked to language or if it represents a domain-general cognitive skill. One possibility is that the ability to process recursive hierarchical structures from sequential stimuli evolved in other cognitive domains and was later exapted for use in language. In other words, the capacity to handle recursive hierarchical structures in temporally ordered stimuli may have initially developed for purposes outside language and was subsequently co-opted for linguistic use.

In literature, numerous studies have provided evidence for the cognitive representations proposed by Dehaene et al. (2015) regarding the processing of temporally distributed stimuli. Several investigations, employing behavioral paradigms and neuroimaging techniques, have explored these cognitive mechanisms (i-v) individually (cf. Dehaene et al., 2015). Interestingly, some of these mechanisms (i.e., i; ii; iii; iv) have been observed in both animals and humans. However, when it comes to the formation of nested tree structure (v), Dehaene and colleagues explain that it necessitates a specific recursive neural code, which remains undiscovered through electrophysiological methods. Moreover, they suggest that (v) might be a cognitive mechanism exclusive to humans, providing insights into the unique nature of human language and cognition. Overall, these studies demonstrate the coexistence of multiple systems in different brain circuits for learning and processing sequential information at varying degrees of complexity and abstraction. However, while the existing body of research has predominantly provided evidence for these cognitive abilities when examined in isolation, only a limited number of studies directed their attention to the comprehensive exploration of the entire journey from sequence to hierarchy. Hence, there remains a notable gap in our understanding of how these intricate processes interact and unfold throughout the entire cognitive continuum. In essence, the current state of research has been more inclined toward dissecting these abilities individually rather than unraveling the dynamic interplay that occurs during the transition from sequence to hierarchy. With respect to this, Dehaene et al. (2015) express the need to comprehend how the brain determines the optimal mechanism model for processing a given sequence. They ponder whether these mechanisms engage in a competition to minimize prediction errors in sensory input until one of them effectively succeeds in predicting it and blocks the others. Alternatively, all systems might operate independently, each striving to capture different aspects of the input sequence. The latter hypothesis gains support from experimental findings indicating that local transition probabilities are extracted independently of coexisting knowledge of the global sequence (Bekinschtein et al., 2009; Wacongne et al., 2011). Dehaene et al. (2015) acknowledge that future studies will be necessary to elucidate this aspect.

Crucially, some recent studies have further explored the cognitive mechanisms at the core of the transition from sequence to hierarchy, investigating the relationship between different cognitive mechanisms at play in encoding sequential stimuli and addressing some relevant research questions (Radulescu et al., 2019; Planton et al., 2021; Schmid et al., 2023; Vender et al., 2023).

Radulescu and colleagues (2019) investigated the factors that drive the shift from memorizing specific items and statistical patterns to forming more abstract representations when exposed to sequentially arranged stimuli, while also exploring the mechanisms that lies at the heart of this transition. Based on the work of Gomez and Gerken (2000), they identified two distinct forms of rule induction: item-bound generalizations and category-based generalizations. “An item-bound generalization is a relation between perceptual features of items, e.g. a relation based on physical identity, like ba-ba (ba follows ba), or “add – ed”. Category-based generalization operates beyond the physical items; it abstracts over categories (variables), e.g. Y follows X, where Y and X are variables taking different values. In natural language, the grammatical generalization that a sentence consists of a Noun-Verb-Noun sequence is based on recognizing an identity relation over the abstract linguistic category of noun (which can be construed as a variable that takes specific nouns as values).” (Radulescu et al. 2019, p. 109-110). Specifically, Radulescu and colleagues inquired whether these forms of encoding are different outcomes of a single mechanism or outcomes of two separate mechanisms. If they are indeed products of the same mechanism, do the two types of generalizations represent stages in a phased process that gradually shifts from lower-level item-bound generalization to a higher-order abstract generalization, or do they result from an abrupt switch between separated mechanisms? Additionally, what initiates this change in the form of encoding? Based on the hypothesis put forward by Aslin and Newport (2012), which proposed that statistical learning serves as the underlying mechanism for both item-bound and category-based generalizations, Radulescu and colleagues (2019) introduced a novel entropy-based model, based on an information-theoretic perspective, to consistently explain how a single mechanism can produce two distinct forms of generalization, the type of context cue distribution that lead to the different forms, and the reasons why the same mechanism can yield different outcomes. These fundamental questions had not been addressed by previous research. The fundamental idea behind their model is that input complexity, measured by the information-theoretic concept of entropy, triggers the shift from item-bound to category-based generalizations. Essentially, entropy quantifies the complexity of a set of items, influenced by both the number of items and their frequency distribution: Entropy rises with an increase in the number of items and with a more uniform frequency distribution among them. In this context, entropy can also be understood as the uncertainty or surprise regarding the occurrence of specific items or their configurations. Both the number of items and their frequency distribution contribute to this uncertainty. An additional factor is crucial to their model, which posits that rule induction functions as an encoding mechanism: Channel capacity (Shannon, 1948). Channel capacity refers to the maximum amount of entropy that can be transmitted through a channel within a given time frame. Radulescu and colleagues hypothesized that cognitive channel capacity is modulated by factors such as attention, memory capacity, and pattern-recognition capacities. Hence, according to them, the encoding mechanism governing sequence processing is naturally and gradually driven by the brain's sensitivity to input complexity (entropy) and interacts with the brain's limited encoding capacity (channel capacity) (Radulescu et al. 2019). Therefore, according to Radulescu and colleagues, based on the level of input complexity and the limited encoding capability (i.e., channel capacity), various methods of information encoding are required to handle the complexity of the input. Specifically, their model predicts that as the complexity of the input increases, the brain is more likely to move away from single item-bound computations and infer abstract rules to form category-based generalizations. To test their model, the researchers conducted two artificial grammar learning (AGL) experiments with adults. These experiments focused on how input complexity affects rule induction. The results confirmed that as the complexity of the input increased, participants were more inclined to make category-based generalizations. In other words, Radulescu and colleagues' finding supports the hypothesis that higher entropy in input data triggers a stronger inclination towards rule generalization: Their study demonstrated that the tendency to form abstract rules increases with the complexity of the input. This observation supports the notion that higher entropy prompts the brain to abstract and generalize rules more effectively. Importantly, moreover, the entropy model proposed by Radulescu and colleagues (2019) explains how both item-bound and category-based generalizations can arise from the same cognitive process, driven by input complexity and the brain's encoding capacity. By modeling the gradual transition from item-bound to category-based generalizations as a function of input complexity and finite brain capacity, Radulescu and colleagues' research provides a comprehensive understanding of the cognitive mechanisms underlying the transition from sequence to hierarchy, which, as they state, represents a fundamental ability at the core of language learning.

Planton et al. (2021) aimed to shed light on the cognitive mechanism that lies at the heart of the capacity to build recursive nested structures in sequential stimuli. Indeed, as they explain, when dealing with linearly arranged stimuli, humans are not only able to detect sequential features by exploiting statistical learning mechanisms, but they are also able to create more abstract representations, such as nested recursive structures. In other words, humans possess the ability to organize sequences into a hierarchical structure of smaller chunks within larger chunks, in a recursive way (Planton et al., 2021). Based on this evidence, Planton and colleagues tested the hypothesis according to which human can form complex recursive abstract representation even when exposed to binary sequences, that is, sequences composed of only two different symbols (e.g., A and B). Specifically, according to them, they would manage to do that by exploiting a cognitive strategy based on an abstract internal language – i.e. language of thought (LoT) - featuring a recursive compression mechanism. Indeed, according to their hypothesis, participants, when exposed to sequences of stimuli, would spontaneously recode them in an abstract form exploiting an internal LoT. Specifically, the proposed language is intended as a systematic framework capable of encoding any arbitrary combinations of nested repetition and alternation structures. In essence, this language consists solely of two basic commands – i.e, "same" and "change"- and their recursive embeddings. Such a language would enable the integration of simple primitives into intricate nested patterns or recursive rules (Planton et al. 2021). As the authors explain, creating a Language of Thought (LoT) model for sequence representation requires choosing a set of rules or operations that enable the (lossless) recoding of any sequence. This language would prioritize an abstract depiction of sequences by maximizing nested repetitions. It follows that sequence length and complexity clearly become two distinguished measures. Indeed, compressing the stored information into a more concise form would enhance working memory capabilities. Their hypothesis suggests that the mental complexity of a sequence is directly related to the length of its shortest representation in the proposed internal language. To test thus hypothesis, they carried out five experimental studies using binary sequences. As the authors explain, binary sequences, due to their simplicity, provide a controlled environment to study sequence memory without the confounding variables present in more complex sequences like language or music. Despite their simplicity, binary sequences can generate complex patterns that require complex mental encoding strategies. Understanding how these sequences are encoded can shed light on broader cognitive processes involved in memory and learning. In their experimental studies, Planton and colleagues presented participants with binary sequences of visual or auditory stimuli in a violation detection task. After an initial exposure phase where participants learned the sequences, they were tested with sequences that included single-item deviations. Participants were tasked with quickly identifying whether a presented sequence contained a violation. To systematically vary the complexity of these sequences, Planton and colleagues exploited the formal language they developed composed of a limited set of primitive instructions. These instructions allowed them to generate different binary sequences and measure their complexity in terms of Kolmogorov complexity, which is defined by the length of the shortest possible description (or program) that can produce the sequence. Hence, in their study, the complexity of each binary sequence was determined by the minimal number of instructions required in the formal language to describe it. The study also aimed to determine whether participants' memory performance was better explained by this compression mechanism or by simpler statistical learning processes. Hence, to separate the effects of compression from statistical learning, the researchers measured the Shannon surprise of each deviant item in the sequences. Shannon surprise quantifies the uncertainty of observing a specific item given the history of previous items, reflecting the degree of statistical learning. Importantly, Shannon surprise is independent of the overall sequence complexity. The study found that both the complexity of the sequences (as defined by Kolmogorov complexity) and the Shannon surprise of the deviant items were significant predictors of participants' performance. This indicates that participants used both compression and statistical learning to process the sequences. Specifically, across five different experiments involving sequences of varying lengths in both auditory and visual modalities, consistent evidence was found that a significant portion of the variation in sequence encoding performance (measured by the ability to detect sequence violations) was explained by the length of the shortest possible description of the sequence in the proposed formal language (i.e., LoT complexity). Interestingly, however, this effect was not observed for very short sequences (6 items) but was most pronounced for longer sequences, in particular for the longest ones (16 items). The authors attribute this result to differences in working memory demands. Indeed, as they explain, the number 6 falls within the typical range for items that can be stored in working memory without compression, which is about 7±2 items (Miller, 1956; Mathy & Feldman, 2012). Therefore, participants could have solved the violation detection task by storing each 6-item sequence in working memory without compression. Similarly, 8-item sequences could have been stored as a flat series of "chunks," which are considered the units of encoding in working memory, without any recursive embedding. Overall, the increasing need for compression explains why the predictive power of LoT complexity grew with sequence length. Furthermore, follow-up analyses revealed that the complexity measure derived from the language they designed better predicted the degree of psychological complexity compared to other sophisticated approaches available in the literature. “Our results support the idea that the inclusion of such a feature is essential to explain human behavior when working memory capacity is exceeded and compression is most beneficial. The fact that we reached such a conclusion using the simplest type of temporal sequences (binary sequences) and a simple deviant detection task (rather than the more demanding recall, completion or production tasks using in the previous literature) is consistent with Fitch's "dendrophilia hypothesis" [Fitch, 2014] which states that "humans have a multi-domain capacity and proclivity to infer tree structures from strings" even in the simplest cases" (Planton et al. 2021, p. 27). In conclusion, Planton et al. (2021) offer a compelling framework for understanding how humans encode and remember binary sequences. By demonstrating that both compression and statistical learning contribute to sequence memory, the study provides valuable insights into the cognitive processes underlying the processing of sequential sequences. However, as Schmid (2023) observed, while the study provided strong evidence for the use of compression in sequence memory, it did not measure the degree of compression achievable by participants. Sensitivity to sequence complexity, as indicated by performance, does not necessarily mean that participants compressed the sequences to the maximum possible extent or that the formal language used in the study perfectly mapped onto participants' mental operations (Schmid, 2023). With this regard, Planton and colleagues acknowledged a limitation in their study, noting that approximately half of the minimal expressions for the sequences they used involved only two hierarchical levels, i.e., a single level of recursive embedding. Hence, they explained that further research is required to determine whether human participants consistently find deeper levels of embedding advantageous, especially when processing short sequences. As they observed, increasing the hierarchical depth might entail an additional cognitive load, making it beneficial only in particular contexts, such as more complex learning tasks or longer sequences (Planton et al. 2021). Another open issue, as Planton et al. (2021) explained, is related to the domain-general or domain-specific nature of the cognitive mechanisms they explored. Indeed, Planton and colleagues proposed that mental compression of sequences occurs at an abstract level, focusing on relationships between items rather than at the sensory level. Their approach effectively predicted the psychological complexity of both tone and visual sequences, implying an abstract symbolic representation. However, as the authors explained, it is debated whether temporal sequence encoding involves a universal mechanism or separate modality-specific systems. One hypothesis they put forward is the presence of an auditory-specific system. Visual sequences might be converted into auditory representations before compression. This would be in line with the results they observed in their study regarding the lower performance and slower responses in visual tasks compared to auditory ones. The authors hence concluded by stating that future research should explore and compare further this cognitive mechanism in different sensory modalities, or using cross-modal transfer, and brain imaging to better understand the sensory and cognitive mechanisms involved.

Two other studies which investigated the cognitive mechanisms involved in the transition from sequence to hierarchy are Schmid et al. (2023) and Vender et al. (2023). These authors have thoroughly investigated the interplay between sequence and hierarchy through the Artificial Grammar Learning (AGL) paradigm. They exploited an artificial grammar that, for reasons detailed in Section *4.1.*, serves as an optimal tool for this investigation: the Fibonacci grammar (Fib). Anticipating what will be discussed in Section *4.1.*, Fib is a simple recursive rewrite system comprising just two symbols (0 and 1) and two rewriting rules (0→1; 1→01) [13]. The recursive application of these rules generates strings of potentially infinite length[14]. Crucially, when Fib strings are parsed sequentially, from left to right, some points can be predicted through low-level transitional probabilities applied to the string. Notably, however, the distribution of points in the string is aperiodic. This means there is no linear function capable of predicting when a point will occur, making it impossible to use simple strategies like detecting recurring patterns in order to predict all the points in the sequences (Schmid, 2023). Interestingly, however, different types of points can be potentially predicted through the formation of abstract recursive hierarchical representations. For the features discussed in detail in Section *4.1.*, Fib strings provide an optimal testing pool for sequential statistical learning and the formation of recursive hierarchical abstract representations. Importantly, moreover, they also offer the opportunity to

[13] 0 rewrites as 1; 1 rewrites as 01.

[14] An instance of Fib string: 0110110101101101011010110110101101.

disentangle these two mechanisms and check their possible interaction (cf. Section *4.4*).

In light of the results from these interesting studies (Dehaene et al., 2015; Planton et al, 2021; Radulescu et al. 2019; Schmid et al., 2023; Vender et al., 2023) our goal will be to explore further the cognitive mechanisms underlying the processing of recursive structures arising from temporally ordered stimuli. Specifically, our investigation will pay particular attention to the mechanisms involved in sequential statistical learning and the associated formation of chunks, their categorization, and finally, the representation of recursive hierarchical structures. Crucially, we will seek to clarify their potential interactions. But that is not all. We also aim to investigate potential differences across various sensory domains, delving into the intricate relationship between these cognitive mechanisms and the realm of perception, seeking a comprehensive understanding of their interplay. As we have seen in section *2.2.*, the abilities to process sequential and static hierarchical structures rely on different neural circuits, as confirmed by Fitch and Martins (2014). Regarding the relationship between these types of cognitive mechanisms and perceptual reality in which we are immersed, we can observe that static hierarchical structures are primarily present in the visual domain. For example, when we see an image, our visual system organizes the static information in a hierarchical fashion, enabling us to perceive the entire image, which is composed of numerous hierarchically organized pixels, contributing to our comprehensive perception of the visual scene. “The eye is the only organ that gives simultaneous information concerning space in any detail.” (Lashley, 1951, p.128). Certainly, we cannot do the same in the auditory sphere since auditory stimuli require a temporal dimension for their unfolding. Hierarchical structures arising from sequential stimuli, on the other hand, are predominantly auditory or motoric, think of music, language, and complex motor actions. However, this type of structure can also regard the visual dimension. Consider, for example, the processing of a video. This certainly involves the representation of hierarchical structures from sequential stimuli in the visual sphere. Hence, unlike what occurs with static hierarchical structures, hierarchical structures arising from sequential stimuli can be conveyed through both the visual and auditory sensory domains.

Starting from this observation, the question arises spontaneously: What could be the case of recursive hierarchical structures arising from sequentially ordered stimuli? First of all, would we find evidence of learning and processing of these structures in different sensory domains (i.e. auditory, visual, and tactile domains)? Or are these structures domain-specific since linked to the cognitive domain of language and musical key tone processing (cf. Section *2.1.2.*)? On the contrary, if we would find evidence of learning and processing these structures in different sensory domains, would we find differences between the visual and auditory sensory domains in processing this type of structures? Could it be that hearing has an advantage over sight at processing recursive hierarchical structures arising from sequential, temporally fading stimuli, given that hearing is the specialized domain for processing this type of structure, since we find these structures both in the syntax of language and the key change modulation in music? Vision might be stronger in processing recursive hierarchical structures arising from static, spatially distributed stimuli than the sequential counterpart. What can we say about touch, then? "The shape of an object impressed on the skin can scarcely be detected from simultaneous pressure, but the same shape can readily be distinguished by touch when traced on the skin with a moving point or when explored by tactile scanning. The temporal sequence is readily translated into a spatial concept." (Lashley, 1951, p.128). As observed by Lahsley (1951), touch would thus appear to be more suitable for processing hierarchical structures arising from sequential (i.e. temporally ordered) stimuli compared to static (i.e. spatially distributed) ones.

2.3. Establishing experimental foundations for empirical investigation of the ability to form recursive hierarchical abstract representation

In the second part of this chapter, we will explore the methodologies for studying the formation of recursive hierarchical abstract representations from sequential fading input. We will begin with the concept of implicit learning, focusing on how structured information is implicitly acquired from the environment. Then, we will provide an overview of the main paradigms available for this investigation.

2.3.1. Implicit Learning

In everyday life, we accomplish tasks and fulfill actions spontaneously, without effort, and unaware of the procedure and the subcomponents underlying them. Driving the car, riding the bike, practicing our favorite sport, playing an instrument, or, even more accessible, walking, …*speaking*! All these actions have been internalized so that we do not need to think about them during their repetitions. When driving our car, it is highly improbable that we think about every little movement to run it properly. We do not need to stay focused on the action. We could go through with it and at the same time listen to the radio, sing, or have a conversation with our passengers. Likewise, when having a conversation, we do not think about sentence constructions, verbs - nouns agreements, inversions… We convey ideas and thoughts naturally and spontaneously, with no effort. This is a different process from, for example, learning poetry, the musical scale, the list of European capitals, recalling the names of the wind roses, or the First World War event sequence. Some people would probably say to be better at the first kind of these tasks; on the contrary, others could feel stronger in the second type. These two types of actions are different. Firstly, they are acquired in different manners, and, interestingly, their learning process relies on (at least in part) different neural correlates in the brain. The first set of actions falls under the term *implicit knowledge*, whereas the second is representative of what is called *explicit knowledge*. Over the past 40 years, precisely with the publication of Ullman's Declarative/Procedural Model Theory, also known as the DP Model (Ullman et al., 1977), the term *implicit learning* has often been used interchangeably with *procedural learning*; on the opposite, *explicit learning* has become equivalent to *declarative learning*. Simplifying, we could say that we build up skills through implicit learning and end up (implicitly) *knowing how* to perform them. On the other hand, during the explicit learning process, we gain *knowledge*, and, to go on with the parallelism, we could say it is more like *knowing that* (Goldberg, 2014). To recapitulate, we have said that unawareness, unintentionality, and the inability to express the content verbally, are typical peculiarities of both the implicit learning process and the phase in which we use this knowledge to perform tasks. We end up knowing more than what we can tell (Nisbett, Wilson, 1977). As we will see, the investigation of the capacity to implicitly acquire structured information from the environment has a long tradition. More recently, it has been exposed to a renewal of interest that has flourished in several studies. One of the reasons why the investigation of implicit learning is still an actual and live research field may be that even after decades of inquiry and the critical findings achieved by scholars during these years, it remains an open problem to precisely understand how implicit learning works. It is still beyond our knowledge which are the precise cognitive mechanisms that underly it. The recent renewed interest in the topic has indeed been reawakened after the development of Ullman's Declarative/Procedural Model theory, which highlighted the importance of the procedural memory system in the acquisition of rule-based information and its fundamental role in the implicit acquisition of grammar in language as well as other sensorimotor and cognitive skills. More precise insight into the cognitive mechanisms and correlates underpinning implicit learning might bring numerous benefits in various research fields, among which linguistics. Indeed, it is widely believed that this cognitive ability is fundamental for the acquisition and processing of language. Notably, a better understanding of the mechanisms underlying implicit learning could be beneficial not only to shed light on the mechanisms that stand at the base of language acquisition and processing but also for a more precise insight into some of the possible causes at the base of developmental disorders, such as specific language impairments. A deeper understanding of implicit learning mechanisms constitutes a fascinating challenge, and the goal is not trivial at all.

2.3.1.1. Same objectives, different research traditions: Realigning Implicit Learning and Statistical Learning

In the literature investigating the ability to implicitly acquire structured information from the input, we found many rifts. Too often, we came across bifurcations and divisions between different lines of investigations and between scholars. Over the years, several researchers have reached impressive results, adding new pieces to the complex research puzzle on language acquisition. However, the divisions we find in the literature have also produced a negative effect. We often see different lines of research that, despite investigating very closely related issues, have never communicated with each other. Fifteen years ago, scholars started to notice a significant problem in the literature: the presence of two wholly separated lines of research, which use different terms and occupy different spaces but investigate the very same phenomena, although from different perspectives: *Statistical Learning* (SL) and *Implicit Learning* (IL). "Recent evolution of research on both IL, initially aimed at studying rule abstraction in complex situations, and SL, initially focused on word segmentation, suggests that the two lines of research explore the same domain-general incidental learning processes" (Perruchet and Pacton, 2006, p.237). Starting from these observations, Perruchet and Pacton (2006) first suggested that Implicit Learning and Statistical Learning are two approaches to virtually the same phenomenon, and, in the same year, Conway and Christiansen proposed to recombine them under the term *Implicit Statistical Learning* (Christiansen, 2019). This fact highlights the importance of the issue: in 2019, some scholars dedicated a Special Issue to this topic[15]. "The aim of the special issue is to facilitate the development of a shared understanding of research questions and methodologies, to provide a platform for discussing similarities and differences between the two strands, and to encourage the formulation of joint research agendas. We then introduce the new contributions solicited for this special issue and provide our perspective on the agenda setting that results from combining these two approaches". With these words, the two editors opened the Special Issue (Rebuschat, Monaghan, 2019). One of the points on which all the scholars participating in the Special Issue agreed upon was the urgency of realigning the two approaches: only by developing a joint research agenda, integrating the two perspectives of investigation, and taking into consideration what has been discovered over the years by both IL and SL, the research on Implicit Statistical Learning will be prolific, and it could lead to a better understanding of the complex mechanisms that stand at the foundation of human language.

[15] Topics in Cognitive Science (2019) Vol.11 Issue 3. https://doi.org/10.1111/tops.12438.

Where do Statistical Learning and Implicit Learning have their roots? Which are the divergences and the commonalities between these two fields of research? It is a common idea that Statistical Learning has a much younger tradition than Implicit Learning. On one side, the term Implicit Learning was introduced by Arthur Reber, who carried out Artificial Grammar Learning (AGL) experiments (Reber, 1967; 1969). On the other side, the Statistical Learning stream of research began in the 1980s, but the most influential and well-known experiment was the one conducted by Saffran and colleagues in 1996, which provided new evidence for babies' extraordinary abilities to deal with statistical information in the linguistic input. However, it is interesting to notice, as Christiansen pointed out, that the two traditions can be traced back to previous research, having a longer pedigree (Christiansen, 2019). Indeed, the precursor of Statistical Learning and Implicit Learning has been Erwin Allen Esper, in 1925, who carried out the first Artificial Grammar Learning experiment, intending to investigate the role of statistic information in forming grammatical categories. Esper was a real pioneer in the field. However, most of his work has gone unnoticed during those years; probably, this was partly due to the fact that Esper was a behaviorist (Christiansen, 2019). Later, in 1957, George Miller started to investigate rules formation in his project Grammarama, working within the Formal Language Theory (FLT) (Christiansen, 2019). Subsequently, it happened that the literature split into two separate branches: on one side, Implicit Learning continued to investigate the role of implicitness and memory in this kind of learning process by using mainly the AGL paradigm introduced by Reber, but also other paradigms such as the Probability Learning task, and the Serial Reaction Time task. On the other, Statistical Learning focused more on investigating the ability to uncover the structure of the input by exploiting its distributional properties. Most of these works deployed artificial languages. Within the Statistical Learning approach, exciting experiments have been conducted with babies, and this constitutes an element of novelty with respect to previous research, which had mainly investigated adults' abilities.

Despite their common origins and the similarities of their research agenda, the two approaches show some differences in the way they tackle the issue. Firstly, in their investigation, IL and SL focus on two slightly different computational capacities: on one side, the Implicit Learning literature investigates the ability to select chunks, mainly focusing on rule abstraction and basic learning and memory processes. On the other side, Statistical Learning research focuses on the capacity to exploit transitional probabilities to determine chunk boundaries (Arnon, 2019; Christiansen, 2019; Perruchet, 2019; Perruchet, Pacton, 2006; Rebuschat, Monaghan, 2019). Statistical Learning has always been more focused on investigating the ability to track statistical information in the input. As we have seen in Chapter 1, at the level of syntax, SL has investigated phenomena such as the frequency of syntactic structures, transitional probabilities between words for the segmentation of phrases, dependencies for phrase boundary identification, distributional cues for the formation of syntactic categories, and transitional probabilities between adjacent and nonadjacent dependencies, among the others. Scholars belonging to the Implicit Learning stream investigate chunk-based learning, whereas scholars investigating Statistical Learning study probabilistic learning (Christiansen, 2019). In addition to this, we can notice a slight divergence in their research focus: while Implicit Learning focused more on what mechanisms are involved in the learning process, Statistical Learning focused on the kind of structure that can be learned (Christiansen, 2019). Secondly, from the point of view of syntax acquisition, the two streams of research have traditionally used different tools, and their research focused on slightly different aspects. The Implicit Learning paradigm has mainly employed the Artificial Grammar Learning paradigm (AGL), testing the learnability of formal grammars; hence its paradigm was strictly interconnected with the Formal Language Theory (FLT). Numerous experiments in IL have also combined the AGL paradigm with the Serial Reaction Time Task (SRT). On the other hand, experiments in SL have mainly been conducted by creating sets of artificial sentences, which were constructed in such a way to provide an optimal research environment for the investigation of a specific, circumscribed phenomenon. However, especially in the more recent years, several scholars belonging to the SL research field have started to use tools and methodologies employed initially by IL research (Perruchet, Pacton, 2006). Specifically, some experiments have been carried out using the AGL paradigm (cf. Saffran and

Wilson, 2003) or the Serial Reaction Time task (cf. Hunt, 2002; Hunt and Aslin, 2001).

However, as we said, besides the divergences, it is essential to note that these two approaches share several commonalities and that, in the end, they investigate the very same phenomena, despite analyzing them from different angles. Nevertheless, the two approaches never conflated into a unique line of research (Christiansen, 2019). In support of the fact that the two lines of research remained largely separated over the years, we can observe that Statistical Learning and Implicit Learning studies have always been presented at different international conferences and published in different journals[16].

To summarize, we have seen that in psycholinguistics research, besides the Statistical Learning approach, we find another stream of research that focused on the investigation of humans' abilities to extract information from complex stimuli in the environment implicitly: Implicit Learning. We have seen that these two investigations lines have always remained largely separated, despite studying closely related issues. Many scholars point out that this bifurcation in the literature constitutes a severe problem for science to progress. The problem might be solved by developing a joint research agenda between the two strands of SL and IL. As explained above, when trying to bring together SL and IL, we came across a significant divergence: the Implicit Learning line of research mainly focused on the formation of chunks, whereas Statistical Learning primarily investigated the computation of statistical information. However, as Perruchet and Pacton (2006) suggested, the divergence is not entirely insurmountable, and two solutions can be advanced: the first is that statistical computation and chunk formation might be seen

[16] "Research on Implicit Learning […] tended to fall within the purview of cognitive psychology, appearing in journals such as Journal of Experimental Psychology: Learning, Memory and Cognition, Journal of Experimental Psychology: General, and Quarterly Journal of Experimental Psychology" (Christiansen, 2019, p.470). As opposed, studies within the field of statistical learning have been published in journals such as Cognitive Psychology, Journal of Memory and Language, and Cognition (Christiansen, 2019).

as two subsequent steps during the language acquisition process. Specifically, statistical computations might lead to the formation of chunks. In other words, according to this theory, chunk formation is based on preliminary statistical analyses. "Typically, chunk boundaries are defined as the points where the predictability of successive or spatially contiguous elements is the lowest" (Perruchet and Pacton, 2006, p.235). Next to this possibility, a different solution might be the following: "[…] chunking is a primitive process the result of which amounts to simulating statistical computations" (Perruchet and Pacton, 2006, p.235). In other words, "[…] the formation of chunks is the only effective process, with the sensitivity to statistical structure being a by-product of this process" (Perruchet and Pacton, 2006, p.235). As the authors explain, this last theory is observable in two computational models: the Competitive Chunking model (Servan-Schreiber and Anderson, 1990) and PARSER (Perruchet and Vinter, 1998). "In PARSER, for instance, the chunks are formed from the outset on a random basis, as a natural consequence of the capacity-limited attentional processing of the incoming information. These chunks are then forgotten or strengthened according to the laws governing associative memory" (Perruchet and Pacton, 2006, p.235).

In conclusion, future research should consider and explore these hypotheses further to bring together IL and SL research. The unification of these two lines of investigation will pave the way for novel research questions, which might lead us to a better understanding of the complex mechanisms that stand at the foundation of human language. Indeed, several exciting issues still await an explanation, and numerous research possibilities would arise from creating a joined research agenda between Implicit Learning and Statistical Learning. In the present work, we aim to consider and bring together the results and methodologies provided by different research approaches to investigate the ability to implicitly acquire recursive embedded structures in different sensory modalities.

In the next section, we will focus on the AGL paradigm, and the different type of tasks traditionally used within this research paradigm. Specifically, we will dwell on the Serial Reaction Time task, which constitutes the methodology that we will exploit for our investigation, as can be seen in Chapter 5.

2.3.2. Artificial Grammar Learning (AGL)

Artificial Grammar Learning (AGL) is a commonly utilized paradigm in cognitive science. It has found widespread application in cognitive psychological research, primarily for assessing implicit learning and the acquisition of structural regularities. Furthermore, AGL has played a crucial role in psycholinguistic studies, offering a means to investigate the underlying mechanisms involved in human language acquisition and processing. The primary objective of AGL studies is to explore whether exposure to strings generated according to specific grammatical rules results in the implicit acquisition of those grammatical structures. Typically, in an AGL study, the symbols of the strings generated by the grammar are encoded onto stimuli that are presented to participants. The choice of the type of stimuli can vary, with some studies using tones, colors, geometric shapes, or letters. However, as noted by Pothos (2007), most studies tend to use letter strings. As anticipated in the previous section, experiments employing the AGL paradigm have been conducted since the 1920s in Esper's work. In the late 1950s, George A. Miller (1967) also employed this approach, generating short strings using a simple artificial grammar to examine the learning process associated with rule-based strings in contrast to randomly generated ones. However, the modern incarnation of the AGL paradigm was first introduced in 1967, when the American cognitive psychologist Arthur S. Reber utilized AGL to investigate implicit learning. Over the last few decades, numerous researchers have embraced and applied this methodology extensively.

The utilization of the Artificial Grammar Learning (AGL) paradigm in linguistic studies offers a multitude of advantages, as elucidated by Phillips (2017) and Compostella (2019). Firstly, AGL studies transcend the confines of human language, making it a versatile tool for investigating pattern learning abilities, even in non-human animals[17]. Another significant advantage lies in its applicability for examining the behavior of non-verbal populations, such as individuals with non-verbal autism, as well as infants who are yet to attain full linguistic competence[18]. Moreover, AGL studies prove well-suited for research involving speakers of diverse languages, obviating the need to design distinct protocols for each linguistic group. This uniformity in methodology facilitates cross-linguistic investigations. Furthermore, the use of an artificial grammar within AGL experiments enables the generation of strings that have not been previously encountered by the subjects under scrutiny. This eliminates potential interferences arising from semantics or pragmatics, fostering a controlled and rigorous research environment.

Various paradigms can be used in AGL research. The selection of a particular task over another is closely tied to the research's objectives. Indeed, each paradigm comes with its own set of strengths and weaknesses, making it more or less appropriate depending on the research's goals. Researchers can modify a paradigm in several ways to address different research questions. The two most widely used tasks are the *Forced Choice paradigm* and the *Serial Reaction Time task*.

In the **Forced Choice paradigm**, an experiment unfolds in two stages: initially, participants encounter a series of strings generated according to the rules of an artificial grammar (training phase). They are instructed to pay attention to these

[17] One noteworthy study worth mentioning is "A Framework for the Comparative Study of Language" conducted by Uriagereka, Reggia, and Wilkinson in 2013. In this study, the researchers delved into the intriguing question of whether animals possess the capacity to identify complete recursion, a distinctive feature of context-free grammar. Their objective was to ascertain whether the ability to recognize full recursion is exclusive to human language or if it also exists within non-human animals.

[18] Among the other, we find "Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge," conducted by Gomez and Gerken in 1999. In this research, infants were exposed to artificial grammar sequences. The researchers assessed whether infants displayed sensitivity to the underlying patterns using the method known as *head-turn preference procedure*. Another relevant study is "Statistical Learning by 8-Month-Old Infants," conducted by Saffran, Aslin, and Newport in 1996. In this study, infants were exposed to artificial grammar sequences. Results showed that infants succeeded in segmenting words by applying statistical learning strategies.

strings and try to notice whether there are patterns or regularities in them. This exposure period is long enough to facilitate memorization. Subsequently (test phase), participants are presented with additional strings, previously unseen. Their task is to classify these strings as either grammatical or ungrammatical, based on what they have encountered in the training phase. Indeed, some of these test strings are generated using the same rules as the training phase (grammatical strings), while others are generated randomly (ungrammatical). During their choices, participants rely on a sense of whether the string "feels right" or "does not feel right." Experiments can slightly differ in their test phase. In some cases, participants are presented with two strings at a time and are asked to choose which one is grammatical or which one they prefer (dual forced choice paradigm). In other cases, instead, strings are presented one by one, and participants have to say whether the string is grammatical (single forced choice paradigm). One limitation of this paradigm, as pointed out by Phillips (2017) and Compostella (2019), is that determining whether a string is grammatical or not requires conscious thought. Indeed, participants need to recall and apply in the test phase the structures learned in the training phase. Thus, in the test phase, participants need to make their knowledge of the grammar explicit or consciously accessible for evaluation and classification of the strings. Additionally, even in the training phase, participants are aware that there may be regularities or patterns in the stimuli to which they are exposed, since they are specifically instructed to pay attention to that. This might lead to potential difficulties in assessing whether the eventual learning has an implicit nature. Determining whether knowledge is conscious, or unconscious is not always straightforward, and it can pose a significant challenge. Importantly, however, different possibilities are available to check whether participants have developed conscious strategies during the task. Dienes et al. (1995), investigated the extent to which participants in forced-choice AGL experiments were conscious of their acquired knowledge, considering two criteria: (1) whether participants had metaknowledge of the knowledge they have acquired, and (2) whether participants had intentional control over the knowledge they have acquired. Regarding the first criterion, they found that participants lacked metaknowledge of their knowledge. As observed in other experiments, when asked whether they had discovered a rule or could explain the type of rule, participants were unable to provide satisfactory answers. In many cases, they reported making choices based on sensations or guessing, despite performing above chance levels on the task. Concerning the second criterion, they discovered that participants exercised conscious control in judgments and decisions. However, these findings should be interpreted critically, as the criteria for defining "consciousness" itself are not entirely clear (Compostella, 2019; Dienes et al. 1995). In conclusion, Dienes et al. (1995) underscored the importance of selecting appropriate criteria for investigating consciousness. They noted that different criteria, such as those related to metaknowledge and intentional control, can yield varying results in studies. This variability arises because these criteria reveal different aspects of knowledge and its application. Since different criteria of consciousness provide different insights due to their focus on distinct facets of knowledge, it is crucial to thoughtfully choose criteria that align with the specific research goals (Dienes et al. 1995).

The **Serial Reaction Time task** (SRT), introduced by Nissen and Bullemer (1987), is a widely adopted approach for investigating Implicit Statistical Learning (ISL) within the AGL paradigm. With this task, it is possible to tackle the capacity to subconsciously discern patterns and rules that are present in strings. More precisely, in this experimental framework, participants encounter sequences of stimuli that adhere to specific underlying rules, in other words, they are generated according to the rules of a specific artificial grammar. Being unaware of the presence of underlying rules, participants are instructed to respond to stimuli as swiftly and precisely as possible by pressing designated keys in response to the stimuli. If learning occurs, it is expected to find a reduction in reaction times (RTs) and/or an improvement in accuracy rates as the task unfolds. The Serial Reaction Times paradigm presents several advantages in contrast to the forced choice paradigm (Compostella, 2019; Phillips 2017). In addition to the previously mentioned aspect, wherein participants remain unaware of the presence of an underlying grammar, there is a reduced likelihood of participants making arbitrary decisions, a scenario more plausible when evaluating whether a string adheres to a grammar in the forced choice paradigm. Phillips (2017) highlights that distinguishing between a participant randomly pressing buttons during a serial reaction time task and one making random choices in a forced choice task is more straightforward. A participant displaying a high error rate coupled with low reaction times could signal a tendency toward random choices. In the forced choice paradigm, instead, there are no discernible cues to detect random selections. A high error rate in this context could be attributed to either random choices or a lack of implicit learning. However, despite the advantages just mentioned, it is important to remember that even during the SRT task, phenomena might take place that can interfere with the performance of RTs and accuracy, making it more complex to interpret the results in terms of effective implicit learning. For example, RTs might also decrease because of a habituation effect to the task. Conversely, they might increase because of a fatigue effect due to the prolonged duration of the task, perhaps accompanied by a lowering of accuracy. It is important to remember that RTs and accuracy rates observed in an SRT task may not solely indicate implicit learning; they may also be influenced by other factors (Compostella et al., under review).

2.3.3. Formal Language Theory and the Chomsky hierarchy

Formal language theory (FLT) is a field of study which was initiated by Noam Chomsky in the 1950s, with the goal of systematically studying the computational basis of human language. His framework has been very successful and over the years has come to play a key role not only in linguistics but also in other disciplines. In fact, to date, FLT still plays a major role in linguistics theories while also representing the basis of the theoretical foundations in computer science. Moreover, FLT has found application in neuroscience and cognitive science, and, more recently, also in biology (Fitch & Friederici, 2012; Jäger & Rogers, 2012). FLT describes the mathematical and computational properties of several classes of languages. In this framework, a language is understood as a set of expressions, which consist of finite strings of symbols. Strings are produced by the application of a set of rules, i.e., a grammar, over a finite set of symbols, i.e., an alphabet (Hopcroft & Ullman, 1969). More specifically, following Jäger & Rogers (2012), four elements must be specified in order to define a grammar (*G*): (i) a finite set of non-terminal symbols (*NT*); they are the symbols on which the rules of the grammar are applied; (ii) a finite set of terminal symbols (Σ); they appear in the strings of the language and are the result of the application of the rules of the grammar on the non-terminal symbols; (iii) a specific non-terminal symbol that is called a start symbol (S); (iv) a finite set of rules (R). Given two elements α and β, belonging to Σ and/or NT, rules have the following form: $\alpha \rightarrow \beta$ (i.e., α can be replaced by β). Specifically, a grammar G is said to generate a string ϖ if and only if it is possible to start from S and produce ϖ through a finite set of rule applications. The set of sequences produced to go from S to ϖ is called the derivation of ϖ. The set of all strings that can be generated by G is called *language of G* and is written as *L(G)* (Jäger & Rogers, 2012). As it was originally formulated (Chomsky, 1956), the Chomsky hierarchy proposes four nested levels of grammars ordered by their complexity. At every level correspond a specific automaton, which can *generate* the strings of their relative languages. Automata are abstract representations of computational system (Fitch & Friederici, 2012). Automata can recognize certain strings taken as input and reject others, depending on their computational power. Starting from the higher level, where there are the most powerful grammar and automata, to the lower level, where we find the least powerful ones, at the higher stage of the hierarchy we find *Type 0* languages, also *called recursively enumerable languages*. They are generated by the *unrestricted grammar* and the corresponding automaton is the Turing Machine. Immediately below (*Type 1*) are c*ontext-sensitive languages*, generated by *context-sensitive grammar* (CSG), and the corresponding automaton is the *linear bounded automaton*. After that, we find *context-free* languages (*Type 2*) generated by context-free grammar (CFG), and the corresponding automaton is the pushdown automaton (PDA). Finally, we find (*Type 3) regular languages*, generated by regular grammar (also called finite-state grammar, FSG), and the corresponding finite-state automaton (FSA)[19]. Every class in the Chomsky hierarchy can be effectively generated by the class above it, which means that Type-0 grammars encompass all grammars from Type-3 to Type-1 as well.

[19] For a more detailed analysis, we refer to Chomsky (1956); Fitch & Friederici (2012); Hopcroft, Ullman (1969); Jäger & Rogers (2012).

Type 3 languages are too weak to describe human language. Indeed, regular languages are suitable for basic matching tasks, such as finding specific words in text. However, they cannot capture multiple long-distance dependencies and recursive structures, which are on the other hand present in natural languages (Chomsky 1956; 1957). Finite-state automata can recognize only simple, long-distance dependency (e.g., ab*a) (Fitch & Friederici, 2012). At the opposite extreme, also recursively enumerable languages are inadequate for describing human language. Turing machines, as compared to finite-state automata, have a storage tape of unbounded length and thus an unlimited memory (Fitch & Friederici, 2012). They are extremely powerful and can describe a wide range of computational processes, but at the same time they are too flexible and unrestricted for modelling human language. As a matter of fact, they can generate both valid and invalid sentences. Moreover, their over flexibility may lead to ambiguity and lack of predictability in the sentence structure. Last but not least, they cannot model the hierarchical and compositional structure of human language, since they do not impose structured constraints.

Hence, where do natural language and thus human computational powers fall within the Chomsky hierarchy? The prevailing consensus among researchers in this field is that human languages necessitate "mildly context-sensitive" grammars (MSCGs). Indeed, human language exhibits at least context-free complexity but does not exceed context-sensitive complexity. Empirical evidence shows that some languages display limited crossing dependencies, which surpass the weak generative power of context-free grammars (CFGs) but do not necessitate the full capabilities of context-sensitive grammars (CSGs). Specifically, crossing dependencies have been observed in Dutch (Figure 4) and Swiss German (Huybregts, 1976; 1984; Shieber, 1985). Hence, human language would require grammars possessing a level of computational power that extends just beyond what can be captured by context-free grammars (CFGs) (Fitch, Friederici, 2012). Prompted by the recognition that context-free grammars (CFGs) were insufficient to capture the full range of syntactic phenomena observed in natural languages, particularly those involving crossing dependencies, scholars in the mid-1980s started to propose the earlier mildly context-sensitive (MCS) grammar formalisms.

One of the pioneering formalisms to emerge during this period was the Tree Adjoining Grammar (TAG), introduced by Joshi (1985). Another influential formalism developed around the same time was the Combinatory Categorial Grammar (CCG), proposed by Steedman (1985).

Context-free grammars (CFG) are commonly used for modelling human languages due to their balance of expressiveness and practicality. A fundamental difference between pushdown automaton (PDA) and finite-state automaton (FSA) lies in the fact that the latter have more memory than the former. Memory is defined as the number of symbols the automaton can rely on when determining the next symbol. In FSA, transitions between states are determined only by the current state and the input symbol. They operate in a sequential manner, processing input symbols one at a time. On the contrary, pushdown automata (PDA) have more memory, because they have got a pushdown stack that allows them to temporarily store and retrieve symbols. Indeed, a PDA can interact with the stack by popping and pushing symbols onto the stack as it reads the input. Hence, in a PDA, every transition is determined by the current state, the input symbol it reads, and the symbol it pops from the stack. The additional memory of the PDA enables it to recognize context-free languages, which have nested and hierarchical structures. Importantly, however, the stack follows a last-in, first-out (LIFO) principle. This means that it can push symbols onto the stack and pop symbols off the stack, but it cannot directly modify symbols that are already on the stack. The transitions in a PDA are primarily determined by the current state, the input symbol, and the symbol *at the top* of the stack. PDA are adept at recognizing languages described by context-free grammars (CFGs), which have greater expressive power compared to finite-state grammars (FSGs). In CFGs, non-terminal symbols can appear on both sides of a rule, allowing for the generation of nested recursive structures. This feature makes them suitable for representing various syntactic structures present in natural languages, such as nested dependencies (Figure 3). However, also CFGs do have limitations. Indeed, they struggle to capture certain intricate aspects of natural language syntax, such as cross-serial dependencies (Figure 4) and some long-distance dependencies, while also movement and sentence transformation (e.g., from active to passive) present a level of complexity that is beyond what CFGs can effectively handle (Fitch, Friederici, 2012). These more complex structures can, on the contrary, be generated by context-sensitive grammars (CSGs) and be recognized by linear bounded automata (LBA). The peculiarity of the LBA is that they have a tape on which they can move left or right. Moreover, during each transition, they can both read and write symbols on the tape. In a nutshell, the key distinction between PDA and LBA is that the latter can overwrite symbols on the tape, which means they can change the contents of the tape as they process the input. This ability to both read and write symbols on the tape, modifying the symbols on the tape in response to the current state and the input symbol as well as move the tape head in both directions, allows LBAs to recognize complex context-dependent rules and non-local dependencies typical of context-sensitive languages.

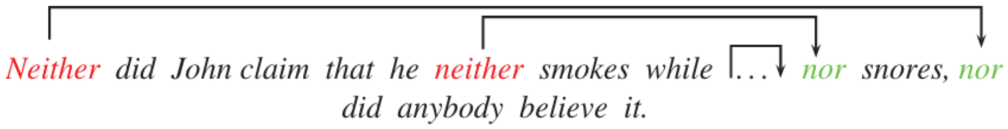

Neither did John claim that he neither smokes while ... nor snores, nor
did anybody believe it.

Figure 3. Example of nested dependency in English. Taken from Jäger & Rogers (2012), p. 1960.

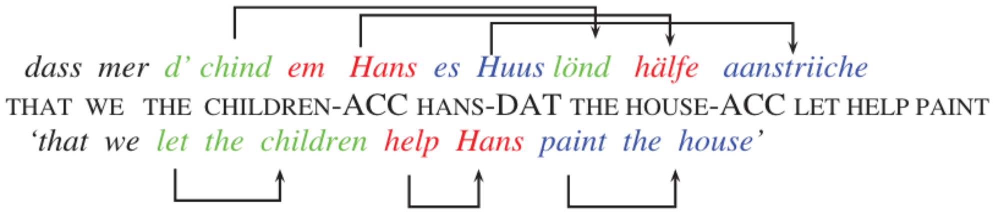

dass mer d' chind em Hans es Huus lönd hälfe aanstriiche
THAT WE THE CHILDREN-ACC HANS-DAT THE HOUSE-ACC LET HELP PAINT
'that we let the children help Hans paint the house'

Figure 4. Example of cross-serial dependencies in Swiss German. Taken from Jäger & Rogers (2012), p. 1960.

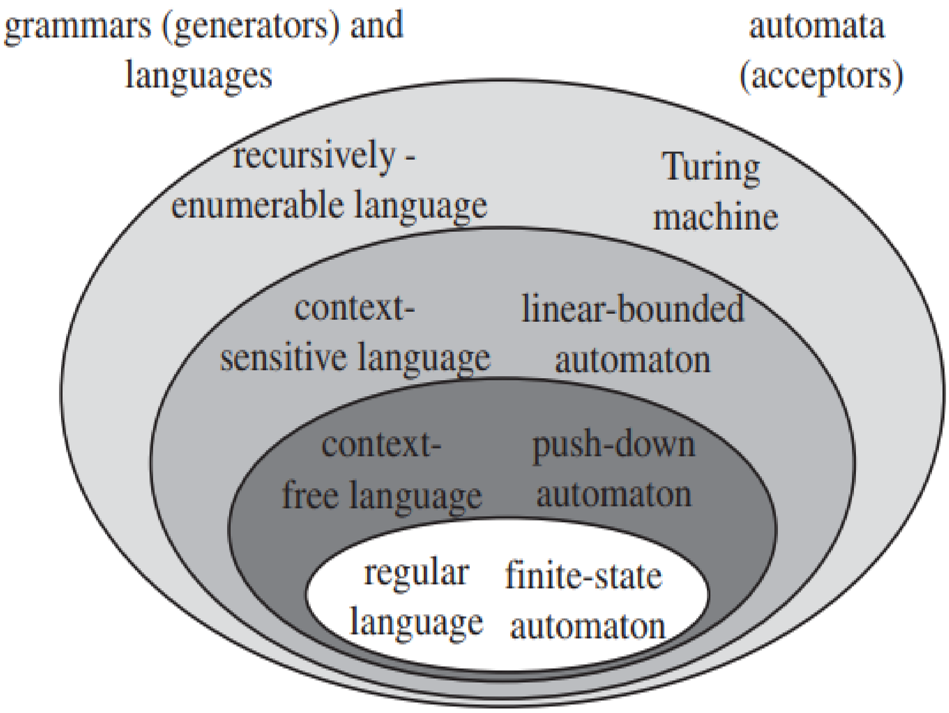


Figure 5. The Chomsky hierarchy with languages, grammars, and automata. Taken from Fitch, Friederici, 2012).

2.3.4. Formal complexity vs. cognitive complexity

In this section, a special focus will be laid on the concept of complexity, analyzing and confronting on one side *computational complexity in formal grammars and automata* and, on the other, *computational complexity in the human brain*. Several studies seem to suggest that it is important not to be misled by taking for granted a one-to-one correspondence between formal complexity as represented in the Chomsky hierarchy and cognitive complexity. Indeed, results of recent studies suggest that the position of grammar in the Chomsky hierarchy is not the only factor to consider when determining the complexity of processing by the human brain. (Chesi, Moro, 2014). Other factors might affect cognitive processing and first empirical pieces of evidence seem to confirm that cognitive and formal complexity are not two sides of the same coin (Bach et al., 1986; Christiansen, Chater, 1999; Christiansen, MacDonald, 2009; de Vries, Christiansen, Petersson, 2011; Fitch, Friederici, 2012; Öttl et al., 2015; Uddén et al., 2012).

Evidence suggesting that there is not a direct match between complexity as outlined in the Chomsky hierarchy and cognitive complexity stems from at least two sources. As we mentioned in the previous section, each level within the Chomsky hierarchy has the capability to generate the grammars of the lower levels. In other words, Type-0 grammars not only encompass Type-1 and Type-2 grammars but also include Type-3 grammars. The same is true for the respective automata. A Turing machine can accept both context-sensitive languages, context-free languages, and regular languages. On the opposite, grammars and automata that are at lower levels cannot generate and recognize those at higher levels. For example, a push-down automaton cannot recognize context-sensitive languages. As we have already explained, human language faculty are often found in the literature to correspond to “mildly context-sensitive" grammars. Crucially, however, the fact that we possess mildly context-sensitive capacities does not mean that our cognitive system is able to process all the languages that in the Chomsky hierarchy are generated at lower levels than the mildly context-sensitive grammars. In the same vein, each of these abstract categories of automata, even the less powerful class of finite state automata (FSAs), includes a multitude of automata that surpass the capabilities of any human being (Fitch, Friederici, 2012). Hence, stating that human languages require grammatical structures with at least the capabilities of context-free grammars does not necessarily imply that the human brain can encompass every conceivable context-free or finite-state grammar. For example, a phone book represents a finite list, which can be easily encapsulated by a simple finite-state automaton (FSA), with an assigned state for each name-number pair. Significantly, though, a phone book can be quite extensive. Take, for instance, the telephone directory of a large city such as Manhattan. Although the Manhattan phone book is a (long) finite list that can be captured by a simple finite state automaton (FSA), with one state for each name/number combination, this list is far too vast for any human to manage. (Fitch, Friederici, 2012). "Whatever class of computational systems natural language entails, it will always be some subset of the categories of automata described in FLT" (Fitch, Friederici, 2012, p. 1938). The second source of evidence that cognitive complexity does not reflect formal complexity as outlined in the Chomsky hierarchy comes from some interesting psycholinguistic studies in which learning of nested dependencies with that of cross-serial dependencies has been tested and compared. Nested dependencies are structures that can be generated by context-free grammars and recognized by push-down automata, whereas cross-serial dependencies by context-sensitive grammars and linear-bounded automata, respectively. Being context-sensitive grammars at a higher level in the Chomsky hierarchy than context-free grammars, cross-serial dependencies are therefore formally more complex than nested dependencies. The question is thus the following: are cross-serial dependencies perceived as more complex and therefore processed with more difficulty than nested dependencies by the human cognitive system? In the literature, it has been hypothesized that for both humans and animals, languages that are higher up in the Chomsky hierarchy are more complex to process than those at lower levels (Fitch, Hauser, 2004; Friederici et al., 2006). However, in reality, the correspondence between formal complexity and empirically testable cognitive complexity is far from obvious. Rather, it is at best suggestive (Öttl et al., 2015). To our knowledge, the first study that has addressed this issue is that of Bach and colleagues (Bach et al., 1986). They examined and confronted, by means of acceptability judgments and accuracy in paraphrase comprehension, subjects' performance on nested and cross-serial dependencies. Specifically, they tested linguistic performances on Dutch cross-serial dependencies of the type in *a*, and German center-embedded (nested) constructions, of the type in *b*, with the same number of dependency levels.

a. Jeanine heeft de mannen Hans de paarden helpen leren voeren.
   Joanna has the men Hans the horses helped teach feed.
   ENG: Joanna helped the men teach Hans to feed the horses.
b. Johanna hat die Manner Hans die Pferde futtern lehren helfen.
   Joanna has the men Hans the horses feed teach helped.
   ENG: Joanna helped the men teach Hans to feed the horses.'

What they found was that natural language sentence of the type in *a* with cross-serial dependencies were perceived as easier than sentences with structures of the type in *b*, with center-embedded dependencies. Thus, this result goes in the opposite direction from what was assumed by proponents of a parallelism between formal complexity and cognitive complexity. Bach and colleagues, however, did not formulate any hypothesis about what might have been the cause of their result: why did subjects perceive structures with nested center-embedded dependencies as more complex than those with cross-serial dependencies? Based on their findings, Bach et al. (1986) challenged the effectiveness of stack-based parsing algorithms; however, they did not propose a theory of linguistic complexity to account for the differences in complexity. Some years later, Joshi (1990) offered theory which showed that taking into account derivational generative power along with weak and strong generative power can be crucial for understanding why certain structures may not align with expected cognitive processing difficulties (Joshi, 1990). Weak generative power refers to the types of strings a grammar can generate, while strong generative power pertains to the types of structural descriptions a grammar can generate. Derivational generative power, on the other hand, concerns the complexity of the derivation process (i.e., the steps or rules needed to generate a structure). In his paper, Joshi (1990) investigates why crossed dependencies in languages like Dutch are processed with less cognitive difficulty compared to nested dependencies in languages like German. He introduces the concept of the Embedded Push-Down Automaton (EPDA) to model these dependencies more effectively than the traditional Push-Down Automaton (PDA). The EPDA, aligned with the Tree Adjoining Grammar (TAG) framework, can handle both crossed and nested dependencies by allowing partial interpretations during the parsing process. This model helps explain why crossed dependencies involve fewer intermediate steps in their derivational process: once an element and its dependent are processed, they can be directly linked without needing to revisit previously processed elements. In contrast, nested dependencies require the parser to manage multiple levels of embedding, increasing the number of intermediate steps as elements are pushed and popped from the stack more frequently. Therefore, by considering derivational generative power, researchers can gain a more nuanced understanding of the relationship between syntactic complexity and cognitive processing. This approach helps explain why certain structures that are more complex in terms of formal grammar may not necessarily be more difficult for the human brain to process, and vice versa. An important study that builds on this direction and goes further by offering a cognitive theory that directly explains the computational differences observed by Bach et al. (1986), considering factors such as cognitive memory costs and more biologically tuned parameters, was proposed by Edward Gibson. In his paper *Linguistic complexity: locality of syntactic dependencies*, Gibson (1998) presents an original theory on the relationship between the mechanisms of sentence processing and available computational cognitive resources. The Syntactic Prediction Locality Theory (SPLT) comprises two key elements: an integration cost component and a component related to the memory cost incurred when keeping track of necessary syntactic elements. Memory cost can be measured in terms of the number of syntactic categories required for the input string to result in a grammatical sentence. Crucially, both memory costs and integration costs are significantly influenced by *locality.* Specifically: *"(1) the longer a predicted category must be kept in memory before the prediction is satisfied, the greater is the cost for maintaining that prediction; and (2) the greater the distance between an incoming word and the most local head or dependent to* *which it attaches, the greater the integration cost"* (Gibson, 1998, p.1). Gibson's theory provided explanations for numerous phenomena that had previously lacked adequate understanding or explanation. Among the other, the theory showed that processing nested center-embedded dependencies would require higher memory load than processing cross-serial dependencies, and this would account for the result indicating that the former are cognitively more demanding than the latter. Indeed, "[…] the categories that are predicted first are associated with the most memory cost, so satisfying these first results in lower complexity for cross-serial dependencies than for nested dependencies" (Gibson, 1998, p.50).

It is important to highlight that Bach et al.'s (1986) results have been replicated and confirmed by numerous subsequent studies. Other interesting results that go in the same direction as those trumped by Bach and colleagues are those of Christiansen and Chater (1999), who tested both humans and artificial neural networks on nested and cross-serial dependencies, with three levels of dependency. Chesi and Moro (2014) also contended that formal complexity, as outlined in the Chomsky hierarchy, does not seamlessly translate into an indicator of cognitive complexity. They proposed that factors unrelated to a grammar's position in the hierarchy play a role in influencing the cognitive processing costs. With this regard, they proposed a definition and quantification of complexity based on two factors: time and space. Specifically, time complexity (i.e., hierarchy), is defined as the quantity of computational states traversed; in other words, it refers to the level of structural embedding. Space complexity (i.e., locality), instead, refers to the quantity of items stored and retrieved, in other words, it refers to the intervening elements within a filler-gap dependency that have to be stored in memory. Interestingly, they observed that these two factors can be differentiated not just in terms of computation but that the difference held also at the neurological level. Indeed, they reported that specific brain regions engaged in hierarchical syntactic processing and the formation of non-local dependencies exhibit increased activity as hierarchical depth increases, such as in the embedding of relative clauses. This heightened activity also occurs when dependencies necessitate additional working memory, as for long dependencies in which several constituents intervene in the structure between filler and gap. Further confirmation of the fact that formal complexity as represented in the Chomsky hierarchy and cognitive complexity are not directly linked has been provided also by other recent AGL studies (de Vries et al., 2011; Öttl et al., 2015; Uddén at al., 2012). Uddén et al. (2012) repeatedly exposed participants over a period of nine days to strings of letters featuring cross-serial or nested dependencies. Results showed that participants successfully learned both type of structures. Importantly, however, consistent with Bach et al. (1986), they found a processing advantage towards cross-serial over nested dependencies. As Öttl et al. (2015) pointed out, however, a possible limitation of Uddén and colleagues' study lies in the fact that strings were visually presented, being thus the temporal sequential dimension absent, which is nevertheless an important feature in natural languages processing. Moreover, the set of stimuli they employed was rather limited (Öttl et al., 2015). De Vries et al. (2011) further investigated the issue by adopting a more natural experimental setting by testing participants in a SRT task in which they were exposed to sequences of auditory stimuli. Results confirmed those found by Uddén at al. (2012): subjects displayed better performances in processing cross-serial dependencies than nested dependencies. However, as Öttl et al., (2015) pointed out, although they created a more naturalistic setting than the one used by Uddén and colleagues, the set of stimuli they used was quite small. Taking these limitations into consideration, Öttl et al., (2015) wanted to investigate deeper in the issue, by increasing the size of stimuli and thus creating a more naturalistic setting. As in de Vries et al. (2011), they presented subjects with sequences of auditory stimuli. Results in this case did not support a processing advantage for the cross-serial dependencies over the nested dependencies: after only one hour of exposition to stimuli, participants learned both the two types of dependencies. Despite not having found any specific advantage, Öttl and colleagues' result confirmed the hypothesis according to which cognitive complexity does not reflect formal complexity. Indeed, participants did not display greater cognitive difficulty when processing cross-serial dependencies compared to nested dependencies, as we should expect in case there was a one-to-one correspondence between formal complexity, as defined in the Chomsky hierarchy, and cognitive complexity. In conclusion, as we have seen in this section, considering derivational generative power alongside weak and strong generative power can clarify why certain syntactic structures do not always align with anticipated cognitive processing difficulties (Joshi, 1990). Importantly, incorporating biological aspects of computation, such as cognitive memory limitations and constraints, offers further insight into these discrepancies (Gibson, 1998). This approach underscores that formal complexity intended as weak and strong generative power alone does not fully capture cognitive processing challenges.

2.3.5. AGL with grammars belonging to the Chomsky hierarchy

In cognitive science and psycholinguistics, FLT has been extensively used to investigate the ability to implicitly acquire and process structures containing patterns and regularities, both in humans and animals. Indeed, numerous AGL studies have investigated the ability to process strings generated by grammars belonging to the Chomsky hierarchy. Most AGL studies have focused on investigating the learnability of two types of grammars: finite-state grammars and context-free grammars (Fitch, Friederici, 2012). Finite-state grammars have been employed already in the pioneering work of Reber (1967), which to this day remains one of the most cited and famous works that used finite-state grammar. He demonstrated that adult subjects succeeded in learning the regularities of this grammar, without being previously informed of the presence of underlying rules. Reber’s work played a foundational role in the study of implicit learning and the investigation of cognitive processes involved in language acquisition and pattern recognition. Numerous works since Reber's have continued to investigate the learnability of finite-state grammars, testing different populations, such as children, adults, but also animals, and transmitting strings by means of different sensory stimuli (Christiansen, Conway, 2005; Gomez, Gerken, 1999; Saffran, Wilson, 2003, among the others). These works were primarily focused on testing statistical learning abilities. Indeed, as we saw in Chapter 1, statistical learning is a fundamental cognitive skill that plays a crucial role in various cognitive activities, including but not limited to language acquisition.

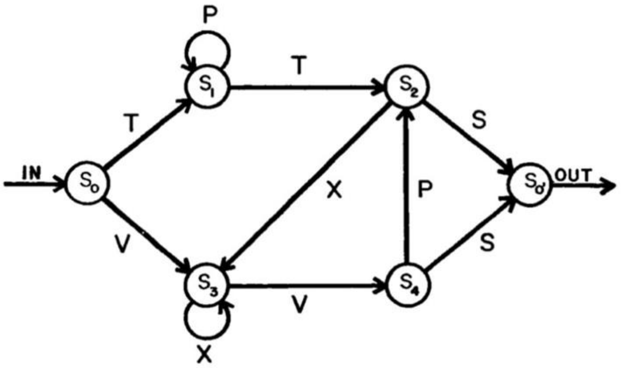


Figure 6. The finite-state grammar used in Reber (1967). Picture taken from Reber, 1967, p.856.[20]

On the other hand, more recently, several studies in the domain of artificial grammar learning (AGL) have begun to utilize context-free languages to investigate the supra-regular hypothesis (Fitch, Friederici, 2012). Indeed, as we have seen in this chapter, finite-state grammars have proven insufficient for capturing many syntactical phenomena found in human natural languages, and it is thus widely believed that humans possess mildly context-sensitive computational capacities. Among different context-free languages, the primary focus of most studies has been on A^nB^n. In the next section, we will review some studies which have employed this language. Moreover, we will address the frequent misinterpretations of results related to A^nB^n stringsets, which have led to significant confusion in the literature, causing several problems.

[20] Instances of grammatical strings are: VXVS, TPPTS, VXXVPS, …

Instances of ungrammatical strings are: TPTPS, VXPS, …

2.3.5.1. Testing the learnability of the A^nB^n language

To our knowledge, the first study which tested the learnability of the A^nB^n language is Fitch & Hauser (2004). In this study, the two scholars tested and compared the acquisition of two different languages in two species. Specifically, they exposed a group of college undergraduates and a group of cotton-top tamarins to auditory consonant-vowels syllables featuring the regularities of two artificial languages: the regular language $(AB)^n$ and the context-free language A^nB^n. Results were interesting: the group of students succeeded in learning the regularities of both two languages, while the cotton top tamarins group learned the regular language $(AB)^n$ without instead showing any signs of learning with respect to A^nB^n. From this result, Fitch & Hauser (2004) concluded that, being tamarins unable to process and learn the supra-regular language A^nB^n, this should count as evidence supporting the supra-regular distinctiveness hypothesis (Fitch, Friederici, Hagoort, 2012) and concluded that tamarins are thus unable to process simple phrase structures (Fitch & Hauser, 2004; Fitch, Friederici, 2012). While Fitch & Hauser (2004) intended *phrase structure grammar* as *supra-regular grammar*, some scholars, on the other hand, mistakenly made an association between *phrase structure grammar* as reported in Fitch & Hauser (2004), and *recursive grammar*, thus erroneously inferring that Fitch & Hauser's study carried evidence about the (in)ability to process and acquire recursive structures (Fitch, Friederici, 2012). This was a regrettable misinterpretation, as Fitch & Hauser's paper neither made any conclusions about nor mentioned recursion. Their explicit focus was on the supra-regular computational ability, which does not have association with recursion (Fitch, Friederici, 2012). "Fitch and Hauser […] report that tamarin monkeys are not capable of recursion. Although the monkeys learned a nonrecursive grammar, they failed to learn a grammar that is recursive. Humans readily learn both. The lack of recursion in tamarins may help to explain why animals did not evolve recursive language, but it leaves open the question of why they did not evolve nonrecursive language." (Premack, 2004, p. 318). Also Corballis (2007) misinterprets Fitch & Hauser's result, mistakenly interpreting it as evidence of recursive processing abilities. "Fitch and Hauser found that tamarins had little difficulty distinguishing the FSG sequences, but could not master the CFG sequences, where *n* was either 2 or 3. They concluded that tamarins were therefore unable to process recursive sequences." (Corballis, 2007, p.699). At top of that, in the literature we also find studies that having the specific goal of testing recursion have tested the learnability of A^nB^n, without creating an experimental design able to disentangle recursion from different parsing strategies. Gentner et al., (2006) carried out a study entitled *Recursive syntactic pattern leaning by songbirds*, in which they tested the learnability of A^nB^n in a group of starlings. "Here we show that European starlings (Sturnus vulgaris) accurately recognize acoustic patterns defined by a recursive, self-embedding, context-free grammar. They are also able to classify new patterns defined by the grammar and reliably exclude agrammatical patterns. Thus, the capacity to classify sequences from recursive, centre-embedded grammars is not uniquely human." (Gentner et al., 2006, p.1). In general, the misinterpretation has become so widespread that we even find in the literature the statement that "The A^nB^n language […] is generally assumed to be recursive because new sentences can be formed by successive insertion into the frame AXB, for example AB, AABB, AAABBB and so on" (Marcus, 2006, p.1117).

In this subsection we will explain (i) why mastery of A^nB^n cannot be taken as evidence for recursion and (ii) what can we infer from mastery of A^nB^n.

In the language A^nB^n, the number of instances of A matches precisely with the number of occurrences of B. Examples of strings belonging to this language include $A^1A^2A^3B^1B^2B^3$; $A^1A^2A^3A^4A^5B^1B^2B^3B^4B^5$; $A^1A^2A^3A^4A^5A^6A^7A^8B^1B^2B^3B^4B^5B^6B^7B^8$; …

The first important point to consider is this: for every finite subset of strings of the type A^nB^n, there is a finite-state automaton (FSA) that can recognize these strings. Even if the underlying pattern suggests a context-free language, FSAs are capable of handling any finite set of strings (Chomsky, 1959; Langendoen, 1975). This means that simply recognizing a finite set of strings of the language A^nB^n does not prove that the system has supra-regular computational power. Chomsky (1959) showed that for finite sets of A^nB^n, the complexity of the language can be approximated by a regular grammar. This approximation works because the set of strings is finite, and FSAs are perfectly capable of handling finite sets. In other words, if the recognition of A^nB^n is limited to finite samples, an FSA could also perform this task. Hence, the true test of supra-regular computational power requires demonstrating the ability to generalize beyond finite examples (i.e., for all *n*). Interestingly, however, even in cases where the ability to generalize beyond finite examples is attested, we cannot be sure that this was achieved using a recursive procedure. Different strategies are available to fully process the language A^nB^n, recognizing and thus generalizing to non-finite sets of strings. All of these strategies require computational abilities beyond finite-state automata, but only some of them involve a recursive strategy. It remains to be shown which of the various supra-regular strategies subjects exploit in learning the grammar. Specifically, there are as many as three different cognitive strategies (Corballis, 2007; Fitch & Friederici, 2012; O'Donnell et al., 2005). We can see these three strategies in Figure 7. Following Fitch and Friederici (2012), the most straightforward approach is (a): the 'tally-and-evaluate' method. This method ascertains the quantities of 'a's and 'b's within the string and approves it only if their counts match. This approach requires supra-regular computational abilities and produces a single hierarchical level (Figure 7a). An alternative possible strategy (b) produces a 'nested' or 'center-embedded' arrangement. This strategy can be carried out by a pushdown automaton because it associates each 'b' with the most recently encountered 'a' (Figure 7b). The third approach (c) involves the formation of crossed dependencies, and it cannot be executed using a single pushdown stack (Figure 7c) (Fitch & Friederici, 2012, p. 1941). Consequently, this necessitates a grammar beyond the capabilities of context-free grammars. However, contrary to the assertion by Fitch and Friederici (2012), it is not accurate to say that such structures require at least a context-sensitive grammar. A Tree Adjoining Grammar (TAG) can precisely handle these dependencies between two sets of elements without being as powerful as a context-sensitive grammar (CSG).

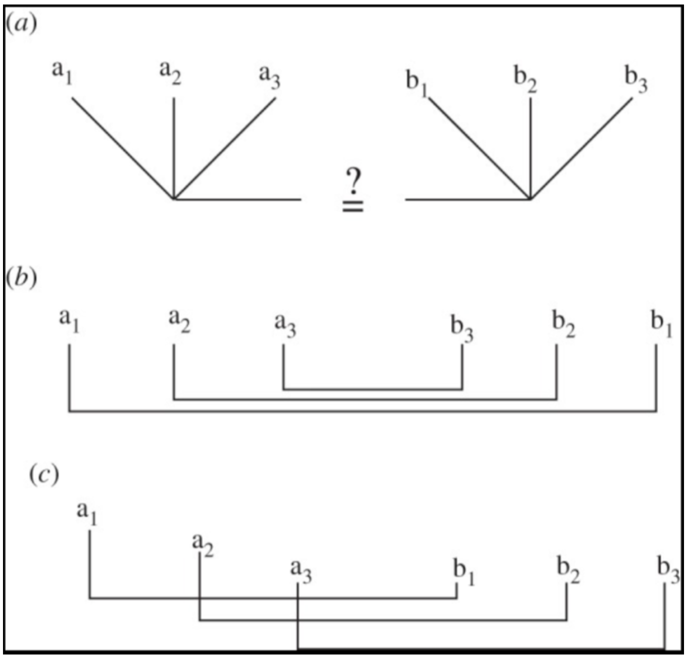


Figure 7. The three possible strategies to recognize the A^nB^n language, as explained in Fitch and Friederici, 2012. Taken from Fitch, Friederici, 2012, p.1941.

In conclusion, if a system successfully recognizes the strings of A^nB^n by exploiting one of the three strategies presented above (a, b, or c), we can say that the system has supra-regular computational power. It follows that the system possesses an auxiliary working memory compared to that of a finite-state automaton, and this could be in the form of a counter tape or a pushdown stack. Importantly, this tells us nothing about whether it applies a recursive procedure or not. In other words, to fully recognize A^nB^n, which means applying the rules to an indefinite number of novel sequences and showing robust performance across a wide variety of new and increasingly complex sequences, different strategies are available. All of them require the automata to have computational abilities beyond finite-state, but this does not guarantee the use of a recursive strategy. Indeed, using a recursive strategy is only one of the possibilities (Fitch & Friederici, 2012). Specifically, the only approach that requires a recursive strategy is (b). Hence, it has to be shown which of the various supra-regular strategies the parser exploits in mastering.

To more effectively and precisely study the ability of recursion, the first step would be to create an experimental design that establishes agreement dependencies featuring embedded recursion between the As and Bs. One reason several studies, including those by Fitch and Hauser (2004) and Gentner et al. (2006), have not demonstrate that participants used a recursive ability when recognizing strings of A^nB^n language is that these studies did not incorporate an experimental design with agreement dependencies between the As and Bs (Ferrigno et al. 2020). This lack of dependency meant the experiments did not strongly test for recursion. As Ferrigno and colleagues (2020) explain, in a sentence like “The mouse[A1] the cat[A2] chased[B2] ran[B1],” each “A” phrase (such as “The mouse[A1]” and “the cat[A2]”) needs to be correctly paired with a corresponding “B” phrase (like “chased[B2]” and “ran[B1]”). Because such dependencies are absent in A^nB^n strings, participants might have used non-recursive strategies to judge grammaticality or differentiate stimuli that followed the rule from those that did not. Crucially, Perruchet and Rey (2005) as well as de Vries et al. (2008) have investigated the hypothesis that participants exposed to A^nB^n language may use strategies other than recursion. These studies included critical test trials necessary for demonstrating recursion. In these trials, participants were presented with violations of the A^nB^n language due to the dependency structure (e.g., A1A2A3A4B3B4B1B2) rather than the number or order of As and Bs. Using methods similar to those of Fitch and Hauser (2004), these studies found that humans did not identify these trials as grammar violations, suggesting that they were likely using alternative strategies such as tracking A-B switches or counting (Ferrigno et al. 2020). A recent interesting study by Ferrigno and colleagues (2020) tested recursion abilities using the A^nB^n language by creating an experimental design that effectively tested recursion abilities, disentangling it from other possible computational strategies, by inserting agreement dependencies between the As and Bs. Ferrigno et al. (2020) employed a cross-population design that included a nonlinguistic sequence generation task. This task aimed to see if participants could generalize item groupings exploiting a center-embedded, recursive strategy. The study comprised four groups of participants: children, U.S. adults, adults from a Bolivian indigenous group lacking reading skills and formal mathematics abilities, and monkeys. All human participants intuitively created recursive structures from ambiguous training data, whereas monkeys required extra training to reach comparable outcomes. These results suggest that the capability to employ recursive hierarchical strategies is an inherent aspect of human cognition, emerging early in development and present across various cultures. Although this skill is not unique to humans, indicating that nonhuman animals can also produce and recognize new sequences with recursive, hierarchical, and center-embedded structures, monkeys did not initially apply abstract hierarchical organization. However, with additional exposure, two out of three monkeys eventually learned to generalize and construct new center-embedded sequences.

In the next section, we will focus on a crucial aspect to consider in the empirical study of the ability to deal with recursion: the distinction between algorithmic properties and representational abilities.

2.3.6. Understanding the empirical aspects of recursion: Algorithmic properties and representational abilities

In the first part of this chapter, we addressed a longstanding problem in the field of linguistics: the absence of a definitive and unequivocal definition of recursion. We clarified the concept of recursion and discussed its role in the processing and acquisition of human language syntax. In this section, we turn our attention to another issue frequently encountered in the literature, one of paramount significance for empirical investigations into this phenomenon.

One of the primary challenges in empirically studying the ability to handle recursive embedding lies in distinguishing between algorithmic properties and representational abilities (Martins, 2012; Lobina, 2011). Traditional definitions of recursion have sometimes focused on algorithmic characteristics, while at other times, they emphasized which structures could be categorized as recursive (Martins, 2012). However, as Martins (2012) aptly notes, while definitions that emphasize the procedural (i.e., algorithmic) aspect can serve as an initial step in delineating the phenomena under investigation, they may not provide the most suitable framework for empirical research. "In spite of the pervasiveness of structures that can be modelled using recursive algorithms or rule sets, not all of them will be represented as such. This means that the amount of recursion in a structure will only be relevant for an observer to the extent that he can decode it meaningfully. […] not all activities that can be synthesized with recursive processes are perceived as structurally meaningful by the observers. Hence, the ability to produce recursive structures and the ability to decode them do not necessarily come together […]" (Martins, 2012, p. 2057). In other words, a definition centered on algorithms does not guarantee that the cognitive process of representing and processing recursion aligns with it (Lobina, 2011; Martins, 2012). Furthermore, understanding how a computation is implemented, especially within the intricate realm of the human brain, often remains elusive to external observers until meaningful behavioral correlations emerge (Martins, 2012). Therefore, verifying whether a structure generated by a recursive algorithm is indeed represented recursively by human cognition presents an exceedingly complex challenge. Given our inability to peer inside this cognitive "black box," it follows that definitions primarily focused on algorithmic properties may not offer the most relevant framework for empirical research. While focusing on structures and outputs generated by recursive algorithms is a commendable starting point, it may not suffice. To gain insights into how human cognition represents recursive structures in behavioral experimental tasks, it is more prudent to focus on the distinctive signatures of recursion (Martins, 2012). For this reason, Martins (2012) suggests that "[…] the key empirical test for recursion is the ability to represent dependency relationships that were not previously defined, or to represent information within hierarchical levels not previously 'available'. What this ability presupposes is the knowledge (or expectation) that all nodes within a hierarchy can behave similarly and can display the same properties relatively to the way they interact with the nodes 'above' and 'below'." (Martins, 2012, p. 2058).

Hence, in conclusion, Martins (2012) proposes that a definition centered on representational abilities, such as the capacity to represent self-similarity across hierarchical levels, offers a more promising approach for investigating the cognitive ability to deal with recursive hierarchical structure. With this perspective, the focus shifts from mere process-oriented descriptions to the ability to represent different hierarchical dependencies within the same set of rules. Subjects capable of such representation demonstrate the potential to generalize and generate new levels of embedding beyond what is specified a priori, whether in the algorithm or in the input. Hence, the presence of distinctive behavioral signatures becomes paramount, as they serve as indicators of the cognitive processes at play.

2.4. Conclusion

This chapter began by exploring the concept of recursion, a central idea in Chomsky's theory of linguistics. Despite its significance, recursion was not clearly and universally defined for many years. Interest in recursion surged with the publication of the influential paper "The Faculty of Language: What is it, Who has it, and How did it Evolve?" by Hauser, Chomsky, and Fitch in 2002. This paper hypothesized that recursion might be the central and unique feature of the human language faculty, distinguishing between the faculty of language in the broad sense (FLB) and the narrow sense (FLN). The FLB includes various components like sensory-motor and conceptual-intentional systems, while the FLN, they proposed, is solely comprised of recursion and is unique to humans. Despite the impact of Hauser et al.'s work, their paper did not offer a precise definition of recursion, often linking it to the concept of discrete infinity. This has led to a proliferation of varied and sometimes conflicting definitions in subsequent research, causing considerable terminological confusion. In the chapter we aimed to providing a clear definition of recursion. We started by observing that recursion is eminently a formal notion. Indeed, the concept is used not only in linguistics but also in computer science. Hence, we clarified the definition of recursion in computer science, distinguishing it from iteration. We have seen that in computer science, a recursive algorithm solves a problem by calling itself with a smaller instance of the same problem, whereas an iterative algorithm uses loops to repeat a set of instructions until a condition is met. Although any problem solvable by a recursive algorithm can also be addressed with an iterative algorithm, there are significant distinctions regarding clarity, conciseness, and memory utilization. Recursive algorithms often mirror the problem's structure more closely, but they can be less efficient in terms of memory usage due to the stack space consumed by recursive calls. After that, we provided a clear definition of recursion in cognitive science and linguistics. Recursion is intended as the embedding of elements within elements of the same type. The chapter made a clear distinction between iteration without embedding, iteration with embedding, and recursion. Iteration without embedding involves discrete steps that are independent of each other, while iteration with embedding can generate hierarchical structures by creating dependencies among constituents. Recursive embedding allows for the construction of new hierarchical levels without additional rules. Specifically, we also distinguished between different types of recursion: tail recursion and nested recursion. Importantly, we tailed back to linguistics phenomena each of these different concepts, providing linguistic example for the different types of recursion, as well as for the different types of iteration (with and without embedding). In summary, the first part of the chapter provided a comprehensive overview of recursion, defining it clearly in both computer science and linguistics, and distinguishing it from iteration. We are convinced that the first fundamental step in proceeding to the investigation of the cognitive ability to form recursive hierarchical abstract representations is to be clear in mind what are the defining characteristics of a structure as such. This comprehensive overview of the concept of recursion allowed us to draw the following conclusions: Firstly, recursion is not exclusively tied to the concept of discrete infinity. While a recursive embedding algorithm can generate an unlimited number of sentences from a finite set of elements, other algorithms, like iteration (with or without embedding), can achieve the same. For instance, a sentence can be extended with no upper bound to its length by iterating a finite set of nouns. Secondly, not all hierarchical structures in human syntax are recursive. Hierarchy does not imply recursiveness. Recursive embedding is just one of the many structural phenomena in human syntax. Thirdly, the presence of long-distance dependencies does not necessarily indicate recursion. Only nested recursion produces sentences with long-distance dependencies, while tail recursion does not. Non-recursive algorithms, like iterative embedding, can also create long-distance dependencies. Lastly, there is a significant distinction between how recursive algorithms are treated in computer science versus natural language. In computer science, recursive algorithms can be transformed into iterative versions. That is, problems solved using recursive algorithms can be solved also using iterative algorithms. However, this parallelism does not translate well into natural language (Parker, 2006). Indeed, in natural language, semantics provides the necessary information to build structure, which would not be apparent otherwise from the strings of sentences. Hence, in language, processing which require recursion cannot be correctly solved my means of iteration. The strict ordering required in recursion, absent in iteration, helps distinguish these forms in natural language. An iterative processing solution of sentence structures does not capture their complex meanings. Semantics provide the necessary information to identify the correct structure, which cannot be determined solely from the string. It follows that a linguistic system without semantics would not need recursion; if there is no meaning to convey, iteration would suffice for the syntax of the communication system (Parker, 2006).

After this clarification of the concept of recursion, we have critically examined the hypothesis by Hauser, Chomsky, and Fitch (2002) that recursion is a unique feature of human language, not found in other cognitive domains or non-human species. We explored various perspectives on this hypothesis, revealing a complex and nuanced understanding of recursion's role in human cognition. Firstly, the hypothesis posits that recursion is a defining and universal trait of human language, distinguishing it from other cognitive processes and non-human communication systems. However, our investigation has uncovered several challenges to this view. While recursion is a foundational aspect of linguistic theory, its presence and significance in everyday language usage may not be as pervasive as initially claimed. Evidence suggests that deeply nested recursive structures are rare in both spoken and written language. Authors like Karlsson (2010) and Verhagen (2010) argue that recursion might be less central to everyday linguistic practice than Hauser et al. (2002) proposed. Moreover, the possibility that languages exist without recursion, as illustrated by the Pirahã language, challenges the idea that recursion is essential for all linguistic systems. This language uses alternative methods to convey complex ideas without relying on recursive structures. Similarly, Kinsella (2010) argues that recursion is not the sole feature that defines human language. Other linguistic features, such as structure-dependence and duality of patterning, also contribute to the uniqueness of human language and operate independently of recursion. In addition to linguistic considerations, we examined the role of recursion in non-linguistic cognitive domains. Recursion appears in various cognitive processes such as numerical reasoning, navigation, and music. However, the extent to which recursion is necessary or optional in these domains remains debated. For instance, in navigation, recursive strategies could be beneficial but are not the only possible approaches. In other cognitive domains, however, there are clear instances of necessary non-linguistic recursion, such as in music (Bach's embedded key changes), visual perception, social cognition, and theory of mind (Parker, 2006). Overall, our exploration suggests that while recursion is at play in human language, it is not necessarily unique to it. The presence of recursion in other cognitive domains and its optional nature in some contexts imply that recursion might have evolved from broader cognitive capacities rather than being a language-specific trait.

After having clarified the concept of recursion, we offered a comprehensive exploration of the intricate relationship between linear order, hierarchy, and their interplay in human language. We began by delving into the historical debate within linguistic theory, specifically focusing on Kayne's work, which challenged the prevailing assumption and argued for a rigid connection between hierarchical structure and linear order, emphasizing the crucial role of linear order in syntax. Our discussion expanded to the broader cognitive implications of this debate, emphasizing the necessity of considering the linear, temporal dimension for a better understanding of the mechanisms at the core of human language. Combining the insights on the significance of sequentiality in language with our exploration of the role of recursion in human language, a fundamental claim emerged: the ability we want to investigate, which play a role in language as well in music, is the ability to process recursive hierarchical structures from sequential (i.e. temporally ordered) sequences of stimuli. Subsequently, we introduced Dehaene and colleagues' (2015) taxonomy, offering a framework to understand the diverse cognitive representations at play during the processing of sequential stimuli, with increasing complexity and abstraction. Moreover, we reviewed some studies which have investigated the cognitive mechanisms driving the processing of sequential sequences of stimuli, moving from low-level statistical computations to the formation of abstract structured representations.

The second part of the chapter started with an introduction into implicit learning and its distinction from explicit knowledge. As we have seen, implicit learning enables us to master various skills effortlessly, without conscious thought. We discussed the cognitive mechanisms underlying this process and their importance in fields such as language acquisition. The chapter also explored the division between Statistical Learning (SL) and Implicit Learning (IL) research, despite their commonalities. We emphasized the need for a shared research agenda to bridge this gap. After that, we have introduced the Artificial Grammar Learning (AGL) paradigm, a critical paradigm in cognitive science and psycholinguistics. AGL is used to assess implicit statistical learning and the acquisition of structural regularities. It offers a versatile approach for investigating pattern learning in various populations and languages. Specifically, we have covered two primary AGL paradigms: the Forced Choice paradigm and the Serial Reaction Time task, highlighting the advantages and challenges of both. Subsequently, we have explored Formal Language Theory (FLT) and the Chomsky hierarchy, a framework often used to investigate the computational foundations of human language. We have seen that regular languages are too simplistic for capturing human language intricacies, while recursively enumerable languages are overly flexible. The consensus is indeed that human languages fall within "mildly context-sensitive" grammars, just beyond context-free grammars. Then, we have explored the relationship between formal complexity, as defined in the Chomsky hierarchy, and cognitive complexity, shedding light on important distinctions. While it is commonly assumed that a direct match exists between the Chomsky hierarchy's formal complexity and cognitive complexity, recent studies challenge this notion. The evidence strongly suggests that cognitive processing capabilities are influenced by factors beyond formal grammatical hierarchy, challenging our assumptions about how the human brain manages linguistic structures. After that, we delved into AGL studies investigating the learnability of grammars within the Chomsky hierarchy. As we have seen, many of them focused on finite-state and context-free grammars, investigating the learnability of languages such as A^nB^n. Crucially, however, we have observed that results from studies on the learnability of A^nB^n stringsets have caused confusion in the literature, with some mistakenly inferring recursion from these findings. Indeed, learning A^nB^n does not conclusively demonstrate recursion. Various cognitive strategies are involved in mastering A^nB^n making it challenging to determine the specific cognitive mechanism applied. In this vein, we have instead considered recent studies that have investigated recursion using A^nB^n language, effectively disentangling it from other possible strategies (Ferrigno et al. 2020).

Lastly, in the chapter we differentiate between algorithmic properties and representational abilities of recursion. Indeed, it is not enough that the language being tested was generated through a recursive process. It is important to keep in mind that participants might process and learn the language in question by adopting techniques that do not involve recursive mechanisms (Martins, 2012). Identifying that a structure underwent recursive processing is demanding and entails the need to eliminate potential iterative and hierarchical explanations (Shirley, 2014).

In conclusion, we can say that despite the numerous studies attempting to test recursive structure learning in AGL, both in humans and animals, there is still a lack of sufficient studies providing clear and irrefutable empirical evidence of the ability (or inability) to process structures recursively. Aside from the problem outlined above regarding the widespread lack of precision in defining and circumscribing the object of study, another problem often encountered is that related to the paucity of suitable tools for the study of recursion in non-linguistic domains (Martins, 2012). We therefore feel it is time to abandon the overrated A^nB^n language: There are numerous opportunities to study recursion outside of Chomsky's hierarchy. It is indeed important to note that Chomsky hierarchy is not the sole source from which grammars can be drawn for testing in AGL paradigms. There remain numerous exciting opportunities to investigate recursion in the AGL field, by using different types of grammars and by adopting new empirical approaches (Fitch, Friederici, 2012).

In Chapter 4, we will delve toward the heart of this thesis by presenting a grammar outside the Chomsky hierarchy that, for the reasons we will outline, presents interesting features for the exploration of recursive learning and processing: the Fibonacci grammar. This grammar will form the object of study in

Chapter 5, where we will test its learnability in three different sensory modalities: the auditory, the visual, and the tactile sensory domains.

In the upcoming chapter, our investigation will extend to the dynamic connection between cognition and perception, emphasizing both domain-specific and domain-general aspects of learning. Specifically, we will center our attention on sequential implicit statistical learning and the ability to form recursive abstract representations across various sensory domains. As we have explained, our interest in the present thesis is to shed light on a particular type of recursive hierarchical structure, that is, recursive hierarchical structures arising from temporally ordered stimuli, a cognitive architecture where elements are utilized recursively across various hierarchical levels, leading to the formation of nested structures during the sequential parsing of information. Given that this type of structure can potentially be conveyed through visual, tactile, and auditory stimuli, as we will see in Chapter 5, we aim to investigate the possibility of processing and learning this type of structure across different modalities, shedding light on their similarities and differences. Is the ability to form recursive hierarchical abstract representations from sequentially ordered stimuli a stimulus-independent or modality-based process? Does it consist of a unitary, single mechanism shared across sensory domains, or are there different modality-constrained mechanisms? The investigation of the ability to learn and process this type of structures has not been fully addressed and constitutes an intriguing challenge. Hence, it is a question of significant interest to verify *whether* and *how* learning is affected by different modalities and to observe potential differences or similarities across these three sensory domains. Taking into consideration that recursive hierarchical structures arising from temporally ordered, fading stimuli are peculiar architectures which are present both in language and music (cf. Section *2.1.2.*; *2.2.1.*), and considering that music and language are preferentially conveyed through the auditory perceptual domain, the question arises: Is the ability to form recursive hierarchical abstract representations arising from sequential stimuli a modality-based capacity? Are we better at learning and processing these structures in the auditory domain? Does the acoustic domain have an advantage over the visual and the tactile ones? Indeed, from our observations, it could be possible that recursive hierarchical learning from sequentially presented, fading stimuli is more robust in the auditory domain, whereas recursive hierarchical learning from spatially arranged, static stimuli is more robust in the visual domain. The alternative view is that this ability is stimulus-independent, and through it, we can equally process recursive hierarchical structures arising from sequentially presented stimuli in the visual, auditory, and tactile domains. But even before considering potential differences or similarities across the three domains, will we find evidence that we are capable of learning and processing these structures in the tactile domain? To our knowledge, no study has ever investigated this issue, which remains entirely unexplored to date. In fact, despite finding some studies which have investigated the ability to implicitly learn and process recursive structures in the visual and auditory domains (Martins, 2012; Martins et al; 2014; 2015; 2017), the few studies found in the literature on tactile implicit learning have been limited to investigating low-level statistical regularities in the tactile domain (Abrahamse et al., 2008; 2009; Conway and Christiansen, 2005; Pavlidou & Bogaerts, 2019). We will present an overview of these studies in the following chapter.

3. A Multisensory Voyage through Time and Space Dimensions with Hearing, Sight, and Touch

3.1. Learning through time and space dimensions with hearing, sight, and touch

In the previous chapter, we outlined our objective, which is to thoroughly investigate the mechanisms underlying the learning of recursive hierarchical structures arising from temporally ordered sequences of stimuli. We aim to investigate whether and how the mechanisms outlined in Dehaene et al.'s (2015) taxonomy, which categorizes various cognitive internal representations during the processing of temporally distributed stimuli, interact with each other. Specifically, among these mechanisms, we are interested in focusing on the learning of sequential statistical regularities and the formation of chunks, their categorization, and the representation of these chunks in recursive hierarchical structures. Additionally, we want to investigate whether the process of learning recursive hierarchical structures from sequential sequences of stimuli is possible in three different sensory domains: auditory, visual, and tactile, examining potential similarities or differences in the process. No study has comprehensively explored this topic so far. To begin, we will review studies that have investigated, in different sensory modalities, (i) implicit sequential statistical learning, a fundamental step in the process of hierarchical structure formation, as confirmed by Planton et al. (2021); Schmid et al. (2023) and Vender et al. (2023); (ii) the ability to learn recursive hierarchical structures.

Throughout this chapter, we will first examine studies on sequential implicit statistical learning in visual, auditory, and tactile domains, with a focus on results and insights. We will address whether sequential implicit statistical learning is a domain-specific or domain-general ability and if any sensory domain prevails on the others. Second, we will review studies on learning recursive hierarchical structures in visual and auditory domains. Notably, no study has explored this in the tactile domain. We will assess if existing studies suggest a domain-general ability or reveal domain-specific nuances in acquiring these structures across sensory modalities.

3.1.1. Implicit statistical learning: Domain-general or domain-specific ability?

As discussed in Chapter 1 and 2, numerous studies have delved into implicit detection and acquisition of statistical information, exploring this ability across different sensory modalities and domains. These investigations consistently highlight the adaptability and universality of this cognitive capacity. In essence, this ability transcends sensory boundaries and stimulus variations, playing a significant role in various cognitive mechanisms, including language (cf. Chapter 1). Indeed, implicit statistical learning ability (ISL) has been identified in various sensory modalities and domains (cf. Frost et al., 2015). Notably, research has demonstrated ISL in auditory nonlinguistic input (Creel et al., 2004; Saffran, 2002; Saffran et al., 1999), visual input (Baker et al., 2004; Chun & Jiang, 1999; Edelman, Hiles, Yang, & Intrator, 2002; Fiser & Aslin, 2001; Kirkham et al. 2002), and tactile input (Conway, Christiansen, 2005; Abrahamse et al. 2008; 2009; Pavlidou & Bogaerts, 2019). As Frost et al. (2015) explain, in the field of cognitive science, theories on implicit statistical learning have arisen as possible domain-general cognitive mechanisms challenging the prevailing domain-specific Chomskyan model of language acquisition. "Rather than assuming an innate, modular, and neurobiologically hardwired human capacity for processing linguistic information, SL, as a theoretical construct, was offered as a general mechanism for learning and processing any type of sensory input that unfolds across time and space." (Frost et al. 2015, p. 2). Initially, the concept of domain generality was introduced to counter the notion of language modularity. "[…] its definition therefore implicitly implied "something that is not language specific". Consequently, within this context, "domain" implies a range of stimuli that share physical and structural properties (e.g., spoken words, musical tones, tactile input), whereas "generality" is taken to be "something that does not operate along principles restricted to language learning." (Frost et al, 2015, p.2). Crucially, as Frost et al. (2015) emphasize, this approach outlines what domain generality excludes rather than explicitly providing a definition or delineating its characteristics. Recent perspectives on implicit statistical learning, however, attribute domain generality to a unified learning system that performs consistent computations across stimuli, different domains, but also various species (Frost et al. 2015). Karuza (2014) conducted three experimental studies using functional magnetic resonance imaging (fMRI) to explore the commonalities and potential differences in the mechanisms of statistical learning across various domains. The aim was to distinguish between two possibilities: (i) whether domain-specific perceptual cortices (such as auditory and occipital regions) perform similar computations during learning, or (ii) if perceptual cortices transmit input to domain-general regions, which then execute these computations irrespective of the stimulus modality. The first experiment consisted of a word segmentation task. The second experiment shifted the input modality and spatiotemporal properties, investigating simultaneously presented visuospatial patterns. The third experiment combined sequential auditory and visual modalities. Participants were assigned to one of two matched scenarios. In the auditory scenario, they undertook a word segmentation task akin to the one in the first experimental study. In the visual scenario, participants took part in the same test, but with each syllable replaced by a corresponding shape. Overall, the findings indicated that both auditory and visual statistical learning involve a domain-general network of regions capable of extracting novel structures, regardless of the input modality. The observed activation was not confined to modality-specific perceptual cortices; instead, it engaged the prefrontal cortex, caudate, putamen, and hippocampal/parahippocampal regions, depending on the experimental context. Notably, however, evidence of spatiotemporal structure effects emerged. Specifically, amygdala activation was only observed in the simultaneous visual task in study 2, suggesting the specialization of the amygdala for spatial structure acquisition. According to Karuza (2014), when the brain encounters structured stimuli, it promptly activates a broad network involving frontal, subcortical, and hippocampal regions. As time progresses, this network becomes narrower and more specialized, with the substrates most adept at handling the specific computations needed for the task taking on the processing load. More precisely, she proposes that the prefrontal cortex and basal ganglia collaborate as a circuit ideally suited for maintaining and updating internal representations. On the opposite, medial temporal regions are deemed most effective for calculating the rapid associations between elements, a crucial process in the initial phases of a statistical learning task.

Crucially, however, despite studies attesting that statistical learning is performed in a "domain-general neural region", which execute computations irrespective of the stimulus modality, we find contrasting evidence that raises concerns about this. Indeed, despite its consistent manifestation across diverse sensory modalities, studies comparing this ability across different domains or with varied stimuli very often suggest the existence of modality-specific constraints as well. "The pattern of results across these different studies is intriguingly consistent: contrary to the most intuitive predictions of domain-generality, the evidence persistently shows patterns of modality specificity and sometimes even stimulus specificity." (Frost et al. 2015, p. 3). As highlighted by Frost et al. (2015), studies consistently indicate limited or no transfer of learning across different modalities (Abrahamse et al. 2008; Redington, Chater, 1996; Tunney, Altmann, 1999). Furthermore, there is no indication of correlation across individuals in their ability to detect conditional probabilities across different modalities and stimuli (Siegelman & Frost, 2015). Interesting, moreover, research indicates that alterations in stimulus presentation parameters affect different modalities in distinct ways (Emberson et al., 2011). Two key indicators of domain-specificity, which we find particularly relevant for the purposes of this thesis, are: (i) the presence of domain-specific spatiotemporal structure effects. Multiple studies suggest that the visual system excels in processing statistical information in spatially distributed input, while the auditory system demonstrates an advantage in sequential input processing; (ii) the presence of qualitative differences in studies comparing sequential ISL across different sensory modalities (Conway, Christiansen, 2005).

What is the process and rationale behind a hypothesized domain-general learning mechanism systematically producing such domain-specific effects? Frost et al. (2015) provide a novel theoretical approach to implicit statistical learning. According to them, implicit statistical learning is conceived as a process involving domain-general neurobiological mechanisms dedicated to learning, representing, and processing diverse distributional properties within various input modalities. Unlike a singular learning system, these principles, crucially, are instantiated by separate neural networks situated in distinct cortical areas such as visual, auditory, and somatosensory cortex. Within their framework, domain generality arises primarily due to the instantiation of similar computational principles by neural networks across modalities. Furthermore, domain generality may also occur through the engagement of partially-shared neural networks that influence the encoding of the statistical structure to be learned, or if representations of stimulus inputs from a specific modality are channeled into a multi-modal region for further computation and learning. Consequently, the encoding of internal representations adheres to constraints determined by the unique properties of the input processed in each cortex, leading to modality-specific outcomes in computations despite the invocation of similar computational principles across multiple cortical and subcortical regions. As they explain, the current neurobiological evidence aligns with both of these latter possibilities (Frost et al. 2015). Indeed, taken together, recent neurobiological findings indicate that the recognition of statistical patterns arises from computations conducted within a specific sensory system, and via a neurocognitive system that spans multiple domains as well, influencing or acting upon inputs derived from representations specific to each sensory modality (cf. Frost et al., 2015).

In the forthcoming sections, we will delve into a comprehensive examination of studies addressing two key aspects: the discernment of domain-specific spatiotemporal structure effects. Our scrutiny will be particularly centered on visual and auditory implicit statistical learning within the framework of spatiotemporal constraints. Additionally, our exploration will extend to the analysis of qualitative differences in sequential implicit statistical learning across various modalities. Notably, we will primarily delve into studies probing this cognitive ability within the tactile domain, comparing findings with those in auditory and/or visual spheres.

3.1.2. Exploring domain-specific spatiotemporal structure effects in visual and auditory implicit statistical learning

In the realm of cognitive perception, the human experience is intricately intertwined with temporal and spatial dimensions. As we navigate the intricate landscape of implicit statistical learning across sensory spheres, it becomes imperative to discern whether and how learning through distinct sensory domains is influenced when stimuli are arranged either spatially or temporally. Specifically, we are interested in shedding light on possible domain-specific spatiotemporal structure effects, exploring their impact on learning within the auditory and visual sensory domains. The contemplation of how space and time dimensions are related our experiences in the visual and auditory sensory domains has a rich history. The philosopher Schopenhauer, in particular, delved into this subject, focusing on the fundamental role that space and time play in shaping our perceptions. In his profound reflections, Schopenhauer highlighted the distinctive nature of sensory experiences in the auditory and visual domains. He proposed that perceptions through hearing unfold exclusively in the dimension of time. Conversely, perceptions through sight are primarily rooted in space, yet, intriguingly, they bear a secondary presence in the dimension of time, a temporal quality bestowed upon them through their duration (Kubovy, 1988). "Perceptions through *hearing* are exclusively in *time;* hence the whole nature of music consists in the measure of time, and on this depends not only the quality or pitch of tones by means of vibrations, but also their quantity or duration by means of the beat or time. The perceptions of *sight,* on the other hand, are primarily and predominantly in *space;* but secondarily, through their duration, they are in time also." (Schopenhauer, 1969 [1859], p. 28). The nuanced relationship between space and time within the visual and auditory experiences was further considered by Goodfellow (1934) and Savin (1967), which have proposed that vision is particularly adept at comprehending spatial elements, whereas the perception of temporal duration may find clearer expression through auditory stimuli. (O'Connor, Hermelin, 1972). In the late 1980s, the subject reemerged, rekindling discussions regarding the interplay between visual and auditory perceptual domains and the dimensions of time and space. The analogy "space:time::vision:audition" began to permeate these debates. During this period, numerous scholars deliberated on the legitimacy of a spatial:visual and auditory:temporal dichotomy (cf. Handel, 1988; Kubovy, 1988). On one hand, Handel (1988) refutes this dichotomy. In his examination, he contends that comparing the experiences of seeing and hearing necessitates adopting a more expansive and integrated conception of space and time. Handel argues that clinging to the spatial:visual and temporal:auditory dichotomy is counterproductive. According to his perspective, the auditory and visual realms are inherently characterized by both temporal and spatial dimensions. He emphasizes that events and objects perceived through these senses are likely embedded within a framework shaped by both spatial and temporal changes. Handel (1988) rejects the notion of parceling out space and time to different senses, asserting that it is a mistake to artificially segregate these fundamental elements of perception. Instead, he advocates for a holistic understanding that recognizes the interconnected nature of spatial and temporal aspects in shaping our experiences of both the auditory and visual worlds. On the other hand, Kubovy (1988) engages in a critical examination of Handel's assertion, challenging the notion that we cannot imagine a visual or auditory event that is nonspatial or atemporal, respectively. Kubovy (1988) contends that Handel's statement either begs the question or is fundamentally false. Delving into the distinction between visual events and visual objects or scenes, Kubovy (1988) posits that while the former, by definition, involves change and is therefore temporal, the latter does not necessarily presuppose time. Vision, according to Kubovy, is not inherently temporal; looking presupposes objects located in space but does not necessitate time. Similarly, seeing presupposes objects but not events, emphasizing the potential independence of vision from temporal constraints. Kubovy acknowledges that events may unfold in the visual field, but they are not indispensable for vision. Contrary to vision, Kubovy emphasizes that audition is intimately tied with time, underscoring the inherently temporal nature of auditory perception. He goes on to propose that the analogy rejected by Handel holds partial truth, offering insights derived from his “theory of indispensable attributes”: “a) Space is the province of vision, (b) Vision is not inherently temporal, (c) Audition is intimately tied to time, (d) Audition is not inherently spatial.” (Kubovy, 1988, p.318) However, he concludes that while the space:time::vision:audition analogy is seductive, it is not fully matured. Nonetheless, Kubovy suggests that it is preferable to yield to the allure of this analogy than to succumb to potentially misleading alternative analogies. He contends that in the future, more nuanced analogies will be necessary to capture the intricate interplay of spatial and temporal dimensions in relation to hearing and sight.

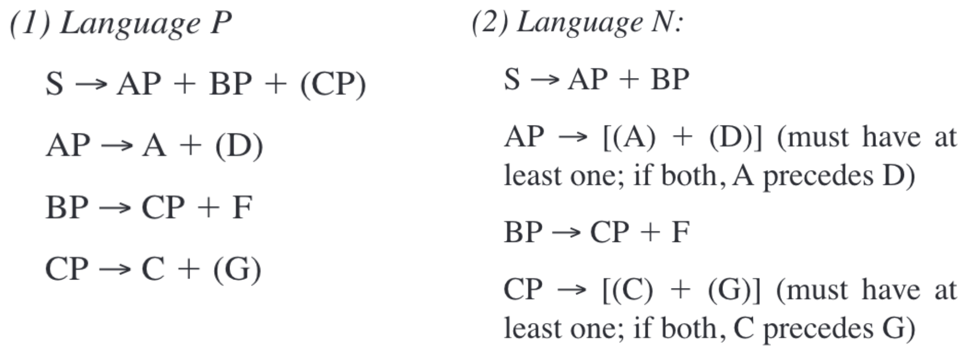
From the debate above, and based on phenomenological reasoning, we can conclude that it is indeed true that vision is more closely linked to the spatial domain, while for hearing time holds greater significance. However, this relationship is not "binding" in the sense that it does not exclude the possibility that vision can also process temporal information, just as hearing can process spatial information. Both vision and hearing can localize stimuli in space, detect movement, perceive rhythm and sequential patterns. Consider, for example, watching a video or discerning a sound source direction. Hence, we agree that it can be misleading to consider vision exclusively within a spatial framework and hearing solely within a temporal framework. "[…] all events, regardless of their sensory modality, contain temporal information that is registered by the brain [...] Time shares this supramodal nature with space [...]" (Repp, Penel, 2002, p. 1085). Crucially, however, we must not overlook the stronger association between the spatial dimension and vision on one hand, and the temporal dimension and hearing on the other. As a matter of fact, in spatial processing, the auditory system must calculate the location of sounds by considering differences in intensity and the arrival time of the sound at each ear. Conversely, the position of visual stimuli is directly mapped onto the retina and subsequently topographically projected into cortical areas. Moreover, it seems almost impossible to imagine a timeless sound or a non-spatial visual perception. On the other hand, it is easier to conceive of a sound without space or a visual scene without the passage of time (Conway, 2005). As noted by Repp and Penel (2002), the recognition that hearing and vision exhibit stronger associations with the temporal and spatial dimensions, respectively, has prompted the hypothesis that these sensory modalities are relatively specialized for temporal and spatial processing, respectively (cf. Freides, 1974; Geldard, 1970; Kubovy, 1988; Näätänen, Winkler, 1999; O'Connor & Hermelin, 1972). As Repp and Penel (2002) correctly explain, empirical evidence to test this claim could come from two sources: "comparisons of the relative sensitivity of each modality to spatial and temporal information, and studies showing dominance of one modality over the other when conflicting spatial or temporal information is presented [...]" (Repp, Penel, 2002, p.1085).

Below, we will review two compelling implicit statistical learning studies that corroborate this hypothesis (Saffran et al., 2002; Conway, Christiansen, 2009).

Saffran's (2002) exploration delved into the acquisition of predictive dependencies, specifically examining conditional probabilities. This inquiry stemmed from the consistent observation of these dependencies in natural languages, prompting the question of whether their utilization is exclusive to language learning or extends to other domains. The study encompassed a series of six artificial grammar learning experiments using the forced-choice task paradigm, which were meticulously designed to probe various facets of the research queries. Two artificial languages were used: Language P, incorporating predictive dependencies, and Language N, lacking predictive dependencies[21] (Figure 8). The participant pool included both adults and children, and the experiments covered linguistic and nonlinguistic auditory and visual learning tasks. The central hypothesis posited that learners exposed to the artificial language P, featuring predictive dependencies, would

[21] In Language P, "[…] dependencies between word categories afforded predictive cues to phrases (e.g., if D is present, A must be present). Language P contains the type of predictive structure found in natural languages. In A phrases, A words can occur without D words, but D words perfectly predict the presence of A words; the same relationship obtains between C words and G words. Similarly, C phrases can occur without F words (as optional units at the ends of sentences; the optional CP was necessary to balance the languages in terms of sentence types), but if an F word is present, a C phrase must precede it. The conditional probability of A|D is 1.0; the same is true of the other within-phrase pairs in the language." Language P, instead, "[…] was characterized by overarching optionality: the presence of one word type never predicted the presence of another. Note, however, that Language N still possesses phrase structure of a sort—the absence of one word type within a phrasal unit predicts the presence of another (e.g., if A is not present, D must be present). Language N contained the same form classes and vocabulary as Language P". (Saffran, 2002, p. 175).

demonstrate enhanced learning outcomes compared to those exposed to language N, without such dependencies. A related hypothesis explored the generalization of these effects, investigating whether the advantages extended beyond linguistic tasks to influence learning more broadly across different cognitive domains.



(1) Language P

S → AP + BP + (CP)

AP → A + (D)

BP → CP + F

CP → C + (G)

(2) Language N:

S → AP + BP

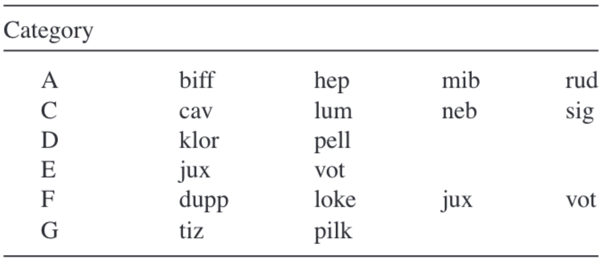
AP → [(A) + (D)] (must have at least one; if both, A precedes D)

BP → CP + F

CP → [(C) + (G)] (must have at least one; if both, C precedes G)

Figure 8. Rules of the two artificial grammars used in the six experimental studies. Picture taken from Saffran, 2002, p. 175. The grammars were adapted from those previously employed by Morgan and Newport (1981) and Saffran (2001).

Experiment 1 involved adult learners. Thew were exposed to string generated by the two artificial languages N and P. Specifically, every letter of the grammar was matched to a range of two to four monosyllabic nonsense words (see Figure 9). Participants were exposed to either Language P or Language N. They listened to a 7-minute recorded block featuring 100 sentences (from either Language P or Language N) repeated four times. After that, they underwent a testing phase. To investigate the impact of predictive dependencies on language learning, participants exposed to both Languages P and N underwent the same test format. Each test item comprised a pair of sentences: one novel grammatical sentence and one ungrammatical sentence. To differentiate between the two groups of language learners, the grammatical items were valid in both languages, and the ungrammatical items were invalid in both languages. This test format allowed for the assessment of rule acquisition in both languages. After listening to each sentence pair, participants were asked to express which sentence sounded more akin to the exposure language. Results indicated that participants exhibited superior learning for Language P, indicating that predictive dependencies enhance the acquisition of sequential auditory language-like stimuli.



Category				
A	biff	hep	mib	rud
C	cav	lum	neb	sig
D	klor	pell		
E	jux	vot		
F	dupp	loke	jux	vot
G	tiz	pilk		

Figure 9. Mapping between grammar letters and monosyllabic nonsense words. Picture taken from Saffran, 2002, p. 176.

Experiment 2 replicated and expanded the findings of Experiment 1, testing the ability to acquire predictive dependencies in child participants. The methodology was the same as Experiment 1. Children demonstrated enhanced learning when exposed to the artificial language with predictive dependencies, thus replicating findings from Experiment 1, and suggesting that the impact of these dependencies might extend to the early stages of language acquisition. Experiment 3 delved into the exploration of the learnability of predictive dependencies in nonlinguistic contexts. The two languages were translated into nonlinguistic sounds, such as ascending buzz, chimes, a chord, ... Notably, participants consistently performed better on Language P than Language N also in this nonlinguistic auditory task. This observation replicated the trend seen in linguistic-like tasks, underscoring the broader influence of predictive dependencies beyond traditional linguistic contexts. Indeed, the findings suggested that predictive dependencies play a significant role in shaping learning outcomes across diverse cognitive domains. Experiment 4 explored visual learning across two modalities, serving as a conceptual replication of Experiments 1 and 3 within the visual domain. Participants encountered either visual nonsense words or visual nonsense shapes during the study. For the nonlinguistic visual condition, Languages P and N were transformed into several shapes. Each "word" represented a unique nonsense shape, such as a red asymmetric oval with yellow dots. These shapes were sequentially presented on the monitor, one at a time. The linguistic visual condition mirrored this process, but instead of shapes, the nonsense words from Experiment 1 appeared in typed capital letters. In the test phase, participants viewed two sequences (made of shapes in the nonlinguistic visual condition or words in the linguistic visual condition) and determined which was more akin to the exposure language, responding through a key press. While predictive dependencies enhanced learning in the auditory domain, as found in Experiments 1-3, they did not exert the same influence in the visual domain. The results suggested that predictive dependencies play a role in auditory but not visual learning, at least within the parameters used in Experiment 4. Interestingly, the absolute performance levels were similar between auditory and visual tasks, but the patterns of performance differed, indicating that predictive dependency cues have a more substantial effect on auditory learning. As Saffran (2002) suggested, this discrepancy might be attributed to the nature of the stimuli and the learners' interpretation. Specifically, the auditory nonlinguistic stimuli in Experiment 3, although devoid of linguistic content, may have been recoded as linguistic entities by the learners. This could not have happened, on the contrary, in the visual tasks in Experiment 4, featuring nonsense shapes without clear linguistic associations. Hence, the possibility remains that predictive dependencies may be operative only in processing language-like stimuli. Therefore, Saffran (2002) aimed to shed more light on the actual possibility that predictive dependencies play a role in learning stimuli outside of language, in other domains. To do so, in Experiment 5, Saffran (2002) compared learning performances between auditory stimuli that would be challenging to verbally label and visual stimuli easy to label. Specifically, for the set of nonlinguistic sounds, stimuli such as drums and bells were used, whereas for visual nonlinguistic shapes stimuli included shapes such as circles, triangles, and hearts. This design aimed to test the influence of ease of verbalization on the modality difference observed in the earlier experiments. If the results from Experiments 1–4 were influenced by the ease of verbalization, we would expect a reversal of the pattern in Experiment 5, with the visual task now exhibiting the effects of predictive dependencies, while not the auditory task. On the other hand, if the previously observed modality effect persisted with the new stimuli, it would suggest that predictive dependencies enhance the learning of sequential stimuli in auditory tasks but not in visual tasks. Result showed that subject showcased superior performances in acquiring language P than language N only in the auditory presentation, replicating the observed pattern of results in the initial four experiments. Crucially, the nature of the materials, whether linguistic or nonlinguistic, did not impact the outcomes. This supports the hypothesis that the constraint in detecting predictive dependencies is not exclusive to language learning. Based on the results observed in Experiment 1-5, the central question which remain to be answered is why predictive dependencies impact sequence learning in the auditory domain but not in the visual domain. Saffran (2002) suggested that the preferential processing of predictive relationships in audition is due to the inherently sequential nature of the auditory world, where sounds are transient and do not persist over time. This is particularly true for linguistic information, but also musical patterns and nonlinguistic sounds, which demand tracking sequences and discerning relationships between events separated in time. In contrast, processing visual scenes mostly require tracking relationships among objects in space, suggesting that visual information is inherently less sequential than auditory information, with exceptions such as signed languages and gestures. Based on these considerations, it is plausible that in a visual task involving simultaneously present predictive dependencies, learners might show an advantage in acquiring language P like in auditory experiments using sequential stimuli. Experiment 6 was designed to test this hypothesis. Experiment 6 served as a conceptual replication of Experiment 4. The same stimuli were used. Importantly, however, in contrast to the sequential presentation of shapes one by one, each "sentence" in Experiment 6 involved simultaneous presentation, with all shapes from the sentence arranged spatially on the screen. The shapes consistently appeared in a specific position on the screen. For example, "A word" shapes were consistently positioned in the upper righthand corner, while "F word" shapes were consistently presented in the middle of the bottom of the screen. This layout, as opposed to a sequentially ordered one, aimed to reduce the likelihood of learners adopting a sequential left-to-right processing strategy. Crucially, learners trained on language P significantly outperformed those trained on language N. This finding corroborated the hypothesis that the visual learning system is more finely tuned to track dependencies among elements simultaneously presented and arranged in space than among elements presented sequentially and arranged in time. Summing up, Saffran (2002) offered significant evidence supporting the idea that individuals can leverage predictive dependencies in AGL experiments across various sensory modalities, albeit with constraints related to the mode of stimulus presentation. Specifically, in the auditory modality, where information is typically sequential and fleeting, sequential presentation triggers effects of predictive dependencies. Similarly, learners can identify and use predictive dependencies in the visual modality, but only when the input is simultaneous, spatially arrayed. As Saffran (2002) explains, it remains uncertain whether these effects stem from inherent perceptual/processing differences or are shaped by experience in each modality.

Conway and Christiansen (2009) investigated the impact of varying presentation formats and rates on implicit statistical learning abilities, focusing on visual and auditory modalities. Three presentation formats were explored: visual input distributed spatially, visual input distributed temporally, and auditory input distributed temporally. Concerning presentation rates, two formats were investigated: a slow and a fast one. To explore how presentation rates and temporal/spatial constraints interact with visual and auditory statistical learning, they employed the AGL paradigm. Participants were exposed to visually or auditorily governed input sequences generated from a finite-state artificial grammar. Based on previous discussions and findings on the topic (cf. Conway, Christiansen, 2009; Saffran, 2002), they predicted optimal learning for visual-spatial and auditory (temporal) conditions, and poorer performance for visual-temporal formats. Concerning the impact of presentation rate on learning, they consider it as an aspect insufficiently studied in statistical learning tasks (Conway, Christiansen, 2009). Despite the scarcity of evidence in the issue, faster presentation rates were expected to accentuate modality constraints, negatively influencing learning in the nonpreferred mode. Hence, they predicted that participants would perform worst in the visuo-temporal condition at a fast presentation rate. The experimental study included an acquisition and a test phase. During the acquisition phase, legal sequences generated by the artificial grammar were utilized, while the test set included both legal and illegal sequences. Legal sequences adhered to the same finite-state grammar rules, while illegal sequences incorporated legal elements followed by illicit transitions and concluding with a legal element. The symbols of the sequences were mapped onto three types of stimuli: visual-temporal, visual-spatial, and auditory.

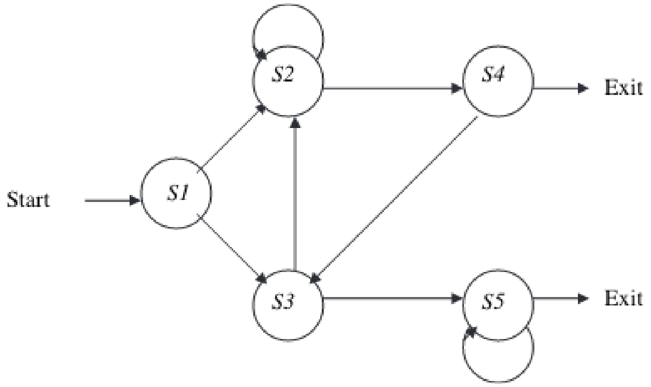


Figure 10. Finite-state artificial grammar used in the experimental study. Figure taken from Conway and Christiansen, 2009, p. 565.

Visual-temporal stimuli comprised sequentially appearing colored squares, each presented for 250 ms (slow) or 125 ms (fast). Visual-spatial stimuli displayed the same-colored squares simultaneously presented in a horizontal row. The temporal duration of the stimuli matched the combined presentation time of the single visual-temporal stimuli. Therefore, the sequence comprised a simultaneous array of squares arranged horizontally from left to right, displayed for a total duration of 1000 ms (250 X 4) in the slow mode, or 500 ms (125 X 4) in the fast mode. Auditory stimuli consisted of sequences of pure tones conveyed through headphones, in which each stimulus had a duration of 250 ms (slow) or 125 ms (fast). Participants were randomly assigned to 12 conditions (3X2 design, i.e., modality X presentation rate format), with six experimental groups undergoing both acquisition and test phases, and six control groups, serving as a baseline, participating in the test phase only. During the acquisition phase, participants performed a match/mismatch task, deciding whether pairs of sequences were the same without giving information about the underlying structure. This was intended to maintain participants' attention high. During the testing phase, participants were asked to categorize each novel sequence based on whether they believed it adhered to the same rules as those encountered earlier. Control participants, not involved in the acquisition phase, underwent an equivalent task. The timing presentation format mirrored that of the acquisition phase. Results for the acquisition phase revealed a notable disparity in performance between modalities, with both the auditory and visual-spatial groups exhibiting significantly superior results compared to the visual-temporal group. Despite this, all six groups performed better than chance level in the match/mismatch task. In this phase, the presentation rate did not influence task performance, in any of the three modalities. As for the testing phase, none of the six groups of control performed above chance levels. This suggests that any observed learning in the experimental groups is attributable to statistical learning taking place in the training phase. Within the experimental group, the auditory group performance was significantly greater than that of the visual-temporal group. Moreover, also the visual-spatial performance exceeded the visual-temporal one. Comparisons of performance between fast and slow rates for each modality/format condition yielded only one significant result for the visual-temporal group, indicating that visual-temporal performance significantly declined at the fast rate compared to the slow rate. Specifically, the visual-temporal group in the fast condition performed the test task at chance levels. Hence, only the auditory and visual-spatial groups exhibited learning at the fast rate, while the visual-temporal group performed no better than chance. Based on these results, the authors concluded that presentation rate affects in different ways statistical learning across modalities/formats. However, to address potential associations between acquisition and test-phase performance, they conducted correlation analyses. Indeed, as Conway and Christiansen (2009) noted, the observed quantitative learning differences across modalities may be explained by a potential association between acquisition-phase and test-phase performance. The superior test-phase performance in the auditory and visual-spatial conditions could be due to certain stimuli being more easily perceived and remembered during the acquisition phase. More perceptible stimuli might have (positively) influenced the learnability of statistical regularities. On the other hand, however, the absence of performance decline in the match/mismatch task at a fast presentation rate for any input conditions, compared to a notable decrease in test performance for the visual-temporal condition, suggests that the test-phase results cannot be solely attributed to acquisition-phase performance. Correlation analyses clarified the various possibilities at play. Results indicated that only the visual-spatial condition showed a statistically significant correlation between acquisition and test-phases performances. This result implied quantitative learning differences between modalities, possibly influenced by differences in perceiving and remembering stimuli in the different conditions. Hence, the authors decided to explore potential qualitative modality effects. To do so, regression analyses were conducted to identify the sources of information extracted in each modality/format condition. They evaluated both legal and illegal test items based on their initial and final anchor strengths (IAS and FAS). These metrics indicates the relative frequencies of the initial and final fragment "chunks" (i.e., bi- and trigrams) that are present in analogous positions in the training items[22]. These analyses revealed notable differences between auditory and visual conditions: auditory learning relied heavily on fragment information at sequence endings, while visual learning was more sensitive to information at sequence beginnings. Overall, these findings suggested that statistical learning is constrained by factors related to presentation modality, rate, and format (spatial vs. temporal distribution). Participants in visual conditions exhibited superior performances related to the extraction of statistical patterns when presented in a spatial format rather than a temporal one. Moreover, visual learning relied more on statistical information at the beginning of input sequences. Conversely, the auditory modality excelled in encoding temporal input, showing heightened sensitivity to the statistical structure at the end of input sequences. Furthermore, modality constraints

[22] “For example, the test item 1-2-1-3-5-2 has an IAS of 4.5 and an FAS of 2.0, indicating that the initial chunks 1-2, 2-1, and 1-2-1 occur frequently in the initial positions of the training set, whereas the final chunks 3-5, 5-2, and 3-5-2 occur slightly less frequently in the final positions of the training set.” (Conway and Christiansen, 2009, p. 572).

were magnified at the fastest presentation rate, notably negatively affecting visual-temporal learning. This suggested that vision struggles to encode temporal regularities, particularly at high presentation rates.

In this section, we have considered domain-specific spatiotemporal structure effects relative to visual and auditory implicit statistical learning. As we have seen, many scholars have explored both theoretically and empirically how learning in diverse sensory domains is affected based on whether stimuli are organized spatially or temporally and discern the nature of this influence. Importantly, we have presented findings from two studies that offer compelling evidence supporting the idea that the visual domain excels in processing spatially presented statistical information, as opposed to hearing that performs better in processing statistical information when stimuli are arranged sequentially (i.e., in the temporal dimension). Saffran (2002) provided convincing proof supporting the ability to acquire predictive dependencies in AGL tasks across various sensory modalities. Crucially, she identified constraints associated with the modality of stimulus presentation. This ability was observed in the auditory modality with sequentially arranged stimuli. In the visual domain, learners could discern predictive dependencies only when the input was simultaneously presented and spatially organized. Conway and Christiansen's study (2009) revealed constraints on statistical learning tied to presentation modality, rate, and format, aligning with Saffran's findings. According to their results, in the visual condition, participants demonstrated enhanced performance in extracting statistical patterns when information was presented spatially. Moreover, the visual domain was more adept at encoding statistical information at the beginning of sequences. In contrast, auditory learning showed superior performance in tracking statistical information in temporal input, being particularly sensitive to sequence endings. Modality constraints in the visual-temporal domain were magnified at faster rates, negatively impacting learning, thus highlighting increased challenges in encoding visual temporal regularities at high speeds.

Crucially, we find no studies that have investigated domain-specific spatiotemporal constraints in the tactile modality. The issue of whether touch is better at processing statistical information when presented in the spatial or temporal dimension has not been empirically addressed in the literature, as far as we know. Unlike the extensive research on implicit statistical learning in the visual and auditory domains, the tactile domain has been significantly underexplored. Only recently have some scholars started shedding light on the capacity to acquire statistical information through touch. As we will see, the few studies on implicit statistical learning in the tactile sphere have focused on learning statistical information from sequentially presented inputs, thus in the temporal dimension. Despite being limited in number, these studies have revealed intriguing results, especially when comparing tactile sequential learning with visual and/or auditory learning. In the next section, we will delve into statistical learning in the tactile domain.

3.1.3. Sequential implicit statistical learning in the tactile sensory domain

"Viewed from phylogenetic and ontogenetic perspectives, the sense of touch plays a central role relative to the other senses. Its fundamental significance to humans derives from its epistemological function, making possible an awareness of surroundings and the consciousness of self. In this way, the sense of touch is sine qua non for thought, action, and consciousness" (Grunwald, 2008, preface). The sense of touch has captured the attention of philosophers and scientists over the years. Back in ancient Greek, philosophers asserted the supremacy of the touch sense, which was often described as the basic sense, the sense prototype. Empedocles, with the word *pagamai,* which means gripper, or flat of the hand, generally referred to the senses (Grunwald, 2008). In the Middle Age, Aristotle, in his *De Anima*, praised the uniqueness qualities of touch sense, which, differently from the other senses, was described as the only one through which we establish a direct contact with the properties of the object of the perception. Hence, the emphasis is on the close contact between the object of the perception and the sense of touch, whose organ, according to the Greek philosopher, is not placed in the skin but in the heart (Grunwald, 2008). The interest in the haptic sense did not diminish in the following years. In the Middle Ages, St. Thomas Aquinas, in his own *De Anima*, emphasized the centrality of touch by claiming that the other senses could not exist without it. From touch (*radix fontalis*), all the other senses originated and are related to (Grunwald, 2008) "[…] the most likely sense-faculty would seem to be touch, the first sense, the root and ground, as it were, of the other senses, the one which entitles a living thing to be called sensitive" (Aquinas, book III, Chapter II, lectio three, § 602, trad.1951). However, it is only with the advent of the 19th century that scientists started to systematically investigate the physiology of haptic perception. The German physiologist and anatomist Ernst Heinrich Weber (1795-1878) carried out the first experiments on the haptic sensory threshold. Interestingly, Weber's aim was not to exclusively focus on the haptic sphere: his studies in the somatosensory domain were intended to find basic principles of perception which could have been later extended to the study of other senses, such as vision (Grunwald, 2008). During the 19th and 20th centuries, the scientific progress permitted to understand the physiological mechanisms of the haptic sense and to identify its anatomy more precisely. Yet, several issues concerning the mechanisms involved in tactile perception remained unexplained. Much remain to be explored, as far as the psychological reality of tactile perception is concerned. This situation is probably due to two causes. Firstly, the complexity and the many facets of the tactile sensory domain has hampered the scientific knowledge to progress. "No other sense exhibits properties so variable in scope or remains so puzzling even today – understood only in terms of its principle features" (Grunwald, 2008, preface). Moreover, the paucity of experimentation carried out by psychologists in the domain of haptic learning and perception, and the consequent shortage of available data, caused a delay in its understanding, as compared to the scientific advance reached in the study of visual and auditory perception and learning.

Very little research has been devoted to implicit statistical learning in the tactile sensory domain. Although it is widely known that the tactile sense can be exploited to acquire information from the environment, and even though the psychophysical and perceptual attributes of the touch sense have been extensively investigated (cf. Craig, Rollman, 1999), surprisingly, very little attention has been focused on exploring the ability to implicitly learn statistically organized information through the tactile sensory domain. (Conway, Christiansen, 2005). Most of the studies conducted so far, both in the research field of implicit learning and statistical learning, have focused primarily on the visual and auditory domains, which has been thoroughly investigated. On the contrary, the sense of touch has been almost completely ignored. Scholars have recently started to provide evidence that we can detect and acquire statistical information tactilely[23]. In the next section, we will focus on studies which have explored tactile implicit statistical learning comparing it with auditory and/or visual learning.

3.1.4. Comparing sequential implicit statistical learning across the tactile, visual, and auditory domains

Conway and Christiansen (2005) investigated whether statistical regularities between elements organized in a sequential input can be detected through touch, vision, and audition. Moreover, they were also interested in verifying which differences, if any, might have occurred in the learning process in the three modalities. They encoded the symbols generated by a finite-state grammar[24] onto visual, tactile, and auditory stimuli, and they presented them to three groups of participants, respectively. Specifically, stimuli were vibrotactile pulses transmitted to the fingers of participants' hands, tones of different frequencies, and black squares that appeared in specific locations on a computer screen. The grammar they employed can generate 23 different sequences of numbers, with a length ranging from three to six elements. The test was divided into two phases: a training phase, in which a total of 12 legal sequences were utilized, each employed twice to create a set of 12 training pairs. Among these pairs, six comprised identical training sequences presented twice (matched pairs), while the remaining six pairs featured two sequences with slight variations (mismatched pairs). Then, in a test phase,

[23] To our knowledge, the only few implicit learning studies conducted in the tactile domain are the following: Abrahamse et al., 2008; 2009; Conway, Christiansen, 2005; Pavlidou, Bogaerts, 2019;

[24] They used the same grammar of Gomez and Gerken (1999).

participants were presented to 10 illegal and 10 novel legal sequences. Legal sequences adhered to the finite-state grammar’s rules, while illegal sequences deviated from the grammar's rules. Specifically, the illegal sequences commenced with legal elements and concluded with legal elements but included multiple illegal transitions within. Thus, the distinction between legal and illegal sequences layed in the statistical relationships among adjacent elements. Participants were divided into two groups: the experimental group, which took part both in the training and in the test phase, and a control group, which skipped the training phase, being exposed only to the test-phase strings. In the test phase, the experimental group was asked to judge pairs of legal and illegal sequences by pressing on “yes” - “no” buttons. Before the training phase, participants in the experimental group were briefed on their involvement in a sensory experiment where they would experience pairs of sequences. Their task was to determine whether each pair was identical and express their decision by pressing a yes/no button. The presentation of each pair occurred randomly and was repeated six times, resulting in a total of 72 exposures. Prior to the test phase, the experimental group participants were informed that the sequences were generated by a set of rules. They were then informed about the upcoming presentation of new sequences. They were informed that some of these would have followed the same generating rules as those in the previous session, while others would have not. The participants' task was to classify each new sequence based on whether it adhered to the same rules or not. The group of participants who did not undergo the training phase (control group) were assigned an identical task. In the tactile condition, the numbers of the sequences were transmitted through vibro-tactile stimuli (150 Hz), generated by five small motors (typically employed in handheld paging devices) which were placed on participants’ fingers. Each number of the grammar was linked to a specific finger stimulation. The duration of each finger pulse was 250 ms, with a 250 ms gap between pulses within a sequence. In the visual condition, black squares were presented on the computer monitor, each appearing in distinct locations denoted by elements 1 to 5, with 1 representing the leftmost and 5 the rightmost location. Hence, a visual stimulus comprised a spatiotemporal sequence of black squares appearing at different locations. Each element was visible for 250 ms, and there was a 250 ms gap between each element. In the auditory condition, the stimuli were composed of pure tones, each representing musical notes such (i.e., C, C#, F, F#, and B). Also in this task, the duration of each element (tone) was 250 ms, with a 250 ms interval between consecutive tones. Results clearly indicated that learning occurred in all three domains, since the experimental outperformed the control groups, in all the three domains. Specifically, in the tactile task, the experimental group correctly classified 62% of sequences, whereas the control group 45% of them; in the auditory task, the experimental group correctly classified 75% of sequences, while the control group 44%; in the visual task, the experimental group correctly classified 63% of the sequences, while the control group the 47%. However, a quantitative difference was detected: participants in the auditory group significantly outperformed those assigned to the visual and tactile groups. Moreover, they also found a qualitative learning difference between the modalities: Tactile learning revealed itself to be more sensitive to the initial information in the strings, while auditory learning tended to be most sensitive to the final information within sequences. Nevertheless, a lingering question remained regarding whether the noted distinctions in learning outcomes stemmed solely from the low-level, perceptual characteristics of the specific stimulus elements employed in the three experiments. For example, auditory stimuli might have been perceptually more salient than tactile or visual stimuli. To investigate further this issue, they developed a second experiment. In this second experiment, they added a pretraining phase in which they assessed the perceptual comparability of stimuli in the three different modalities. Moreover, they modified the training phase to ensure that participants underwent comparable training in the three sensory domains. In addition to this, they also investigated more finely qualitative learning differences. In this experiment, they used the same apparatus as that used in the first experiment, however, a different, more complex finite-state grammar was employed, which generated a wider range of more complex sequences, thus enabling the creation of a more difficult task. Moreover, this grammar was symmetrical with respect to the possibility to have certain bigrams or trigrams in the initial or final position of the sequences. In other words, it contained no biases toward either the beginning or ending aspects of sequences in terms of chunk information availability. Tactile and visual stimuli were identical to those used in the first experiment. Auditory stimuli, on the contrary, were slightly different: they used a set of different melodies to avoid familiar musical notes and to have a narrower frequency range (i.e., the tones had frequencies of 220 Hz; 246.9 Hz; 261.6 Hz; 277.2 Hz; 329.6 Hz). Six groups of participants took part in the experiment and were randomly assigned to 6 different tasks: visual, auditory, and tactile tasks; visual control, auditory control, and tactile control tasks. In the control condition, participants skipped the training phase, hence taking part only in the pretraining and test phases only. The pretraining phase was created to ensure that the stimuli used were suitable, clearly distinguishable, and perceptible. Moreover, in this phase, participants had the opportunity to familiarize with stimuli. A discrimination task was conducted, in which participants were presented with pairs of stimuli within each modality and were required to determine whether these stimuli were identical or different, rating their similitude on a scale from 1 to 7. This test confirmed that stimuli were appropriate, in all three modalities. Indeed, they turned out to be both discriminable and psychologically perceived in comparable ways in the three modalities. In the training phase, participants were exposed to 24 grammatical sequences. Each sequence was paired with a specific bigram fragment. In half of the sequences, the bigram was present into the sequence itself (e.g., 3–5–4–**1**–**2**–3–1 and 1–2). For the remaining half, the bigram was not present in its entirety within the sequence, but its constituent elements were (e.g., **1**–2–**3**–1–4–5–2 and 1–3). In all instances, the bigrams adhered to the rules of the finite-state grammar. Then, a test phase began, which comprised 16 new grammatical sequences and 16 new illegal sequences. Among this last group of sequences, 8 were illegal-initial sequences, created by changing the second and third element from a grammatical sequence, whereas 8 were illegal-final sequences, and were created by changing the third-to-last or second-to-last element from a grammatical sequence. Every illegal sequence was matched with the grammatical sequence from which it originated, ensuring a balanced distribution where each sequence appeared both at the beginning and the end. This resulted in a total of 32 test pairs. In this phase participants were asked to determine if the pair of elements had appeared consecutively within the sequence by pressing a yes/no key. The aim of the training phase was for participants to pay attention to the legal training sequences without being explicitly informed about the statistical regularities present in the sequences. Results indicated that while tactile training performance was slightly lower compared to visual and auditory performances, overall, scores across the three modalities were approximately equal. The test phase aimed to evaluating participants' ability to grasp the statistical patterns within the training set and apply this knowledge to new stimuli during a classification task. Before starting the test phase, participants were informed that in the training phase they were exposed to sequences generated by grammatical rules. In this phase, they were presented with sequences generated by the same rules as those in the training phase, and sequences that were not generated by those rules. They had to indicate whether these sequences were or not generated by the same rules as those in the training phase. Results indicated that only the auditory experimental group outperformed the auditory control group, thus indicating that participants learned the statistical regularities only in the auditory task. Moreover, they found that participants in the auditory modality were better at discriminating statistical regularities in the final part of sequences that in the initial part. Overall, the second experiment, confirmed the presence of both qualitative and quantitative learning differences across the three different sensory modalities. The absence of learning in the tactile and visual modality in the second experiment was attributed to the higher complexity of the second experiment compared to the first one. Indeed, according to the authors, the grammar was presumably too complex and the differences between grammatical and illegal sequences too subtle. Summing up, results from the first experiment confirmed that both the tactile, auditory, and visual modalities can track and acquire statistical regularities in sequentially presented input. Importantly, however, in both the two experiments, they found both quantitative and qualitative learning difference: the auditory modality is superior to both the visual and the tactile modalities in learning statistical regularities when sequentially presented. Moreover, the auditory modality is more sensitive to statistical regularities in the final part than in the initial part of the sequence.

Abrahamse et al. (2008) investigated and compared sequential statistical learning in the visual and the tactile domains through a serial reaction time task. The aim of their study was to verify whether implicit sequential learning is predominantly motor-based, hence independent from the specific modality through which stimuli are presented, or, if the associations between stimuli are formed in a specific perceptual domain, in other words, if the nature of the learning process is stimulus-specific. To answer their research question, they developed a serial reaction time task in which they exposed two groups of participants to visual or tactile stimuli which were encoded onto the symbols of a 12-element sequence in which there were second-order conditional transitions. Participants in the visual condition were exposed to a sequence of rectangles appearing in one out of four possible locations on a computer screen and they had to press one of four specific keys on the keyboard as fast and accurately as they could when they saw the stimulus, based on its position. Participants assigned to the tactile group perceived a vibrotactile stimulus on one of four fingers and, as for the visual condition, they had to respond by pressing on one of four keys on the keyboard, based on the location of the received stimulation. The task consisted of two phases: the first phase (learning phase) in which both groups of participants were exposed to one random block of tactile stimuli to familiarize participants with the task. Then, they were exposed to 11 sequence blocks (where stimuli followed the rules of the grammar), one random block, and one final sequence block, transmitted though visual or tactile stimuli, depending on the group condition. Then, they were exposed to a second phase (transfer phase) in which participants who trained in the tactile condition were switched to the visual condition, and vice versa. This phase was composed of one random block, one sequence block, and one final random block. The authors were firstly interested in verifying whether participants in the two conditions would learn the statistical regularities contained in the sequence blocks. Moreover, they were interested in testing transfer abilities from one to the other modality. Are subjects who have been trained in the tactile condition able to transfer their implicit knowledge when they become tested in the visual domain? Does it also apply the other way around? To assess the degree of sequence learning in the two groups, they compared the average reaction times (RTs) and error percentage in the combined blocks 11 and 13 with those in the random block 12. Transfer, instead, was assessed by comparing the average RTs of the combined random blocks 14 and

16 with the RTs of block 15. Analyzing the trend of RTs and accuracy rates from block 2 to 11, it was observed that the tactile group exhibited generally higher RTs compared to the visual group and lower accuracy rates. Despite this, the trend of RTs between blocks followed a similar pattern in both modalities, confirming the same sequence learning effect in both modalities. Moreover, the analysis confronting the mean RTs of Blocks 11 and 13 with the random block 12, showed that RTs in the random block 12 were significantly higher than those in blocks 11 and 13, in both the visual and tactile groups. Importantly, however, they found group differences indicating that the difference was significantly more pronounced in the visual group. Overall, the researchers construed these findings to indicate successful acquisition of sequential regularities in both the tactile and visual conditions. Nevertheless, despite this achievement, they concluded that tactile sequence learning happened to a lesser degree when contrasted with visual sequence learning. Then, to verify transfer from one to the other modality, they confronted mean RTs of blocks 14 and 16 with those of block 15, in both the two groups. The results indicated that both groups exhibited similar transfer effects. However, they noticed that the visual group was significantly slower in transitioning to the tactile condition compared to the tactile group transitioning to the visual condition (i.e. passage from block 13 to block 14) (see Figure 11). To investigate deeper into this effect, the authors performed a more sophisticated analysis in which they calculated a learning score and a transfer score. The learning score consisted of the difference between the means of block 11 and 13 compared to block 12, while the transfer score was the difference between the means of block 14 and 16 compared to block 11. They found that the learning score of the visual group was significantly higher than the transfer score of the visual to tactile modality. This difference was not found in the comparison between the tactile learning score and the tactile to visual transfer score (Figure 12). The authors interpreted this result as indication of a decrease in performances for the visual to tactile stimuli condition: while the tactile group showed perfect transfer in the visual modality, the visual group was only able to partly transfer sequence knowledge to the tactile modality. The authors commented on their results by claiming that, in contrast with what had been proposed by many scholars, implicit sequential

learning cannot be entirely motor-based. If this were the case, there should have been no differences neither in learning outcomes in the two different modalities nor when transferring from one to the other modalities. This was not found in their experiment. They concluded by supporting the idea according to which different components play a role in sequential implicit learning. Both motor-based and stimulus-specific abilities are involved in the process.

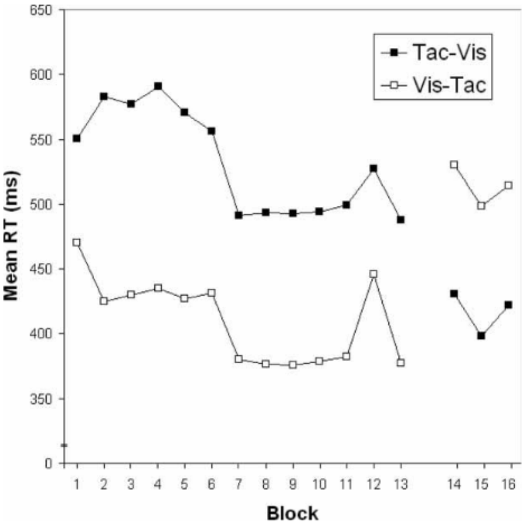


Figure 11. Mean RTs for the two phases (learning phase: block 1-13; and transfer phase: block 14-16) in the two groups (Figure taken from Abrahamse et al. 2008, p. 213).

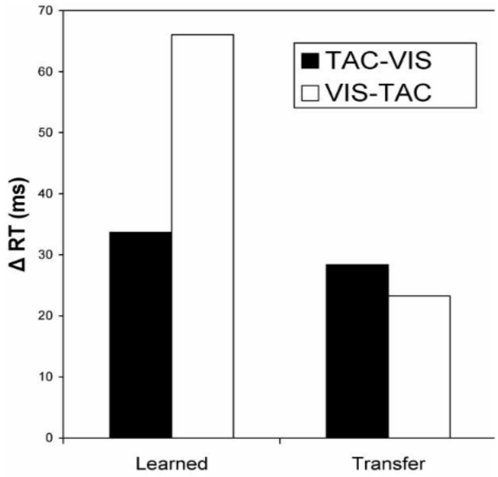


Figure 12. Learning effect and transfer effect for the two groups (Figure taken from Abrahamse et al. 2008, p. 215).

Abrahamse et al. (2009) investigated further the results found in Abrahamse et al., 2008. Firstly, they were interested in examining more in-depth the perfect transfer from the tactile to the visual domain and the partial one for the other way around. Secondly, they wanted to verify whether the presentation of congruent and temporally synchronized visual and tactile stimuli would have enhanced learning. Hence, they developed a new protocol in which they compared three groups: a visual, a tactile, and a bi-modal group. Their interest was verifying whether the bi-modal group would have shown an advantage over the visual-only group. The task consisted of a training phase and a transfer phase. In the training phase, the three groups of participants were exposed to a pseudo-random block of stimuli, 10 sequence blocks, a pseudo-random block and at the end a final sequence block (tot. 13 blocks). Every sequence block contained the same second-order conditional sequence[25] that was repeated nine times. The random blocks contained nine

[25] The SOC used was the following: 242134123143. One sequence block: (242134123143) repeated 9 times.

“An SOC sequence contains no predictive first-order information (all first-order transitions 12, 13, 14, 21, 23, etc., occur equally often), but each first-order transition is followed by a unique position in the sequence (e.g., after transition 12 only position 1 can occur (e.g., after transition 12 only different SOC sequences, with no repetitions. The training phase was followed by the transfer phase, in which participants in the three groups were tested on the transfer to each of the three conditions. Transfer phases consisted of one random block, one sequence block composed of the same SOC sequence used in the training phase repeated 4 times, and a final random block. As far as the training phase is concerned, results confirmed the previous findings according to which sequence statistical learning occurred in both the visual and the tactile conditions, but the tactile group was slower in general, both when compared to the visual and to the bimodal groups. Even accuracy rates were lower for the tactile group as compared to the visual one. Interestingly, the addition of the tactile stimuli to the visual ones did not enhance learning. Indeed, they did not find a significant difference between the visual-only and the bimodal groups. Importantly, however, the time course of learning has been revealed to be the same in the three groups (see Figure 13). By observing the results from the transfer phases, the authors verified whether the knowledge acquired during the training phases would have still been accessible when transferring to different sensory modalities. Results indicated that transfer occurred for all three modalities, with no significant differences between the visual-only and the tactile-only transfer conditions, but with a reliable difference between the bimodal transfer condition and both the visual-only and the tactile-only conditions. Indeed, when switching to the bimodal condition, the tactile training group showed worse transfer scores as compared to the visual-only and the bimodal training groups. This result confirmed the one obtained in the training phase: the bimodal condition of stimuli presentation did not enhance learning. Based on these results, Abrahamse et al. (2009) concluded by claiming that, the differences found between visual and tactile groups do not reflect a difference in sequence learning abilities in the two sensory domains. In fact, there might be a difference in the

position 1 can occur), thus the sequence is only predictive on a second-order level. In comparison, learning of FOC sequences can be based on first-order information about the immediate preceding position. […] In an SOC sequence, an event *t* is predicted by the previous two events, in which P[*t*|(*t*-2), (*t*-1)] is the same for all sequential events." (Du & Kelly, 2013, p.157).

An example of FOC (first-order transition) is: 13234213414. This sequence has been used in Deroost et al., 2010.

expression of the knowledge acquired, that does not correspond to a reduced sequence learning ability in the tactile group. “[…] rather than sequence learning it seems the expression of sequence learning that is impaired with single tactile stimuli compared to single visual stimuli” (Abrahanmse et al. 2009, p.182). As Abrahamse et al. (2009) pointed out, this interpretation is in line with the ideas of other scholars in the field (see Deroost et al. 2009; Frensch et al. 1998; Hoffmann and Koch 1997). The authors concluded by underlying the importance of taking into consideration the following observation when interpreting results from implicit learning studies: results should not be directly taken as an expression of sequence learning; in fact, they reflect the degree of sequence learning combined with the task-dependent constraints for the expression of those knowledge (Abrahamse et al., 2009).

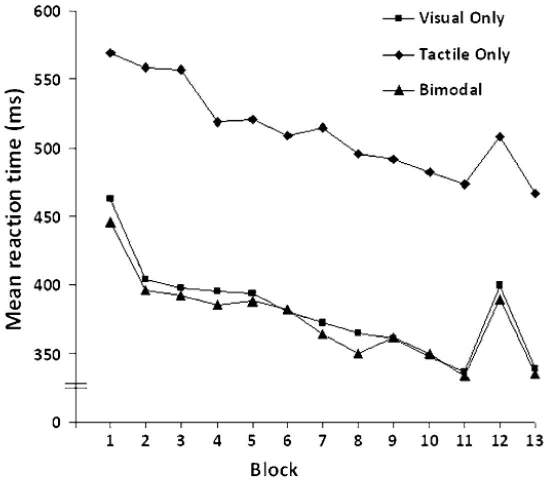


Figure 13. RTs curves for the visual, tactile, and bimodal conditions across blocks in the training phase. Blocks 1 and 12 are random blocks. (Figure taken from Abrahamse et al., 2009, p. 179).

Pavlidou & Bogaerts (2019) have been the first who investigated implicit statistical learning abilities across the visual, auditory, and tactile sensory domains and their relationship with reading competencies in children, using the AGL paradigm. The aim of their study was twofold: on one side, they wanted to verify whether ISL would have occurred in all three sensory domains. Is ISL a unified ability?

Secondly, they aimed at verifying the relationship between ISL and reading and reading-related abilities in typically developing children (basic reading skills, reading fluency, and phonological awareness). For all three modalities, the tasks were composed of two phases. In the training phase, children were exposed to sequences of stimuli encoded onto the strings generated by an artificial grammar[26]. After the training phase, children were informed about the presence of rules in the strings they had just been presented with and, in the subsequent phase, that is the testing phase, they were presented with new strings and they were asked to judge them as grammatical or ungrammatical, based on their feelings concerning how familiar these strings looked to them. For the visual task, stimuli consisted of alien images. Interestingly, in Pavlidou & Bogaerts's visual experiment, the strings were presented all at once, one at a time. In other words, visual sequences were presented simultaneously and spatially. This is a point of departure from the presentation modality that was adopted for the visual tasks in both Conway and Christiansen, 2005 and Abrahamse et al., 2008; 2009. Indeed, in these last studies, the visual sequences were presented in a sequential, temporal manner, in line with the auditory and tactile stimuli presentations. In other words, the symbols of the visual sequences appeared one at a time, in different locations on the screen. Pavlidou & Bogaerts's decision concerning the modality presentation of visual stimuli was based on the observation that the visual domain is more suited to deal with statistical information that is contained in spatially arranged elements, as opposed to the auditory domain, which is better at tracking statistical regularities that are sequentially presented (for further discussion and references see Conway and Christiansen, 2005). For the auditory task, stimuli consisted into 5 different pure tones (261.6 Hz; 277.2 Hz; 349.2 Hz; 370 Hz; 493.9 Hz). Each stimulus had a duration of 500 ms and within each stimulus 100 ms intercurred. Each sequence of stimuli was separated by an interval of 1700 ms after which appeared a fixation cross on the screen. As far as the tactile task is concerned, stimuli consisted of a vibration that was transmitted to one out of four possible fingers of one hand (thumb, index, middle, and ring fingers). The vibrations lasted 500 ms and were presented every 100 ms. Vibrations were produced and transmitted through an

[26] They used the same grammar of Knowlton and Squire (1996).

innovative device composed of a main wireless body, silicon finger sensors, and a control panel (see Pavlidou & Bogaerts, 2019, p.6, Box 1). Surprisingly, results indicated that above-chance performance occurred in the visual and tactile but not in the auditory task. This result is in contrast with Conway and Christiansen's results, which showed better learning performances in the auditory domain than in the tactile and visual domains. (Conway and Christiansen, 2005). The second important result is that they did not find correlations between learning performances in the three different sensory domains. They thus concluded by suggesting some degree of modality specificity in the learning process. Regarding the relationship between ISL abilities and reading skills, the authors found a statistically significant correlation only between phonological awareness and ISL in the visual domain. Other slight correlations that although did not reach any significance were found between ISL in the visual domain and fluent reading and basic reading skills as well as between ISL in the auditory domain and basic reading skills, reading fluency, and phonological awareness. Importantly, no performance correlations across modalities were found. This is another important piece of evidence that made the authors be more in favour of the existence of modality constraints.

Summing up, few studies have been conducted so far in the realm of tactile implicit statistical learning. Yet, it is interesting to note that the few studies conducted in the field have provided evidence for the fact that both children and adults can learn sequential statistical information tactilely. However, sequential statistical learning performances in the tactile domains have revealed themselves as being worse as compared to those in the visual and auditory domain in adults (Conway and Christiansen, 2005; Abrahamse et al., 2008; 2009), whether in children tactile sequential statistical learning has turned out to be less effective than visual spatial statistical learning but more powerful than auditory sequential statistical learning (Pavlidou & Bogaerts, 2019).

3.1.5. Is the auditory domain superior in sequential implicit statistical learning? Evaluating the Auditory Scaffolding hypothesis

In the previous sections, we have seen that various studies have focused on investigating implicit statistical learning abilities in different sensory domains. What emerges from the overall picture of these studies is that (i) this ability is present in the auditory, visual, and tactile domains; (ii) there are domain-specific differences. Specifically, regarding statistical learning in the temporal (i.e., sequential) dimension, several studies have found that the auditory domain excels, showing an advantage over the visual domain (Saffran, 2002; Conway, Christiansen, 2005; 2009). Crucially, however, when we examine studies that have compared this capacity in the tactile domain, we find two contrasting pieces of evidence. On one hand, Conway and Christiansen (2005) have found that hearing has an advantage over the tactile domain in acquiring sequential statistical information. On the contrary, Pavlidou & Bogaerts (2019) have found the opposite, namely that the tactile domain has an advantage over the auditory one.

Based on the evidence that adult subjects perform most effectively sequential statistical learning through the auditory sensory domain, rather than with sight, and taking into consideration the fact that the nature of sound is fundamentally temporal and sequential, Conway et al. (2009) formulated the *Auditory Scaffolding Hypothesis*. According to this hypothesis, exposure to sound plays a crucial role in the development of cognitive abilities related to temporal and sequential patterns. The hypothesis suggests indeed that sound serves as a cognitive support or "scaffolding" for the development of general abilities related to recalling, producing, and learning sequential information. It follows that the absence of sound exposition during early development may disrupt the formation of sequencing skills. In other words, deafness may negatively impact the developments of cognitive functions related to sequential information. “Although it is common to consider deafness as affecting the sense of hearing alone, we argue that because sound is the primary gateway to understanding temporal and sequential events, auditory deprivation may result in significant disturbances on a wide range of other tasks.” (Conway et al., 2009, p. 276). To test this hypothesis, Conway et al. (2009) examined sequencing skills in two distinct groups of children. One group was composed of deaf children with cochlear implants (CIs); the other one consisted in an age-matched hearing group. They evaluated children's motor sequencing abilities by means of a fingertip tapping task. In one variation of the task, children were instructed to rapidly tap their thumb and index finger together. In another version, they were asked to swiftly tap the tip of their thumb, after that the index, middle, ring, and pinky finger, following this specified order. Results indicated that deaf children with CIs performed less effectively than the control group. Notably, as they highlighted, the deaf children did not exhibit impairments in various non-sequencing tasks, such as tactile perception and visual-spatial memory. In addition to the fingertip tapping task, they assessed children's visual sequential learning abilities though an AGL task, in which sequences of colored squares, generated by an artificial grammar, were sequentially displayed on a touch-sensitive screen. The task required the children to remember and reproduce the sequence of colors, by tapping in the right order the panels in which squared appeared. After this initial phase, the test phase began, and children were exposed to some sequences generated by the same artificial grammar and other sequences generated by a different grammar. The task for participants was the same as that in the initial phase. As the authors explain, since each color corresponded uniquely to a specific position on the screen, a child might have recalled a sequence of locations, a sequence of colors, or both. The results indicated that normal-hearing children demonstrated a significantly higher sequence learning score compared to the deaf children, with the latter group showing limited improvement. Moreover, a smaller percentage of deaf children exhibited the effects of implicit sequence learning compared to their normal-hearing counterparts. Overall, both the fingertip tapping task and the sequential learning AGL tasks revealed that deaf children exhibit atypical motor and visual sequence learning compared to age-matched normal-hearing children. Based on these results, Conway et al. (2009) suggested that early deafness led to secondary disruptions in non-auditory sequencing skills, thus corroborating the auditory scaffolding hypothesis. The authors concluded by asserting that, while it is evident that the absence of sound impedes the acquisition of spoken language, auditory deprivation also hinders the proper development of non-auditory sequencing cognitive functions. Two possible mechanisms are suggested to account for this phenomenon. One potential explanation is that exposure to sounds would offer an opportunity for automatic imitation (i.e., vocal rehearsal) of the auditory input, whether vocally or covertly. “Imitating what is heard gives a discrete verbal label to a continuous auditory signal, providing anchor points for learning associations among the discrete symbols (i.e., words). Under this “embodied” account, hearing thus recruits vocal rehearsal processes that presumably strengthen the development of domain-general implicit sequence learning abilities.” (Conway et al., 2009, p. 278). Alternatively, another possible mechanism would rely on the notion that all environmental input inherently contains "modality-neutral" information alongside the modality-specific signal itself. Unlike vision, sound may carry specific higher-level patterns of information associated with serial order and temporal change. “Under this view, hearing is the primary gateway for perceiving high-level sequential patterns of input that change over time (rather than over space). The development of fundamental sequence learning mechanisms would thus be delayed when this type of input is unavailable, as is the case in deafness.” (Conway et al., 2009, p. 278). Summing up, Conway et al. (2009) claim that sound is crucial for developing cognitive processes related to temporal and sequential behavior. According to them, exposure to sound aids in encoding and manipulating sequential information, while a lack of auditory stimulation hinders these skills. The study highlights the broader impact of sound on cognition beyond auditory perception, with implications for neurocognitive development across various populations.

Crucially, as pointed out by Giustolisi et al. (2022), several studies have found evidence contrasting with the findings in Conway et al. (2009). (cf. Giustolisi & Emmorey, 2018; Hall et al., 2018; Terhune-Cotter et al., 2021; von Koss Torkildsen et al., 2018).

Hall et al. (2018) contests two aspects of Conway et al.'s (2009) study. Firstly, as they rightly point out, in the population of deaf children tested by Conway and colleagues, the period of auditory deprivation largely overlaps with that of language deprivation. Indeed, they tested deaf children born to hearing parents who were exposed to sounds through cochlear implants. So, the effects found in Conway et al. (2009) could be attributable to the lack of exposure to sound, while also to the lack of exposure to language. These two things should be disentangled. Secondly, as rightly pointed out by the authors, the task they used may not be the most suitable for investigating sequential statistical learning abilities. Indeed, in this task, children with a high working memory span are expected to exhibit no learning effects. This is because their performance would already be at its maximum for both familiar and unfamiliar sequences. In other words, they can accurately remember both familiar and unfamiliar sequences. Therefore, any detectable learning effects would likely be observed only in children who struggle to remember the unfamiliar sequences presented to them. Hall et al. (2018) expanded upon Conway and colleagues' study by conducting two experimental investigations. They not only tested the two groups of children examined in Conway et al. (2009) (i.e., hearing children and deaf children born to hearing parents) but also included a third group: Deaf children with no delay in language exposure (i.e., children born to deaf parents and exposed to sign language from birth). In the first experiment, they replicated Conway's AGL task and found no results in any of the three groups of children. In contrast, in the second experiment, when they tested children with a classic serial reaction time task, they found evidence of learning in all three populations.

Results consistent with those of Hall et al. (2018) and contrasting with the auditory scaffolding hypothesis are provided by Terhune-Cotter et al. (2021) and von Koss Torkildsen et al. (2018). Both studies explored learning abilities in deaf children compared to hearing controls, demonstrating comparable performance between the two groups in implicit statistical learning tasks.

Compelling evidence that contradicts the auditory scaffolding hypothesis emerges also from the study by Giustolisi et al. (2022), who carried out an experimental study to verify whether sequential rule learning is hindered in children with congenital deafness. Their study aimed to assess whether a lack of hearing experience impedes learning sequential patterns, as suggested by the auditory scaffolding hypothesis. However, unlike previously conducted studies on the topic, they tested this ability at a more complex and abstract level. Indeed, instead of investigating the ability to acquire finite-state grammar statistical information, they also investigated the acquisition of sequential nested and crossed dependencies. The research involved 15 deaf adult participants (Italian Sign Language signers) and 15 hearing adults. They took part in a visual artificial grammar learning task, comprising sequences of stimuli generated by grammars of increasing computational complexity (from finite-state to mildly context-sensitive grammars). Specifically, the grammar tested were the following: the regular grammar (AB^{n}A); a context sensitive grammar (Mirror grammar); a mildly context-sensitive grammar (Copy grammar). The sequences of symbols generated by the grammars were encoded onto colorful decorated squares that were sequentially presented. Every square remained visible on the screen till the complete sequence was presented. Then, the whole sequence disappeared, and participants were asked to give an answer before the appearing of a new sequence. Each participant was tested on the three target grammars in a randomly determined order. The procedure for each grammar comprised two distinct phases: exposure and testing phase. Throughout the exposure phase, participants were presented with 30 grammatical sequences with N values of 2, 3, and 5. In the testing phase, participants encountered a total of 87 strings, comprising 36 grammatical ones (including N = 2 and N = 3, as well as extensions to N = 4 and N = 6) and 51 ungrammatical ones (N = 2, 3, 4, 6). Hence, test stimuli included sequences of the same length as those in the exposure phase and sequences of different lengths, enabling testing for rule generalization. Ungrammatical strings encompassed sequences featuring a missing element or incorrect category membership. Participants were tested on their ability to accept grammatical strings and reject ungrammatical foils in the testing phase (they were tasked with determining whether the sequence aligned with the same schema observed in the exposure phase. They conveyed their judgment by pressing a yes/no key on a keyboard).

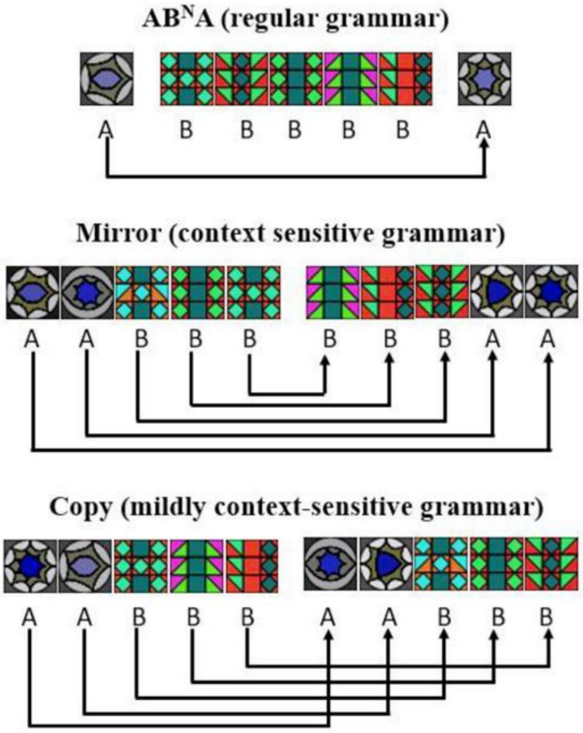


Figure 14. Instances of grammatical sequences for the three grammars with N=5. Picture taken from Giustolisi et al., 2022, p.8.

Results showed that both the deaf and hearing groups demonstrated proficiency in learning each of the three grammars, correctly accepting novel grammatical sequences and rejecting ungrammatical foils. The authors also focused on participants' ability to generalize rules to stimuli of novel lengths. As explained above, during the exposure phase, participants were exposed to sequences of N = 2, 3, or 5. In the testing phase, sequences of N = 4 and sequences of N = 6 were introduced. Both groups exhibited the ability to generalize to N4 sequences across all three grammars. However, while both deaf and hearing participants demonstrated the ability to generalize to strings of N = 6 in the regular AB^nA grammar, only the hearing participants displayed N = 6 generalization in the two supra-regular grammars. Moreover, comparing the hearing and the deaf groups, the former surpassed the latter in certain aspects: Better performances were found concerning the rejection of ungrammatical strings in the regular AB^nA grammar and the supra-regular Copy grammar. As the author explained, however, it is important to note that both groups demonstrated mastery of the three grammars. Deaf participants consistently performed equally well in generalization relative to recognition of grammatical sequences. Hence, the marginal decrease in performance observed among deaf participants, as opposed to their hearing counterparts, was not ascribed to differences in rule extraction abilities between the two groups. As a matter of fact, the authors conducted further Bayesian analyses to shed more light on the results. Importantly, these analyses confirmed that deaf participants' results were caused by actual learning of the specific target grammar. In other words, participants did not adopt alternative strategies. As the authors proposed, the difficulty in generalizing to N = 6 might indicate that these sequences were too long for the deaf population to be tracked. Moreover, Giustolisi et al. (2022) proposed that difficulties could have stemmed from potential interference during the encoding of stimuli. Specifically, verbal rehearsal strategies may have impacted sequence learning task performance. "[…] verbal rehearsal strategies may have a relevant impact on sequence learning tasks performance. Sequence tracking may be more difficult for the deaf population due to visual stimulus interference with their verbal coding strategies. Hearing participants may have implemented some form of verbal (vocal) recoding to track the incoming sequence […] Deaf participants attempting to implement such verbal encoding would suffer from interference, since verbal recoding of the experimental stimuli would need to use the same visual channel as their signed language." (Giustolisi et al. 2022, p. 18). The authors concluded by suggesting that future research could explore this hypothesis, potentially comparing hearing and deaf participants using nonvisual stimuli, such as tactile stimuli. Overall, Giustolisi and colleagues' findings provided clear evidence against the auditory scaffolding hypothesis, especially considering that the deaf participants were all born deaf, with the majority never using a cochlear implant (Giustolisi et al., 2022).

In this section, we have presented and discussed Conway et al.'s (2009) *Auditory Scaffolding Hypothesis*. This theory underscores the pivotal role of exposure to sound in shaping cognitive abilities related to temporal and sequential patterns. Specifically, it proposes that sound serves as a cognitive support or "scaffolding", fostering the development of general capacities involved in recalling, producing, and learning sequential information. Conway et al. (2009) explain that the auditory scaffolding hypothesis finds support through two sets of evidence: (i) individuals congenitally deaf display non-auditory sequencing abilities; (ii) there are modality-specific constraints observed in hearing populations. Specifically, in support of (i), explain “[…] recent findings suggest that deaf children have disturbances on exactly these same kinds of tasks that involve learning and manipulation of serial-order information.” (Conway et al. 2009, p. 275). In support of (ii), on the other hand, they elaborate “[…] normal hearing adults do best on sequencing tasks when the sense of hearing, rather than sight, can be used.” (Conway et al. 2009, p.275).

Crucially, however, as we have seen, many studies have refuted (i), providing evidence that deaf populations succeed in learning domain-general sequential information. This contrasts with what was hypothesized by Conway and colleagues. (cf. Giustolisi et al. 2022; Hall et al., 2018; von Koss Torkildsen et al., 2018; Terhune-Cotter et al., 2021). Regarding (ii), which concerns modality constraints in hearing populations, we find confirmation that hearing has an advantage in processing sequential stimuli compared to vision, as evidenced by many studies (cf. Conway et al., 2009; Saffran, 2002). However, this is not the complete picture. Indeed, by introducing a third variable into the equation, we observe a change in perspective. Specifically, when comparing auditory and tactile domains in processing sequential statistical information, we find conflicting evidence (cf. Conway, Christiansen, 2005, for auditory superiority; Pavlidou, Bogaerts, 2019, for tactile superiority). In other words, while it is confirmed that the auditory domain has an advantage over the visual domain in processing sequential statistical information, the auditory superiority is not confirmed in the comparison with the tactile domain. In other words, it is still unclear whether hearing has superiority over touch in processing sequential statistical information. We believe that further studies should investigate this issue by comparing, through paradigms as similar and comparable as possible, this ability to acquire sequential statistical information in the tactile and auditory domains. Indeed, this will be one of the objectives of the present investigation, which will explore domain-specific constraints in the processing of sequential structures across different sensory domains (Chapter 5).

In the preceding sections, we delved into sequential implicit statistical learning across various sensory domains, elucidating both domain-general and domain-specific mechanisms at play. Our exploration encompassed studies in visual, auditory, and tactile domains, prompting us to question whether any specific domain holds an advantage in processing sequential statistical information. The auditory scaffolding hypothesis (Conway et al., 2009) was scrutinized, leading to the emergence of data problematic for the theory. Numerous studies demonstrated that exposure to sound might not be as pivotal as previously thought for the development of cognitive abilities tied to temporal and sequential patterns, particularly evident in deaf populations proficiently processing domain-general sequential statistical information (cf. Giustolisi et al, 2022). Despite this, uncertainties persist regarding whether the auditory domain excels above all others in processing sequential statistical information. Our conclusion rests on the confirmed advantage of the auditory domain over the visual domain in sequential processing (cf. Conway, Christiansen, 2009; Saffran, 2002), with the comparison to the tactile domain still to be verified.

As explained in Chapter 2 of this thesis, our goal extends beyond investigating processes and mechanisms related to the processing and acquisition of low-level transitional regularities. We also aim to explore more abstract representations in the processing of sequential structures (cf. Dehaene et al., 2015). Specifically, we want to shed light on the mechanisms underpinning the cognitive ability to deal with recursive hierarchical structures arising from sequential input—a mechanism at play in human language but also in music (cf. Section *2.1.2.*; *2.2.1.*). The overarching goal of this thesis is to illuminate the mechanisms involved in the formation of recursive hierarchical abstract structures arising from sequential, temporally ordered, fading stimuli.

Turning our attention to the next section, we will specifically shift our focus to recursion. Specifically, we are interested in investigating the representational features of recursion, that is, the capacity to represent and apply self-similarity across hierarchical levels (cf. *Section 2.3.6.*). Having thoroughly explained the mechanisms of recursion in Chapter 2, our objective is now to determine whether recursion is a domain-specific or domain-general ability. Do we possess the capacity to process and form recursive structures in different sensory domains? If so, are there differences across these domains? While the literature on this topic includes AGL studies conducted in the visual and auditory sensory domains, it is noteworthy that no study has explored recursion in the tactile domain. These studies will be reviewed in the upcoming section.

3.1.6. Visual and auditory recursive hierarchical learning

In this section, we review some interesting experimental studies which investigated recursion in the visual domain (Martins, 2012; Martins et al., 2014; 2015), and in the auditory domain (Martins et al., 2017).

Martins (2012) devised a novel assessment task named the visual recursion task (VRT) to gauge individuals' proficiency in conceptualizing visuo-spatial hierarchies as recursive structures and applying these conceptualizations to generate subsequent levels of embedding. The VRT method draws inspiration from geometrical self-similar fractals, generated through recursive embedding rules over multiple iterations. In other words, they created self-similar visual patterns by iteratively applying the same rules across multiple hierarchical levels. In the task, participants were exposed to the initial three iterations of a fractal structure and were subsequently required to identify the correct fourth iteration from two options (Figure 15). As the authors explain, successful performance necessitated acquiring categorical knowledge about constituent elements, recognizing hierarchical structures, detecting similarities in the disposition between elements across levels, and applying abstract rules to extend one level beyond the given. To differentiate between recursion and embedded iteration, a non-recursive control task was introduced. This control task involved iterative processes that embedded constituents within fixed hierarchical levels without generating new levels. The task procedure was the same as that in the VRT task (Figure 16).

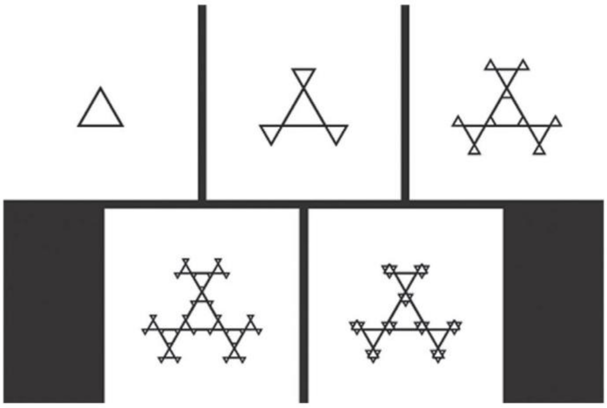


Figure 15. Instances of trial from the visual recursion task[27]. (Figure taken from Martins, 2012, p. 2061)

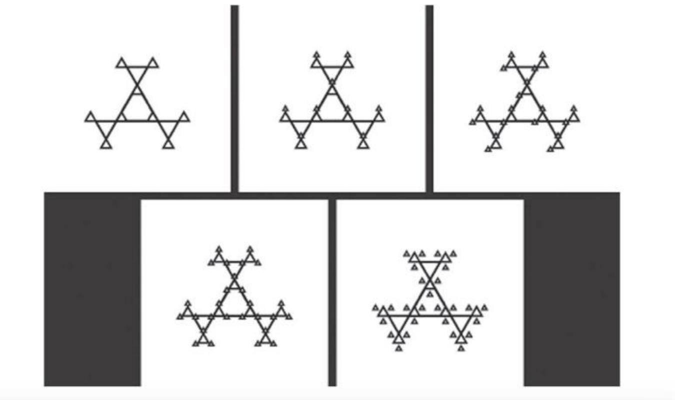


Figure 16. Trial from the visual hierarchical task: The task procedure mirrors that of the visual recursion task. (Figure taken from Martins, 2012, p. 2061).

Two groups of participants took part in the tasks. Fluid intelligence and working memory were also measured. The findings indicated that visual recursion had lower accuracy and longer response times compared to embedding iteration. Fluid intelligence emerged as the most reliable predictor for both tasks, but the predictive power of verbal working memory was higher for visual recursion, while spatial

[27] The top row displays the initial three stages of a fractal creation process. Subsequently, the participant is tasked with identifying, from the images presented in the bottom row, the one that accurately represents the fourth iteration.

working memory was more influential for embedded iteration. Based on this result, the authors concludes that it remains to be verified whether verbal processing resources are essential for recursive representations in the visual domain or not. "[…] the next empirical question is whether verbal processing resources are a necessary condition for recursive representations in the visual domain or whether they are recruited when available, given that they enhance reasoning in non-linguistic domains" (Martins, 2012, p. 2061).

Martins et al. (2014) delved into the exploration of the human ability to discern well-formed visuospatial hierarchical structures. Their research centered on implementing rules that either carried out transformations within a hierarchical level or produced additional self-similar hierarchical levels. Two tasks, the Visual Recursion Task (VRT) and the Embedded Iteration Task (EIT), were employed to scrutinize the cognitive processes associated with the representation of visuospatial hierarchies. The tasks were adapted from those used in Martins (2012). Both tasks involved exposing participants to a set of figures constituting a generative process, followed by a forced-choice phase concerning subsequent further iterations (cf. Martins, 2012), necessitating the extraction of simple rules from initial iterations for predicting subsequent transformations. In VRT, each iterative step generated a new hierarchical level following a spatial rule analogous to previous levels. Conversely, in EIT, new elements were iteratively embedded within an existing hierarchical level without generating new levels. As a control measure, the researchers introduced a 'similarity task' (Positional Similarity Visual Task — PSVT), where participants matched a target visuospatial hierarchy with two alternatives. The correct alternative matched one of the three previously presented images. During four sessions inside a 3 Tesla MRI scanner, participants engaged in VRT, EIT, and PSVT stimuli, with an event-related design for randomizing stimuli across sessions. In the VRT, participants were informed that new elements would have been added at each step to create new hierarchical levels, following a spatial rule constant across levels. Conversely, in the EIT, they were instructed that new elements would have been added to an existing hierarchical level according to a predictable spatial rule. Martins et al. (2014) hypothesized that the brain utilizes distinct resources when processing hierarchies, depending on whether it employs a "fractal" (recursive, generating new levels) or a "non-fractal" (hierarchical within level) cognitive strategy. Behavioral responses (reaction times and accuracy rates) and neural circuits activated by the tasks were analyzed and compared. Results indicated lower accuracy rates in EIT compared to VRT and PSVT, with faster responses observed in VRT. Brain imaging results unveiled key findings: Both the within and between levels rule processes activated a bilateral network (the dorsal stream) involving visual association areas, fronto-parietal circuits related to spatial reasoning, and regions like the inferior frontal gyrus (IFG). This supported the notion that Broca's area might be generally involved in maintaining online information or rules supporting iterative processes, rather than being specifically involved in recursive tasks. Recursive processes generating new hierarchical levels activated brain areas usually involved in the integration of categorical and spatial information. Specifically, it activated regions within the parieto-medial temporal pathway (PMT), including the posterior cingulate cortex (PCC) and retrosplenial cortex (RSC), along with projections to the medial temporal cortex (MTL). These regions are known for their roles in the formation of cohesive representations, integrating spatial and semantic information and are also associated with episodic memory. The importance of the MTL in processing spatial, linguistic, and social hierarchies has been underscored in prior research. Furthermore, activations were identified in the anterior portions of the superior and middle temporal gyri (STG and MTG, respectively), usually associated with the retrieval of abstract categories. Taken together, these findings emphasized the critical role of episodic memory and the integration of both spatial and categorical information. As the authors explained, the intriguing aspect lies in the fact that the visuo-spatial hierarchies used in the study did not inherently convey "semantic" information. Regarding this, the authors put forth the hypothesis that representing hierarchical dependencies may require the retrieval of "semantic" information of a more abstract nature. Within-level iterative rules showed more specific activation of brain areas involved in spatial domains, involving the dorsal stream, dorsal fronto-parietal network (FPN), IFG, and basal ganglia. Interestingly, Broca's area appeared more active in within-level computations than recursive ones. The results suggested that Broca's area does not exhibit specific activation in processing cross-level hierarchical integration.

Instead, it seems to play a broader role in the storage and maintenance of rule-based iterative information, possibly involving working memory processes. Additionally, these findings proposed that recursive embedding serves as a more memory-efficient approach for generating complex hierarchies. In summary, Martins et al. (2014) proposed that the brain employs distinct resources for processing hierarchical structures, depending on whether a "fractal" (generating new levels) or a "non-fractal" (hierarchical within level) cognitive strategy is applied. Recursive mechanisms activated brain areas associated with the integration of abstract semantic and spatial information, while within-level iterative rules correlated more strongly with working memory abilities. The study concluded by suggesting future research across different domains to verify if domain-specific, localized computational processes are needed for the creation of hierarchical structures.

Martins et al. (2015) aimed to investigate a widely held hypothesis suggesting that the capacity to form and utilize recursive representations in processing hierarchical structures is contingent upon language abilities. If this holds true, linguistic resources should inevitably come into play when representing recursion in non-linguistic domains. Hence, the primary objective of Martins et al. (2015) was to directly explore whether verbal resources are essential for acquiring and applying recursive rules in the visual domain. As the authors explained, some scholars posited a close association between the evolution of language and the emergence of recursion. A notable hypothesis asserts that recursion constitutes a domain-specific linguistic computational system, independent from other interacting systems (Hauser et al., 2002). According to this view, while the use of recursive rules might be present in non-linguistic domains, such applications could hinge on a previously evolved system, relying on language faculties. Conversely, Pinker and Jackendoff (2005) proposed that recursion's utilization in certain domains, such as visual perception, can occur autonomously of language. Overall, regarding the relationship between human language and recursion, there are three logically plausible scenarios (Martins et al., 2015):

- Hypothesis 1: The capacity for creating recursive representations is specific to language and is executed by a dedicated linguistic module for recursion.

Representation of recursion in the other domains relies on language and thus utilizes linguistic resources.

- Hypothesis 2: The capacity for generating recursive representations is not language specific but domain general. There exists a unified cognitive system responsible for recursion that can be engaged by multiple domains, without language holding a primary role.
- Hypothesis 3: The ability to construct recursive representations is specific to multiple domains but extends beyond language. In other words, each cognitive domain can access its own dedicated system for implementing recursive representations, independent of other domains.

As Martins et al. (2015) pointed out, so far, most studies that have investigated recursion have done so in the linguistic domain. It is noteworthy, however, that some studies have started exploring this ability beyond the linguistic domain, such as in vision, as demonstrated in the studies conducted by Martins and colleagues mentioned earlier. As we have seen, Martins and colleagues discovered that, in contrast to non-recursive iterative processes, visual recursive abilities showed only a weak correlation with specifically visual resources, such as non-verbal intelligence, spatial short-term memory, and spatial working memory. However, they exhibited a strong correlation with recursive planning tasks (Martins et al., 2014) and verbal working memory processing component (Martins, 2012). Nevertheless, Martins et al. (2015) clarified that this latter finding does not necessarily imply that visuo-spatial recursion relies on resources specific to verbal processing. Instead, this correlation may be influenced by a third variable shared by both domains, such as cognitive resources involving the central executive. Interestingly, moreover, Martins et al. (2014) showed that visual recursion does not selectively activate perisylvian language areas when compared to a simple iterative task. However, as Martins et al. (2015) explained, these findings were correlational and thus require confirmation through methods that manipulate the capacity to utilize linguistic resources in order to have a more accurate view. To shed more light on the issue, in Martins et al. (2015) participants were tasked with completing a Visual Recursion Task (VRT) amidst verbal interference. If verbal rehearsal of digits negatively impacts the processing of recursive hierarchies in the visual domain, it would provide evidence supporting the hypothesis that language has a major role in the representation of recursion in non-linguistic contexts. On the other hand, if the ability to represent visual recursion remains unaltered when linguistic resources are restricted, it would bolster the notion that the visual domain can directly tap into the cognitive system of recursion, regardless of language. In their investigation, the authors employed a dual-task paradigm to examine whether the utilization of verbal resources is a prerequisite for representing recursion in the visual domain. The methodology comprised executing a primary task (specifically, a visual recursion task) either independently or concurrently with a secondary interference task. If the performance in the primary visual recursion task diminishes when a secondary verbal interference task is introduced, it implies that verbal resources are essential for solving visual recursion. However, as they explained, it is essential to consider that a decline in visual recursion task performance in concomitance of verbal interference could also be attributed to general attention constraints. To address this possibility, they incorporated a nonverbal motor interference task in their experimental study. Participants underwent four experimental sessions, with each session consisting of 12 trials: (i) Visual recursion task (VRT) in the absence of a secondary task; (ii) VRT with a motor task interference; (iii) VRT with low-load verbal task interference; (iv) VRT with high-load verbal task interference. In the VRT task, the stimuli and methodology used were the same as those described in Martins (2012). Hence, it regarded visual fractals generation. As we have mentioned, participants engaged in the Visual Recursion Task (VRT) either in isolation or alongside one of three interference tasks: motor interference, low-load verbal interference, and high-load verbal interference. During the motor task, participants viewed a sequence of six simultaneously presented pictures representing finger-tapping movements. Participants were instructed to repeatedly perform the sequence tapping their own fingers and then to press a button when ready to transition to the VRT task. Throughout the VRT trial, participants were asked to continuously replicate the sequence using only their right hand, without utilizing other cognitive (e.g., verbal) or physical resources aside from their fingers. Following their response to the VRT trial, participants were then prompted to type the motor sequence they performed.

The verbal interference task utilized the digit span methodology (i.e., it tapped into verbal working memory). Participants were visually exposed to a series of digits and were required to verbally repeat the sequence while simultaneously undergoing a VRT trial. Following each trial, participants were prompted to type the sequence on the keyboard. Under the low-load verbal task condition, participants were tasked with memorizing a randomly generated sequence up to 6 digits, aligning with the content load in the motor task. In the high-load verbal task condition, participants faced the challenge of memorizing a sequence up to 7 digits. The findings indicated that participants demonstrated the ability to acquire principles involved in the recursive generation of visuo-spatial hierarchies and apply this structural knowledge to various recursive instances. Notably, performance exhibited improvement with practice and did so even in the absence of response feedback, strongly supporting the presence of rule induction or a generalization mechanism. Secondly, the study revealed comparable high-performance levels in the Visual Recursion Task (VRT) both without interference and when coupled with secondary verbal or motor tasks. In other words, a lack of interference from either motor or verbal secondary tasks on the visual recursion task was observed. As the authors explained, this strongly implies that the capacity to comprehend and apply principles governing the creation of recursive self-similar visual hierarchies in the spatial dimension remains unaffected by secondary motor or verbal tasks. Intriguingly, the correct rehearsal of concomitant verbal or motor material appeared to enhance, rather than diminish, performance in the visual recursion task. The authors postulated that the presence of a secondary task might compel participants to consciously focus on the primary task, potentially setting the stage for them to adopt a more rigorous and analytical cognitive approach. Overall, the authors concluded by suggesting that these findings cast doubt on the viability of Hypothesis 1. Instead, their results align more with the assertions of Hypotheses 2 or 3, both of which posit that recursion can be conceptualized autonomously from language. The inquiry into whether recursion represents a unified, domain-general cognitive system (Hypothesis 2) or functions as a combination of multiple distinct, domain-specific modules (Hypothesis 3) emerges as an intriguing avenue for future research.

Martins et al. (2017) aimed to deepen our understanding of human cognitive recursion in non-linguistic domains: the auditory domain. To do so, they carried out two experimental studies, investigating the representation of musical fractals. The investigation had a twofold objective: firstly, to determine if adults, including both musicians and non-musicians, can portray hierarchical relationships in the auditory domain. This involved assessing and comparing their capacity to induce and implement iterative rules within the same level and recursive rules across different levels in structured tonal sequences (Experiment 1). Secondly, the study aimed to explore whether the ability to represent recursion in the auditory domain aligns with similar tasks in the action and visual domains. This investigation aimed to discern whether constructing recursive representations relies on domain-specific resources or a domain-general cognitive framework (Experiment 2). In Experiment 1, they explored whether humans possess the ability to represent recursion in the auditory domain. Similarly to prior research in vision (cf. Martins, 2012; Martins et al., 2014; 2015), they implemented recursive rules on sequences of tones and assessed participants' capacity to discern these rules. In a two-alternative forced-choice paradigm, participants experienced three steps of a recursive process producing auditory fractals (pure tone sequences). Subsequently, they were tasked with distinguishing between a correctly generated fourth step in the same process and a foil (as in similar previous tasks in the visual domain). In addition to this task, referred to as the Auditory Recursion Task (ART), a control task, the Auditory Iteration Task (AIT), was devised. The methodology and stimuli employed in the AIT mirrored those utilized in the ART. In both tasks, participants were instructed to attentively listen to the initial three iterations and envision the sound of the fourth iteration. Subsequently, they were required to identify the correct fourth iteration from two options. The critical divergence between the two tasks laid in the procedure governing iteration generation. Indeed, while the AIT shared hierarchical, sequential, and iterative aspects with ART, it did not involve recursive procedures in generating hierarchical structures. In fact, in the ART, each iteration incorporated novel tonal elements recursively within distinct hierarchical levels, featuring varying tone durations at each step. Conversely, in the AIT, elements were integrated within a consistent, single hierarchical level (maintaining the same tone duration), and no additional levels were added to the structure. Notably, participants were not explicitly educated on the concepts of recursion or iteration. Instead, they had to implicitly discern these regularities while exposed to examples of stimuli. Both musicians and non-musicians were tested. Specifically, thirty non-musicians participated in the Auditory Recursion Task, and a distinct group of 24 non-musicians engaged in the Auditory Iteration Task. Additionally, 20 musicians undertook the Auditory Recursion Task. Stimuli in the Auditory Recursion Task were constructed as an auditory equivalent to visual fractals, inspired by Mandelbrot (1977). As Martin et al. (2017) explain, in the visual domain, hierarchical levels are represented by constituent size, with larger constituents dominant and smaller subordinate ones. The transformation rule (generator) captures the spatial arrangement of subordinate elements relative to the dominant. Martins and colleagues' (2017) auditory fractals were crafted using auditory features akin to these parameters: note duration and pitch indicated hierarchical level, with longer and lower-pitched notes signifying dominance over higher and shorter ones. Tone space was the parameter modulated by the generator. Specifically, for each dominant tone in one iteration, they introduced three new subordinate tones, shorter in duration and higher in pitch. These subordinate-note contours followed a particular pattern (ascending or descending) and were at a specific pitch distance from the dominant tone. This constituted the recursive rule operating over different hierarchical levels (Figure 17). The target stimulus was created through four iterations (Figure 18). The first iteration featured a low-pitch pure tone (i.e., the initiator). The second iteration retained the initiator tone and added three new tones based on a specific rule (i.e., the generator). This rule manipulated pitch contour (ascending or descending), pitch interval between successive tones (four or eight semitones), and pitch interval between consecutive levels (four or eight semitones). The same generator was applied across all hierarchical levels, ensuring constant pitch and rhythmical relations between dominant and subordinate elements, resulting in a hierarchical self-similar structure. They generated four successive iterations of 24 distinct types of auditory fractals. For each of them, they created (i) a well-formed fourth continuation of the first three iterative steps and (ii) an ill-formed continuation, that is, a 'foil' stimulus, achieved by applying a different generator to the third iterative step. The foils fell into three categories: (i) positional, (ii) odd, and (iii) repetition (Figure 19).

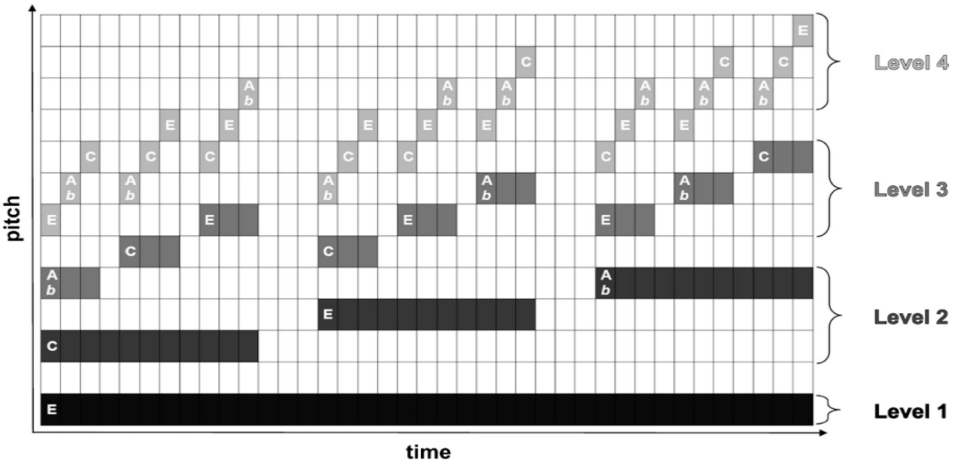


Figure 17. Illustrative instance of a tonal auditory fractal. (Figure taken from Martins et al., 2017, p. 35)[28].

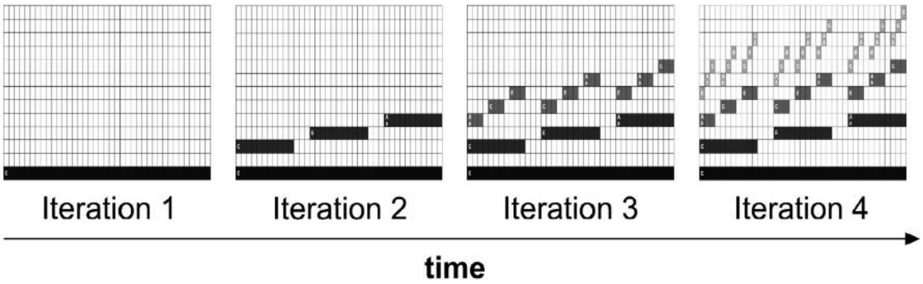
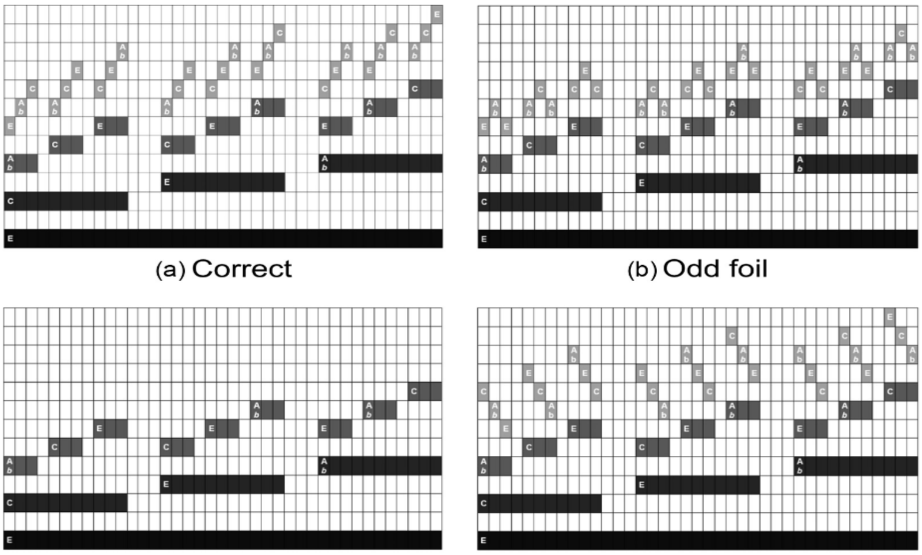


Figure 18. Recursive process creating an auditory fractal. (Figure taken from Martins et al., 2017, p. 35)[29].

[28] As explained by Martins et al. (2017), this auditory arrangement exhibits a hierarchy comprising four levels, distinguished by varying shades of gray. At the bottom is the dominant level (Level 1), featuring a protracted, low-pitch note lasting 7.3 seconds. The second level (Level 2) is crafted from three notes, each lasting slightly less than one-third of the duration of the dominant note (Level 1), interspersed with brief silent pauses. These three notes ascend sequentially, maintaining a fixed pitch interval between each pair. Level 3 is generated using the same principle, introducing sets of three notes at a specific pitch interval in relation to a dominant note (i.e., every note at Level 2).

[29] The process involves the addition of a new hierarchical level at each step in the process, illustrated by a lighter shade of gray in the figure. This new level consists of notes with shorter duration and higher pitch.



(a) Correct (b) Odd foil

(c) Repetition foil (d) Positional foil

Figure 19. Different categories of the fourth iteration (Figure taken from Martins et al., 2017, p. 36)[30].

The results demonstrated that participants, irrespective of their musical background, successfully grasped recursive rules governing the creation of auditory fractals and applied these rules productively. This happened in the absence of feedback or explicit instructions. They consistently rejected incorrect continuations of recursive processes across the three different types of foil categories. As the authors explained, this implies that participants did not rely on a single, simple auditory heuristic to solve the task. When comparing performance between the two tasks, overall accuracy was similar, with participants performing well in both AIT and ART. However, interesting differences emerged as well: (i) the accuracy learning

[30] The *repetition* foil (c) consists of a duplication of the third iteration. In both the *odd* (b) and *positional* (d) foils, a new hierarchical level is introduced, yet the contour of this level does not align with the pattern established in prior iterations. In the *odd* (b) foils, the final note in each set of three matches the pitch of the initial note in that set, disrupting the projected directional flow (whether ascending or descending). Conversely, in *positional* foils (d), the directionality remains consistent in every triplet, but it diverges from the directionality of other hierarchical levels.

curve was steeper in ART compared to AIT; (ii) while participants exhibited consistent rejection of all foil categories in ART, they struggled with rejecting *odd* foils in AIT. With respect to this, the authors explained that the principle inferred in ART enabled participants to equally reject all three foil types, whereas AIT performance showed less consistency across foil categories, hinting at a potentially larger role for heuristic strategies, though not exclusively, in dealing with this task. Considering musicians versus non-musicians, the authors found that, while non-musicians demonstrated performance above chance in ART, their performance was relatively modest (71%), especially when compared to their ability to perform a similar task in the visuo-spatial domain (>84%). Notably, the musicians' accuracy level (84%) closely resembled that in the visuo-spatial recursion task of the general population (Martins, 2012). Based on this finding, the authors suggested that expertise and practice effects may exert a more significant influence in the auditory domain than in the visual domain. In Experiment 2, the authors delved deeper into the nature of the rule induced in ART. In this pursuit, they explored the connection between accuracy in ART and that in other non-auditory recursive tasks: The Tower of Hanoi task (ToH) and the Visual Recursion Task (VRT). Establishing a strong correlation between ART, ToH and VRT would provide supporting evidence that ART engages with aspects specific to recursion. However, as the authors explained, it is important to take into consideration that, being ART an auditory task, the specific skill to perceive musical tone structure is expected to play a role as well. For this reason, to quantify the general effects of auditory and musical expertise, the authors incorporated in the experiment a control Auditory Iteration Task (AIT), a Visual Iteration Task (VIT), and a Melodic Memory Task (MMT). Additionally, they considered the number of years of musical training undergone by the participants. Hence, with Experiment 2, which included ART, AIT, ToH, VRT, VIT, MMT, and taking into account the number of years of musical training, the researchers examined whether, even after accounting for effects related to visual and auditory processing, ART exhibited a correlation with different recursive tasks. If such a correlation were established, it would substantiate the proposition that the ART task specifically engages the capacity to represent recursion in the auditory domain. Moreover, this finding would bear significance in addressing the broader question of whether recursion is a domain-general or domain-specific cognitive function. A total of 40 participants took part in this experiment. Differing from the approach in Experiment 1, the inclusion criteria for participants in this study encompassed individuals with diverse levels of musical expertise, ranging from those with no musical training to others with up to 16 years of musical training. Each participant took part in all tasks. Concerning the individual tasks, they utilized the exact same versions of ART and AIT as employed in Experiment 1. As for the Visual Recursion Task (VRT) and Visual Iteration Task (VIT), these tasks were modified versions detailed elsewhere (cf. Martins, 2012). The foil categories mirrored those used in ART and AIT (i.e., *odd*, *repetition*, and *positional*). As for the Tower of Hanoi (ToH), it consists in a visuo-motor task that involves the hierarchical movement of disks across pegs to complete puzzles, adhering to well-defined rules. Crucially, this task is optimally approached using a recursive strategy. Regarding the Melodic Memory Task (MMT), it aims to evaluate participants' memory for short melodies. In this task, participants listened to pairs of brief melodies (comprising 10 to 17 notes) and were tasked with determining whether the two melodies shared an identical pitch interval structure or not. Experiment 2 replicated the earlier findings from Experiment 1, confirming that humans possess the capacity to represent recursion in the auditory cognitive domain. Furthermore, the study reiterated the influential role of musical training as a significant predictor of performance in both the Visual Recursion Task (VRT) and the Auditory Recursion Task (ART), alongside the ability to discern changes in melodic contour (MMT). The second key discovery unveiled two critical aspects: (i) performances in ART demonstrated to be correlated with those in VRT and the Tower of Hanoi (ToH) task. This strongly indicate that while performance in ART is contingent on general capacities for processing auditory stimuli, it crucially aligns with other recursive abilities as well; (ii) in contrast, there were no discernible specific correlations between the Auditory Iteration Task (AIT) and Visual Iteration Task (VIT). On the opposite, AIT exhibited a robust connection with other auditory measures, suggesting that the processing of simple iteration in the auditory domain depends more on resources specific to that modality, without extending across diverse cognitive domains. Hence, with these two experimental studies, Martins et al. (2017) demonstrated that (i) human possess the ability to process and represent recursion in the non-linguistic auditory domain; and (ii) there is a strong correlation between this ability and the same ability in the action sequencing and visual domains. As the authors explained, it follows that, despite domain-specific constraints, the ability to construct recursive representations may be implemented through a more abstract mechanism. Cumulatively, Martins and colleagues' (2017) results strongly indicate that recursion is a domain-general proficiency.

Summing up, in this section, we have examined studies that investigated the representational capabilities of recursion – defined as the ability to depict self-similarity across hierarchical levels - in both visual and auditory sensory domains (Martins, 2012; Martins et al., 2014; 2015; 2017). It is crucial to note the absence of research exploring recursion in the tactile domain. Experimental studies consistently demonstrated the ability to represent recursion in the visual domain (Martins, 2012; Martins et al., 2014; 2015). Martins et al. (2014) found that the brain employs distinct mechanisms for processing visual hierarchical structures depending on whether a "fractal" (recursive, generating new levels) or a "non-fractal" (iterative, hierarchical within level) cognitive strategy is employed. Specifically, they observed that recursive mechanisms activate brain areas associated with abstract categorical and semantic integration, while within-level iterative rules are more strongly correlated with working memory abilities. Notably, within-level computations exhibited more pronounced activation in Broca's area, suggesting a broader role in storing and maintaining rule-based iterative information, possibly involving working memory processes. Furthermore, Martins and colleagues (2015) challenged the hypothesis that the capacity for creating recursive representations is specific to language and executed by a dedicated linguistic module. Their results suggested that recursion can be conceptualized autonomously from language. The auditory domain also exhibited the ability to depict recursive hierarchical structures, demonstrating that humans possess the capability to process and represent recursion in the non-linguistic auditory domain (Martins et al., 2017). Additionally, Martins and colleagues (2017) established a strong correlation between the ability to deal with recursion in the auditory domain and corresponding abilities in action sequencing and the visual domain. This implies that, despite domain-specific constraints, the construction of recursive representations may be accomplished by a more abstract mechanism. As a whole, Martins and colleagues' findings in the auditory and visual spheres strongly support the notion that recursion is a domain-general proficiency.

Importantly, however, we want to focus on a crucial difference that characterizes the paradigm that has investigated recursion in the auditory sphere compared to that in vision. The studies that have explored the ability to represent recursion in the visual domain seen in this section, utilized a paradigm in which fractal figures were presented. In these figures, the recursive hierarchical structure developed in space. The recursive hierarchical relationships between elements were thus perceived simultaneously by participants, that is, in the spatial dimension. Thus, we could say that these studies have explored the ability to process *static* recursive hierarchical structures in the visual domain. In contrast, studies that have investigated recursion in the auditory domain have explored the ability to process and represent recursive hierarchical structures that developed in the temporal dimension, not in space. Therefore, in this case, recursive hierarchical structures unfolded over time, during listening. The temporal dimension was thus crucial to understanding how elements connected and evolved in the context of a larger structure. Hence, in this case, the ability to process recursive hierarchical structures from sequential stimuli in the auditory modality has been investigated.

3.2. Conclusion

In this chapter, we reviewed studies which investigated sequential implicit statistical learning and the ability to form recursive hierarchical structures in different sensory modalities. Specifically, we have delved into the debate concerning domain-specific versus domain-general aspects of implicit statistical learning. Then, we focused on domain-specific spatiotemporal structure effects and qualitative differences across modalities. Notably, we noticed a lack of studies exploring domain-specific spatiotemporal constraints in the tactile modality. Regarding qualitative differences in sequential implicit statistical learning, we observed that studies predominantly focused on the visual and auditory domains, with recent explorations in the tactile domain confirming the capacity for acquiring sequential statistical information. Overall, these studies revealed that, in the temporal dimension, the auditory domain excels, showing an advantage over the visual domain. However, contrasting evidence emerged when comparing the auditory and tactile domains in processing sequential (i.e. temporal) statistical information (cf. Conway, Christiansen, 2005, for auditory superiority; Pavlidou, Bogaerts, 2019, for tactile superiority). We then reviewed the Auditory Scaffolding Hypothesis (Conway et al. 2009). This theory highlights the crucial impact of sound exposure on shaping cognitive abilities related to temporal and sequential patterns. It suggests that sound acts as a cognitive support or "scaffolding," facilitating the development of general skills in recalling, producing, and learning sequential information. As explained by the authors, the theory is supported by evidence from congenitally deaf individuals displaying non-auditory sequencing abilities and the observation of modality-specific constraints in hearing populations. Crucially, however, we found studies providing evidence that contradict the hypothesis, highlighting deaf populations' success in learning domain-general sequential information (cf. Giustolisi et al. 2022; Hall et al., 2018; von Koss Torkildsen et al., 2018; Terhune-Cotter et al., 2021). Regarding modality constraints in hearing populations, we have established that hearing exhibits an advantage in processing sequential stimuli compared to vision, as supported by consistent studies (cf. Conway et al., 2009; Saffran, 2002). However, introducing a third variable alters the perspective. Indeed, when examining the processing of sequential statistical information in the auditory and tactile domains, conflicting evidence arises (cf. Conway, Christiansen, 2005, for auditory superiority; Pavlidou, Bogaerts, 2019, for tactile superiority). In essence, while it is confirmed that the auditory domain surpasses the visual domain in processing sequential statistical information, this superiority is not consistently observed when comparing it with the tactile domain. Thus, it remains uncertain whether hearing holds an advantage over touch in processing sequential statistical information. Finally, we explored recursion in different sensory domains, finding consistent evidence confirming this ability in the visual and auditory domains. Martins et al.'s (2014) work demonstrated distinct brain mechanisms for visual recursion, with recursive processes activating areas associated with abstract categorical and semantic integration, while within-level iterative rules correlated more strongly with working memory abilities. Notably, Martins et al. (2015) challenged the belief that recursion is a language-specific ability. Furthermore, Martins et al. (2017) broadened the scope by extending recursion's understanding to the auditory domain, establishing a correlation between auditory recursion abilities and corresponding skills in action sequencing and the visual domain. These findings support the notion that recursion is a domain-general proficiency.

However, regarding studies that have investigated recursion, we have highlighted two important aspects to take into consideration. The first concerns a difference in the paradigms utilized to investigate visual and auditory recursion. In visual studies, participants were tested on the ability to process and represent *static* recursive structures *spatially* arranged in fractal figures. In contrast, auditory studies focused on the ability to process and represent dynamic, *sequential* recursive structures unfolding over *time* during listening. Despite the evidence finding a correlation between these two abilities (Martins et al. 2017), it is important to consider that, although both tasks are of a recursive nature, they may partially involve different cognitive skills. As we observed in the case of sequential implicit statistical learning, we cannot exclude the presence of domain-specific spatiotemporal constraints in the ability to process recursion. Secondly, we emphasized the notable absence of research on recursion in the tactile domain. This represents a significant gap in our current comprehension of how recursion operates across different sensory modalities. In essence, we observe a lack of studies that have devised paradigms capable of directly testing and comparing the ability to process and represent recursive hierarchical structures arising from sequentially presented input across the auditory, visual, and tactile sensory domains.

4. A New Methodology for the Investigation of Recursive Structures in Temporally Ordered Fading Sequences

Chapter 2 explored the challenges involved in experimentally investigating recursion from a cognitive perspective. Despite recursion being considered to play a role in several cognitive domains, among which language and music, studying this cognitive skill experimentally presents significant difficulties (cf. Section *2.3.5.1.*). Numerous attempts have been made, but many studies have encountered problems and failed to demonstrate this ability clearly and irrefutably. Finding the appropriate tools to investigate this cognitive ability is a significant challenge, far from trivial. The challenge is often attributable to a recurring issue—the lack of suitable tools for exploring recursion in non-linguistic domains, as observed by Martins (2012). In fact, multiple factors must be considered to design an experiment that can specifically test this ability, ensuring that the results can be attributed to this particular capacity and not to other cognitive mechanisms. Despite the scarcity of studies providing sufficiently clear and irrefutable empirical evidence regarding the ability to build abstract recursive representations, Chapters 2 and 3 have highlighted some recent studies that are interesting in this regard (Ferrigno et al., 2020; Martins et al., 2014; 2015; 2017; Planton et al., 2021; Schmid et al., 2023). Although these studies had slightly different research objectives and questions, they collectively demonstrated that humans are equipped with the cognitive ability to form hierarchical recursive representations outside the language domain (cf. Chapters 2 and 3).

However, considering that in both language and music, hierarchical recursive structures arise from sequentially ordered stimuli, the aim of this thesis is not to demonstrate general recursive ability, but to shed light on the ability to build recursive hierarchical abstract representations from temporally ordered sequences of stimuli. What are the cognitive mechanisms involved in the transition from the sequential to the hierarchical dimension? In Section *2.2.1.*, we observed that while existing research has predominantly focused on analyzing these cognitive abilities individually, only a limited number of studies have explored the comprehensive journey from sequence to hierarchy. Consequently, a significant gap exists in our understanding of how these processes interact and unfold across the entire cognitive continuum (Dehaene et al., 2015). Despite this, recent studies have yielded interesting results on the mechanisms involved in the transition from low-level, item-based computational strategies to the formation of increasingly more compact and structured abstract representations in processing sequentially arranged stimuli. These studies highlighted the different mechanisms underlying this process, also offering hypotheses about why individuals process sequences of items by forming incrementally higher levels of abstraction. (Planton et al., 2021; Radulescu et al., 2019; Schmid et al., 2023). As outlined in the previous chapters, the objective of this thesis is to further illuminate the cognitive mechanisms involved in processing sequential sequences. We aim to shed light on the cognitive mechanisms at work in processing linearly arranged stimuli at increasing degrees of abstraction. Specifically, our goal is to elucidate how cognition derives recursive hierarchical patterns from sequentially presented input and the different cognitive mechanisms involved in the process. This includes investigating how low-level implicit statistical learning relates to the formation of chunks, their categorical abstract representation, and the organization of these chunks into recursive hierarchical abstract representations.

In Chapter 3, we delved into the relationship between cognition and perception, examining the ability to learn sequential statistical information and represent recursion in different sensory domains. Concerning implicit sequential learning, we have seen that it remains unclear whether the auditory domain holds an advantage over the visual and tactile domains. Regarding recursion, studies have demonstrated the ability to process recursion in the auditory and visual domains, with a correlation observed across different sensory spheres, suggesting a domain-general characteristic of recursion. However, no study has investigated this ability in the tactile domain. Driven by concerns about the presumed specificity and uniqueness of this ability in human language (cf. Section *2.1.2*) and the lack of studies comprehensively investigating the ability to create abstract recursive hierarchical representations from sequential stimuli across different sensory domains (cf. Section *3.2.*), we aim to shed light on the nature of this ability. Specifically, we will investigate whether it is a domain-specific or domain-general ability, thereby clarifying the relationship between this cognitive ability and perception. To achieve this, we will examine it across three sensory domains: auditory, visual, and tactile. Hence, in Chapter 5, we will experimentally explore whether and how the human parser derives recursive abstract representations from sequentially arranged fading sequences of visual, auditory, and tactile stimuli. To illuminate the entire process leading to the creation of abstract recursive hierarchical representations, starting from exposure to a simple linear arrangement of stimuli, we require a suitable experimental design that enables us to delve precisely into this process. Specifically, it should allow us to directly study and compare the ability to create recursive abstract representations from sequentially presented stimuli in the three sensory domains, using a task as similar as possible across all three domains. Moreover, it should enable us to further explore the relationship between sequentiality and hierarchy, shedding light on the connection between low-level statistical mechanisms and the formation of recursive abstract representations. By doing so, we aim to clarify both the qualitative and quantitative aspects of this cognitive mechanism, building on the results and unresolved issues from previous studies in this field.

This chapter proposes a novel approach to address these challenges. We will explore the learnability of the Fibonacci grammar (Fib) using the AGL paradigm. Notably, Fib falls outside the Chomsky hierarchy, belonging to a different class of grammars known as Lindenmayer systems (L-systems). Initially developed by Aristid Lindenmayer in 1968 to model biological cell growth in plants, Fib serves as a valuable tool for examining the cognitive mechanisms at the heart of the creation of recursive abstract representation in sequentially ordered stimuli. It distinguishes itself from the typical rewrite grammars that are commonly used by offering intriguing characteristics like self-similarity and aperiodicity that make it ideal for our experimental purposes. The chapter also introduces a cognitive parsing algorithm designed for processing Fibonacci strings. This algorithm relies on hierarchical reconstruction through the recursive application of deterministic transitions between progressively larger embedded chunks.

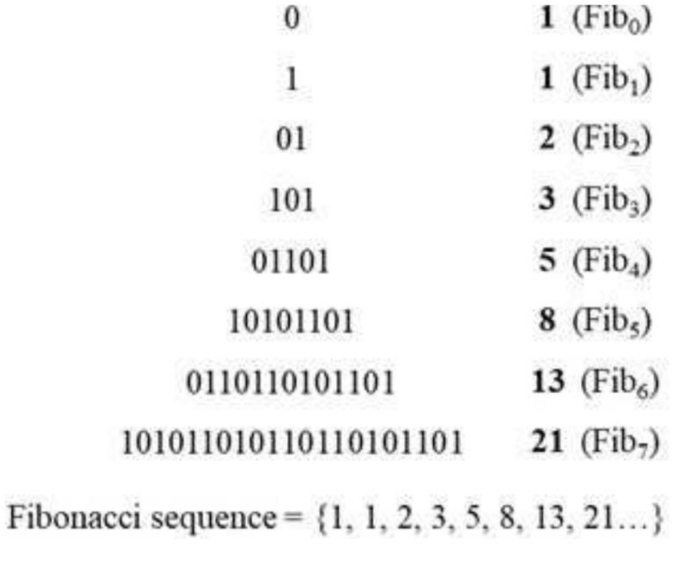
4.1. AGL with Lindenmayer Systems: The Fibonacci Grammar

Lindenmayer grammars (L-grammars), when considered as a framework to be learned within the AGL paradigm, offer multiple benefits as evaluation tools for recursive parsing. Aristid Lindenmayer introduced Lindemayer systems (L-systems) in 1968 as a formal means to describe the growth of algae, building upon Chomsky's research on formal grammars. Subsequently, they have undergone extensive development to simulate real-world plant growth and cellular behavior, serving as a tool to generate fractals, among other applications that encompass the creation of complex structures through straightforward rules[31] (Shirley, 2014). L-systems are non-canonical grammar, they do not belong to the Chomsky hierarchy. As outlined by Krivochen and Saddy (2018), L-systems exhibit several key characteristics and differences in comparison to Chomsky-normal grammars: (i) in L-grammars, there exists no distinction between terminal and non-terminal symbols; all symbols undergo rewriting. Put differently, the rewriting process, governed by the application of grammar rules, does not terminate after a finite number of steps; (ii) rewriting rules are applied simultaneously in L-systems. This stands in contrast to Chomsky-normal grammars, where rules are applied sequentially, following a specific order outlined in the grammar; (iii) L-systems display the property of self-similarity: each generated string can be mapped onto the previous generation. " L-systems are essentially recurrence relations, which means that once the initial state is given, the state of the system at any point is defined as a function of the preceding states". (Krivochen and Saddy, 2018, p.10).[32] Within the realm of L-systems, the Fibonacci grammar stands out as an interesting tool for assessing the creation of abstract recursive hierarchical representations

[31] For more details on L-systems, see: Lindenmayer, 1968; Prusinkiewicz and Lindenmayer, 1990.

[32] For further details on the properties of L-grammars and their comparison with Chomsky-normal grammars, see Krivochen, Saddy, 2018.

arising from sequentially ordered stimuli in the AGL paradigm. It also sheds light on the possible interplay between low-level statistical learning mechanisms and the emergence of recursive hierarchical representations, illuminating the entire continuum from sequence to hierarchy. The Fibonacci grammar, often abbreviated as *Fib*, is a simple rewrite system in which the alphabet is composed of only two symbols: Σ(0;1). Fib is characterized by the following rewrite rules: 0→1; 1→01; (0 rewrites as 1; 1 rewrites as 01). By the application of the rewrite rules, strings of 0s and 1s are generated. Every generated string is called generation$_n$ of Fib. By applying the rules repeatedly, strings of potential infinite length can be generated. Given its generative rules, it follows that Fib is an asymmetric grammar. In each generation, the number of the 0s is different from the numbers of 1s. Specifically, the ratio between 1s and 0s approximates the golden ratio (1.618). Fib derives its name from a unique feature observed in its generation process, namely, the number of digits (1s and 0s) in each row corresponds to a number of the well-known Fibonacci sequence (Figure 20).



0	**1** (Fib$_0$)
1	**1** (Fib$_1$)
01	**2** (Fib$_2$)
101	**3** (Fib$_3$)
01101	**5** (Fib$_4$)
10101101	**8** (Fib$_5$)
0110110101101	**13** (Fib$_6$)
101011010110110101101	**21** (Fib$_7$)

Fibonacci sequence = {1, 1, 2, 3, 5, 8, 13, 21…}

Figure 20. Representation of the Fibonacci grammar. Figure taken from Vender et al., 2023.

The characteristics that make Fib particularly suitable for our research objectives are the following:

(i) as already anticipated, Fib display the property of self-similarity. Due to the recursive nature of the generative process, each new generation results from the concatenation of the two preceding generations (Figure 21). “This means that any generation can be parsed with two consecutive smaller generations that are natural constituents of the grammar. For example, Generation 4 [01101] can be divided into Generations 2 and 3 [[01][101]], which can be further divided into Generations 1 and 2 [[01][1][01]], which can (trivially) be further divided into Generations 0 and 1 [[[0][1]][[1][[0][1]]]].” (Schmid et al., 2023, p.8). Crucially, every generation is a nested embedding of constituents, which mirror the hierarchical structure of Fib. This entails that transitions in Fib structure are preserved at different levels. In other words, the transitions exhibit a scale-free property, wherein the transitional probabilities between low level points remain consistent with those between larger constituents (Schmid et al., 2023). To illustrate, at every generation, there are first- and second-order transitional probabilities between the two symbols of the grammar 0 and 1. Specifically, the first-order transitional probability according to which after 0 there is always 1, that is p(1|0)=1[33]; and the second-order transitional probability according to which after the bigram 11 there is always 0, that is p(0|11)=1[34]; however, after the bigram 01 there can be a 0 or a 1. This means that this transition is probabilistic. Specifically, *p*(1|01)= 0.62, whereas p(0|01)=0.38. Nevertheless, points that are probabilistic at the low levels, can be predicted, becoming hence deterministic, at higher levels. Crucially, indeed, the same type of first- and second-order transitional probabilities that are observable between the two symbols of the grammar hold also between bigger chunks, which can be formed by recursively merging lower-level chunks linked by deterministic transitions, as we will further elucidate in the upcoming section. Being Fib a recursive self-similar grammar in which there is no distinction between terminal and non-terminal symbols, it generates potentially very long sequences of (binary) symbols which can be progressively chunked and compressed through a recursive algorithmic procedure. This property allows us to present strings of varying lengths to

[33] The probability of *1* following *0* is 100%.

[34] The probability of *0* following *11* is 100%.

participants, thereby providing different levels of possible embeddings. This is a particularly interesting feature. In fact, a point left unresolved by previous studies and analyzed only by Schmid et al. (2023) concerns the degree of recursive hierarchical compression that participants can achieve when processing a sequential structure. Planton et al. (2021) left this question open, as they did not analyze the degree of compression participants could reach (Schmid et al., 2023). In other words, they did not measure the extent to which participants can build recursive hierarchical structures by progressively compressing the sequence into abstract representations. In fact, Planton et al. (2021) investigated the ability to use recursive strategies with strings that allowed for a maximum of two hierarchical levels (one single level of embeddings). In contrast, using Fib, we can address this issue. Fib allows us to create binary sequences of symbols that can be compressed and processed into multi-level recursive hierarchical representations. Crucially, we can easily and precisely calibrate the maximum number of possible embeddings in a sequence. This enables us to investigate how many levels of recursive embeddings participants can actually achieve.

(ii) in all strings, points are aperiodic. This means that a parser could not predict the occurrences of every point based on linear functions. In other words, there are no low-level strategies that could be exploited to predict all the points (Vender et al., 2019; 2020; 2023; Schmid et al., 2023). Because of this property, it can therefore be ruled out that a low-level statistical processing strategy could account for learning some type of points, which, on the other hand, might potentially be predicted by exploiting a recursive hierarchical processing strategy, as we will see in more detail below.

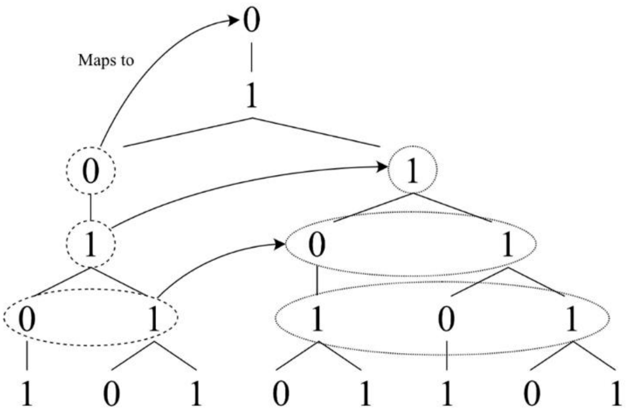


Figure 21. Self-similarity in the Fibonacci grammar. Taken from Vender et al., 2023, p.58.

4.2. Fib's parsing algorithm

In the previous section, we have anticipated that different types of point can potentially be learned (i.e. predicted) by recursively forming increasingly larger hierarchical structures. The purpose of this section is to present a theoretical demonstration of how the hierarchical processing can take place at the cognitive level.

The mechanism by which the parser (human cognition) can reconstruct the hierarchical structure of Fib, incrementally disambiguating and hence predicting points (i.e., disambiguated points, hence *D points*) that at lower hierarchical levels would not have been predictable (i.e., non-disambiguated points, hence *ND points*), is the recursive application of deterministic transitions between increasingly larger embedded chunks (Schmid et al., 2023). As explained above, since Fib is a self-similar grammar, the deterministic transitions between chunks mirror those observed between the two symbols of the grammar *0* and *1*. Crucially, in this parsing strategy, the same deterministic transitions are used both to predict the subset of disambiguated (D) points at each specific level and to form increasingly larger chunks. The fundamental step that is necessary for applying deterministic transitions at various levels is the *categorization of chunks*. Categorization takes place based on two perceptual features: alternation and repetition. Let us now look in detail at how this mechanism works.

The parser is exposed to a sequence of *0*s and *1*s[35]: The first feature of the sequence that the parser may notice is the following: the element *0* can never repeat itself. It is at this point that the parser learns to predict the D points at Level 0 (i.e., 0**1**). The next step is as follows: the parser notices that the element *1* can repeat itself, maximum once. This step coincides with the learning of D points at Level 1 (i.e., 11**0**). It is expected that the parser learns to predict chronologically earlier D points at Level 0 than D points at Level 1. In fact, to predict D points at Level 0, the parser needs to keep track in working memory of only one symbol (i.e., *0*), to predict the next one (i.e., *1*) in the sequence *0**1***. D points at Level 0, indeed, correspond to a first-order transitional probability, where $p(1|0)=1$. In the case of D points at Level 1, on the other hand, the parser needs to keep track of two symbols (i.e., *11*), to predict the following *0,* in the sequence *11**0***. In other words, D points at Level 1 correspond to a second-order transitional probability. Indeed, $p(0|11)=1$. After having acquired the regularities corresponding to D points at Level 0 and Level 1, the parser has the necessary and sufficient information to create two categories: *Category 0*, which groups the elements that can never repeat themselves; *Category 1*, which contains the elements that can repeat themselves, maximum once. At this point, the parser creates the first chunk, combining *0* and *1*: since $p(1|0)=1$, it forms the chunk [01] (Figure 22); once the chunk [01] is created, the parser is locked into the possibility of creating the chunk [110], which it could potentially form by exploiting the second-order transitional probability according to which $p(0|11)=1$. In fact, this would involve the creation of sequentially overlapping chunks. Given the impossibility to simultaneously processing the two overlapping chunks [01] and [110] on the string, the formation of the hypothetical chunk [110] is therefore blocked (Figure 23). At this point the parser learns the rule, which it will also apply to subsequent levels, for forming chunks: chunks, at each

[35] In the three experimental studies that we will present in Chapter 5, subjects were exposed to perceptual stimuli (auditory, tactile, or visual) encoded onto the *0*s and *1*s of the Fibonacci grammar. However, as a matter of convenience, in this section we will talk about *0*s and *1*s symbols, and not the specific stimuli that were used in the three sensory spheres.

level, can only be formed by applying the first-order transitional regularity according to which $p(1|0)=1$, on the elements belonging to the two categories. By applying the rule recursively, the parser can form increasingly larger embedded chunks. Hence, chunks, at each level, have characteristics of ordinal sets, in which the position of the sub-chunks mirrors that of the two symbols in [01] (i.e., the elements belonging to *Category 0* always precedes the elements belonging to *Category 1*). Having created the chunk [01], the string is now decomposed into [01]s and [1]s (Figure 24). The parser notices that the chunk [01] can repeat itself, maximum once. In contrast, [1] can never repeat itself. Therefore, it assigns [01] to *Category 1*, and [1] to *Category 0* (Figure 25). At this point it will be clear to the reader the importance that the categorization of chunks (i.e., labeling) plays in this processing strategy: Categorizing is a necessary step that must be done at each level to proceed with the hierarchical reconstruction. In fact, as we have just seen, at Level 0, [1] was assigned to *Category 1*. Instead, at Level 1, [1] is assigned to *Category 0*, that is, it is labelled as *0*. It follows, therefore, that chunks cannot remain categorized as they were at the previous level, but they need to be re-categorized. At Level 2, the parser has the possibility to predict the [1] following [01][01], i.e., D points at Level 2, by applying the second-order transitional regularity according to which $p(0|11)=1$ (Figure 26). Also at this level, it creates the new chunk [101], applying the first-order transitional probability according to which $p(1|0)=1$ (see Figure 27). The string is now decomposed into the following chunks: [01]s and [101]s (Figure 28). At Level 3, the parser notices that the chunk [101] can repeat itself, at most once, while the chunk [01] can never repeat itself. It therefore places [101] into *Category 1*, and [01] into *Category 0* (Figure 29). At Level 3, the parser learns that after [101][101] there is always [01], applying the second-order transitional regularity according to which $p(0|11)=1$ (Figure 30). At this point it creates the chunk [01101], applying the first-order transitional probability according to which $p(1|0)=1$ (Figure 31). Now, at this level the string is decomposed into the chunks [01101] and [101] (Figure 32). At Level 4, the parser observes that the chunk [01101] can repeat itself, at most once, while chunk [101] can never repeat itself. Therefore, it assigns [01101] into *Category 1* and [101] into *Category 0* (Figure 33). At Level 4, the parser learns that after [01101] [01101] there is always [101], applying the second-order transitional regularity according to which *p*(0|11)=1 (Figure 34). At this point, the parser can create the chunk [10101101], applying the first-order transitional probability according to which *p*(1|0)=1 (see Figure 35). Hence, the string is decomposed into [10101101] and [01101] (Figure 36). At Level 5, the parser notices that [10101101] can repeat itself maximum once, while [01101] can never repeat. Therefore, it assigns the chunk [10101101] into *Category 1* and the chunk [01101] into *Category 0* (Figure 37). At Level 5, the parser learns that after [10101101] [10101101] there is always [01101], applying to second-order transitional regularity according to which *p*(0|11)=1 (Figure 38). At this level the parser can create the chunk [01101101101], applying the first-order transitional probability according to which *p*(1|0)=1 (see Figure 39). For the sake of practicality, we have explained how the parser can predict D points up to Level 5. Of course, the same parsing strategy could be iterated to learn and hence predict D points at levels above 5 as well.

Summing up, in this parsing strategy: (i) binary categorization (labeling) is required at each hierarchical level: The parser cannot proceed with hierarchical reconstruction without doing categorization at each level; (ii) binary categorization is possible due to the fact that the Fib grammar is self-similar: Categorization at *Levels n > 1* is done based on the transitional probabilities between chunks, which mirror those between the symbols *0* and *1*; (iii) the prediction of D points is possible by exploiting the transitional probabilities between chunks. Specifically, by applying the first-order transitional probability according to which *p*(x∈C1|y∈C0) =1 (i.e., the probability of an element *x* belonging to *Category 1* following an element *y* belonging to *Category 0* is 100%). It follows that to reconstruct the hierarchical structure and predict points at different levels of the Fib grammar, it is necessary for the parser to first acquire D points at *Level 0* (0**1**) and *Level 1* (11**0**).

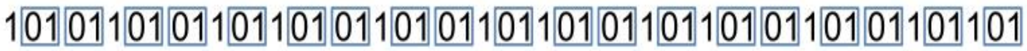


Figure 22. Creation of the chunk [01].

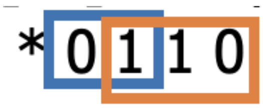


Figure 23. Impossibility to create two overlapping chunks [01] and [110].

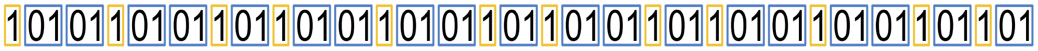


Figure 24. Chunking and parsing the string in [01] and [1].

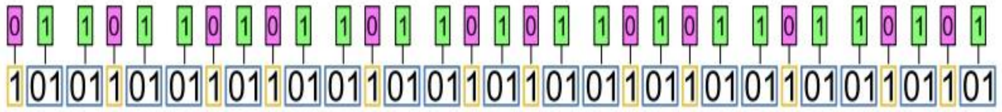


Figure 25. Categorization of the chunk [1] as 0 and the chunk [01] as 1.

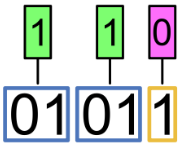


Figure 26. Prediction of the [1] following [01][01], i.e., D points at Level 2.

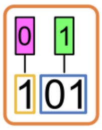


Figure 27. Creation of the chunk [101].

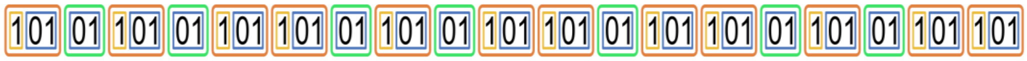


Figure 28. Chunking and parsing the strings in [101] and [01].

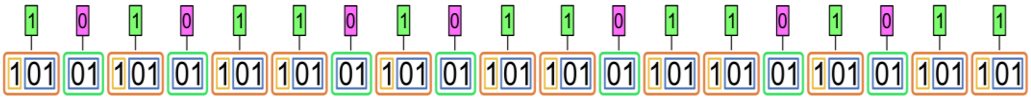


Figure 29. Categorization of the chunk [101] as 1 and the chunk [01] as 0.

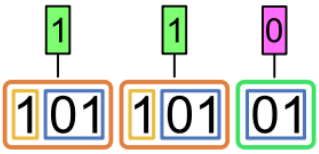


Figure 30. Prediction of the [0] following [101][101], i.e., D points at Level 3.

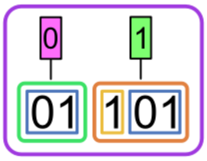


Figure 31. Creation of the chunk [01101].

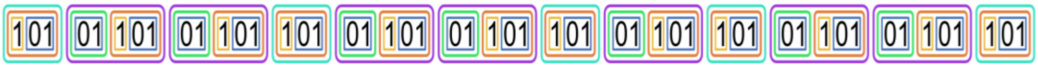


Figure 32. Chunking and parsing the strings in [101] and [01101].

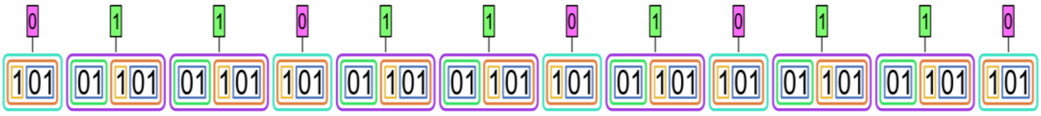


Figure 33. Categorization of the chunk [101] as 0 and the chunk [01101] as 1.

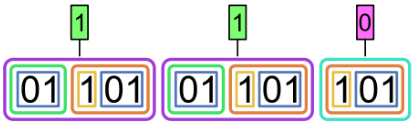


Figure 34. Prediction of the [1] following [01101][01101], i.e., D points at Level 4.

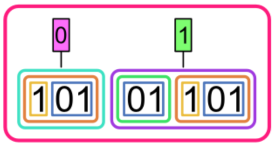


Figure 35. Creation of the chunk [10101101].

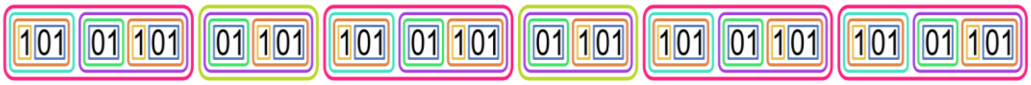


Figure 36. Chunking and parsing the strings in [10101101] and [01101].

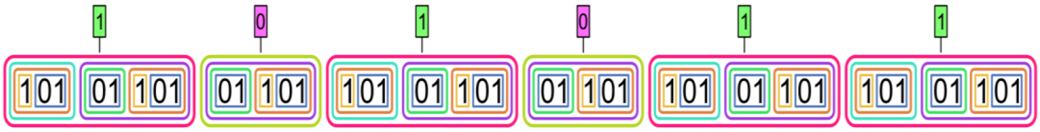


Figure 37. Categorization of the chunk [10101101] as 1 and the chunk [01101] as 0.

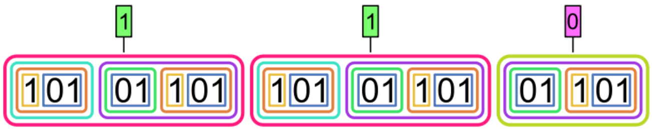


Figure 38. Prediction of the [0] following [10101101][10101101], i.e., D points at Level 5.

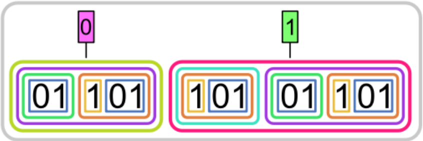


Figure 39. Creation of the chunk [0110110101101].

Here, we think it is important to elaborate on the concept of ND points. In this section, we have emphasized that D points are the predictable points at each level by applying the cognitive parsing strategy explained above. On the contrary, by definition, NDs s are the points that are not predictable at each level considered. However, this does not mean that they are not predictable at all. In fact, the set of ND points at *Level X* corresponds to the totality of points (D + ND) at *Level X+2*. In other words, ND points at *Level X* contain both the points that at *Level X +2* could be predicted (D points at *Level X+2*) and those that cannot be predicted at *Level X +2* (ND *Level X+2*). Similarly, ND points at Level X+2 encompass both points that are predictable at Level X+4 (D points at Level X+4) and points that are not predictable at Level X+4 (ND points at Level X+4). And so forth. (Figure 40).

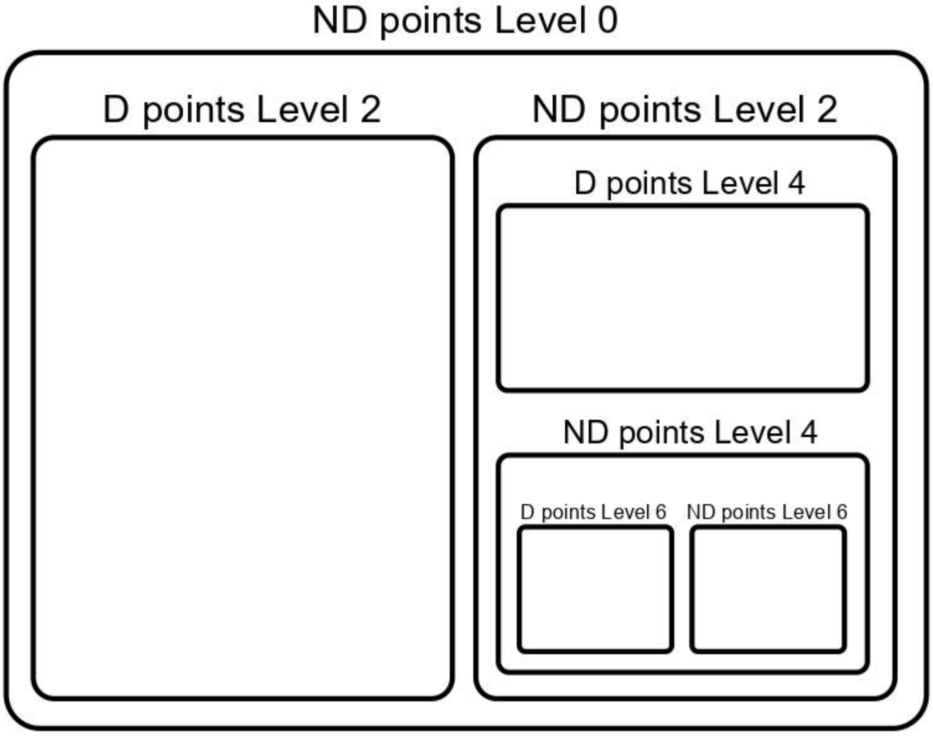


Figure 40. Representation of D and ND points at different hierarchical levels.

4.3. Are there alternative strategies to process and acquire the Fibonacci grammar?

In the previous section, we outlined the cognitive parsing algorithm we proposed for processing and acquiring the regularities of the Fibonacci grammar. Specifically, we provided a theoretical illustration of how recursive hierarchical processing can occur at the cognitive level. The parser might use a recursive application of deterministic transitions between progressively larger embedded chunks. This mechanism enables the parser to reconstruct the hierarchical structure, gradually resolving ambiguity and predicting points that would not have been predictable at lower hierarchical levels.

However, the question naturally arises: Is this the only mechanism available to predict points of increasing complexity in Fibonacci sequences? The answer is no. Theoretically, there exists another possible mechanism. However, it is unlikely that this alternative mechanism could be used if the parser were exposed to a single long sequence of Fib in a Serial Reaction Time task, which is the paradigm we use in the experimental studies presented in Chapter 5. This is due to the temporally fading nature of input stimuli inherent in the Serial Reaction Time task paradigm. Below, we will explain why.

Iteration is the alternative strategy to applying a recursive hierarchical algorithm that could potentially be used to predict points in Fibonacci sequences. Crucially, however, we believe that this strategy could only be used if the experimental design had precise characteristics. This is a fundamental point to consider when preparing an experimental design to investigate recursive ability while excluding other possible strategies. As we have pointed out, it is essential to distinguish between algorithmic properties and representational abilities (cf. Section *2.3.6.*). In an AGL task, it is not sufficient to use a recursive grammar; we must ensure that participants can process the sequences generated by the grammar recursively. More importantly, we must consider whether there are alternative strategies to process the sequence that could produce the same output as a recursive mechanism. As Lobina (2011) explains, all tasks that can be solved recursively can potentially also be solved iteratively. Summing up, adopting a recursive grammar, in this case, the Fibonacci grammar, as an experimental tool does not guarantee that we are testing recursion. We must also choose an experimental design that ensures we are testing recursive ability while excluding other possible cognitive algorithmic mechanisms. As we will see in Chapter 5, we will create our experimental design by combining the Fibonacci grammar with the Serial Reaction Time task paradigm (cf. Section *2.3.2.*). In this way, can we effectively and precisely verify whether participants adopt a recursive procedure in parsing the sequence, excluding other alternative parsing strategies such as iteration.

Below, we will explain the iterative procedure that the parser could potentially adopt if an experimental paradigm, such as the forced-choice paradigm, is created to allow for this parsing strategy. As we explained above, it is implausible that this strategy could be used when participants are exposed to a single long sequence of Fib generation in a Serial Reaction Time task. Crucially, after outlining this formal mechanism, we will explain why we think the iterative strategy is not feasible in a Serial Reaction Time task and why we believe that the recursive cognitive parsing mechanism is the most plausible option. Specifically, we will present the reasons why we consider iteration to be cognitively implausible in our experimental protocol. This will reinforce our hypothesis that if participants succeed in learning points of increasing complexity in the Fib sequence, it would be through the application of the recursive hierarchical parsing mechanism we have proposed, excluding other possible parsing strategies.

The alternative strategy to predict points of increasing complexity in Fib sequences is related to the use of a flat iterative statistical mechanism. In this mechanism, no hierarchical levels are created; instead, individual symbols of the sequence are processed through a purely sequential strategy. If we carefully observe the sequences generated by the Fibonacci grammar, we notice that, at each level, the Disambiguated (D) points are always preceded by a specific sequence of symbols, unlike the Non-Disambiguated (ND) points. Specifically, D points at Level 0 correspond to the 1s following *0*; D points at Level 1 correspond to the 0s following *11*; at Level 2, D points are the 1s following *0101*; at Level 3, D points correspond to the 0s following *1101101*; at Level 4, they are the 1s following

010110101101, and so on. In other words, *p* (1|010110101101) = 1[36]. All D points at higher levels are similarly preceded by sequences of symbols, progressively longer as the level increases. It follows that the D points of the Fibonacci sequence could potentially be predicted based on incrementally longer preceding sequences. For example, an iterative algorithm that the parser could follow to predict the D points at Level 4, processing the sequence from left to right, is outlined in (16), where *n* refers to the specific position of the parser in the sequence at a given time 0, and *n*+1 to the immediately adjacent position in the string proceeding from left to right.

(16) If *n*= 0 and *n*+1= 1 and *n*+2= 0 and *n*+3= 1 and *n*+4= 1 and *n*+5= 0 and *n*+6= 1 and *n*+7= 0 and *n*+8= 1 and *n*+9= 1 and n+10= 0 and *n*+11= 1, then *n*+12= 1.

As we can see in (16), to predict D points at Level 4 by exploiting an iterative strategy, the parser would need to keep in mind 11 elements of the sequence to predict the twelfth element (i.e., D points at Level 4). However, could this strategy be applicable by the human parser when exposed to a Fib sequence? Aligning with what Schmid et al. (2023) argued, we assert that it may not be possible for several reasons.

First of all, the iterative strategy would be implausible, regardless of the type of experimental paradigm used. The literature tells us that human working memory resources limit the number of items we can keep in mind without resorting to chunking strategies. While there are individual variations in working memory resources, despite these discrepancies, the literature agrees that the number of items is less than 10 (Miller, 1965 proposes 7±2 items; Baddeley et al., 1974 and Cowan, 2001 propose 4 items; for issues on working memory capacities and limitations, see also Feigenson, 2011; Kane et al., 2004; Li et al., 2013). Therefore, the iterative strategy could be at work for predicting the D points from L0 to L3, but it seems implausible for predicting the D points at L>3 due to human working memory resource limitations. The exclusion of the possibility that participants use an

[36] the probability of having a 1 after *010110101101* is 100%.

iterative strategy, however, should be experimentally demonstrated rather than simply relying on what the literature says about working memory resources and limitations. For instance, we should conduct additional tests on participants' working memory abilities. Crucially, however, by using a Serial Reaction Time task where a single long sequence corresponding to a full generation of Fibonacci is presented to participants through fading sequences of items, we can disentangle the two possible strategies, recursive and iterative, thereby excluding the possibility of the latter. In fact, as Schmid et al. (2023) explain, in a Serial Reaction Time task, the sequences that would need to be remembered to use an iterative strategy would be overlapping (Figure 41). Therefore, to predict D points without resorting to a hierarchical strategy, the parser would need to simultaneously track different overlapping fading sequences of incrementally greater length (Schmid et al., 2023). Moreover, the parser must effectively manage the challenges posed by the similarity of the patterns to correctly identify and separate them. "[…] the sequence being binary, the patterns are distinguishable only by their positional order; the parser must therefore also be able to deal with the interference caused by the similarity in the patterns' elements." (Schmid et al. 2023, p. 23). Finally, the patterns that enables the prediction of D points would need to be retained in memory for a relatively long time, comprising the response-to-stimulus interval and the time frame required to respond to the trial (Schmid et al., 2023; cf. Section *5.1.3.*). All of this would result in a strategy that is reasonably implausible to sustain for the human parser.

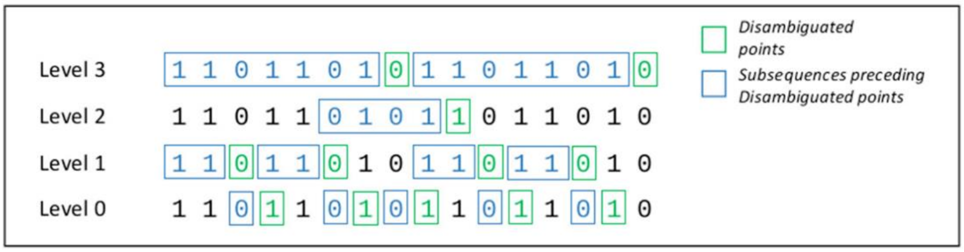

Level 3 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0
Level 2 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0
Level 1 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0
Level 0 1 1 0 1 1 0 1 0 1 1 0 1 1 0 1 0
Disambiguated points
Subsequences preceding Disambiguated points

Figure 41. Representations of the subsequences preceding disambiguated points at different hierarchical levels. Linear subsequences required to predict D points at each level overlaps. Figure taken from Schmid et al. 2023 p. 23.

The hypothesis that the human parser might exploit this purely sequential strategy to learn the regularities of the Fibonacci generations is therefore implausible due to issues related the different concurring factors we have illustrated above. For this reason, we regard the recursive hierarchical parsing algorithm presented in the previous section as the only cognitively plausible mechanism that the human parser could leverage to predict originally indeterministic points in a Fibonacci string in a Serial Reaction Time task paradigm. Hence, if the human parser successfully predicts increasingly complex points in a Serial Reaction Time task, we can conclude that it has genuinely applied the recursive cognitive strategy detailed in the preceding section.

4.4. Previous studies with the Fibonacci grammar

In this section, we will briefly present in chronological order the AGL studies with the Fibonacci grammar that have been carried out so far.

Shirley (2014) conducted a series of experiments to investigate the processing of complex sequences generated by two Lindenmayer grammars (L-grammars): the Fibonacci grammar and the XOR grammar. The aim of her research was to address the debate about the computational abilities of the human brain in supporting hierarchical cognitive systems like language and music, particularly the need for recursive processing. She carried out seven experimental studies within the Artificial Grammar Learning (AGL) paradigm, primarily using the two-alternate forced-choice task, while also recording electrical activity with the EEG method. Shirley's findings suggest that human adults can develop and retain a lasting representation of the Fibonacci L-grammar within the AGL paradigm. Crucially, her results indicate that participants were unlikely to rely solely on low-level mechanisms for accurate performance in the AGL experiments, and supported previous studies finding that context-free structures can be learned independently of semantics or contextual information. The EEG analyses did not provide clear evidence of participants' awareness of response errors, but spectral analysis suggested that different cognitive mechanisms were at play for the processing of the Fibonacci grammar and the XOR grammar.

Geambaşu et al. (2016) conducted two experiments using the Fibonacci grammar. The experiments aimed to test participants' ability to detect rhythms generated by these grammars using kick and snare drum sounds, encoding the symbols of the Fibonacci grammar. The experiments consisted of an exposure phase where participants listened to sequences following the grammar and a subsequent test phase where participants had to determine if test sequences matched the grammar. Two foils grammars were created: Swap and Mirror. The Swap foil was created by taking a sequence from the Fibonacci sequence and swapping two randomly adjacent symbols within the string. On the other hand, the Mirror foil was generated by cutting the Fibonacci sequence in half and then mirroring the first half to replace the original second half. In experiment 1, participants were instructed to listen to a 3-minute-long rhythmic sequence carefully. In the test phase, participants listened to 36 test sounds, and their task was to determine whether each test sound followed the same rhythm as the one they heard during the listening phase. Participants were divided into two conditions: the Mirror condition and the Swap condition. In both conditions, participants listened to the sequences of the Fibonacci grammar for 3 minutes. During the test phase, participants in both conditions had to discriminate between 10-second-long grammatical (Fib) and ungrammatical sequences (Mirror or Swap sequences, depending on their condition), deciding whether the sequences they heard in the test phase matched the rhythm of the sequences from the listening phase. In Experiment 2, participants received more detailed instructions, and the conditions remained the same. The participants listened to the same Fibonacci grammar sequence for 3 minutes during the exposure phase and then had to discriminate between grammatical and ungrammatical sequences in the test phase. At the group level, both experiments did not show clear evidence that participants were able to learn the Fibonacci grammars and discriminate them from the ungrammatical sequences (mirror and swap foil grammars). However, at an individual level, five participants in Experiment 2 were able to correctly identify grammatical and ungrammatical strings above chance level, especially those with musical training. The researchers suggested that the difficulty some participants had in discriminating between grammatical and ungrammatical sequences might be due to the foil grammars being too similar to the target grammar. They recommended further research with optimized foil grammars and different paradigms like Serial Reaction Time or EEG to investigate participants' cues and error detection points. Additionally, they highlighted the importance of specific instructions and participant age as potential factors affecting performance in such tasks and proposed addressing these issues in future experiments to better understand the role of rhythm detection in learning complex grammatical structures.

In Geambaşu et al. (2020), the primary objective was to explore whether adult participants could effectively learn and process sequences featuring the Fibonacci grammar. The experiment was similar to that carried out by Geambaşu et al. (2016). Sequences were composed of two distinct drum sounds, each lasting 200 milliseconds - a kick sound and a snare sound. The experimental setup involved an exposure phase where participants were familiarized with the Fibonacci-grammatical drumming sequence, followed by a test phase where participants were presented with various test sequences. These test sequences included both grammatical sequences that adhered to the Fibonacci grammar and ungrammatical sequences designed to share surface properties with the grammatical ones. The ungrammatical sequences were meticulously created so that both the test and foil sequences had the same number of elements, equal duration, and maintained similar surface properties. Hence, the construction of pseudo-Fib foil sequences aimed to maximize the similarity in surface properties between the grammatical and ungrammatical sequences while ensuring that discrimination between them was not overly obvious. The test sessions involved two different sets of instructions for participants, depending on the task paradigm employed - Yes/No or 2AFC. Results showed that, at the group level, participants were able to discern the grammar and distinguish between the grammatical and ungrammatical test sequences in both the Yes/No and a two-alternative forced choice task (2AFC). While their performance was significantly better than chance, it did not reach the high levels of accuracy seen in Shirley (2014), which transmitted the two symbols of the grammar through syllables-stimuli. As the authors explain, several factors could potentially account for this discrepancy in performance. Specifically, they point to the potential greater complexity of processing recursive grammars in a non-linguistic, musical context. Indeed, they propose that the speech stimuli used by Shirley (2014) might have facilitated the process of structural learning and recursive processing, as compared to musical stimuli. The researchers emphasized the need for future research to incorporate real-time measures, such as electrophysiological recordings or serial reaction time tasks, to gain a deeper understanding of how participants process complex grammars. In conclusion, this study demonstrated that adult participants were capable of detecting and discriminating the rules of the Fibonacci grammar, whose symbols were trasnmitted through drum sounds, though their performance was not as robust as in previous studies involving linguistic stimuli.

In Vender et al. (2019), the researchers conducted an AGL study in monolingual and bilingual children, both with and without dyslexia. They used a modified Simon task, a modified version of the classic Serial Reaction Time task, where the order of stimuli followed the rules of a Fibonacci grammar. The stimuli consisted of blue and red squares encoding the two symbols of the grammars, which were visually presented on a screen. Specifically, red stimuli corresponding to the 0s of the grammar were presented on the left side of the screen, whereas blue square encoding the 1s of the grammars were presented to the right side. Participants were asked to press the 1 key on the keyboard to answer to the red square, and the 0 key for the blue square. Importantly, every sixth item, the stimulus appeared on the opposite side of the screen (i.e., incongruent item). This effect was intended to maintain children attention high and make the presence of grammatical rules in the stimuli more subliminal. The goal was to assess whether participants implicitly learned the grammar's low-level regularities of Fib (after 0 there is always 1; after 11 there is always 0) and to examine group differences, particularly in the context of bilingualism and dyslexia. The study involved four groups of 10-year-old children: Italian monolingual typically developing children, bilingual typically developing children with Italian as a second language (L2), Italian monolingual dyslexic children, and bilingual dyslexic children with Italian L2. The results of the study revealed that all groups, including dyslexic children, showed evidence of implicit learning. They became increasingly sensitive to the grammar's regularities, leading to faster RTs and improved accuracy in predictable trials. However, group differences were observed, with bilingual children performing better overall than monolinguals, and dyslexic children being less accurate than the control group. In conclusion, the study found that all groups, including dyslexic children, were capable of implicit learning of the grammar's low-level regularities. Bilingualism seemed to confer advantages even for dyslexic children, while dyslexia was associated with lower accuracy, likely due to processing limitations. Overall, the findings suggested that bilingualism could be beneficial for linguistically impaired individuals, emphasizing the importance of supporting bilingualism in such populations. The study also proposed avenues for further research to explore the precise nature of implicit learning and its relationship to hierarchical structure.

Building on the research questions left open in Vender et al. (2019), Vender et al. (2020) tested a group of ten-year-old children with a modified Simon task, as in Vender et al. (2019), this time exposing them to sequences of stimuli governed by the rules of the Fibonacci grammar and the foil Skip grammar. The stimuli were the same as those used in Vender et al. (2019), that is, blue and red squares appearing on the right or left side of the screen rispectively, with an incongruent item every six items. The study had two main objectives. Firstly, they wanted to ascertain whether the children could discern low-level statistical regularities within the sequences, confirming the results found in Vender et al. (2019). Secondly, beyond recognizing simple low-level regularities, they were interested in verifying whether children could detect more complex structural patterns. In order to do that, they observed how children responded to specific points, called k-points, which in Fib hold special significance, as they allow, from a purely formal perspective, full reconstruction of the hierarchical structure of Fib (Krivochen, Phillips, Saddy, 2018). K-points are the 1s that follow the bi-gram 01 in Fib. Interestingly, as explained by the researchers, these points are not predictable by exploiting low-level statistical computations, but they could be predicted if the parser had access to the hierarchical structure of the grammar. Crucially, however, in Skip, these points do not have any structural importance. Hence, if participants learned the hierarchical structure of Fib, they expected to find increasingly better performances on k-points in Fib, and then, moving to Skip, they expected participants'performance to abruptly decline. Children did manage to learn the simplest low-level rule of the Fibonacci grammar (i.e., after 0 there is always 1), confirming what had been found in Vender et al. (2019). However, they encountered difficulty with more complex sequential statistical regularities (i.e., the regularity according to which after 11 there is always 0). The researchers hypothesized that this might have been due to the experiment's fewer trials compared to previous ones and because this rule was less central to understanding the structure of the Fibonacci grammar. Interestingly, the study provided compelling evidence that the children were sensitive to the structure of the Fib sequences. They appeared to pay particular attention to k-points in the Fibonacci string. In fact, RTs on k-points in Fib became progressively lower, while a rise occurred in the transition to Skip, as expected. This suggested that the children were not merely discerning sequential statistical patterns; they were arguably basing their increased capacity of prediction of non-deterministic points on some sort of hierarchical processing.

Schmid et al. (2023) explored whether participants could process binary sequences as nested structures. To investigate this, they tested adults' ability to learn the properties of sequences generated by the Fibonacci grammar through a Serial Reaction Time (SRT) task. They encoded the two symbols of the grammar onto blue and red squares, as in Vender et al. (2019; 2020). However, in contrast to previous studies, stimuli were always presented in the centre of the screen, to avoid the confounding congruency factor introduced by the Simon task. As Schmid et al. (2023) explained, Fib is a recursive rewrite system that generates aperiodic self-similar sequences with a hierarchical nature. Due to the self-similarity property, the transitions between elements at the lower level mirror those between elements at the higher level. Importantly, each level contains transitions that are either deterministic or probabilistic (i.e., disambiguated vs. non-disambiguated points). Crucially, however, the probabilistic transitions at one level are nested within deterministic transitions at the higher hierarchical level. This property, if exploited by participants, would allow for a reduction in the number of probabilistic transitions through the recursive embedding of deterministic ones. The researchers had two main predictions: First, they hypothesized that as participants engaged with the sequence, they would gradually reconstruct the underlying nested hierarchical structure. This should manifest as an increasing ability to anticipate predictable points in the sequence (disambiguated points) that were ambiguous at a lower level as compared to non-disambiguated points that would, on the opposite, show higher RTs and/or lower accuracy. The study supported this prediction, showing that participants indeed displayed a progressive ability to anticipate disambiguated points, which showed a steeper reduction of RTs compared to non-disambiguated points at the same level. Specifically, at levels 0, 1, 2, and 3, participants showed a more pronounced reduction in RTs on disambiguated points compared to their non-disambiguated counterparts at the same levels. At levels 0 and 1, participants became more accurate on disambiguated points, while the accuracy for non-disambiguated points decreased. At levels 2 and 3, both disambiguated and non-disambiguated points underwent a decrease in accuracy. Importantly, despite the decrease in accuracy at levels 2 and 3, the study's predictions were not invalidated. At level 2, indeed, the decrease in accuracy was significantly more pronounced for non-disambiguated points compared to disambiguated ones. Moreover, at level 3, the accuracy was overall higher for disambiguated points. The authors suggested that the decrease in accuracy rates at higher levels could potentially be attributed to factors like participant boredom due to the simplicity of the task. Nevertheless, the study's findings supported the notion that participants were progressively building and learning the hierarchical structure of the sequences, up to the third level. Indeed, at level 4, they found no significant results anymore. Schmid et al. (2023) carried out a second analysis in which they aimed to delve deeper into how participants processed the Fibonacci string. They wanted to verify whether participants became increasingly better at predicting disambiguated point by representing the nested structure of the grammar instead of exploiting a flat statistical learning process. To do so, they checked whether participants not only recognized disambiguated points at different levels but also were sensitive to the higher-level structure in which these points appeared. To investigate this, they examined what they termed "structural contexts" within the sequence. They distinguished between two conditions: one where a disambiguated point appeared at a higher level inside a constituent following a deterministic transition, which they termed a "non-ambiguous structural context," and the other where it appeared in a constituent following a probabilistic transition, known as an "ambiguous structural context." Their hypothesis was that if participants truly grasped the hierarchical structure of the sequence, disambiguated points occurring within a non-ambiguous structural context should be processed faster than those within an ambiguous structural context. At Levels 1 and 3, they observed that participants displayed a significant processing advantage for points occurring in non-ambiguous structural contexts compared to those in ambiguous structural contexts. This suggested that participants were progressively learning and processing the constituent structure. However, when they looked at level 2, they found no significant effect of structural context in either reaction times or accuracy. They considered various explanations for this lack of effect at level 2, including the possibility that certain points at this level had already been learned very early in the experiment. They explained that these points might have reached a performance plateau, making it difficult to detect the influence of structural context. Summing up, the study's findings, especially those at levels 1 and 3, ruled out the possibility that participants were merely memorizing preceding subsequences. Instead, it suggested that participants were indeed organizing the input in a hierarchical manner representing nested constituents. Wrapping up, this second analysis provided further evidence that participants were processing the Fibonacci grammar as a nested structure through hierarchical processing. Self-similarity played a crucial role in processing the Fibonacci sequence, by contributing to the reduction of unpredictability and guiding the human parser in the reconstruction of its hierarchical structure.

Vender et al. (2023) aimed to shed light on the relationship between the cognitive development of hierarchical representations and their linearization that stands at the basis of language processing. The paper addressed the following questions: (i) To what extent is language processing based on sequential versus hierarchical learning? (ii) Can independent cognitive biases be identified for sequential and hierarchical learning, and how do they interact? The study explored whether these two modes of learning are independent or intertwined and sought to define the algorithm by which humans derive structure from linear order. The overarching goal was to uncover the link between sequential and hierarchical computations, suggesting a cognitive bias that shifts from the horizontal axis to the vertical one.

Vender et al. (2023) went beyond establishing a connection between linear and hierarchical representations, aiming to identify a potentially domain-general cognitive bias in humans for projecting sequentially-ordered symbol arrays into graph-like structures. In this study, they used a modified Simon Task with sequences generated by the Fibonacci grammar and its modifications (Skip and Bif) to investigate how participants implicitly learned the statistical regularities of these grammars and whether they exhibited hierarchical learning, with a special focus on those points that have structural significance in Fib (k-points), as in Vender et al., (2020). They carried out two experimental studies. In Study 1, they presented participants with a sequence of three Skip blocks followed by three Fibonacci (Fib) blocks. The main goal was to investigate whether learning k-points (i.e., the 1 following 01) in Fib could be solely attributed to statistical distributional statistics or if it involved a hierarchical style of computation. The logic behind this design was to compare the learning effects of k-points (011) in Fib to the learning effects of the sequence "010" in Skip. Indeed, in Skip 010 are more frequent than 011, while the opposite holds for Fib. If learning k-points was solely based on statistical frequencies, one would expect similar learning effects for 011 in Fib and 010 in Skip. On the contrary, if participants succeeded in learning k-points in Fib while not showing an improvement on 010 in Skip, this would suggest that the parser's ability to predict k-points was not merely a result of statistical sampling from the string. In the second experiment, the researchers aimed to compare learning effects on 011 in Fib and Bif. The strings generated by these two grammars are superficially similar but have a different hierarchical structure. Importantly, in Fib, k-points are significant to reconstruct the hierarchical structure, while in Bif, they have no structural importance. Comparing learning of these two grammars allowed the researchers to gain insights into the nature of hierarchical learning and the strategies employed by the parser to predict k-points. Overall, the results of both experiments provided evidence not only for statistical sequential learning but also for hierarchical learning of the Fibonacci grammar, suggesting that the results observed were not due to low-level statistical effects and confirming the presence of hierarchical reconstruction in parsing the Fibonacci strings. With this study, the authors investigated the cognitive foundations of language, focusing on the ability to group symbols into chunks, categorize them, and establish a linear order relation in a bidimensional space. They provided interesting insights on the nature of the interplay between the capacity to build hierarchical representations and sequential statistical learning, by exploring the intricate relationship between precedence and containment during the processing of Fib string. Specifically, the authors argued that precedence and containment are not opposing ways of processing a temporally ordered sequence; instead, they are interdependent implementations within a bidimensional computational space. They proposed that humans possess a possibly domain-general disposition to introduce a vertical computation axis while processing symbol sequences horizontally. This cognitive bias, labeled the *Bootstrapping Principle,* is seen as a cognitive source of the hierarchy-based computations in natural language. In this view, the vertical axis represents a distinct instantiation of the same abstract mathematical relation of linear ordering (i.e., reflexive, asymmetric, and transitive), interpreted as containment as well as precedence. The construction of this space was suggested to be primarily determined by the labeling requirements (i.e., the categorization of chunks emerging as output of statistical sequential learning). Specifically, they suggested that the parser reinterprets precedence as containment and applies a labeling algorithm based on this reinterpretation. "[…] once the parser has reached the knowledge that the natural chunks in a Fib-string are 01 and 1, there is a natural trigger for the parser to re-analyze the relation of precedence between subsequences of that string as a relation of containment between the sets corresponding to those subsequences. […] If $x < y$ within a chunk, then $x \subseteq y$." (Vender et al., 2023, p. 78). Hence, the labeling algorithm emerged as a solution for mapping precedence into containment. The authors suggested that formal properties of Fib-generations, particularly self-similarity, acted as triggers for associating precedence with containment. In summary, the study proposed that hierarchy is projected from linear order, with both relations being interpretations of the same abstract mathematical relation of linear ordering.

4.5. Conclusion

In Section *2.3.5.1.*, we discussed the possible limitations of using A^nB^n stringsets as a measure of recursion and the scarcity of studies that have adequately demonstrated recursive capabilities. We emphasized the need for clearer definitions and suitable tools to study recursion across sensory domains. To address this challenge, in the present chapter we introduced the Fibonacci grammar (Fib), a grammar belonging to the Lindenmayer systems (L-systems). Fib's unique properties, including self-similarity and aperiodicity, make it a suitable tool for investigating the formation of recursive hierarchical abstract representations arising from sequential stimuli. However, we have emphasized that to study recursion using Fib and harness its unique properties, it is crucial to implement and tailor experimental designs that address the challenges inherent in studying the cognitive abilities to deal with recursive processes. We also proposed a recursive parsing algorithm for processing Fibonacci strings. Moreover, we outlined the reasons why we believe it might be the only mechanism compatible with human cognitive resources for predicting originally indeterministic points in Fibonacci sequences in a Serial Reaction Time task. We concluded the chapter by presenting the main results of the studies which have investigated the learnability of the Fibonacci grammar so far.

In the following chapter, we will present our experimental studies, applying the AGL paradigm to test the learnability of the Fibonacci grammar, further exploring the cognitive mechanisms involved in processing these structures. Specifically, we will present an original methodology to investigate the ability to implicitly form recursive hierarchical abstract representations arising from sequentially arranged fading stimuli in three different sensory modalities: the visual, the auditory, and tactile domains. In our experimental studies, we will expose participants to sequences generated by the Fibonacci grammar, presenting the two symbols of Fib through different types of sensory stimuli, and testing their performances through SRT tasks. In all three studies, the stimuli will consist of sequences of temporally ordered elements, with the sequential dimension being primary. The linear order of symbols is closely linked to the hierarchical structural representations participants may form to simplify processing and anticipate points of increasing complexity in the sequence. Given the fact that the ability to form recursive hierarchical abstract representations from sequentially ordered stimuli is a cognitive ability at work in both language and music (cf. Section *2.1.2.*; *2.2.1.*) and considering that music and language are preferentially conveyed through the auditory perceptual domain, the question arises whether this cognitive ability is modality based. Are we better at learning and processing these structures in the auditory domain? Does the auditory domain have an advantage over the visual and the tactile ones? Indeed, given the results in the literature that we presented in the preceding chapters, the ability to form recursive hierarchical abstract representations from sequential stimuli might be stronger in the auditory domain than in other sensory domains. The alternative hypothesis is that this ability is stimulus-independent, and, on these grounds, we could form recursive hierarchical abstract representations from sequential stimuli in the visual, auditory, and tactile domains in a very similar way.

5. The Present Study: AGL with the Fibonacci Grammar. Investigating the Formation of Recursive Hierarchical Abstract Representations Arising from Sequential Fading Stimuli in the Auditory, Tactile, and Visual Sensory Domains

In this chapter, we present the results of three AGL studies that investigated the ability to process and represent recursive hierarchical structures arising from sequentially presented fading input in three different sensory modalities: auditory, tactile, and visual domains. We designed three Serial Reaction Time tasks in which three groups of participants were exposed to the same sequence of binary stimuli featuring Fib rules. The sequence was conveyed through different types of stimuli (i.e., two pure tones of different amplitude in the auditory condition; two colorful squares for the visual condition; and two vibrotactile stimuli for the tactile condition). For the reasons we explained in Section *4.1.*, the Fibonacci grammar is particularly well-suited for investigating the representation of recursive hierarchical abstract representations arising from sequential stimuli. Importantly, it allows us to study the transition from linear to recursive hierarchical processing, shedding light on the interaction of different mechanisms involved in the process, with varying degrees of abstraction, as detailed in previous chapters. Thus, the paradigm we used in these studies, enabled us to examine the interplay between sequential implicit statistical learning mechanisms and the formation of abstract recursive hierarchical representations in three different sensory domains. Crucially, to our knowledge, this is the first experimental study ever conducted that has explored this ability through a paradigm capable of directly comparing performances in all three sensory spheres. The goal of this study was to understand whether the ability to form recursive hierarchical abstract representations from sequential stimuli is present in all three sensory modalities and to identify any similarities or differences among them. Specifically, we wanted to verify whether participants exploited the cognitive parsing mechanism presented in Section *4.2.* to acquire the regularities of the Fibonacci string in the three different sensory modalities. To do that, we measured reaction times and accuracy rates in correspondence to every D (disambiguated) point and ND (non-disambiguated) point at each hierarchical level of the Fibonacci grammar. Specifically, D points are those points that can be predicted by acquiring low-level transitional regularities - as in the case of D points at level 0 and 1- or by tracking transitional regularities between increasingly larger chunks, hence forming recursive hierarchical representation - as in the case of D points at levels ≥ 2 - (cf. Section *4.2.*). In case of learning, we expected to find better performances on D points than ND points within levels, in terms of increasingly shorter reaction times throughout the task. As explained in Section *4.2.*, by definition, ND points are those that elude predictability at each examined level. However, this does not imply absolute unpredictability. To clarify, the set of ND points at Level X encompasses all points (both D and ND) at Level X+2. In simpler terms, ND points at Level X include both those that could be predicted at Level X+2 (D points at Level X+2) and those that remain unpredictable at Level X+2 (ND at Level X+2). The same holds true for the higher levels (cf. Section *4.2.*). For this reason, we did not rule out the possibility of finding signs of learning also for the category of ND points. However, we anticipated that if evidence of learning were to emerge within a level for both D and ND points, we would expect the latter to be characterized by generally higher reaction times (RTs) compared to D points. Additionally, we would expect the decrease in RTs for ND points to start later within the blocks compared to D points.

Regarding accuracy, it will not be considered the primary indicator of learning. Instead, we will base the significance of our findings on reaction times, as we believe they are a more accurate measure in cases of learning effects compared to accuracy. As we will discuss in the following sections, due to the simplicity of the task, we expect response accuracy to be nearly at ceiling levels, as found in Schmid et al. (2023). Therefore, given the simplicity of the task, we will not rely on accuracy results to determine significance. However, we still consider it important to analyze accuracy, as it might provide relevant insights for the overall interpretation of the results. High rates of inaccuracy, for instance, could indicate that participants responded randomly or did not complete the task diligently, thereby highlighting potential outliers. Additionally, even if we expect accuracy levels to be very high, they could still show trends that align with or diverge from the reaction time results. If both reaction time and accuracy results exhibit similar trends (e.g., increased accuracy alongside decreased reaction times in cases of learning), it would strengthen and confirm the findings based on reaction times. However, we must account for other factors that might affect accuracy levels, such as boredom or fatigue caused by the task's length.

As explained in *Section 4.2.*, in Fib strings, specific points can be predicted by leveraging low-level statistical information (Disambiguated points at Level 0 and 1). Specifically, these points can be predicted by means of low-level conditional statistics applied to the sequence of symbols. As we have seen, D (disambiguated) points at Level 0 coincide with a first-order transitional regularity where p (1|0) = 1, while D points at Level 1 represent a second-order transitional regularity where p (0|11) = 1. Consistent with previous studies exploring domain-specific differences in processing low-level sequential statistical learning (cf. Section *3.1.4.*), we expected to find that D points at Level 0 and Level 1 were learned in all three sensory modalities. Indeed, previous studies have found evidence of this ability in both the auditory, tactile, and visual domains. However, differences have also been identified. Notably, there is agreement among studies that the auditory domain outperforms the visual domain in processing sequential statistical information, while evidence regarding the auditory-tactile comparison is inconclusive (cf. Section *3.1.4.*). Consequently, we predicted better performance in learning these points in the auditory sphere compared to the visual sphere, but we did not have specific hypotheses for the tactile domain. Moreover, we were interested in investigating whether the tactile domain would have exhibited advantages or disadvantages compared to the other two sensory domains. Since D points at Level 0 are computationally less complex than those at Level 1, we expected the former to be learned chronologically earlier than the latter. Indeed, as highlighted in Section *4.2.*, to predict D points at Level 0, the parser needs to keep track of only one element (0) to predict the next one (1) in the bigram 01, representing a first-order transitional regularity where p (1|0) = 1. In contrast, for D points at Level 1, two symbols (11) need to be held in working memory to predict the next one (0) in the trigram 110, being a second-order transitional regularity where p (0|11) = 1.

As for the ability to predict D points at Levels ≥ 2, the cognitive parsing algorithm, outlined in Section *4.2.*, relies on the construction of increasingly larger hierarchical abstract representations formed through the application of a recursive algorithm. This mechanism, the only strategy deemed cognitively plausible to predict points of different complexity in Fib sequences, involves transitioning from the sequential dimension to the hierarchical one (cf. Section *4.3.*). The transition is achieved by identifying and projecting low-level statistical regularities across different hierarchical levels. To elaborate, according to the Fib cognitive parsing algorithm proposed in Section *4.2.*, at each hierarchical level, the parser must chunk symbols, categorize them, and recursively form larger embedded chunks based on their distributional properties. Crucially, in this cognitive algorithm, the distributional properties between the two symbols of the grammar 0 and 1 are projected across the board due to the self-similarity property of Fib. As discussed in Section *3.1.6.*, in the literature we find studies that have explored and confirmed the ability to create recursive hierarchical representations in the auditory and visual spheres (Martins et al., 2014; 2015; 2017). The results of these studies suggest that the ability to represent recursive abstract structures is a domain-general cognitive skill. Indeed, a correlation has been found between auditory recursion abilities and corresponding skills in action sequencing and the visual domain (Martins et al., 2017). However, as highlighted in Section *3.1.6.*, visual studies conducted so far have investigated the ability to represent recursive structures in fractal images, that is, in *static* figurative representations where recursion was *spatially* displayed. In contrast, given the temporal nature of sound, studies that have investigated recursion in the auditory domain have examined this ability in the context of *sequential* fading auditory sequences, thus in the *temporal* dimension. To our knowledge, no study has explored the ability to form recursive hierarchical abstract structures from sequentially arranged stimuli in the visual domain. Crucially, moreover, no study has ever investigated the ability to represent recursive hierarchical structures in the tactile domain. It follows that, at the moment of our investigation, we had no information about the potential ability to process and represent recursion through touch. Our study, therefore, is the first to aim at investigating the ability to process and represent recursive hierarchical structures arising from sequentially presented fading input, through a directly comparable paradigm across the visual, auditory, and tactile sensory domains. The goal of our study was to shed light on possible domain-general and domain-specific constraints in the process. Regarding the auditory domain, in line with previous studies, we expected to find evidence of this ability. On the contrary, we had no specific expectations for the visual and tactile domains. We asked ourselves whether evidence of this ability would have been found in these two domains as well. And, if so, if there would have been domain-specific differences. Having observed that the auditory domain outperforms the visual domain in processing sequential implicit statistical information (cf. Section *3.1.2.*) and considering that the ability to represent recursive hierarchical structures in our paradigm is closely intertwined with sequential implicit statistical learning abilities (cf. Section *4.2.*), we hypothesized the auditory domain to have an advantage over the visual one in creating recursive hierarchical representations from sequentially arranged sequences of fading stimuli. The question remained, however: What would be the outcome in the tactile domain?

Wrapping up, the final goal of our study was to shed light on the ability to form recursive hierarchical abstract representations from sequentially arranged fading stimuli in the auditory, tactile, and visual sensory domains. Specifically, we aimed to explore the relationship between sequential implicit statistical learning and the formation of recursive hierarchical representations. Moreover, we were interested in assessing possible domain-specific constraints in the process.

5.1. Method

5.1.1. Participants

Thirty-one subjects took part in the Auditory Study (21 females and 10 males, mean age= 24.96 SD=6.21); thirty-five subjects took part in the Tactile Study (23 females and 12 males, mean age= 26.66 SD= 3.22); thirty-one subjects took part in the

Visual Study[37] (7 male, 24 female). Their ages ranged from 21 to 37 years (*M* = 24.76 *SD* = 6.27). Participants of all the three studies were volunteers recruited through announcements at the University of Verona. They had normal or corrected-to-normal vision, and they did not suffer from any neurological, speech, learning, hearing disorder. They presented correct use and functioning of upper limbs. The three studies were approved by the local ethics committee at the University of Verona and conducted in accordance with the standards specified in the 2013 Declaration of Helsinki. Informed written consent was obtained from all participants. Each participant was provided with a reimbursement of €5.

5.1.2. Materials

In all three experiments, participants were exposed to the same sequence of stimuli, whose pattern was determined by the rules of the Fibonacci grammar. Specifically, the string was composed of 534 stimuli generated by Fib (from generation 14) divided into 3 blocks of 178 stimuli each (blocks 1-3).

In the Auditory Study, stimuli consisted of two pure tones (sine wave) of the same amplitude but of different frequencies generated by the Audacity® software version 3.0.0. Stimulus A had a frequency of 333 Hz, whereas Stimulus B of 286 Hz (Conway & Christiansen, 2005). The stimuli were transmitted through Bluetooth V5.0 bone conduction headphones (Tayogo ® S2 14 x 4.5 x 13 cm; 35 grams), whose ear hooks were correctly positioned around participants' ears so that the transducers sat comfortably outside of their ear and just above their temple bones. Participants were asked whether the volume of stimuli was adequate. The $[0]_s$ of the Fib grammar were transmitted through Stimulus A (333 Hz), while the $[1]_s$ through Stimulus B (286 Hz). The task was conducted using the DMDX Automode software version 6.0.0.4.

[37] Method and part of data from the Visual Study have already been presented in a different paper (Compostella et al., under review), in which we investigated the interaction between implicit statistical learning and the cognitive bias known as alternation advantage in serial reaction time tasks. In the present work, we analyze part of the same data to investigate a different issue, hence adopting a different approach in the analysis.

In the Tactile Study, stimuli were created using Audacity® software version 3.0.0. and consisted of two pure tones (sine wave) with a frequency of 120 Hz, an amplitude of 0.8, and a duration of 1000 ms. The two tones were used to generate light vibro-tactile impulses which were transmitted to the participants' thumbs via the same pair of Bluetooth V5.0 bone conduction headphones used in the Auditory Study (Tayogo ® S2 14 x 4.5 x 13 cm; 35 grams). The headphones were placed in contact with the fingertips of participants' thumbs and were held firmly by two elastic latex bands that wrapped the headphones around the fingers. The intensity of the impulses was modulated to be well perceived by participants, but at the same time light, not annoying, and not audible. Fib's $[0]_s$ were transmitted through vibro-tactile stimuli to the right thumb, the $[1]_s$ through vibro-tactile stimuli to the left thumb. The experiment was conducted using the DMDX Automode software version 6.0.0.4. (Forster & Forster, 2003).

In the Visual Study, stimuli consisted of blue and red squares (dimensions 1012x536 pixels, BMP files), sequentially presented one at a time, to the right or left of a computer screen. The pattern of stimuli was determined by the Fib grammar. Fib's 1s and 0s were associated and transmitted as blue squares and red squares, respectively. Red squares always appeared to the left side of the computer screen, while blue squares to the right. The task was run in DMDX Automode version 6.3.1.4 software (Forster & Forster, 2003). The methodology employed in the Visual Study mirrors that used in Schmid et al. (2023) and Vender et al. (2019; 2020; 2023), with the exception that, unlike the studies by Vender and colleagues, this study did not include the presence of incongruent stimuli (See Section *4.4.* for further details).

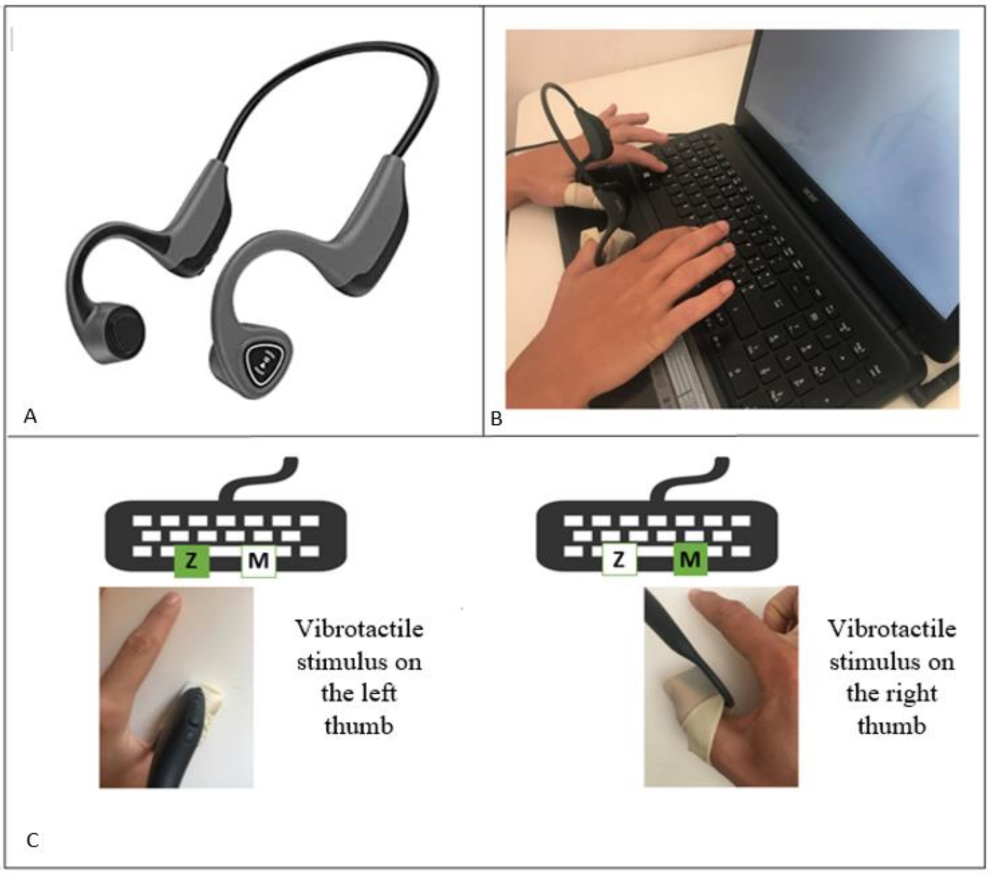

A
B
Z
M
Vibrotactile stimulus on the left thumb
Z
M
Vibrotactile stimulus on the right thumb
C

Figure 42. (A) Bone-conduction headphones used in both the tactile and auditory studies; (B) Apparatus of the Tactile Study: PC and bone-conduction headphones on participant's thumbs; (C) Vibrotactile stimulus on participant's thumbs and respective response buttons on the keyboard.

5.1.3. Procedure

In all three studies, participants were tested individually in the LaTeC (Language, Text and Cognition) Laboratory at the University of Verona, in a dimly lit and soundproof testing room. They were not informed that the sequence of stimuli followed the rules of an artificial grammar, as in Vender et al. (2019; 2020; 2023); and Schmid et al. (2023). Participants were informed that they would have been exposed to a binary sequence of stimuli and instructed to respond to the two stimuli by pressing specific keys on a computer keyboard as quickly and accurately as possible. Only at the end of the experiments, participants were informed that the sequence of stimuli was not random, and they were asked if they had noticed any patterns. To get participants acquainted with the task, they started with a familiarization phase, which comprised eight trials that did not adhere to the Fib grammar rules. In this phase, they received on-screen feedback indicating whether their responses were correct or incorrect. After completing the familiarization phase, participants were given the opportunity to ask any questions they may have had. If no questions were asked, the testing phase started, and no feedback was provided to participants. The task lasted about 20 minutes for all three studies.

In the Auditory Study, participants were informed that they would have been exposed to a sequence of two auditory stimuli of different frequency and were asked to respond to Stimulus A by pressing as fast and accurately as they could the [z] key button on the keyboard and to Stimulus B by pressing [m]. Each trial began with a fixation cross that appeared in the center of the screen for 500 ms, followed by a 250 ms delay before the transmission of one of the two tones (Stimuls A, or Stimulus B). The tones were transmitted to both ears through headphones and had a duration of 1000 ms, regardless of participants' response time. If participants did not provide a response within this time window, a new fixation cross appeared in the center of the screen. The timing started with the beginning of tone transmission and ended when the participant provided a response by pressing a key.

In the Visual Study, participants were given instructions that they would have seen red and blue squares appearing on the screen and were asked to respond as fast as possible to stimuli by pressing on the computer keyboard the [z] key for red squares and the [m] key for blue squares. Each trial began with a fixation cross that appeared in the center of the screen for 500 ms, followed by a 250 ms delay before the appearance of the red or blue square. Red squares were always displayed on the left side of the screen, while blue squares were displayed on the right side. The squares remained visible for 1000 ms, regardless of participants' response time. If participants did not respond within this time window, the stimulus disappeared, and a new fixation cross appeared in the center of the screen. The timing started with the square's appearance and ended when the participant gave an answer by pressing the key.

In the Tactile Study, participants were informed that they would have been exposed to a sequence of vibrotactile stimuli transmitted to the right or left thumb and were asked to respond to stimuli by pressing specific keys on the computer keyboard, trying to be as accurate and as fast as possible. Specifically, they were required to press the [z] key on the keyboard when they perceived the stimulus on their left thumb, whereas [m] key for the stimulus on their right thumb. Each trial began with a fixation cross that appeared in the center of the screen for 500 ms, followed by a 250 ms delay before the appearance of the vibrotactile stimulus to the left or to the right thumb. The stimulus lasted for 1000 ms, regardless of participants' response time. If participants did not provide a response within this time window, the stimulus ended, and a new fixation cross appeared in the center of the screen. The timing started with the beginning of the vibrotactile stimulus and ended when the participant pressed the [z] or [m] on the keyboard.

5.2. Data analysis

We analyzed RTs and accuracy rates across blocks (1-3) comparing disambiguated (D) and non-disambiguated (ND) points at every level (1-6) in each modality (auditory, tactile, visual). As outlined in Chapter 4, D points are the points within Fib strings that can be predicted at each level of Fib's hierarchical structure, if the parser exploited the cognitive parsing algorithm proposed in Section *4.2.* This cognitive parsing strategy specifically entails the recursive application of first- and second-order transitional regularities between progressively larger embedded chunks (cf. Section *4.2.*). Comparing the two types of point (D vs. ND) allows us to finely evaluate the presence of learning, avoiding potential problems inherent in considering only D points. In fact, in the case where a progressive decrease in RTs on D points across blocks was observed would not guarantee us that learning has taken place. Other factors, such as habituation to the task effect, might have played a role. The habituation to the task effect is a phenomenon that leads to a decrease in reaction times, irrespective of whether statistical learning is present in the task. In serial reaction time task protocols, participants may naturally adapt to the task over time. As they engage in the task, their reaction times may undergo changes. Initially, responses might be slower, but through practice, participants could improve their speed as they become more familiar with the task. Likewise, potential differences in RTs and in the trends of accuracy rates on D points across modalities might be related to factors intrinsic to the individual modalities which may operate differently in the three sensory domains. Comparing learning in different modalities through different experimental tasks should be taken cautiously (Abrahamse et al., 2009). Indeed, cross-modalities differences do not directly reflect learning differences but might be the result of a combination of learning and modality-dependent constraints. In our protocol, instead, we can investigate and compare learning across three different sensory modalities through a task that is as similar and directly comparable as possible. In fact, the only variable among the three tasks lies in the nature of the stimuli used to transmit the sequences of Fib. Apart from this, everything else is identical. By comparing the differences in the trajectory of RTs curves for D points and ND points between blocks, we have an index of learning that we can then compare across the three modalities, assessing the magnitudes. This would serve as a clear and accurate indicator of potential learning and any differences across different sensory spheres. As for accuracy, as explained in Section *5.2.*, it will not be taken as our primary indicator of learning. Instead, we will focus on reaction times to determine significance, as we believe they provide a more precise measure of learning effects. Indeed, given the simplicity of the task, we expect accuracy to be at ceiling, as observed by Schmid et al. (2023). Thus, we will not use accuracy results to determine significance. However, analyzing accuracy remains important for a comprehensive understanding of the results. High levels of inaccuracy could suggest random responses or a lack of careful task completion, identifying potential outliers. Furthermore, even though we anticipate very high accuracy levels, they may still undergo changes throughout the task, possibly revealing trends that either match or differ from the reaction time results. Should both reaction times and accuracy exhibit parallel trends (such as increased accuracy and decreased reaction times in case of learning), it would reinforce and validate the conclusions drawn from reaction times data. However, we must consider that factors like boredom or fatigue from the length of the task could also affect accuracy levels.

5.2.1. Data analysis plan

Considering the varying frequencies of 0s and 1s in the string (cf. Section *4.1.*), our approach involved not directly comparing 0s and 1s. Instead, we chose to examine them individually, assessing differences in reaction times (RTs) and accuracy rates at each level. The comparison specifically looked at points that could be predicted at a specific level (D points) and those that could not be predicted (ND points). To verify if there were learning differences between disambiguated (D) and non-disambiguated (ND) points at each level, we analyzed and compared RTs and accuracy rates in correspondence to every instance of these two points in each block, in the three modalities. Data were analyzed with a series of linear mixed effects regression models using lme4 and lmerTest (Bates et al., 2015; Kuznetsova et al., 2017) in R (R Core Team 2022).

To check whether the decrease of RTs for D and ND points was statistically significant, whether there were significant differences in the trend across blocks between the two types of point, and whether these differences were modulated by the specific modalities, we ran a series of Linear Mixed Models with *RTs* as dependent variable, *Block* (1-3), *Point_Level$_n$ $_{(1\text{-}6)}$*, and *Modality* (auditory, tactile, and visual) as independent variables, and *Subject* as random intercept. *Point_Level$_n$* is a discrete variable that contrasts disambiguated (D) and non-disambiguated (ND) points, differently operationalized depending on the level taken into consideration (1-6). As explained in Section *4.2.*, levels refer to the different hierarchical levels of Fib's structure. In this analysis, only correct responses were taken into consideration. For accuracy, we conducted a series of Generalized Mixed Models based on binomial distribution (Jaeger 2008) with *Accuracy* as dependent variable, *Block* (1-3), *Point_Level$_n$ $_{(1\text{-}6)}$*, and *Modality* (auditory, tactile, visual) as independent variable, and *Subject* as random intercept. Then, to unpack the significant interaction we found in both the analysis of RTs and accuracy rates, we ran post-hoc tests with Tukey correction of p-values (emmeans()-function in R).

The present analysis allows us to: (i) observe whether there are differences in the trends of RTs and accuracy rates across blocks between disambiguated (D) and non-disambiguated (ND) points within levels, in the three sensory modalities;

if participants manage to predict D points by exploiting the cognitive parsing strategy explained in Section *4.2.*, we expect to find a significative difference in the trend of RTs, possibly accompanied by a difference in accuracy rates across blocks between D and ND points, within levels (i.e., level 0 D vs. ND; level 1 D vs. ND; …). Specifically, we expect D points to have a steeper and/or earlier decrease in RTs across blocks, and shorter RTs than ND ones as soon as they are predicted. As for accuracy, as we explained, we expect it to be at ceiling throughout the task for both D and ND points due to the test's simplicity. However, performance improvements might be reflected not only in decreased reaction times but also in a slight increase in accuracy rates or a more pronounced progressive increase across blocks for D points compared to ND points. However, it is important to consider that other factors, such as boredom or fatigue due to the length of the task, might impact accuracy levels. For this reason, as we have explained, we will primarily consider RTs to determine significance.

Hence summing up, we expect the trend in RTs (and possibly accuracy rates) across blocks to be modulated by the type of point (D vs. ND); if this is the case, (ii) determine up to which level the difference holds; (iii) verify whether the different trend between D and ND points across blocks is modulated by the type of modality (auditory, visual, tactile), comparing their magnitudes. In other words, the present analysis will allow us to compare learning in the three sensory domains, by checking whether the changes in RTs (and possibly in accuracy rates) across blocks between D and ND points are modulated by the sensory modality and observing up to which level the potential difference holds.

5.3. Results

Analysis 1: Deterministic Vs. Non-Deterministic points within Level 0 in the Auditory, Tactile, and Visual studies

At Level 0, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this level, D points correspond to those 1 that follow 0 (0**1**); ND points to those 1 that follow 01 (01**1**). Results are reported in Table 1 and 4, respectively. As observable in Figure

43, RTs in the visual modality are considerably lower than those in the auditory and tactile ones. In all three modalities RTs for D points are significantly shorter than those for ND points, in every block. However, both those for D points and ND points decreased across blocks. Specifically, the slope for the former is steeper than that of the latter in the auditory and visual modalities, whereas, in the tactile modality, RTs for D and ND points diminished across blocks in a similar way. From the LMM analysis, we found a main effect of *Block* (χ2 =792.32, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Point_Level_0* (χ^2 =359.70, df = 1, *p* < .001), with participants being faster on disambiguated (D) than non-disambiguated (ND) points. *Modality* was significant (χ^2 =500.05, df = 2, *p* < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_0*Block* interaction was significant (χ^2 =103.78, df = 2, *p* = < .001), indicating that RTs across blocks were modulated by the type of point (D vs. ND). *Point_Level_0*Modality* was significant χ^2 =10.55, df = 2, *p* = < .01), meaning that the differences in RTs between D and ND points were modulated by the modality. *Block*Modality* was significant (χ^2 =232.90, df = 4, *p* = < .001), meaning that RTs across blocks were modulated by the type of modality. The interaction *Point_Level_0*Block*Modality* was also significant (χ^2 =40.70, df = 4, *p* = < .001): The difference in the trend of RTs between D and ND points across blocks were modulated by the modality. Post-hoc comparisons revealed a significant decrease in RTs for D points in the auditory modality from Block 1 to Block 2; from Block 1 to Block 3; from Block 2 to Block 3; RTs for ND points decreased significantly from Block 1 to Block 2, and from Block 1 to Block 3. RTs on D points were significantly faster than those on ND points in all three blocks. In the tactile modality, a significant decrease in RTs was found on both D and ND points from Block 1 to Block 2, from Block 1 to Block 3, from Block 2 to Block 3. RTs on D points were significantly shorter than those on ND points in all three blocks. In the visual modality, we found a significant decrease in RTs on D points from Block 1 to Block 2; from Block 1 to Block 3; from Block 2 to Block 3; On the contrary, ND points did not decrease significantly across blocks. Overall, these results indicate that, in line with our expectations, at Level 0 RTs for D points decreased along the blocks in all three modalities. The decrease occurred as early as the transition between Block 1 and Block 2 and then continued in Block 3, in all three modalities. Interestingly, in the auditory and tactile studies, the decrease in RTs also occurred for ND points (the decrease began already in the transition between the first and second block and continued in the third block), as opposed to the visual sphere, where ND points did not decrease. Crucially, however, in both the auditory and tactile spheres, from the pairwise comparison in post-hoc analyses between Block 1 and Block 3, we observed that ND points decreased to a lesser extent than D points. Moreover, RTs on D points were significantly faster than those on ND points in all three blocks, in all three modalities (see Table 2). These results clearly confirm the acquisition of the first-order transitional regularity according to which $p(0|1)=1$, excluding the hypothesis that the observed decrease in RTs on D points is attributable to an effect of habituation to the task (see *Data Analysis* Section). Indeed, not only D points displayed a decrease of RTs along the blocks, but we also found a difference in the trend of RTs between the set of D and ND points. Specifically, since we found significant lower RTs on D points than ND ones already in Block 1, we confirm that this regularity has been learnt within the first block of the task, in all three modalities. Crucially, moreover, it is important to highlight that we also found learning differences among modalities: As observed by comparing D points between Block 1 and Block 3 within modalities, the decrease was more pronounced in the auditory sphere (β=85.11; SE=3.12) than in the tactile (β=45.29; SE=2.92) and visual (β=27.14; SE=3.08) ones. Moreover, from pairwise comparisons between D and ND points in the third block (i.e., at the end of the task), we observe a greater difference between the two types of points in the auditory sphere (β=-122.90; SE=3.64) than in the tactile (β=-66.7; SE=3.43) and visual sphere (β=-73.85; SE=3.58). These findings are in line with our hypotheses: as observed in previous studies, the auditory domain displayed an advantage over the visual one in sequential statistical learning. In addition to this, our findings also suggest an advantage for the tactile sphere over the visual one. Despite these domain-specific learning differences, we observed that RTs in the visual study were overall lower than those in the auditory and tactile ones, in each block, for both D points and ND points (see Table 3). A potential explanation for this result may be attributed to a general processing advantage for the visual sphere, independent from learning, possibly linked to domain-internal factors such as faster communication channels connecting visual input processing and motor output.

	Block 1	Block 2	Block 3
RTs Disambiguated Points Auditory	622.17 (148.59)	560.38 (156.90)	538.84 (154.51)
RTs Non-Disambiguated Points Auditory	695.70 (117.56)	671.41 (135.64)	664.69 (129.40)
RTs Disambiguated Points Tactile	638.20 (108.52)	622.17 (119.78)	594.37 (110.80)
RTs Non-Disambiguated Points Tactile	695.51 (100.49)	682.07 (107.69)	660.90 (101.05)
RTs Disambiguated Points Visual	282.26 (108.52)	266.21 (103.30)	255.70 (114.99)
RTs Non-Disambiguated Points Visual	338.55 (98.21)	336.11 (96.17)	330.12 (100.57)

Table 1. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 0 in each Modality (Analysis 1).

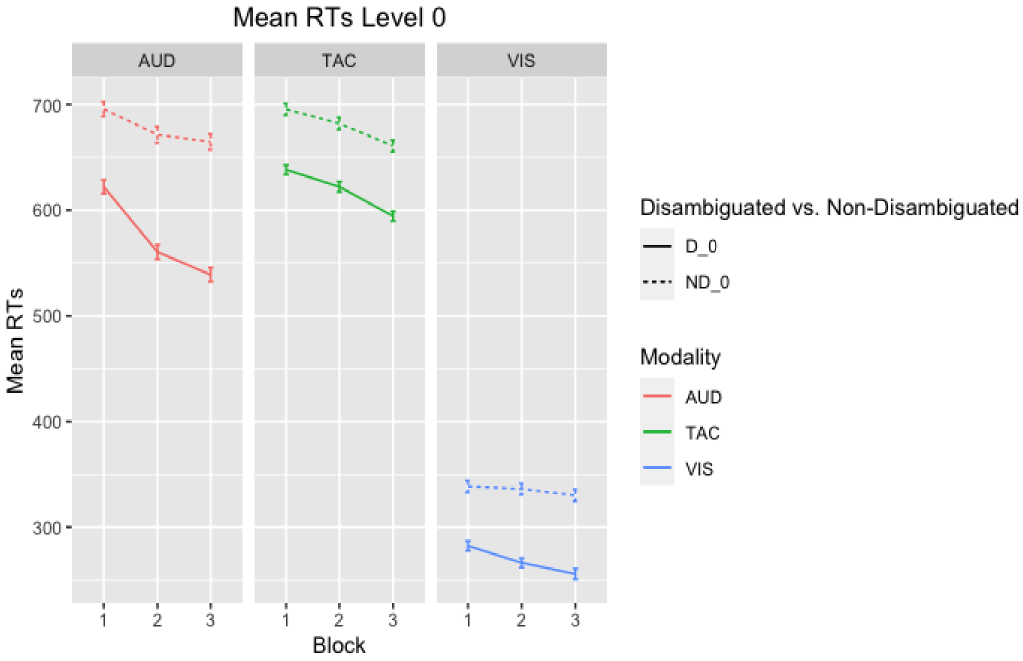


Figure 43. Mean RTs for D and ND points by block at Level 0 in the three studies (Analysis X). Error bars denote the 95% confidence interval. D_0 = Disambiguated points at Level 0; ND_0 = Non-Disambiguated points at Level 0; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	*SE*	*t*	*p*
Point_Level_0*Block| Auditory	Block 1 D – Block 2 D	61.75	3.13	19.73	<.0001
	Block 1 D – Block 3 D	85.11	3.12	27.30	<.0001
	Block 2 D – Block 3 D	23.36	3.11	7.52	<.0001
	Block 1 ND - block 2 ND	26.09	4.16	6.27	<.0001
	Block 1 ND – Block 3 ND	33.04	4.18	7.90	<.0001
	Block 1 D – Block 1 ND	-70.83	3.73	-18.97	<.0001
	Block 2 D – Block 2 ND	-106.49	3.63	-29.36	<.0001
	Block 3 D – Block 3 ND	-122.90	3.64	-33.78	<.0001
Point_Level_0*Block| Tactile	Block 1 D – Block 2 D	15.51	2.93	5.29	<.0001
	Block 1 D - Block 3 D	45.29	2.92	15.53	<.0001
	Block 2 D – Block 3 D	29.78	2.93	10.18	<.0001
	Block 1 ND – Block 2 ND	10.97	3.90	2.812	0.0136
	Block 1 ND - Block 3 ND	35.41	3.91	9.05	<.0001
	Block 2 ND – Block 3 ND	24.44	3.86	6.33	<.0001
	Block 1 D - block 1 ND	-56.9	3.47	-16.37	<.0001
	Block 2 D – Block 2 ND	-61.4	3.43	-17.92	<.0001
	Block 3 D - Block3 ND	-66.7	3.43	-19.48	<.0001
Point_Level_0*Block| Visual	Block 1 D – Block 2 D	16.81	3.07	5.47	<.0001
	Block 1 D - Block 3 D	27.14	3.08	8.80	<.0001
	Block 2 D – Block 3 D	10.33	3.09	3.34	0.0024
	Block 1 D - block 1 ND	-55.63	3.61	-15.42	<.0001
	Block 2 D – Block 2 ND	-68.72	3.56	-19.32	<.0001
	Block 3 D - Block3 ND	-73.85	3.58	-20.61	<.0001

Table 2. Summary of significant LMM coefficients and contrasts on RTs (Point_Level_0 * Block | Modality) (Analysis 1).

		β	*SE*	*t*	*p*
Modality*Block*Point_Level_0	Block 1 D AUD - VIS	342.75	18.4	18.67	<.0001
	Block 1 D TAC - VIS	359.15	98.3	20.16	<.0001
	Block 2 D AUD – TAC	-62.64	17.8	-3.51	0.0019
	Block 2 D AUD - VIS	297.80	18.4	16.23	<.0001
	Block 2 D TAC - VIS	360.44	17.8	20.23	<.0001
	Block 3 D AUD – TAC	-56.22	17.8	-3.15	0.0060
	Block 3 D AUD - VIS	284.78	18.4	15.52	<.0001
	Block 3 D TAC - VIS	341.00	17.8	19.14	<.0001
	Block 1 ND AUD - VIS	357.94	18.6	19.28	<.0001
	Block 1 ND TAC - VIS	360.38	18.0	20.00	<.0001
	Block 2 ND AUD - VIS	335.58	18.5	18.11	<.0001
	Block 2 ND TAC - VIS	353.13	18.0	19.62	<.0001
	Block 3 ND AUD - VIS	333.83	18.5	18.01	<.0001
	Block 3 ND TAC - VIS	333.90	18.0	18.55	<.0001

Table 3. Summary of significant LMM coefficients and contrasts on RTs (Modality * Block * Point_Level_0) (Analysis 1).

Summarizing, at Level 0, from the analysis of RTs we found that D points were learned in all three modalities, already in Block 1. Moreover, we found a domain-specific distinction in learning: in the auditory domain we found a significantly higher performance in learning compared to the tactile and visual domains. In turn, the tactile sphere proved to be superior to the visual sphere. Despite this learning advantage, we found that reaction times were consistently faster in the visual domain compared to auditory and tactile domains. This suggests a general processing advantage for visual information, independent from learning, possibly due to faster communication pathways between visual input and motor responses.

As for accuracy, as observable in Figure 44, D points were more accurate than ND points, in every block, in all the three modalities. In the auditory and tactile studies, the accuracy of D points increases across blocks, while that of ND ones decreases. In the visual study, on the other hand, that of both D and ND points decreases along the task, but the latter to a greater extent. The GLMM analysis revealed a main effect of *Block* ($\chi2 = 24.91$, df = 2, $p < .001$), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_0* was also significant ($\chi^2 = 23.94$, df = 1, $p < .001$), meaning that D points were significantly more accurate than ND points. *Modality* was significant ($\chi^2 = 32.15$, df = 2, $p < .001$), meaning that there were significant differences in accuracy rates in the three modalities. The *Point_Level_0*Block* interaction was significant ($\chi^2 = 19.40$, df = 2, $p < .001$), indicating that the trend for accuracy rates across blocks was modulated by the type of point. The *Block*Modality* interaction was significant ($\chi^2 =62.58$, df = 4, $p < .001$) too, meaning that the trend for accuracy rates across blocks was modulated by the modality. The interaction *Point_Level_0*Block*Modality* was significant ($\chi^2 =13.75$, df = 4, $p < .01$): the difference in the trend of accuracy rates between D and ND points across blocks was modulated by the modality. In the auditory modality, post-hoc comparisons showed a significant increase in accuracy on D points from Block 1 to Block 3; from Block 2 to Block 3. On the contrary, accuracy rates on ND points did not change significantly across blocks. In the tactile modality, post-hoc comparisons reported a decrease in accuracy rates on D points from Block 1 to Block 2, followed by a significant increase from Block 2 to Block 3. On the contrary, accuracy on ND points decreased significantly from Block 1 to Block 3. In the visual domain, results indicated a significant decrease in accuracy rates on both D and ND points from Block 1 to Block 2, from Block 1 to Block 3. This could have been caused by a fatigue effect, linked to increased cognitive effort associated with learning this regularity in the visual sphere. Crucially, however, in both the auditory, tactile, and visual studies, we found that accuracy rates on D points were higher than those on ND points in all three blocks (see Table 5). Overall, these results are in line with our hypothesis and confirm what we have found for RTs: at Level 0, D points are processed differently from ND points. Specifically, concerning accuracy, an advantage of D points over ND points is observed. This advantage is evident both in terms of an increase in the accuracy rates of the former in the presence of a decrease (as in the tactile modality) or a stable trend (as in the auditory modality) in the latter's accuracy, and in the overall higher accuracy rates of D points compared to ND points in all blocks and across all three modalities. Despite this, post-hoc comparisons also indicated differences between modalities. Specifically, accuracy rates in the tactile domain were higher than those in the auditory domain on D points in Block 1. Accuracy rates in the visual modality were higher than those in the auditory and tactile ones on D points in Block 1 and on ND points in Block 1 and 2. They were higher than those in the auditory modality on D points in Block 2 and on ND points in Block 3 (see Table 6). Hence, overall, these results do confirm the presence of domain-specific differences in sequential statistical learning abilities. Specifically, the auditory and tactile domain proved to be superior to the visual sphere. Crucially, moreover, the fact that accuracy rates on D points were higher in the visual sphere than tactile and auditory ones in the first block, is also in line with results on RTs. Specifically, it reinforces the hypothesis of a general processing advantage (e.g., superior communication channels connecting visual input processing and motor output), independent from learning, for the visual sphere. Indeed, in this specific case, the visual sphere turned out to be more accurate than the auditory and tactile ones only at the beginning of the task, when the learning effects were still in their early stages and had not yet become prominently evident.

	Block 1	Block 2	Block 3
Accuracy Disambiguated Points Auditory	0.94 (0.23)	0.96 (0.19)	0.97 (0.15)
Accuracy Non-Disambiguated Points Auditory	0.90 (0.30)	0.91 (0.29)	0.89 (0.31)
Accuracy Disambiguated Points Tactile	0.97 (0.17)	0.95 (0.21)	0.97 (0.16)
Accuracy Non-Disambiguated Points Tactile	0.92 (0.26)	0.90 (0.30)	0.90 (0.31)
Accuracy Disambiguated Points Visual	0.99 (0.08)	0.98 (0.14)	0.97 (0.17)
Accuracy Non-Disambiguated Points Visual	0.98 (0.14)	0.96 (0.20)	0.94 (0.23)

Table 4. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 0 in each Modality (Analysis 1).

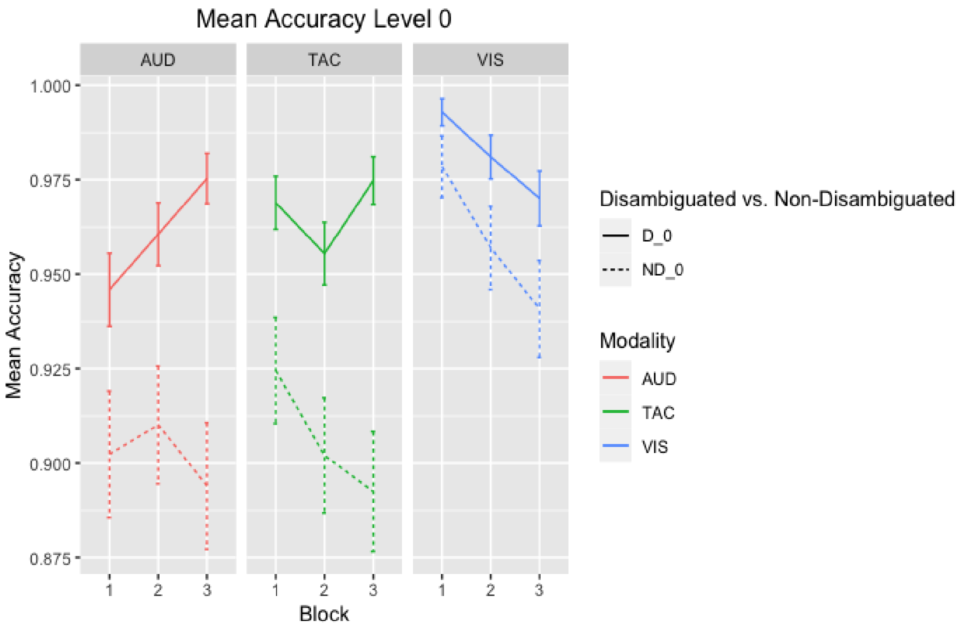


Figure 44. Mean accuracy rates for D and ND points by block at Level 0 in the three studies (Analysis 1). Error bars denote the 95% confidence interval. D_0 = Disambiguated points at Level 0; ND_0 = Non-Disambiguated points at Level 0; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	*SE*	*t*	*p*
Point_Level_0*Block| Auditory	Block 1 D – Block 3 D	-0.85	0.17	-4.96	<.0001
	Block 2 D – Block 3 D	-0.50	0.18	-2.77	0.0155
	Block 1 D - block 1 ND	0.69	0.14	4.89	<.0001
	Block 2 D – Block 2 ND	0.94	0.15	6.20	<.0001
	Block 3 D - Block3 ND	1.64	0.17	9.68	<.0001
Point_Level_0*Block| Tactile	Block 1 D – Block 2 D	0.40	0.16	2.53	0.0304
	Block 2 D – Block 3 D	-0.63	0.17	-3.75	0.0005
	Block 1 ND – Block 3 ND	0.45	0.14	3.19	0.004
	Block 1 D - block 1 ND	1.02	0.16	6.30	<.0001
	Block 2 D – Block 2 ND	0.95	0.14	6.82	<.0001
	Block 3 D - Block3 ND	1.70	0.16	10.54	<.0001
Point_Level_0*Block| Visual	Block 1 D – Block 2 D	1.01	0.30	3.36	0.0022
	Block 1 D - Block 3 D	1.49	0.28	5.22	<.0001
	Block 2 D – Block 3 D	0.48	0.20	2.36	0.0484
	Block 1 ND – Block 2 ND	0.74	0.24	3.09	0.0056
	Block 1 ND - Block 3 ND	1.10	0.23	4.77	<.0001
	Block 1 D - block 1 ND	1.14	0.32	3.54	0.0004
	Block 2 D – Block 2 ND	0.88	0.21	4.17	<.0001
	Block 3 D – Block 3 ND	0.75	0.17	4.29	<.0001

Table 5. Summary of significant GLMM coefficients and contrasts on Accuracy (Point_Level_0 * Block | Modality) (Analysis 1).

		β	*SE*	*t*	*p*
Modality*Block*Point_Level_0	Block 1 D AUD – TAC	-0.86	0.31	-2.72	0.0177
	Block 1 D AUD - VIS	-2.22	0.39	-5.64	<.0001
	Block 1 D TAC - VIS	-1.37	0.40	-3.45	0.0016
	Block 2 D AUD - VIS	-0.86	0.34	-2.51	0.0327
	Block 1 ND AUD - VIS	-1.77	0.36	-4.93	<.0001
	Block 1 ND TAC - VIS	-1.25	0.36	-3.50	0.0013
	Block 2 ND AUD - VIS	-0.92	0.33	-2.81	0.0139
	Block 2 ND TAC - VIS	-0.83	0.32	-2.58	0.0265
	Block 3 ND AUD - VIS	-0.77	0.32	-2.40	0.0428

Table 6. Summary of significant GLMM coefficients and contrasts on Accuracy (Modality * Block * Point_Level_0) (Analysis 1).

Summarizing, at Level 0, the analysis of accuracy rates confirmed what has been found for RTs. D points have been processed differently than ND points, in all three modalities, and the difference was observable already in Block 1. Overall, these results confirm the acquisition of D points at Level 0 already within the first block. Furthermore, we confirm the disadvantage of the visual sphere in acquiring this sequential statistical regularity compared to the auditory and tactile spheres.

Analysis 2: Deterministic Vs. Non-Deterministic points within Level 1 in the Auditory, Tactile, and Visual studies

At Level 1, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this level, D points correspond to those 0 that follow 11 (11**0**); ND points to those 0 that follow 01 (01**0**). Results are reported in Table 7 and 10, respectively. As observable in Figure 45, RTs in the visual modality are considerably shorter than those in the auditory and tactile ones. In all three modalities, RTs for D points are shorter than those for ND points, in every block, except in Block 1 in the visual modality, where the RTs for D points are slightly higher than those for ND points. In the auditory and tactile modalities, RTs for both D and ND points decrease across blocks. Despite this, the slope for the former is steeper than that of the latter, especially in the auditory modality. Instead, in the visual mode, only RTs for D points descend along blocks, as opposed to those for NDs, for which no substantial descent occurs. From the LMM analysis, we found a main effect of *Block* (χ2 =582.59, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Point_Level_1* (χ^2 =165.0, df = 1, *p* < .001), with participants being faster on disambiguated (D) than non-disambiguated (ND) points. *Modality* was significant (χ^2 =466.31, df = 2, *p* < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_1*Block* interaction was significant (χ^2 =76.38, df = 2, *p* = < .001), indicating that RTs across blocks were modulated by the type of point (D vs. ND). *Point_Level_1*Modality* was significant (χ^2 =111.16, df = 2, *p* = < .001), meaning that the differences in RTs between D and ND points were modulated by the modality. *Block*Modality* was significant (χ^2 =214.69, df = 4, *p* = < .001), meaning that RTs across blocks were modulated by the type of modality. The interaction *Point_Level_1*Block*Modality* was significant (χ^2 =22.23, df = 4, *p* = < .001), meaning that the differences in the trend of RTs between D and ND points across blocks were modulated by the modality. In the auditory modality, post-hoc comparisons revealed a significant decrease in RTs for D points from Block 1 to Block 2; from Block 1 to Block 3; from Block 2 to Block 3. In the tactile modality, we found a significant decrease in RTs on D points from Block 1 to Block 2, from Block 1 to Block 3, from Block 2 to Block 3. In the visual modality, we found a significant decrease in RTs on D points from Block 1 to Block 2; from Block 1 to Block 3 (see Table 8). Crucially, however, as observable from pairwise comparisons between Block 1 and Block 3, the decrease was sharper in the auditory modality than the tactile and visual one (Table 8), confirming a learning advantage for this sensory domain over the tactile and visual ones. In the auditory and tactile studies, RTs on ND points decreased as well. However, in the former, the decrease was significant already in the passage from Block 1 to Block 2, whereas in the latter we found a significant decrease only from Block 1 to Block 3. Moreover, as observable from pairwise comparisons between Block 1 and Block 3, the decrease occurred to a greater extent in the auditory domain than in the tactile one (Table 8). In contrast, ND points did not decrease in the visual domain. RTs for D points were lower than those for ND points in all three blocks in the auditory and tactile spheres, while in the visual sphere the difference occurred only in blocks 2 and 3. Moreover, as we observed from pairwise comparison between D and ND points in the third block, the difference between the two types of point was greater in the auditory sphere than in the tactile and visual spheres (Table 8). Overall, these results are in line with our predictions and with results observed at Level 0. Specifically, they indicate that D points at Level 1 were learned in all three sensory domains, but with domain-specific differences: from the trend of RTs on D points, we observed that the auditory modality displayed an advantage over the tactile and visual ones. Moreover, the tactile domain displayed better performances compared to the visual one. Importantly, D points were learned earlier in the auditory and tactile modalities (Block 1) than the visual modality (Block 2), as observed comparing D and ND points within blocks. Comparing the three modalities, we also found that RTs in the auditory modality were significantly faster than those in the tactile modality on D points in Block 2 and 3. RTs in the tactile modality were faster than those in the auditory modality on ND points in Block 1. Moreover, as observed at the previous level, RTs in the visual study were overall shorter than those in the auditory and tactile study, for both D and ND points, in all blocks (see Table 9). This suggests a general processing advantage, independent from learning, for the visual sphere.

	Block 1	Block 2	Block 3
RTs Disambiguated Points Auditory	654.91 (145.48)	583.23 (153.46)	567.32 (150.81)
RTs Non-Disambiguated Points Auditory	715.29 (118.79)	700.65 (120.79)	677.96 (122.47)
RTs Disambiguated Points Tactile	653.59 (102.12)	635.53 (107.39)	617.09 (110.45)
RTs Non-Disambiguated Points Tactile	665.79 (98.02)	662.94 (99.49)	654.00 (102.08)
RTs Disambiguated Points Visual	315.54 (103.54)	301.46 (106.25)	296.42 (117.65)
RTs Non-Disambiguated Points Visual	309.57 (97.48)	317.14 (106.11)	309.85 (104.70)

Table 7. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 1 in each Modality (Analysis 2).

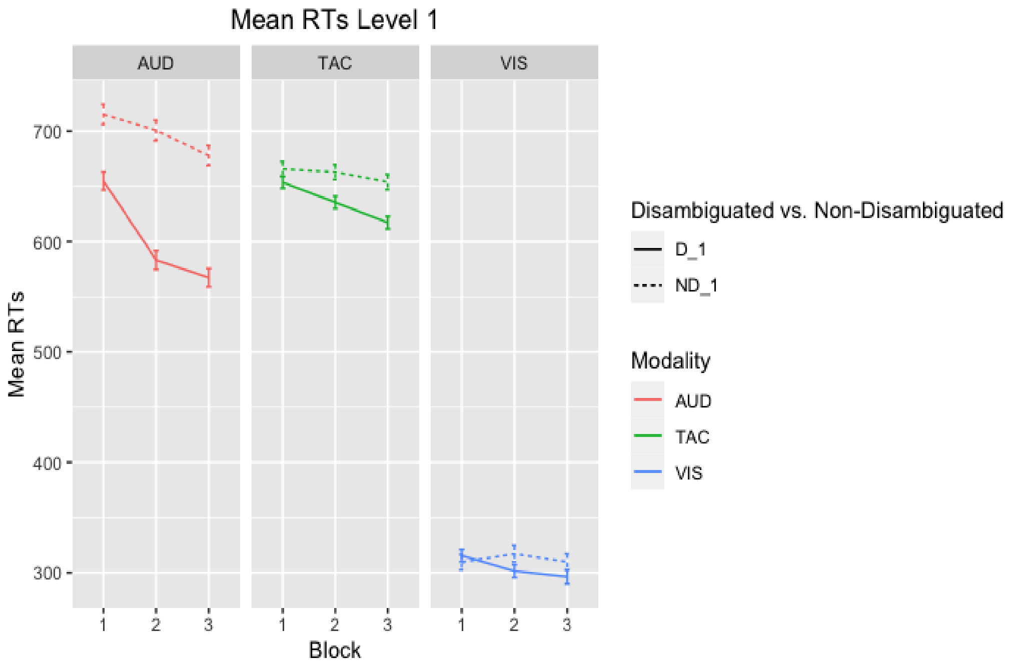


Figure 45. Mean RTs for D and ND points by block at Level 1 in the three studies (Analysis 2). Error bars denote the 95% confidence interval. D_1 = Disambiguated points at Level 1; ND_1 = Non-Disambiguated points at Level 1; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	***SE***	***t***	***p***
Point_Level_1*Block| Auditory	Block 1 D – Block 2 D	70.65	3.84	18.38	<.0001
	Block 1 D – Block 3 D	87.17	3.82	22.80	<.0001
	Block 2 D – Block 3 D	16.51	3.81	4.34	<.0001
	Block 1 ND - block 2 ND	19.71	5.32	3.71	0.0006
	Block 1 ND – Block 3 ND	39.20	5.22	7.51	<.0001
	Block 2 ND – Block 3 ND	19.49	5.21	3.74	0.0005
	Block 1 D – Block 1 ND	-59.92	4.65	-12.88	<.0001
	Block 2 D – Block 2 ND	-110.87	4.63	-23.96	<.0001
	Block 3 D – Block 3 ND	-107.90	4.50	-24.00	<.0001
Point_Level_1*Block| Tactile	Block 1 D – Block 2 D	17.28	3.59	4.82	<.0001
	Block 1 D - Block 3 D	36.02	3.58	10.05	<.0001
	Block 2 D – Block 3 D	18.74	3.60	5.21	<.0001
	Block 1 ND - Block 3 ND	12.99	4.66	2.79	0.0147
	Block 1 D - block 1 ND	-12.15	4.16	-2.92	0.0035
	Block 2 D – Block 2 ND	-27.24	4.16	-6.54	<.0001
	Block 3 D - Block3 ND	-35.18	4.15	-8.47	<.0001
Point_Level_1*Block| Visual	Block 1 D – Block 2 D	14.21	3.76	3.78	0.0005
	Block 1 D - Block 3 D	19.04	3.77	5.05	<.0001
	Block 2 D – Block 2 ND	-14.51	4.33	-3-35	0.0008
	Block 3 D - Block3 ND	-12.44	4.33	-2.87	0.0041

Table 8. Summary of significant LMM coefficients and contrasts on RTs (Point_Level_1 * Block | Modality) (Analysis 2).

		β	*SE*	*t*	*p*
Modality*Block*Point_ Level_1	Block 1 D AUD - VIS	339.60	18.4	18.49	<.0001
	Block 1 D TAC - VIS	340.00	17.8	19.07	<.0001
	Block 2 D AUD – TAC	-53.77	17.8	-3.01	0.0090
	Block 2 D AUD - VIS	288.16	18.4	15.42	<.0001
	Block 2 D TAC - VIS	336.94	17.8	18.90	<.0001
	Block 3 D AUD – TAC	-51.55	17.8	-2.89	0.0129
	Block 3 D AUD - VIS	271.47	18.4	14.79	<.0001
	Block 3 D TAC - VIS	323.03	17.8	18.12	<.0001
	Block 1 ND AUD - TAC	47.37	18.1	2.61	0.0277
	Block 1 ND AUD - VIS	405.01	18.7	21.68	<.0001
	Block 1 ND TAC - VIS	357.65	18.1	19.77	<.0001
	Block 2 ND AUD - VIS	379.51	18.7	20.33	<.0001
	Block 2 ND TAC - VIS	349.66	18.1	19.34	<.0001
	Block 3 ND AUD - VIS	366.93	18.6	19.69	<.0001
	Block 3 ND TAC - VIS	345.77	18.1	19.13	<.0001

Table 9. Summary of significant LMM coefficients and contrasts on RTs (Modality * Block * Point_Level_1) (Analysis 2).

Summarizing, at Level 1, we found that RTs for D points decreased across blocks in all modalities. The decrease was already observable in the transition from the first to the second block and then continued in the third block, indicating that the second-order transitional regularity according to which *p*(0|11)=1 was learnt in the initial blocks of the task. Specifically, as observed comparing D and ND points within modalities, D points at Level 1 were learnt already in Block 1 in the auditory and tactile modalities, whereas in Block 2 in the visual modality. The combination of this observation with the trends observed in the RTs curve for D points in all three modalities confirms that the auditory sphere outperformed the tactile and visual spheres. Additionally, the tactile sphere surpassed the visual sphere in learning this sequential statistical regularity. This is in line with what has been found at Level 0. However, as observed at L0, also at L1 we consistently observed overall faster reaction times in the visual domain than in the auditory and tactile domains. This suggests a broader processing advantage for visual information, unrelated to learning, possibly stemming from quicker communication pathways between visual input and motor responses.

As for accuracy, as observable in Figure 46, D points were more accurate than ND points, in every block, in all the three modalities. In the auditory study, the accuracy of both D and ND points increases across blocks. On the contrary, in the tactile and visual studies, accuracy rates slightly decrease along the task. The GLMM model failed to converge, indicating difficulties in obtaining a satisfactory estimation of the model parameters. The inability to achieve convergence implies that the estimated parameter values may not be reliable or reflective of the true underlying relationships within the data. The present convergence issue may stem from the complexity of the model. For this reason, we proceeded by (i) investigating the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to check whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_1* (Disambiguated at level 1 vs. Non-disambiguated at Level 1) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* ($\chi2$ =16.67, df = 2, p < .001), indicating the presence of significantly different accuracy rates between modalities. *Point_Level_1* was also significant (χ^2 =255.67, df = 1, *p* < .001), meaning that D points were significantly more accurate than ND points. The *Point_Level_1*Modality* interaction was significant (χ2 =41.57, df = 2, p <.001): the difference in accuracy rates between modalities was modulated by the type of point. Post-hoc comparisons showed that D points in the visual modality were more accurate than D points in the auditory modality. ND points in the visual modality were more accurate than ND points in the tactile and auditory modalities. ND points in the tactile modality were more accurate than ND points in the auditory modality (see Table 12). For analysis (ii), we ran separated models, splitting data according to *Modality*. Hence, the effect of *Point_Level_1* and *Block* was investigated in the three separated datasets (Auditory, Tactile, Visual). To check whether there were differences in accuracy rates between disambiguated and non-disambiguated points at level 1 within the three modalities, we conducted three GLMM models (one in each modality) with *Accuracy* as dependent variable, *Block* (1-3) and *Point_Level_1* (Disambiguated at level 1 vs. Non-disambiguated at Level 1) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory modality, the GLMM model revealed a main effect of *Block* (χ2 =22.093, df = 2, p < .001), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_1* was also significant (χ^2 =63.417, df = 1, *p* < .001), meaning that D points were significantly more accurate than ND points (96% vs. 83%). The *Point_Level_1*Block* interaction was significant (χ2 =8.994, df = 2, p <.05). Post-hoc comparisons reported a significant increase in accuracy on D points from Block 1 to Block 3; from Block 2 to Block 3. Accuracy rates on ND points increased significantly from Block 2 to Block 3. Accuracy rates on D points were significantly higher than those on ND points in all three blocks (See Table 11).

In the tactile modality, the analysis revealed a main effect of *Block* (χ2 =6.27, df = 2, p < .05), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_1* was also significant (χ^2 = 5.32, df = 1, *p* < .05), meaning that D points were significantly more accurate than ND points (96% vs. 93%). The *Point_Level_1*Block* interaction was not significant, indicating that the trend of accuracy rates across blocks was not modulated by the type of point. Being the interaction not significant, we ran a second GLMM with accuracy as dependent variable, *Block* (1-3) and *Point_Level_1* (Disambiguated at level 1 vs. Non-disambiguated at Level 1) as independent predictors, and *Subject* as random intercept. We found a main effect of *Block* (χ2 =16.48, df = 2, p < .001), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_1* was significant (χ^2 = 33.217, df = 1, p< .001), indicating that D points were significantly more accurate than ND ones. We then compared the two models with the anova()-function in R, without finding any significant result. Therefore, we failed to reject the null hypothesis, meaning that the two models did not differ. We ran post-hoc tests on the simpler model ((accuracy ~ Point_Level_1+Block + (1 |Subject)). Results indicated a significant decrease in accuracy rates from Block 1 to Block 2, and from Block 1 to Block 3. Accuracy on D points was higher than that on ND points (see Table 11).

In the visual modality, the analysis revealed a main effect of *Block* (χ2 =14.43, df = 2, p < .001), indicating that accuracy decreased across blocks. *Point_Level_1* was significant (χ^2 = 9.61, df = 1, *p* < .01), indicating that D points were significantly more accurate than ND points (98% vs. 97%). The *Point_Level_1*Block* interaction was also significant (χ^2 = 8.26, df = 2, *p* < .05): accuracy rates across blocks were modulated by the type of point. Post-hoc comparisons showed that accuracy rates for D points decreased significantly from Block 1 to Block 3, whereas those for ND points did not change significantly across blocks. Accuracy rates on D points were significantly higher than those on ND points in Block 1 and Block 2 (See Table 11).

	Block 1	Block 2	Block 3
Accuracy Disambiguated Points Auditory	0.94 (0.24)	0.95 (0.21)	0.98 (0.15)
Accuracy Non-Disambiguated Points Auditory	0.83 (0.38)	0.80 (0.40)	0.87 (0.34)
Accuracy Disambiguated Points Tactile	0.97 (0.17)	0.95 (0.21)	0.96 (0.19)
Accuracy Non-Disambiguated Points Tactile	0.95 (0.21)	0.92 (0.27)	0.93 (0.26)
Accuracy Disambiguated Points Visual	0.99 (0.09)	0.99 (0.12)	0.97 (0.16)
Accuracy Non-Disambiguated Points Visual	0.97 (0.145	0.97 (0.18)	0.97 (0.16)

Table 10. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 1 in each Modality (Analysis 2).

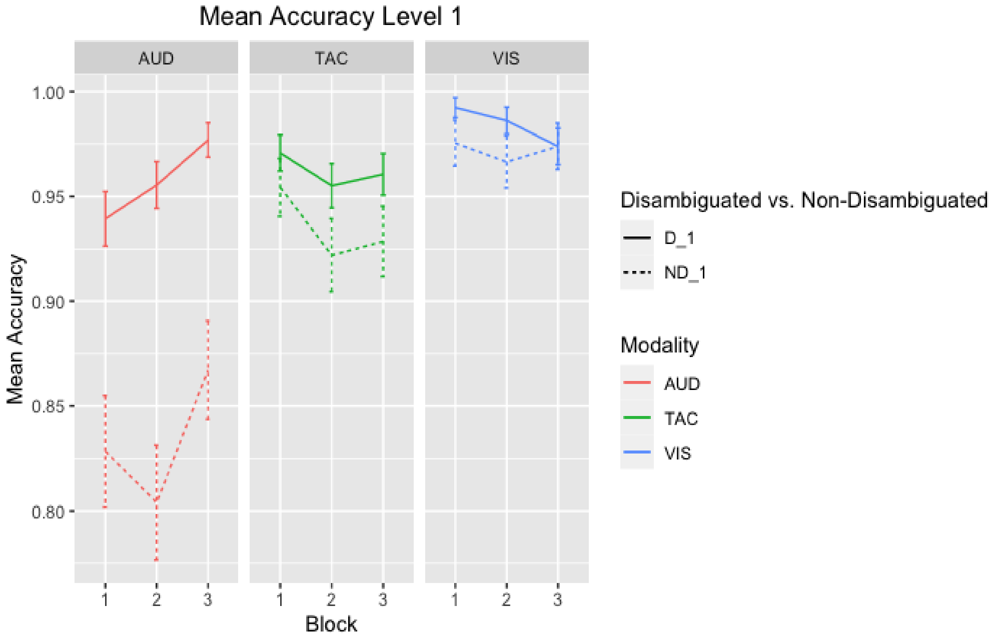


Figure 46. Mean accuracy rates for D and ND points by block at Level 1 in the three studies (Analysis 2). Error bars denote the 95% confidence interval. D_1 = Disambiguated points at Level 1; ND_1 = Non-Disambiguated points at Level 1; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	*SE*	*t*	*p*
Point_Level_1*Block| Auditory	Block 1 D – Block 3 D	-1.02	0.22	-4.69	<.0001
	Block 2 D – Block 3 D	-0.69	0.23	-3.03	0.0070
	Block 2 ND – Block 3 ND	-0.49	0.14	-3.53	0.0012
	Block 1 D - block 1 ND	1.21	0.15	7.96	<.0001
	Block 2 D – Block 2 ND	1.72	0.16	10.55	<.0001
	Block 3 D - Block3 ND	1.92	0.21	9.02	<.0001
Point_Level_1+Block| Tactile	Block 1 – Block 2	0.62	0.15	3.98	0.0002
	Block 1 – Block 3	0.47	0.16	3.02	0.0071
	D - ND	0.7	0.12	5.76	<.0001
Point_Level_1*Block| Visual	Block 1 D – Block 3 D	1.26	0.35	3.56	0.0011
	Block 1 D - block 1 ND	1.19	0.38	3.10	0.0019
	Block 2 D – Block 2 ND	0.92	0.30	3.05	0.0022

Table 11. Summary of significant GLMM coefficients and contrasts on Accuracy (Point_Level_1 * Block |Auditory; Point_Level_1 + Block |Tactile; Point_Level_1 * Block |Visual;) (Analysis 2).

		β	*SE*	*z*	*p*
Modality *Point_Level_1	D AUD - VIS	-1.21	0.30	-4.07	0.0001
	ND TAC - AUD	-1.45	0.26	-5.48	<.0001
	ND AUD - VIS	-2.17	0.29	-7.51	<.0001
	ND TAC - VIS	-0.72	0.29	-2.45	0.0379

Table 12. Summary of significant GLMM coefficients and contrasts on Accuracy (Modality * Point_Level_1) (Analysis 2).

Summarizing, regarding accuracy, at Level 1 we observed that only in the auditory sphere accuracy rates on both D points and ND ones increase along blocks. In contrast, in the tactile and visual studies, accuracy on D points decreased. Once again, in line with previous analyses, this supports the primacy of the auditory sphere in sequential statistical learning. Interestingly, however, in all three modalities we found that accuracy rates on D points were overall higher than those on ND ones, confirming that the two types of points were processed differently.

Analysis 3: Deterministic Vs. Non-Deterministic points within Level 2 in the Auditory, Tactile, and Visual studies

At Level 2, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this level, D points correspond to those 1 that follow the chunks [01][01]; ND points to those 1 that follow [1][01]. According to the cognitive parsing strategy outlined in Section *4.2.*, in order to predict D points at Level 2, it is necessary to have first learned the first- and second-order regularities corresponding to D points at Level 0 and Level 1. In fact, as we have explained, the parser must first form two categories of points, create chunks by leveraging the first-order transitional regularity acquired at Level 0, and track distributional information between chunks by exploiting the second-order regularity acquired at Level 1. Therefore, learning to predict D points at Level 2 is computationally more complex than learning sequential statistical information at previous levels. It requires a higher degree of abstraction and the projection of the acquired sequential statistical information (D points at Level 0 and Level 1) onto the hierarchical axis (cf. Section *4.2.*). Having acquired both D points at Level 0 and Level 1 in all three modalities, the parser potentially possesses the necessary information to proceed with hierarchical learning at Level 2. Consistent with previous studies (cf. Section *3.1.6.*), we anticipate that D points at Level 2 will be learned in the auditory domain. Conversely, we do not have specific expectations regarding the tactile and visual domains. To our knowledge, no study has investigated the learning of recursive hierarchical structures in the tactile domain. Furthermore, studies examining hierarchical recursive learning in the visual domain have focused on the acquisition of recursive structures arising from spatially distributed stimuli. In contrast, we lack information on the ability to acquire recursive hierarchical structures arising from sequential input in the visual domain (cf. Section *3.1.6.*; *3.2.*). However, since this ability is closely tied to sequential statistical learning, and considering that previous studies have demonstrated the auditory domain to outperform the visual domain in this regard, we expect that if D points are learned in the visual domain as well, the learning performance will be superior in the auditory domain (in terms of steeper RTs curves decreasing along blocks and/or early acquisition of the regularity within blocks).

Results are reported in Table 13 (RTs) and Table 16 (accuracy). As observable in Figure 47, RTs in the visual modality are considerably shorter than those in the auditory and tactile ones. Moreover, we can observe that in all three modalities, RTs for D points decrease across the blocks. From the LMM model, we found a main effect of *Block* (χ2 =95.54, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Point_Level_2* (χ^2 =4.38, df = 1, *p* < .05), with participants being faster on disambiguated (D) than non-disambiguated (ND) points. *Modality* was significant (χ^2 =552.38, df = 2, *p* < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_2*Block* interaction was significant (χ^2 =18.90, df = 2, *p* = < .001), indicating that RTs across blocks were modulated by the type of point (D vs. ND). *Block*Modality* was significant (χ^2 =34.76, df = 4, *p* = < .001), meaning that RTs across blocks were modulated by the type of modality. However, the interaction *Point_Level_2*Block*Modality* was not significant, meaning that the differences in the trend of RTs between D and ND points across blocks were not modulated by the modality. To explore the nature of the interactions more finely, first (i) we assessed the interaction between type of point and modality to verify whether the differences in RTs between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to check whether the trend of RTs across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a LMM model with *RTs* as dependent variable, *Point_Level_2* (Disambiguated at level 2 vs. Non-disambiguated at Level 2) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of Modality (χ2 =538.10, df = 2, p < .001), indicating the presence of significantly different RTs between modalities. *Point_Level_2* was also significant (χ2 =6.64, df = 1, p < .01), meaning that D points were significantly faster than ND points. The *Point_Level_2*Modality* interaction was significant (χ2 =9.54, df = 2, p <.01): The difference in RTs between modalities was modulated by the type of point. Post-hoc comparisons showed that both D and ND points were faster in the visual modality than in the auditory and tactile ones (Table 15). This result is in line with previous findings (analyses at Level 0 and Level 1), indicating a general processing advantage, independent from learning, for the visual modality over the auditory and tactile modalities. To conduct analysis (ii), we subsequently splitted data according to *Modality*. The effect of *Point_Level_2* and *Block* was investigated in the three separated datasets (Auditory, Tactile, Visual). To verify if there were learning differences between disambiguated and non-disambiguated points at level 2 in each modality, we ran three (one for each modality) LMM models with *RTs* as dependent variable, *Block* (1-3) and *Point_Level_2* (Disambiguated at level 2 vs. Non-disambiguated at Level 2) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory modality, we found a main effect of *Block* (χ2 = 68.34, df = 2, p < .001), indicating that RTs became shorter across blocks. *Point_Level_2* was not significant (χ^2 = 3.20, df = 1, *p* =0.073), indicating the absence of significant differences between disambiguated points (D) and non-disambiguated (ND) points (672.38ms vs. 684.67ms, respectively). The *Point_Level_2*Block* interaction was significant (χ^2 = 13.81, df = 2, *p* = < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant decrease in RTs for D points from Block 1 to Block 2; from Block 1 to Block 3. On the contrary, RTs for ND points did not significantly decrease. RTs on D points were significantly faster than those on ND points in Block 2 and Block 3 (see Table 14). Since RTs on D points decreased in the passage from Block 1 and Block 2 and the difference between D and ND points became evident in Block 2, we conclude that D points at Level 2 were learnt within Block 2. This is in line with our hypothesis: Being D points at Level 2 computationally more complex than D points at Level 0 and Level 1, they were learned later in the task. Moreover, this result is also in line with our cognitive parsing algorithm hypothesis. Indeed, as discussed in Section *4.2.*, in order to predict D points at Level 2, the parser would need to have previously acquired the first- and second-order regularities corresponding to D points at Level 0 and 1, respectively.

From the LMM analysis in the tactile modality, we found a main effect of *Block* (χ2 =103.86, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Point_Level_2* (χ^2 =5.05, df = 1, *p* < .05), with participants being faster on non-disambiguated (ND) than disambiguated (D) points (675.68 ms vs. 681.40 ms, respectively). The *Point_Level_2*Block* interaction was not significant, indicating that RTs across blocks were not modulated by the type of point (D vs. ND). Being the interaction not significant, we ran a second LMM with RTs as dependent variable, *Block* (1-3) and *Point_Level_2* (Disambiguated at level 2 vs. Non-disambiguated at Level 2) as independent predictors, and *Subject* as random intercept. We found a main effect of *Block* (χ2 =129.514, df = 2, p < .001), indicating a significant decrease in RTs across blocks. *Point_Level_2* was marginally significant (χ^2 =3.856, df = 1, p= .049), indicating that ND points were faster than D ones. We then compared the two models with the anova()-function in R, without finding any significant result. Therefore, we failed to reject the null hypothesis, meaning that the two models did not differ. We ran post-hoc tests on the simpler model ((RTs~ Point_Level_2+Block + (1 |Subject)). Results indicated a significant decrease in RTs from Block 1 to Block 2, from Block 1 to Block 3, and from Block 2 to Block 3 (see Table 14). From this result, we conclude that D (and ND) points were learned in the tactile modality, specifically, in Block 2.

From the LMM analysis in the visual modality, we found a main effect of *Block* (χ2 =16.28, df = 2, p < .001), indicating that RTs significantly decreased across blocks. *Point_Level_2* was also significant (χ^2 =4.58, df = 1, *p* < .05), indicating that D points were overall faster than ND ones (334.08 ms vs. 336.25 ms). The *Point_Level_2*Block* interaction was significant (χ^2 =10.83, df = 2, *p* < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant decrease in RTs on D points from Block 1 to Block 3. On the contrary, ND points did not decrease significantly across blocks. RTs for ND points were faster than D points in Block 1, whereas the opposite was observed in Block 3, where RTs for D points were significantly shorter than those for ND points (Table 14). Hence, from these results, we confirm that D points at Level 2 were acquired in the visual modality, specifically in Block 3.

Overall, at Level 2, we found that D points were learned in all three modalities. However, notable domain-specific differences emerged: in the auditory and tactile modalities, acquisition occurred in Block 2, while in the visual modality, it took place in Block 3. Furthermore, when comparing the magnitude of the decrease in reaction times (RTs) from Block 1 to Block 3 across the sensory domains, we noted that the decline in RTs on D points was more pronounced in the auditory sphere compared to the tactile and visual spheres (Table 14). Upon closer examination of the graph and RTs data for D points in the tactile and auditory domains, it becomes evident that RTs decreased to a similar extent across the blocks, showing comparable trends in Block 1 and Block 3. The key distinction lies in the fact that in the auditory domain, the decrease is more substantial already from Block 1 to Block 2, while in the tactile domain, it occurs more prominently from Block 2 to Block 3. These results align with our initial hypotheses regarding the auditory domain's superiority over the visual one. Interestingly, we also discovered that the tactile domain exhibits an advantage over the visual one in sequential hierarchical learning, displaying a trend similar to that observed in the auditory domain.

	Block 1	Block 2	Block 3
RTs Disambiguated Points Auditory	699.74 (122.80)	661.05 (147.43)	657.91 (136.74)
RTs Non-Disambiguated Points Auditory	688.38 (107.19)	689.03 (110.78)	676.70 (114.43)
RTs Disambiguated Points Tactile	699.14 (102.30)	684.89 (113.04)	660.58 (101.30)
RTs Non-Disambiguated Points Tactile	688.94 (96.91)	677.56 (98.46)	661.35 (100.73)
RTs Disambiguated Points Visual	342.29 (100.3)	335.04 (90.41)	325.03 (96.60)
RTs Non-Disambiguated Points Visual	331.84 (93.95)	337.81 (104.78)	338.71 (106.50)

Table 13. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 2 in each Modality (Analysis 3).

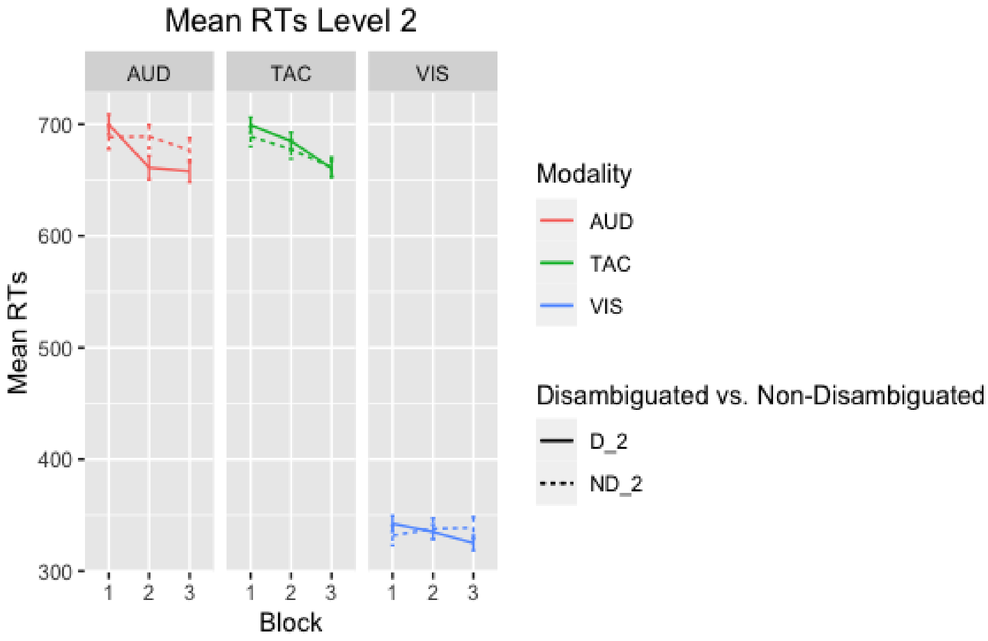


Figure 47. Mean RTs for D and ND points by block at Level 2 in the three studies (Analysis 3). Error bars denote the 95% confidence interval. D_2 = Disambiguated points at Level 2; ND_2 = Non-Disambiguated points at Level 2; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		ß	*SE*	*t*	*p*
Point_Level_2*Block|	Block 1 D – Block 2 D	36.82	5.52	6.67	<.0001
Auditory	Block 1 D – Block 3 D	42.10	5.52	7.62	<.0001
	Block 2 D – Block 2 ND	-18.69	6.32	-2.96	0.0367
	Block 3 D – Block 3 ND	-17.06	6.41	-2.66	0.0078
Point_Level_2+Block|	Block 1 – Block 2	10.7	3.12	3.42	0.0019
Tactile	Block 1 - Block 3	34.6	3.13	11.09	<.0001
	Block 2 – Block 3	24.0	3.09	7.76	<.0001
Point_Level_2*Block|	Block 1 D – Block 3 D	17.43	4.32	4.03	0.0002
Visual	Block 1 D - Block 1 ND	10.96	5.12	2.14	0.0323
	Block 3 D - Block 3 ND	-12.54	4.99	-2.51	0.0121

Table 14. Summary of significant LMM coefficients and contrasts on RTs (Point_Level_2 * Block | Auditory; Point_Level_2 * Block | Tactile; Point_Level_2 * Block | Visual) (Analysis 3).

		ß	*SE*	*t*	*p*
Modality *Point_Level_2	D AUD - VIS	340.52	17.4	19.52	<.0001
	D TAC - VIS	352.11	16.9	20.80	<.0001
	ND AUD - VIS	347.38	17.6	19.74	<.0001
	ND TAC - VIS	345.40	17.1	20.22	<.0001

Table 15. Summary of significant LMM coefficients and contrasts on RTs (Modality * Point_Level_2) (Analysis 3).

In summary, at Level 2, we observed the acquisition of D points in all three modalities, revealing domain-specific differences. The auditory and tactile domains exhibited significantly higher proficiency in learning D points at this level compared to the visual domain. Specifically, D points were acquired in Block 2 in the auditory and tactile domains, while in Block 3 in the visual domain. Upon examining the trend in reaction times (RTs) across blocks, we observed a similar RTs curve for D points in the auditory and tactile domains. These findings collectively indicate the superiority of the auditory and tactile domains over the visual domain in sequential hierarchical learning. However, in line with results at previous levels, despite this learning advantage, we consistently noted overall quicker reaction times in the visual domain compared to the auditory and tactile domains. This hints at a general processing superiority for visual information, irrespective of learning, potentially due to faster communication channels between visual input and motor responses.

As for accuracy, as observable in Figure 48, D points are more accurate than ND points in the third and final block, in all the three modalities. In the auditory study, the accuracy of D points is higher than that of ND points in all the three blocks. Moreover, the former increasingly increases across Block, whereas the opposite trend is observable for the latter. In the tactile and visual studies, despite the general lowering of accuracy rates along the task, it is observed that D and ND points have opposite and mirror-like behavior in the transition from the second to the third block. The former become more accurate, while the opposite happens for the latter. The GLMM model failed to converge, indicating difficulties in obtaining a satisfactory estimation of the model parameters. Since the present convergence issue may stem from the complexity of the model, we reduced it and proceeded by (i) investigating the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point; (ii) investigating the interaction between type of point and block within the individual modalities, to check whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. In analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_2* (Disambiguated at level 2 vs. Non-disambiguated at Level 2) and Modality (Auditory, Tactile, Visual) as independent variables with full interaction, and Subject as random intercept. The analysis revealed a main effect of *Modality* ($\chi2$ =12.31, df = 2, p < .01), indicating the presence of significantly different accuracy rates between modalities. *Point_Level_2* was also significant ($\chi2$ =21.38, df = 1, p < .001), meaning that D points were significantly more accurate than ND points. However, the *Point_Level_2*Modality* interaction was not significant, meaning that the difference in accuracy rates between modalities was not modulated by the type of point. Hence, we ran a second model with *Accuracy* as dependent variable, Modality (Auditory, Tactile, Visual) as independent variables, and *Subject* as random intercept. Since Modality was significant (χ2 =16.57, df = 2, p < .001), we ran on this model post-hoc comparisons, which showed that accuracy rates in the visual modality were higher than those in the auditory modality (β =-1.023; SE =0.27; z =-3.86; p =0.0003), and tactile modality (β =-0.842; SE = 0.26; z =-3.24; p=0.0034). This result is in line with what has been found at previous levels, indicating a general processing advantage, independent from learning, for the visual modality. For analysis (ii), we splitted data according to *Modality* and investigated the effect of *Point_Level_2* and *Block* in the three separated datasets (Auditory, Tactile, Visual). To check whether there were differences in accuracy rates between disambiguated and non-disambiguated points at level 2 in the three modalities, we conducted three GLMM models (one in each modality) with *Accuracy* as dependent variable, *Block* (1-3) and *Point_Level_2* (Disambiguated at level 2 vs. Non-disambiguated at Level 2) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory domain, the analysis revealed no significant effects for *Block*, indicating the absence of significant differences in accuracy rates between blocks. *Point_Level_2* was also not significant, meaning that there were no significant differences for D and ND points (92% vs. 87%). The *Point_Level_2*Block* interaction was instead significant, indicating that the trend of accuracy rates across blocks was modulated by the type of point (D vs. ND). Post-hoc comparisons reveled significantly higher accuracy rates on D than ND points in Block 2 (β =0.51; SE = 0.199; z = 2.567; p =0.0103), and Block 3 (β = 0.8331; SE = 0.187; z = 4.465; p <.0001). This result confirms what has been observed for RTs: In the auditory domain, D points at Level 2 have been acquired in Block 2.

In the tactile domain, the analysis revealed a main effect of *Block* (χ2 =6.17, df = 2, p < .05), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_2* was not significant, indicating that there were no significant differences in accuracy rates between D and ND points (91% vs. 90%, respectively). The *Point_Level_2*Block* interaction was significant (χ2 =7.97, df = 2, p < .05), indicating that the trend of accuracy rates across blocks was modulated by the type of point. Post-hoc comparisons showed a significant decrease in accuracy for D points from Block 1 to Block 2, and on ND points from Block 1 to

Block 3 and from Block 2 to Block 3; accuracy rates for ND points were significantly lower than those for D points in Block 3 (Table 17). This last result is in line with what has been found for RTs, confirming that D points have been learned in the tactile modality.

In the visual modality, the analysis revealed a main effect of *Block* (χ2 =10.80, df = 2, p < .01), indicating that accuracy decreased across blocks. *Point_Level_2* was not significant, indicating that the difference between D and ND points was not significant (96% vs. 95%). The *Point_Level_2*Block* interaction was significant (χ^2 = 8.42, df = 2, *p* < .05). This means that accuracy rates across blocks were modulated by the type of point. Post-hoc comparisons showed that accuracy rates for D points decreased significantly from Block 1 to Block 2, and from Block 1 to Block 3, whereas for ND points they decreased from Block 1 to Block 3 and from Block 2 to Block 3. Accuracy rates for ND points were significantly lower than those for D points in Block 3 (Table 17). We hypothesize that the overall decrease in accuracy rates is a result of the high cognitive load demand required for the acquisition of this regularity. Importantly, the confirmation of D points being more accurate than ND points in Block 3 confirms their acquisition, in line with the analysis of reaction times (RTs).

	Block 1	**Block 2**	**Block 3**
Accuracy Disambiguated Points Auditory	0.91 (0.29)	0.92 (0.26)	0.92 (0.27)
Accuracy Non-Disambiguated Points Auditory	0.89 (0.31)	0.88 (0.32)	0.85 (0.36)
Accuracy Disambiguated Points Tactile	0.93 (0.26)	0.90 (0.30)	0.91 (0.28)
Accuracy Non-Disambiguated Points Tactile	0.92 (0.27)	0.91 (0.28)	0.86 (0.34)
Accuracy Disambiguated Points Visual	0.98 (0.14)	0.95 (0.22)	0.95 (0.21)
Accuracy Non-Disambiguated Points Visual	0.98 (0.15)	0.97 (0.18)	0.92 (0.30)

Table 16. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 2 in each Modality (Analysis 3).

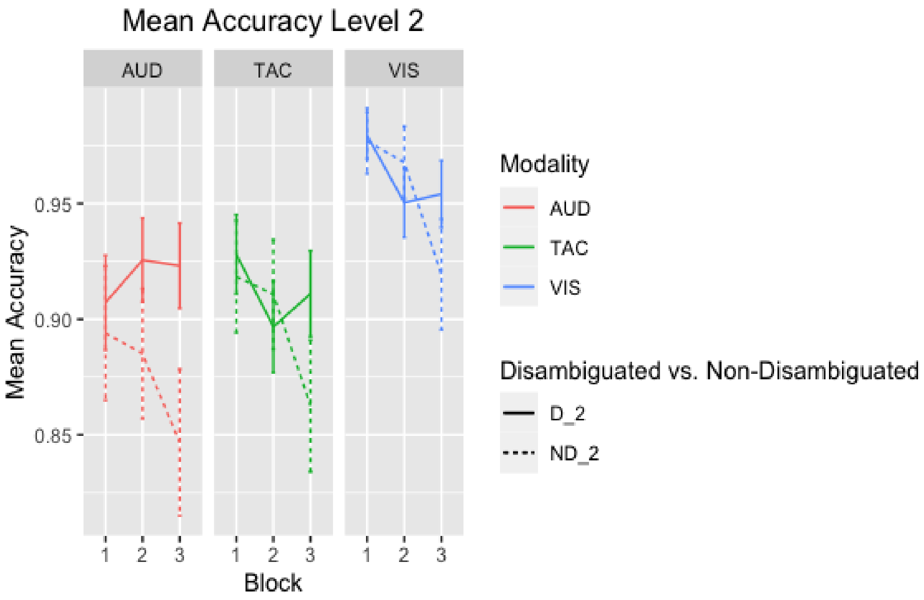


Figure 48. Mean accuracy rates for D and ND points by block at Level 2 in the three studies (Analysis 3). Error bars denote the 95% confidence interval. D_2 = Disambiguated points at Level 2; ND_2 = Non-Disambiguated points at Level 2; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	***SE***	***z***	***p***
Point_Level_2*Block| Tactile	Block 1 D – Block 2 D	0.44	0.18	2.48	0.0346
	Block 1 ND – Block 3 ND	0.66	0.22	3.069	0.0061
	Block 2 ND – Block 3 ND	0.55	0.20	2.73	0.0173
	Block 3 D – Block 3 ND	0.56	0.18	3.12	0.0018
Point_Level_2*Block| Visual	Block 1 D – Block 2 D	0.94	0.30	3.16	0.0045
	Block 1 D – Block 3 D	0.86	0.30	2.84	0.0124
	Block 1 ND – Block 3 ND	1.38	0.36	3.85	0.0003
	Block 2 ND – Block 3 ND	1.0247	0.304	3.371	0.0022
	Block 3 D – Block 3 ND	0.64	0.24	2.69	0.0071

Table 17. Summary of significant GLMM coefficients and contrasts on Accuracy (Point_Level_2 * Block |Tactile; Point_Level_1 * Block |Visual;) (Analysis 3).

Summarizing, result on accuracy rates at Level 2 confirmed what has been found for RTs. Indeed, accuracy rates on D points were higher than those on ND points in Block 2 and Block 3 in the auditory sphere, and in Block 3 in the tactile and visual spheres. As found at the lower levels, we observed that accuracy rates in the visual modality were higher than those in the auditory and tactile modalities. This can be attributable to a general processing advantage, independent from learning, for the visual sphere. Moreover, we observed that accuracy rates in the visual modality decreased along the task on both types of point. This result could be indicative of a higher cognitive effort to learn this regularity in the visual domain compared to the other two domains.

Analysis 4: Deterministic Vs. Non-Deterministic points within Level 3 in the Auditory, Tactile, and Visual studies

At Level 3, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this Level, D points correspond to those 0 that follow the chunks [101][101]; ND points to those 0 that follow [01][101]. Results are reported in Table 18 and 21, respectively. As observable in Figure 49, RTs in the visual modality are considerably lower than those in the auditory and tactile ones. In the auditory modality, the curve of RTs for D points is much steeper than that for ND points. In addition, RTs for D points already drop starting from the transition between Block 1 and Block 2, and then drop further in Block 3. RTs for ND points, on the other hand, only decrease in the transition between Block 2 and Block 3, and still to a lesser extent than for D points. In the tactile modality, RTs for both types of point diminish across blocks, following a similar trend. In the visual modality, instead, neither RTs for D points nor those for NDs seem to decrease along the task. From the LMM model, we found a main effect of *Block* (χ2 =67.81, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Point_Level_3* (χ^2 =4.48, df = 1, *p* < .05). *Modality* was significant (χ^2 =689.55, df = 2, *p* < .001), indicating that there were significant differences in RTs between modalities. The

*Point_Level_3*Block* interaction was significant (χ^2 =9.72, df = 2, *p* = < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). *Block*Modality* was significant (χ^2 =43.13, df = 4, *p* = < .001), meaning that RTs across blocks were modulated by the type of modality. The interaction *Point_Level_3*Block*Modality* was also significant (χ^2 =10.49, df = 4, *p* = < .05): The difference in the trend of RTs between D and ND points across blocks were modulated by the modality. We ran post-hoc tests. In the auditory modality, we found a significant decrease in RTs for D points from Block 1 to Block 2; from Block 1 to Block 3; from Block 2 to Block 3; RTs for ND points decreased as well. Importantly, however, RTs for ND points decreased only from Block 1 to Block 3 and to a lesser extent than D points, as seen by comparing the magnitude of the difference between Block 1 and Block 3 in the two types of point (Table 19). RTs on D points were significantly higher than those on ND points in Block 1, whereas they were significantly shorter in Block 3 (Table 19). In the tactile modality, a significant decrease in RTs on D points was observed from Block 1 to Block 3 (β =13.85; SE =5.47; t =2.53; p =0.0306), whereas RTs for ND points did not change significantly across blocks. However, no significant differences between D and ND points were found in any of the three blocks. In the visual modality, no change was observed across the blocks, either on D or ND points. RTs for D points were significantly higher than those for ND points in Block 3 (β =14.16; SE =6.47; t =2.19; p =0.0286). Despite this, as observed at previous levels, RTs in the visual modality were significantly faster than those in the auditory and tactile modalities in all three blocks, both on D and ND points. As explained before, this might be due to a general processing advantage, independent from learning, for the visual sensory domain over the auditory and tactile ones. In addition to this, comparing modalities, we also observed that RTs in the tactile modality were significantly faster than those in the auditory modality on both D and ND points in Block 1 (Table 20).

Overall, these results indicated that D points at Level 3 were learned in the auditory and tactile modalities, while not in the visual modality. Crucially, however, comparing the auditory and tactile modalities, we observed that RTs on D points in the former modality decreased already in the passage from Block 1 to Block 2, while in the latter modality only in the comparison between Block 1 and Block 3; moreover, comparing the magnitude of the difference between Block 1 and Block 3 in the two sensory spheres, we observed that they decreased to a wider extent in the auditory (β=50.54; SE=6.14; t= 8.23; p<.0001) than in the tactile modality (β =13.85; SE =5.47; t =2.53; p =0.0306). Crucially, however, it is important to note that these results are linked to the fact that RTs on D points in the auditory modality were significantly higher than those in the tactile modality in Block 1. We conclude that D points at Level 3 were learned in both the auditory and tactile modalities in Block 3, while confirming the absence of learning in the visual modality this level.

	Block 1	**Block 2**	**Block 3**
RTs Disambiguated Points Auditory	721.69 (120.35)	697.75 (124.07)	670.56 (134.46)
RTs Non-Disambiguated Points Auditory	704.41 (115.55)	705.09 (115.68)	689.57 (99.95)
RTs Disambiguated Points Tactile	669.05 (101.50)	666.65 (100.20)	656.60 (105.16)
RTs Non-Disambiguated Points Tactile	659.97 (91.37)	657.13 (98.24)	650.03 (97.19)
RTs Disambiguated Points Visual	310.66 (100.90)	317.19 (104.81)	315.56 (105.82)
RTs Non-Disambiguated Points Visual	307.63 (91.27)	317.05 (108.31)	300.82 (102.42)

Table 18. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 3 in each Modality (Analysis 4).

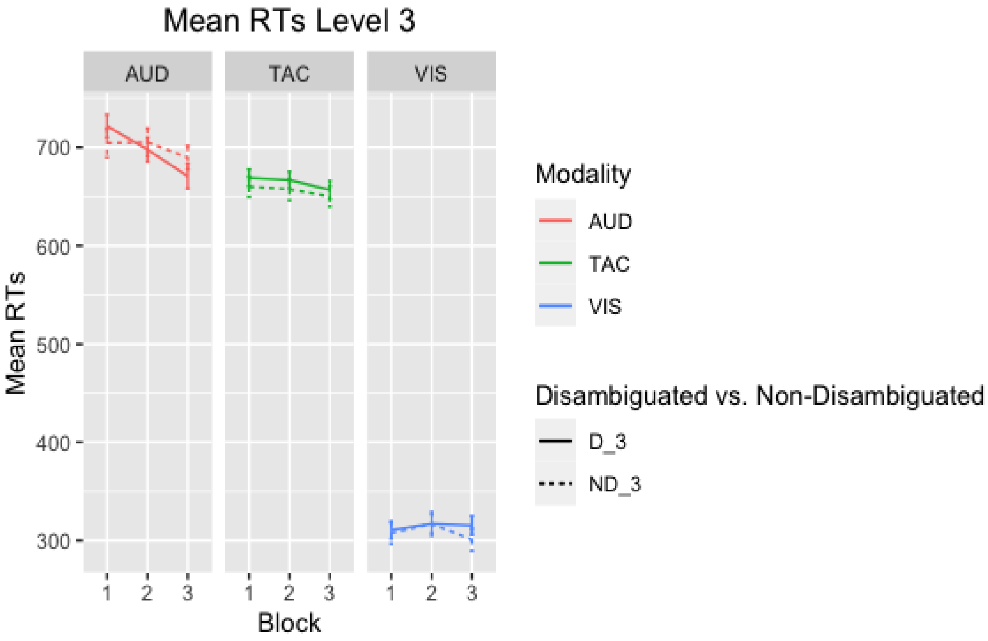


Figure 49. Mean RTs for D and ND points by block at Level 3 in the three studies (Analysis 4). Error bars denote the 95% confidence interval. D_3 = Disambiguated points at Level 3; ND_3 = Non-Disambiguated points at Level 3; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

		β	*SE*	*t*	*p*
Point_Level_3*Block| Auditory	Block 1 D – Block 2 D	25.00	6.27	3.99	0.0002
	Block 1 D – Block 3 D	50.54	6.14	8.23	<.0001
	Block 2 D – Block 3 D	25.54	6.18	4.13	0.0001
	Block 1 ND – Block 3 ND	19.49	7.85	2.48	0.0348
	Block 1 D – Block 1 ND	15.29	7.22	2.12	0.034
	Block 3 D – Block 3 ND	-15.76	6.86	-2.30	0.0216

Table 19. Summary of significant LMM coefficients and contrasts on RTs (Point_Level_3 * Block | Auditory) (Analysis 4).

		β	*SE*	*t*	*p*
Modality*Block*Point_ Level_3	Block 1 D AUD - TAC	50.9	16.7	3.05	0.0080
	Block 1 D AUD - VIS	410.5	17.2	23.86	<.0001
	Block 1 D TAC - VIS	359.6	16.6	21.63	<.0001
	Block 2 D AUD - VIS	380.5	17.2	22.08	<.0001
	Block 2 D TAC - VIS	352.2	16.7	21.15	<.0001
	Block 3 D AUD - VIS	356.5	17.2	20.76	<.0001
	Block 3 D TAC - VIS	342.3	16.6	20.56	<.0001
	Block 1 ND AUD - TAC	44.7	17.4	2.56	0.0306
	Block 1 ND AUD - VIS	399.1	17.9	22.25	<.0001
	Block 1 ND TAC - VIS	354.4	17.3	20.46	<.0001
	Block 2 ND AUD – VIS	382.0	17.8	21.45	<.0001
	Block 2 ND TAC - VIS	343.4	17.2	19.98	<.0001
	Block 3 ND AUD - VIS	386.5	17.7	21.78	<.0001
	Block 3 ND TAC - VIS	349.9	17.2	20.39	<.0001

Table 20. Summary of significant LMM coefficients and contrasts on RTs (Modality * Block * Point_Level_3) (Analysis 4).

Summarizing, at Level 3 we found that learning occurred in the auditory and tactile modalities. Specifically, D points at Level 3 were learned in Block 3 in both modalities. On the contrary, we did not find evidence of learning in the visual modality. Despite these learning differences, we consistently observed faster RTs in the visual domain than in the auditory and tactile domains. This result is in line with what has been observed at previous levels and might indicate a general processing advantage for visual information, independent from learning, possibly stemming from quicker communication channels between visual input and motor responses.

As for accuracy, as observable in Figure 50, in the auditory modality, we observe a similar trend between the two types of point. For both D and ND points accuracy rates slightly decrease from Block 1 to Block 2 and then increase from Block 2 to Block 3. In the tactile, accuracy rates for both D and ND points decrease from Block 1 to Block 2. Accuracy for ND points then increase in the passage from Block 2 to Block 3. In the visual modality, accuracy rates slightly decrease from Block 1 to Block 2 and then increase in Block 3, on both D and ND points. The GLMM model failed to converge. Hence, we reduced the complexity of the model. First, (i) we checked the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to see whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_3* (Disambiguated at level 3 vs. Non-disambiguated at Level 3) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* ($\chi2$ =58.54, df = 2, p < .001), indicating the presence of significantly different accuracy rates between modalities. *Point_Level_3* was also significant ($\chi2$ =3.95, df = 1, p < .05), meaning that ND points were significantly more accurate than D points. The *Point_Level_3*Modality* interaction was not significant; this means that the difference in accuracy rates between modalities was not modulated by the type of point. Hence, we subsequently ran a model with *Accuracy* as dependent variable, *Modality* (Auditory, Tactile, Visual) and *Point_Level_3* as independent variables, and *Subject* as random intercept. *Modality* was significant ($\chi2$ =67.52, df = 2, p < .001), and *Point_Level_3* as well ($\chi2$ =9.85, df = 1, p < .01). Hence, we ran on this second model post-hoc, which showed that accuracy rates in the visual modality were higher than those in the auditory (β=-2.10; SE=0.26; z=-7.96; p =<.0001) and tactile modalities (β=-0.84; SE=0.27; z=-3.15; p =0.0046). Accuracy rates in the tactile modality were higher than those in the auditory one (β=-1.26; SE=0.24; z=-5.31; p =<.0001). For analysis (ii), we splitted data according to *Modality*. Hence, we investigated the effect of *Point_Level_3* and *Block* in the three separated datasets (Auditory, Tactile, Visual). To check whether there were differences in accuracy rates between disambiguated and non-disambiguated points at level 3 in the three modalities, we conducted three separated GLMM models (one for each modality) with *Accuracy* as dependent variable, *Block* (1-3) and *Point_Level_3* (Disambiguated at level 3 vs. Non-disambiguated at Level 3) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory domain, the analysis revealed a main effect of *Block* (χ2 = 9.29, df = 2, p < .01), indicating the presence of significantly different accuracy rates between blocks. Neither *Point_Level_3*, nor the interaction *Point_Level_3*Block* were significant. Being the interaction not significant, we ran a second GLMM with accuracy as dependent variable, *Block* and *Point_Level_3* as independent predictors, and *Subject* as random intercept. We found a main effect of *Block* (χ2 =12.56, df = 2, p < .01), indicating that RTs became shorter across blocks. *Point_Level_3* was significant (χ2 =3.91, df = 2, p < .05). We then compared the two models with the anova()-function in R, without finding any significant result. Therefore, we failed to reject the null hypothesis, meaning that the two models did not differ. We ran post-hoc tests on the simpler model ((accuracy ~ Block + Point_Level_3 + (1 |Subject)). Results indicated a significant increase in accuracy rates from Block 2 to Block 3 (β= -0.49; SE= 0.14; z= -3.53; p = 0.0012). ND points were marginally more accurate than D ones (β= -0.23; SE= 0.12; z= -1.98; p = 0.0480).

In the tactile modality, the analysis revealed a main effect of *Block* (χ2 =11.08, df = 2, p < .01), indicating the presence of significantly different accuracy rates between blocks. *Point_Level_3* was not significant, indicating that there were no significant differences in accuracy rates between D and ND points (93% vs. 95%, respectively). The *Point_Level_3*Block* interaction was not significant, indicating that the trend of accuracy rates across blocks was not modulated by the type of point. Being the interaction not significant, we ran a second GLMM with *Accuracy* as dependent variable, *Block* (1-3) and *Point_Level_3* as independent predictors, and *Subject* as random intercept. We found a main effect of *Block* (χ2 =10.00, df = 2, p < .01), indicating a significant change in accuracy rates across blocks. *Point_Level_3* was significant (χ2 =4.86, df = 1, p < .05). We then compared the two models with the anova()-function in R, without finding any significant result. Therefore, we failed to reject the null hypothesis, meaning that the two models did not differ. We ran post-hocs tests on the simpler model ((accuracy~ Block + Point_Level_3+ (1 |Subject)). Results indicated a significant decrease in accuracy rates from Block 1 to Block 2 (β=0.64; SE= 0.21; z= 3.06; p = 0.0062), and from Block 1 to Block 3(β= 0.53; SE= 0.21; z= 2.51; p = 0.0319). D points were less accurate than ND ones (β=-0.38; SE= 0.17; z= 2.20; p = 0.0274).

In the visual domain, the analysis reported no significant results.

Comparing the three modalities, we found that accuracy rates in the visual modality were higher than those in the auditory and tactile modalities. Moreover, accuracy rates in the tactile modality were higher than those in the auditory one. Overall, at Level 3, the analysis of accuracy rates did not provide any interesting result regarding learning.

	Block 1	Block 2	Block 3
Accuracy Disambiguated Points Auditory	0.81 (0.39)	0.79 (0.41)	0.86 (0.35)
Accuracy Non-Disambiguated Points Auditory	0.85 (0.35)	0.82 (0.38)	0.88 (0.33)
Accuracy Disambiguated Points Tactile	0.95 (0.21)	0.91 (0.28)	0.91 (0.28)
Accuracy Non-Disambiguated Points Tactile	0.95 (0.21)	0.93 (0.25)	0.95 (0.21)
Accuracy Disambiguated Points Visual	0.97 (0.16)	0.96 (0.19)	0.97 (0.17)
Accuracy Non-Disambiguated Points Visual	0.98 (0.14)	0.97 (0.16)	0.98 (0.14)

Table 21. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 3 in each Modality (Analysis 4).

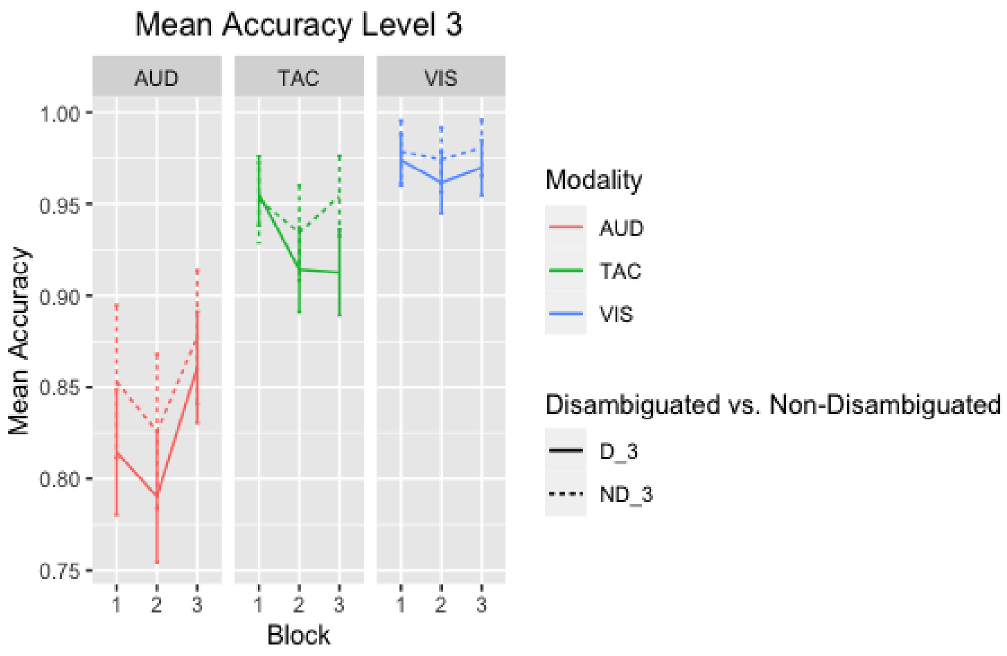


Figure 50. Accuracy rates for D and ND points by block at Level 3 in the three studies (Analysis 4). Error bars denote the 95% confidence interval. D_3 = Disambiguated points at Level 3; ND_3 = Non-Disambiguated points at Level 3; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Analysis 5: Deterministic Vs. Non-Deterministic points within Level 4 in the Auditory, Tactile, and Visual studies

At Level 4, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this level, D points correspond to those 1 that follow the chunks [01101][01101], ND points to those 1 that follow [101] [01101]. Results are reported in Table 22 and 23, respectively. As observable in Figure 51, RTs in the visual modality are considerably lower than those in the auditory and tactile ones. In the auditory and tactile modalities, RTs for D points increasingly decrease across the three blocks, following a similar trend in the two modalities. On the contrary, RTs for ND points do not decrease, either in the auditory or tactile modality. In the visual modality, instead, RTs for D points do not decrease along the task. Those for ND points increase, especially in the transition from Block 2 to Block 3. From the LMM model, we found a main effect of *Block* (χ2 =15.79, df = 2, p < .001), indicating that RTs became shorter across blocks. We also found a main effect of *Modality* (χ^2 =555.28, df = 2, *p* < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_4*Block* interaction was significant (χ^2 =13.56, df = 2, *p* = < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). *Block*Modality* was significant (χ^2 =10.85, df = 4, *p* = < .05), meaning that RTs across blocks were modulated by the type of modality. The interaction *Point_Level_3*Block*Modality* was not significant: The difference in the trend of RTs between D and ND points across blocks were not modulated by the modality. To further investigate the nature of the interactions, (i) we assessed the interaction between type of point and modality to verify whether the differences in RTs between modalities were modulated by the type of point. Secondly, (ii) we explored the interaction between type of point and block within the individual modalities, to verify whether the trend of RTs across blocks were modulated by the type of point, within the single modalities. For analysis (i) we ran a LMM model with *RTs* as dependent variable, *Point_Level_4* (Disambiguated at level 4 vs. Non-disambiguated at Level 4) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =579.82, df = 2, p < .001), indicating the presence of significantly different RTs between modalities. Neither *Point_Level_4* nor the interaction *Point_Level_4*Modality* were significant. Hence, we ran a simpler LMM model with *RTs* as dependent variable, *Modality* as independent variable, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =585.43, df = 2, p < .001). Post-hoc comparisons showed that RTs in the visual modality were faster than those in the auditory (β=348.35; SE=16.7; t=20.82; p<.0001) and tactile modalities (β=345.40; SE=16.2; t=21.27; p<.0001). This is in line with wat has been found at previous levels and could be attributed to a general processing advantage, independent from learning for the visual modality. To conduct analysis (ii), we subsequently splitted data according to *Modality*. Hence, the effect of *Point_Level_4* and *Block* was investigated in the three separated datasets (Auditory, Tactile, Visual). To verify if there were learning differences between disambiguated and non-disambiguated points at level 4 in the three modalities, we ran three LMM models (one for each modality) with *RTs* as dependent variable, *Block* (1-3) and *Point_Level_4* (Disambiguated at level 4 vs. Non-disambiguated at Level 4) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory modality, we found a main effect of *Block* (χ2 = 13.48, df = 2, p < .01), indicating that RTs significantly changed across blocks. The *Point_Level_4*Block* interaction was significant (χ^2 = 11.57, df = 2, *p* = < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant decrease in RTs for D points from Block 1 to Block 3 (β= 28.17; SE= 8.04; t= 3.504; p = 0.0014), and from Block 2 to Block 3 (β= 20.53; SE= 7.78; t= 2.64; p = 0.0229). RTs on D points were significantly shorter than those on ND points in Block 3 (β= -29.56; SE= 9.58; t= -3.09; p = 0.0021).

In the tactile modality, we found a main effect of *Block* (χ2 = 44.80, df = 2, p < .001), indicating that RTs became shorter across blocks. The *Point_Level_4*Block* interaction was significant (χ2 = 12.93, df = 2, p < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant decrease in RTs on D points from Block 1 to Block 3 (β=38.28; SE= 6.05; t= 6.32; p = <.0001), and from Block 2 to Block 3 (β=29.59; SE= 5.91; t= 5.01; p = <.0001). RTs on ND points did not decrease significantly. The difference between D and ND points was significant in Block 3 (β=-24.26; SE= 7.10; t= -3.41; p = 0.0007).

In the visual modality, we ran a LMM with *RTs* as dependent variable, *Block* (1-3) and *Point_Level_4* (Disambiguated at level 4 vs. Non-disambiguated at Level 4) as independent variables with full interaction, and *Subject* as random intercept. The *Point_Level_4*Block* interaction was significant (χ2 = 10.02, df = 2, p < .01), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant increase in RTs for ND points from Block 1 to Block 3 (β=-27.60; SE=9.93; t=-2.78; p = 0.0153). RTs for D points were significantly shorter than ND points in Block 3 (β=-32.03; SE=8.65; t=-3.70; p = 0.0002).

Overall, we found that RTs on D points decreased along the task in both the auditory and tactile spheres following a similar trend along the blocks: the decrease became significant in the transition from Block 2 to Block 3. In contrast, ND points did not decrease. In both modalities, we found a difference in RTs between D points and ND points in Block 3. Hence, we conclude that D points at Level 4 were learned in both the auditory and tactile modalities in Block 3. In the visual study, on the other hand, RTs on D points did not decrease, while those on ND points increased, and the difference between the two types of point was significant in block 3. Hence, we confirm that learning in the visual modality stopped at Level 2. Despite this, we observed, in line with previous level, that RTs in the visual modality were faster than those in the auditory and tactile modalities, suggesting a general processing advantage, independent from learning, for the visual sphere.

	Block 1	Block 2	Block 3
RTs Disambiguated Points Auditory	693.97 (105.57)	691.22 (107.67)	666.06 (115.45)
RTs Non-Disambiguated Points Auditory	678.79 (109.60)	685.29 (116.15)	698.69 (109.45)
RTs Disambiguated Points Tactile	692.56 (100.77)	680.23 (99.23)	653.00 (91.98)
RTs Non-Disambiguated Points Tactile	682.44 (89.50)	673.13 (97.25)	677.02 (114.00)
RTs Disambiguated Points Visual	332.76 (102.27)	338.87 (108.43)	327.45 (92.92)
RTs Non-Disambiguated Visual	330.23 (77.52)	336.09 (98.77)	358.01 (124.37)

Table 22. Mean (SDs) RTs of each block for Disambiguates (D) and Non-Disambiguated (ND) points at Level 4 in each Modality (Analysis 5).

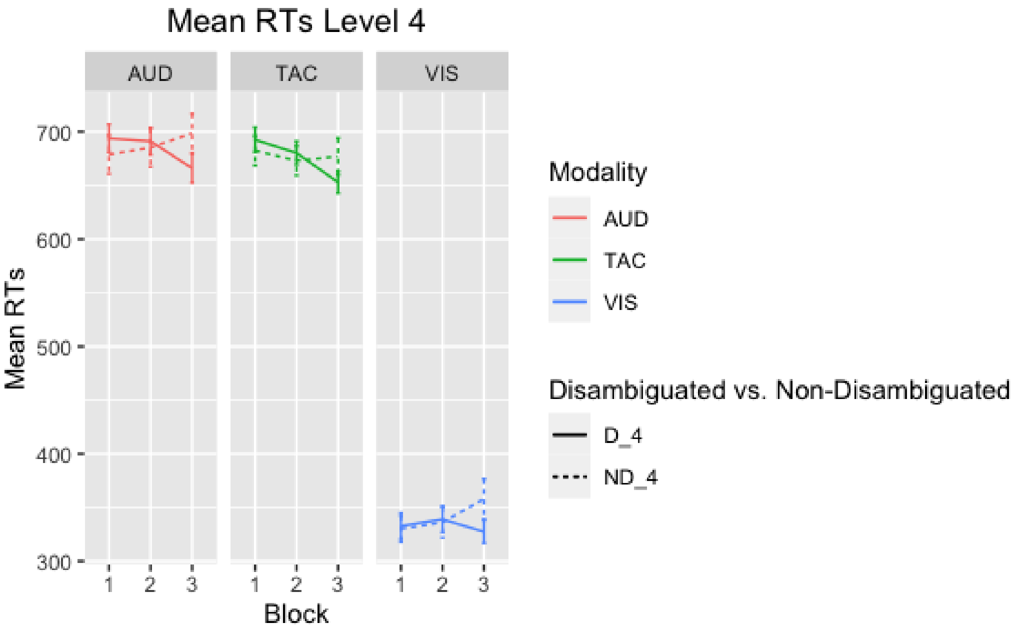


Figure 51. Mean RTs for D and ND points by block at Level 4 in the three studies (Analysis 5). Error bars denote the 95% confidence interval. D_4 = Disambiguated points at Level 4; ND_4 = Non-Disambiguated points at Level 4; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Summarizing, at Level 4 we found that D points were learned in Block 3 in both the auditory and tactile modalities. On the contrary, we did not find any evidence of learning in the visual modality, thus confirming that learning stopped at Level 2. However, in line with findings at previous levels, we found that RTs were consistently faster in the visual domain compared to auditory and tactile domains. This suggests a general processing advantage for the visual domain, independent from learning, possibly due to faster communication channels between visual input and motor responses.

As for accuracy, as observable in Figure 52, in the auditory modality, we observe an increase on D points and a decrease on ND points across blocks. A similar decrease in accuracy rates on ND points is observable also in the tactile and visual modalities. In these latter modalities, accuracy rates on D points decrease as well, albeit to a smaller extent than ND points. The GLMM model failed to converge. Hence, we reduced the complexity of the model. First, (i) we checked the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to see whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_4* (Disambiguated at level 4 vs. Non-disambiguated at Level 4) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* ($\chi 2 = 9.53$, df = 2, $p < .01$), indicating the presence of significantly different accuracy rates between modalities. *Point_Level_4* was also significant ($\chi 2 = 10.88$, df = 1, $p < .001$), meaning that D points were significantly more accurate than ND ones. The *Point_Level_4*Modality* interaction was not significant; this means that the difference in accuracy rates between modalities was not modulated by the type of point. Hence, we subsequently ran a model with *Accuracy* as dependent variable, *Modality* (Auditory, Tactile, Visual) and *Point_Level_4* as independent variables, and *Subject* as random intercept. *Modality* was significant ($\chi 2 = 18.39$, df = 2, p <

.001), and *Point_Level_4* as well ($\chi2$ =11.63, df = 1, p < .001). Hence, we ran on this second model post-hoc comparisons, which showed that accuracy rates in the visual modality were higher than those in the auditory (β=-1.20; SE=0.29; z=-4.19; p =.0001) and tactile modalities (β=-0.90; SE=0.28; z=-3.19; p =0.0040). This result is in line with previous results, suggesting a processing advantage for the visual modality over the other two modalities, independent from learning. For analysis (ii), we splitted data according to *Modality* and investigated the effect of *Point_Level_4* and *Block* in the three separated datasets (Auditory, Tactile, Visual). To check whether there were differences in accuracy rates between disambiguated and non-disambiguated points at level 4 in the three modalities, we conducted three GLMM models (one in each modality) with *Accuracy* as dependent variable, *Block* (1-3) and *Point_Level_4* (Disambiguated at level 4 vs. Non-disambiguated at Level 4) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory modality, we did not find any effect for *Block*, indicating the absence of significantly different accuracy rates between blocks. *Point_Level_4* was also not significant, meaning that there were no significant differences between D and ND points. The *Point_Level_4*Block* interaction was instead significant (χ^2 = 21.64, df = 2, *p*= < .001). Post-hoc comparisons reveled a significant decrease in accuracy for ND points between Block 1 and Block 3 (β=1.60; SE=0.35; z= 4.52; p< .0001), and between Block 2 and Block 3 (β=0.99; SE= 0.29; z=3.44; p =0.0017). Accuracy for D points was significantly higher than that for ND points in Block 3 (β= 1.47; SE= 0.27; z=5.38; p =<.0001).

In the tactile modality, we found that *Block* was not significant, indicating that accuracy rates did not change across blocks. *Point_Level_4* was not significant, indicating that there were no significant differences in accuracy rates between D and ND points (91% vs. 87%, respectively). The *Point_Level_4*Block* interaction was significant ($\chi2$ =7.92, df = 2, p < .05), indicating that the trend of accuracy rates across blocks was modulated by the type of point. Post-hoc comparisons showed a significant decrease in accuracy for ND points between Block 1 and Block 3 (β= 1.21; SE=0.34; z=3.53; p =0.0012); between Block 2 and Block 3 (β=1.12; SE=0.31; z=3.57; p =0.0011). D points were significantly more accurate than ND points in Block 3 (β=0.93; SE=0.26; z=3.57; p =0.0004). In the visual modality, we did not find any significant interaction.

Hence, at Level 4 we found a similar trend in accuracy rates in the auditory and tactile modalities: in both two modalities, we found a decrease on ND points along the task. Accuracy on D points was higher than that on ND points in Block 3. This result is in line with what we found on RTs, confirming that D points at Level 4 were learned in Block 3 in both modalities. In line with results on RTs, in the visual modality no significant interactions were found. As at previous levels, we found that accuracy rates in the visual modality were higher than those in the auditory and tactile ones, suggesting a general processing advantage independent from learning for the visual sensory domain.

	Block 1	**Block 2**	**Block 3**
Accuracy Disambiguated Points Auditory	0.88 (0.33)	0.89 (0.31)	0.91 (0.28)
Accuracy Non-Disambiguated Points Auditory	0.92 (0.27)	0.87 (0.34)	0.74 (0.44)
Accuracy Disambiguated Points Tactile	0.92 (0.27)	0.91 (0.29)	0.90 (0.30)
Accuracy Non-Disambiguated Points Tactile	0.92 (0.27)	0.91 (0.28)	0.80 (0.40)
Accuracy Disambiguated Points Visual	0.97 (0.18)	0.96 (0.19)	0.93 (0.26)
Accuracy Non-Disambiguated Points Visual	0.99 (0.08)	0.98 (0.14)	0.90 (0.30)

Table 23. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 4 in each Modality (Analysis 5).

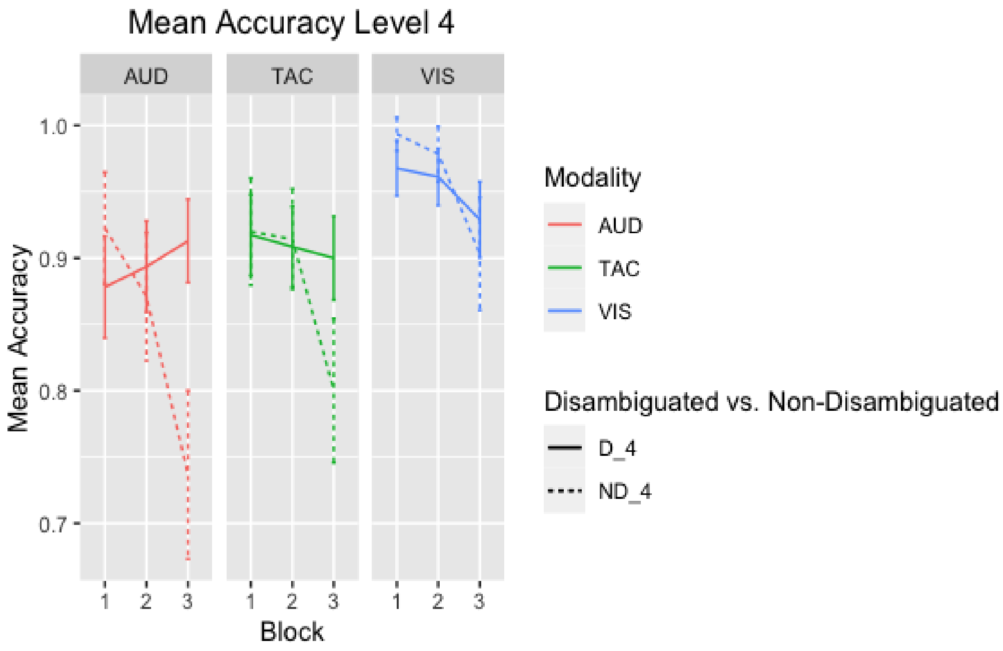


Figure 52. Mean accuracy rates for D and ND points by block at Level 4 in the three studies (Analysis 5). Error bars denote the 95% confidence interval. D_4 = Disambiguated points at Level 4; ND_4 = Non-Disambiguated points at Level 4; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Summarizing, results on accuracy rates confirm what has been found for RTs: D points at Level 4 were learned in the auditory and tactile modalities in Block 3, while we did not find any significant effect in the visual modality, confirming that learning in this sphere stopped at Level 2. Despite this, accuracy rates in the visual modality were higher than those in the auditory and tactile modalities. This result is in line with previous findings, indicating a general processing advantage for the visual domain, independent from learning.

Analysis 6: Deterministic Vs. Non-Deterministic points within Level 5 in the Auditory, Tactile, and Visual studies

At level 5, we compared RTs and accuracy rates in correspondence to every instance of D and ND points in each block, in the three modalities. At this Level, D points correspond to those 0 that follow the chunks [10101101][10101101]; ND points to those 0 that follow [01101] [10101101]. Results are reported in Table 24 and 25, respectively. As observable in Figure 53, RTs in the visual modality are considerably shorter than those in the auditory and tactile ones, while those in the tactile modality are slightly lower than those in the auditory modality. In the auditory and tactile studies, RTs for D points start to decrease in the passage from Block 2 and Block 3. However, only in the tactile modality we observe a difference between D and ND points: RTs for the former are higher than the latter in Block 1 and Block 2, while in Block 3 we observe an inverse pattern. From the LMM model, we found that the factor *Block* was not statistically significant, indicating no difference in RTs between blocks. We found a main effect of *Modality* (χ^2 =653.50, df = 2, p < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_5*Block* interaction was not significant, indicating that RTs across blocks were not modulated by the type of point (D vs. ND). *Block*Modality* was not significant, meaning that RTs across blocks were not modulated by the type of modality. The interaction *Point_Level_5*Block*Modality* was not significant: The difference in the trend of RTs between D and ND points across blocks were not modulated by the modality. To further investigate the nature of the interactions, (i) we assessed the interaction between type of point and modality to verify whether the differences in RTs between modalities were modulated by the type of point. Secondly, (ii) we explored the interaction between type of point and block within the individual modalities, to verify whether the trend of RTs across blocks were modulated by the type of point, within the single modalities. For analysis (i) we ran a LMM model with *RTs* as dependent variable, *Point_Level_5* (Disambiguated at level 5 vs. Non-disambiguated at Level 5) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =762.59, df = 2, p < .001), indicating the presence of significantly different RTs between modalities. Neither *Point_Level_5* nor the interaction between *Point_Level_5*Modality* were significant. Hence, we ran a simpler LMM model with *RTs* as dependent variable, *Modality* as independent variable, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =793.01, df = 2, p < .001). Post-hoc comparisons showed that RTs in the visual modality were faster than those in the auditory (β=390.1; SE=15.3; t=25.42; p<.0001) and tactile modalities (β=348.5; SE=14.9; t=93.3; p<.0001). This result is in line with result at previous level, suggesting a general processing advantage independent from learning for the visual sphere. RTs in the tactile modality were shorter than those in the auditory one (β=41.6; SE=14.9; t=2.792; p=0.0173). To conduct analysis (ii), we splitted data according to *Modality*, exploring the effect of *Point_Level_5* and *Block* in the three separated datasets (Auditory, Tactile, Visual). In the tree modalities, we ran a LMM with *RTs* as dependent variable, *Block* (1-3) and *Point_Level_5* (Disambiguated at level 5 vs. Non-disambiguated at Level 5) as independent variables with full interaction, and *Subject* as random intercept.

In the auditory modality, the analysis revealed no significance.

In the tactile modality, we found a main effect of *Block* (χ2 =12.09, df = 2, p < .01), indicating that RTs became shorter across blocks. The *Point_Level_5*Block* interaction was significant (χ2 = 9.03, df = 2, p < .05), indicating that RTs across blocks were modulated by the type of point (D vs. ND). Post-hoc comparisons reported a significant decrease in RTs on D points from Block 1 to Block 3 (β=21.79; SE= 7.17; t= 3.04; p = 0.0069), and from Block 2 to Block 3 (β=21.41; SE= 7.21; t= 2.97; p = 0.0086). RTs on ND points did not change significantly across blocks. The difference between D and ND points was significant in Block 3 (β=-17.70; SE= 8.01; t= -2.21; p = 0.0273).

As expected, we found no significance in the LMM in the visual modality.

Overall, the only significant effect we found at this level was in the tactile modality. Specifically, we found a significant decrease in RTs on D points in the transition from second to third block, and from the first to the third block. Moreover, RTs on D points were faster than those on NDs in Block 3. Neither in the auditory study nor in the visual one did we find significant interactions. These results suggest that D points were acquired in the tactile modality in Block 3. On the contrary, no learning effects were found in neither the auditory nor the visual modality, confirming thus that learning in the auditory modality stopped at Level 4, while learning in the visual modality stopped at Level 2. However, looking at the graph, we noted that the trend of RTs on D points in the auditory and tactile modalities were similar. Indeed, in the auditory modality there was a decrease in RTs from Block 2 to Block 3 as well, although not significant. Thus, we do not rule out the possibility that the absence of learning effects in the auditory modality was due to insufficient exposure to the string. At this level, as at previous levels, RTs in the visual modality were faster than those in the auditory and tactile modalities, suggesting a general processing advantage for the visual modality, independent from learning. Furthermore, we found that RTs in the tactile modality were lower than those in the auditory modality.

	Block 1	**Block 2**	**Block 3**
RTs Disambiguated Points Auditory	705.14 (117.98)	705.4 (119.49)	686.39 (102.97)
RTs Non-Disambiguated Points Auditory	704.50 (110.17)	694.47 (95.38)	653.30 (81.00)
RTs Disambiguated Points Tactile	663.35 (96.22)	663.12 (103.10)	643.85 (97.43)
RTs Non-Disambiguated Points Tactile	653.30 (81.00)	648.28 (90.26)	659.36 (96.44)
RTs Disambiguated Points Visual	306.55 (86.31)	315.37 (109.61)	301.35 (104.05)
RTs Non-Disambiguated Points Visual	309.82 (101.09)	319.54 (106.76)	300.02 (100.35)

Table 24. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 5 in each Modality (Analysis 6).

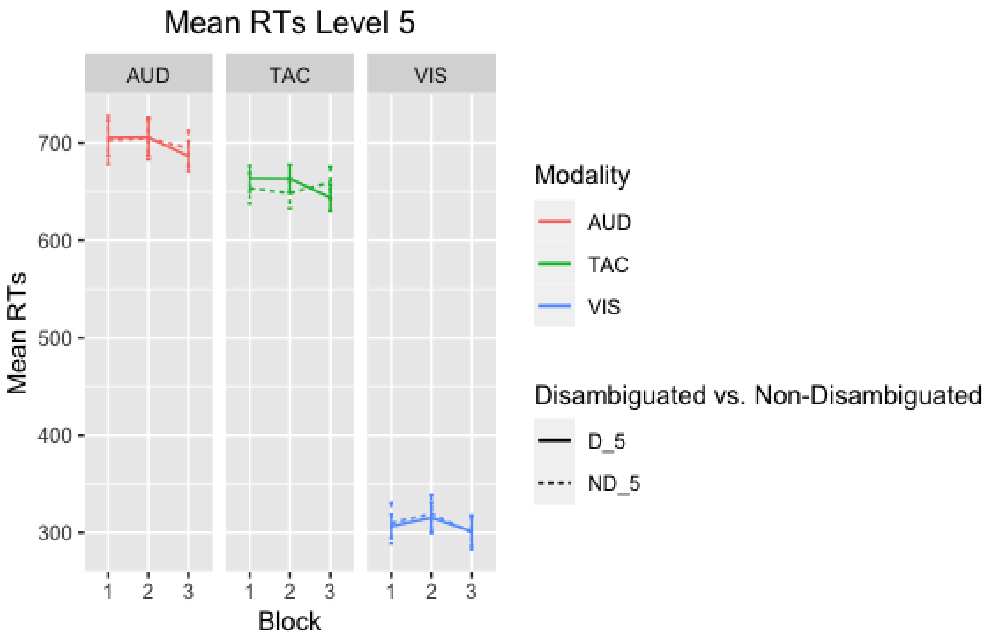


Figure 53. Mean RTs for D and ND points by block at Level 5 in the three studies (Analysis 6). Error bars denote the 95% confidence interval. D_5 = Disambiguated points at Level 5; ND_5 = Non-Disambiguated points at Level 5; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Summarizing, at Level 5 we found that D points were learned in Block 3 in the tactile modality. On the contrary, we did not find significant effects neither in the auditory nor visual modalities, meaning that learning stopped at Level 4 in the auditory modality and at Level 2 in the visual modality. Crucially, however, examining the RTs graph, it became evident that the patterns of response times (RTs) for D points in both auditory and tactile modalities shared similarities. Specifically, the auditory modality exhibited a decrease in RTs from Block 2 to Block 3, although this reduction did not reach statistical significance. Consequently, we cannot dismiss the possibility that the lack of discernible learning effects in the auditory modality may be attributed to insufficient exposure to the string. Again, as found at previous levels, RTs in the visual modality were overall shorter than those in the tactile and auditory modalities, suggesting a general processing advantage, independent from learning.

As for accuracy rates at Level 5, as observable in Figure 54, accuracy in the auditory modality is lower than that in the tactile modality, which in turn is lower than that in the visual modality. In the auditory modality, we observe an increase in accuracy for both D and ND points in the transition from Block 2 to Block 3. The same trend can be observed for D points in the tactile study. The GLMM model failed to converge. Hence, we reduced the complexity of the model. First, (i) we checked the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to see whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_5* (Disambiguated at level 5 vs. Non-disambiguated at Level 5) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and Subject as random intercept. The analysis revealed a main effect of *Modality* (χ2 =30.59, df = 2, p < .001), indicating the presence of significantly different accuracy rates between modalities. Neither *Point_Level_5* or the *Point_Level_5*Modality* were significant. Hence, we ran a model with *Accuracy* as dependent variable, *Modality* (Auditory, Tactile, Visual) as independent variable, and *Subject* as random intercept. *Modality* was significant (χ2 =43.72, df = 2, p < .001). We ran on this second model post-hoc comparisons, which showed that accuracy rates in the visual modality were higher than those in the auditory (β=-2.18; SE=0.35; z=-6.23; p <.0001) and tactile modalities (β=-0.92; SE=0.36; z=-2.54; p = 0.0298). This result is in line with those found at previous levels. Moreover, accuracy rates in the tactile modality were higher than those in the auditory one (β=-1.26; SE=0.29; z=-4.37; p <.0001). For analysis (ii), we splitted data according to *Modality*, investigating the effect of *Point_Level_5* and *Block* in the three separated datasets (Auditory, Tactile, Visual). We did not find any significant interaction, in any of the three modalities.

	Block 1	Block 2	Block 3
Accuracy Disambiguated Points Auditory	0.86 (0.35)	0.83 (0.37)	0.89 (0.32)
Accuracy Non-Disambiguated Points Auditory	0.81 (0.39)	0.86 (0.34)	0.96 (0.19)
Accuracy Disambiguated Points Tactile	0.95 (0.22)	0.93 (0.26)	0.96 (0.20)
Accuracy Non-Disambiguated Points Tactile	0.96 (0.19)	0.94 (0.23)	0.95 (0.22)
Accuracy Disambiguated Points Visual	0.98 (0.13)	0.97 (0.18)	0.98 (0.14)
Accuracy Non-Disambiguated Points Visual	0.97 (0.18)	0.98 (0.13)	0.98 (0.13)

Table 25. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 5 in each Modality (Analysis 6).

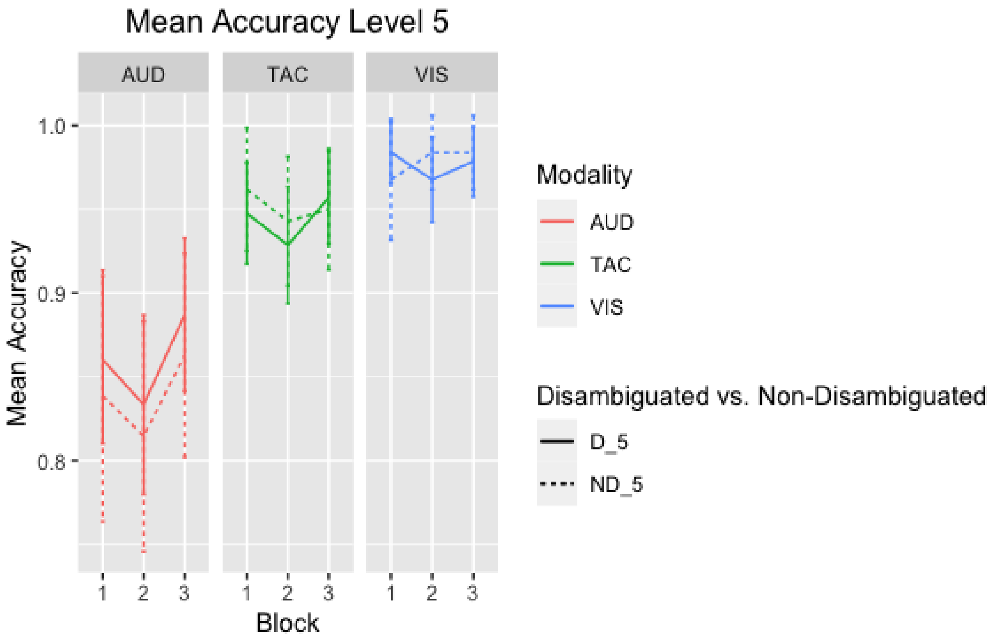


Figure 54. Mean accuracy rates for D and ND points by block at Level 5 in the three studies (Analysis 6). Error bars denote the 95% confidence interval. D_5 = Disambiguated points at Level 5; ND_5 = Non-Disambiguated points at Level 5; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Summarizing, at Level 5, in the analysis of accuracy rates we did not find any significant interaction, within any of the three modalities. The only significant effect found was in the comparison between modalities: accuracy rates in the visual modality were higher than those in the auditory and tactile modalities. This result is in line with previous findings, indicating a general processing advantage for the visual domain, independent from learning.

Analysis 7: Deterministic Vs. Non-Deterministic points within Level 6 in the Auditory, Tactile, and Visual studies

We controlled for possible learning differences between disambiguated and non-disambiguated points at Level 6, by analyzing and comparing RTs and accuracy rates in correspondence to every instance of disambiguated and non-disambiguated point at Level 6, in each block, in the three modalities. At this level, D points correspond to the 1s which follow the chunks [0110110101101] [0110110101101]; ND points at Level 6 are the 1s that follow the chunks [10101101] [0110110101101]. Results are reported in Table 26 (RTs) and 27 (accuracy). From Figure 55, we do not observe any interesting trend in the trend of RTs across the blocks: no correlation appears between type of point (D vs. ND) and block. Instead, the difference in RTs in the three modalities is evident: RTs in the visual modality are overall shorter than both those in the auditory and tactile modalities, while we do not observe any differences among the latter. From the LMM model, we observed the following results: Factor *Block* was not statistically significant, indicating no difference in RTs between blocks. We found a main effect of *Modality* (χ^2 =402.98, df = 2, p < .001), indicating that there were significant differences in RTs between modalities. The *Point_Level_6*Block* interaction was not significant, indicating that RTs across blocks were not modulated by the type of point (D vs. ND). *Block*Modality* was not significant, meaning that RTs across blocks were not modulated by the type of modality. The interaction *Point_Level_6*Block*Modality* was not significant, meaning that the difference in the trend of RTs between D and ND points across blocks were not modulated by the modality. To have reconfirmation of the absence of correlations within individual modalities, we splitted data according to *Modality*, exploring the effect of *Point_Level_6* and *Block* in the three separated datasets (Auditory, Tactile, Visual). As expected, we found no significance, in any of the three modalities, except for the *Block* factor in the visual modality, which was found to be significant: RTs increased significantly along the task (χ^2 =9.86, df = 2, *p*=0.007208). We then assessed the interaction between type of point and modality to verify whether the differences in RTs between modalities were modulated by the type of point. To do this, we ran a LMM model with *RTs* as dependent variable, *Point_Level_6* (Disambiguated at level 6 vs. Non-disambiguated at Level 6) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =529.69, df = 2, p < .001), indicating the presence of significantly different RTs between modalities. Neither *Point_Level_6* nor the interaction between *Point_Level_6*Modality* were significant. Hence, we ran a simpler LMM model with *RTs* as dependent variable, *Modality* as independent variable, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =590.65, df = 2, p < .001).

Post-hoc comparisons showed that RTs in the visual modality were faster than those in the auditory (β=345.41; SE=16.5; t=20.93; p<.0001) and tactile modalities (β=341.05; SE=16.0; t=21.32; p<.0001). This result is in line with previous findings, indicating a general processing advantage for the visual modality, independent from learning.

	Block 1	Block 2	Block 3
RTs Disambiguated Points Auditory	670.33 (120.94)	693.12 (137.39)	692.48 (113.94)
RTs Non-Disambiguated Points Auditory	691.18 (90.05)	667.45 (90.28)	708.53 (102.21)
RTs Disambiguated Points Tactile	678.40 (84.47)	679.39 (106.20)	681.76 (125.46)
RTs Non-Disambiguated Points Tactile	688.42 (96.81)	666.61 (87.04)	669.88 (94.60)
RTs Disambiguated Points Visual	332.46 (81.15)	336.65 (98.31)	369.02 (130.39)
RTs Non-Disambiguated Points Visual	326.92 (72.31)	335.53 (99.77)	337.68 (110.57)

Table 26. Mean (SDs) RTs of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 6 in each Modality (Analysis 7).

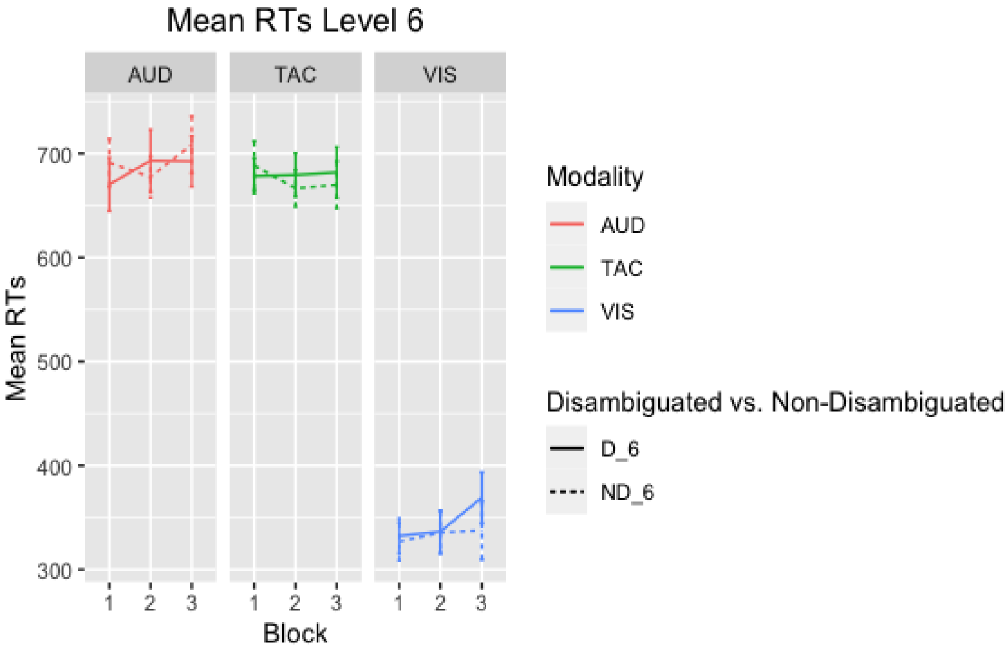


Figure 55. Mean RTs for D and ND points by block at Level 6 in the three studies (Analysis 7). Error bars denote the 95% confidence interval. D_6 = Disambiguated points at Level 6; ND_0 = Non-Disambiguated points at Level 6; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Summarizing, at Level 6, we did not find any significant result concerning learning. The only significant result found concerned the fact that RTs in the visual modality were overall shorter than those in the tactile and auditory modalities, indicating the presence of a processing advantage, independent from learning.

As observable in Figure 56, accuracy in the visual modality is higher than that in the auditory and tactile modalities. In general, we observe a decrease in accuracy rates in all three modalities, which is most noticeable for D points. The GLMM model failed to converge. For this reason, we reduced the complexity of the model. First, (i) we checked the interaction between type of point and modality to verify whether the differences in accuracy rates between modalities were modulated by the type of point. Then, (ii) we investigated the interaction between type of point and block within the individual modalities, to see whether the trend of accuracy rates across blocks were modulated by the type of point, within each modality. For analysis (i) we ran a GLMM model with *Accuracy* as dependent variable, *Point_Level_6* (Disambiguated at level 6 vs. Non-disambiguated at Level 6) and *Modality* (Auditory, Tactile, Visual) as independent variables with full interaction, and *Subject* as random intercept. The analysis revealed a main effect of *Modality* (χ2 =17.62, df = 2, p < .001), indicating the presence of significantly different accuracy rates between modalities. *Point_Level_6* was significant (χ2 =6.08, df = 1, p < .05. However, the *Point_Level_6*Modality* interaction was not significant. Hence, we ran a model with *Accuracy* as dependent variable, *Modality* (Auditory, Tactile, Visual) and *Point_Level_6* as independent variables, and *Subject* as random intercept. *Modality* was significant (χ2 =23.40, df = 2, p < .001), and *Point_Level_6* was significant as well (χ2 =18.07, df = 1, p < .001). We ran on this second model post-hoc comparisons, which showed that accuracy rates in the visual modality were higher than those in the auditory (β=-1.56; SE=0.32; z=-4.79; p <.0001) and tactile modalities (β=-1.20; SE=0.32; z=-3.72; p = 0.0006). This result is, again, in line with previous findings. In analysis (ii), we splitted data according to *Modality*, investigating the effect of *Point_Level_6* and *Block* in the three separated datasets (Auditory, Tactile, Visual). In the auditory study, we found a significant effect of *Block* (χ^2 =22.10, df = 2, *p*=<.001): accuracy rates decreased from Block 1 to Block 3 (β=1.53; SE= 0.35; z= 4.38; p <.001), and from Block 2 to Block 3 (β=0.94; SE=0.28; z= 3.33; p = 0.0025). In the tactile study, we found a significant effect of *Block* (χ^2=25.10, df =2, *p*=<.001) and *Block*Point_Level_6* (χ^2 =12.57, df = 2, *p*=<.01): accuracy rates decreased on D points from Block 1 to Block 3 (β=1.59; SE=0.42 ; z= 3.78; p=0.0005), and from Block 2 to Block 3 (β= 1.88; SE= 0.46; z= 4.10; p = 0.0001). D points were statistically less accurate than ND ones in Block 3 (β= -2.37; SE= 0.64; z= -3.69; p = 0.0002). In the visual study, we found a significant effect of *Block* (χ^2=11.36, df =2, *p*=<.01): accuracy rates decreased from Block 1 to Block 3 (β=2.93; SE= 1.05; z= 2.80; p=0.0142), and from Block 2 to Block 3 (β=1.68; SE=0.58; z=2.90; p =0.0104).

Overall, results on accuracy rates indicated a decrease in accuracy in all the modalities. Moreover, as at previous levels, accuracy rates in the visual modality were higher than those in the tactile and auditory modalities, confirming the presence of a general processing advantage independent from learning.

	Block 1	**Block 2**	**Block 3**
Accuracy Disambiguated Points Auditory	0.91 (0.28)	0.87 (0.34)	0.68 (0.47)
Accuracy Non-Disambiguated Points Auditory	0.93 (0.25)	0.87 (0.34)	0.85 (0.35)
Accuracy Disambiguated Points Tactile	0.91 (0.28)	0.93 (0.25)	0.72 (0.45)
Accuracy Non-Disambiguated Points Tactile	0.93 (0.26)	0.89 (0.31)	0.96 (0.20)
Accuracy Disambiguated Points Visual	0.99 (0.10)	0.98 (0.14)	0.88 (0.33)
Accuracy Non-Disambiguated Points Visual	1.00 (0.00)	0.98 (1.14)	0.95 (0.22)

Table 27. Mean (SDs) accuracy rates of each block for Disambiguates (D) and Non-disambiguated (ND) points at Level 6 in each Modality (Analysis 7).

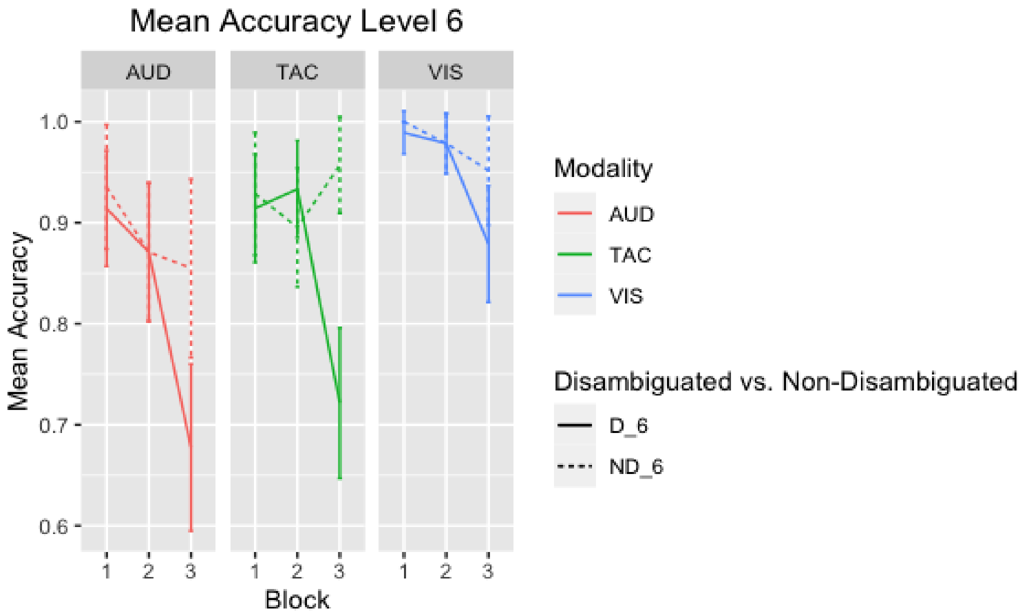


Figure 56. Mean accuracy rates for D and ND points by block at Level 6 in the three studies (Analysis 7). Error bars denote the 95% confidence interval. D_6 = Disambiguated points at Level 6; ND_6 = Non-Disambiguated points at Level 6; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

Analysis 8: Deterministic Vs. Non-Deterministic points within Level 7 in the Auditory, Tactile, and Visual studies

After conducting the analysis at Level 6, we went on to conduct the same analyses at Level 7. From the LMM model, we found that, at this level, the only significant effect was *Modality* (χ^2 =420.68, df = 2, p < .001), indicating that there were significant differences in RTs between modalities. As observable in Figure 57, in line with what has been observed at previous levels, the visual modality displayed lower RTs than the auditory and tactile modalities. To delve deeper into this effect, we ran a simpler LMM model with *RTs* as dependent variable, *Modality* as independent variable, and *Subject* as random intercept. We ran post-hoc tests on this simpler model, which showed that RTs in the visual modality were faster than those in the auditory (β=389.6; SE=15.3; t=25.43; p<.0001) and tactile modalities (β=345.05; SE=16.0; t=21.32; p<.0001). Moreover, RTs in the tactile modality were faster than those in the auditory (β=43.8; SE=14.9; t=2.93; p<.05). To have reconfirmation of the absence of correlations within individual modalities, we splitted data according to *Modality*, exploring the effect of *Point_Level_7* and *Block* in the three separated datasets (Auditory, Tactile, Visual). As expected, we found no significance, in any of the three modalities. Hence, we confirm that learning effects stopped at Level 5.

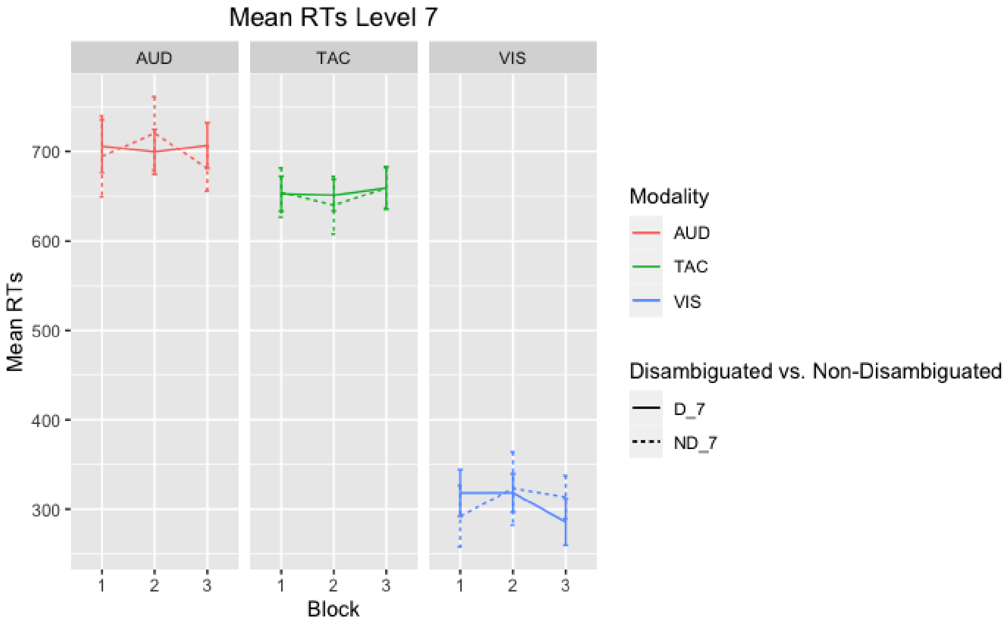


Figure 57. Mean accuracy rates for D and ND points by block at Level 7 in the three studies (Analysis 8). Error bars denote the 95% confidence interval. D_7 = Disambiguated points at Level 7; ND_7 = Non-Disambiguated points at Level 7; AUD = Auditory modality; TAC = Tactile modality; VIS = Visual modality).

In summary, our findings at Level 6 did not reveal any significant learning effect. However, we observed consistently faster RTs in the visual domain compared to both the auditory and tactile domains. This outcome aligns with previous findings and suggests a general processing advantage for the visual sphere, independent from learning, over the auditory and tactile spheres.

5.4. General discussion

In this chapter, we presented the findings from three AGL studies investigating the formation of recursive hierarchical abstract representations that emerge from sequentially presented, temporally fading input across auditory, tactile, and visual domains. Employing three distinct Serial Reaction Time tasks, participants across three groups were exposed to an identical sequence of binary stimuli governed by Fib rules. The stimuli varied across modalities: two pure tones for auditory, two colorful squares for visual, and two vibrotactile stimuli for tactile conditions. Our objective was to ascertain participants' proficiency in capturing Fibonacci string’s regularities across sensory modalities, leveraging the cognitive parsing mechanism detailed in Section *4.2*. Reaction times and accuracy rates were measured in correspondence to each D (disambiguated) and ND (non-disambiguated) point at every level (cf. Section *4.2.*). In all three studies the string corresponded to Fib generation 14, divided into 3 blocks of 178 stimuli each. Given the different frequency of the 0s and 1s in the string (cf. Section *4.1.*), we did not compare 0s and 1s directly, but rather focused on the 0s and 1s separately, going to see at each level whether and what the differences in RTs and accuracy rates were between the points that could be predicted at that level (D points) and those that could not be (ND points).

Hypothesizing learning, we expected to find a more pronounced decrease of reaction times on D points than ND points, throughout the task, possibly accompanied by increased accuracy rates. As explained in the data analysis section, we relied principally on RTs to determine significance, considering accuracy as a secondary measure to ensure comprehensive results and validate trends observed in reaction times. While hypothesizing a difference in the trend of RTs between D and ND points within levels, we did not rule out the possibility to find a decrease of RTs on ND points as well. Indeed, as we explained in Section *4.2.*, by definition, NDs are the points that are not predictable at each considered level. Crucially, however, this does not mean that they are not predictable at all. In fact, the set of ND points at *Level X* corresponds to the totality of points (D + ND) that are analyzed at *Level X+2*. In other words, ND points at *Level X* contain both the points that at

Level X +2 could be predicted (D points at *Level X+2*) and those that cannot be predicted at *Level X +2* (ND *Level X+2*). It therefore follows that the decrease in RTs on ND points at *Level X* may be attributable to the fact that part of these points (i.e., D points at *Level X +2*), are actually predicted. As we explained at the beginning of this chapter, if that were the case, we would expect to find D points at Level X displaying generally lower RTs than ND points at Level X, as the set of NDs include points that are potentially predictable at higher levels, being thus computationally more difficult to predict. Additionally, for the same reason, we would expect ND points started decreasing later in the task compared to D points. Fib strings are particularly well-suited for probing the formation of recursive hierarchical structures, exploring the shift from linear to recursive processing, and unraveling interactions among distinct mechanisms at varying abstraction levels (as elaborated in Chapter 4). This paradigm enabled us to shed light on the interplay between sequential implicit statistical learning and the formation of abstract recursive hierarchical representations in diverse sensory realms, directly comparing participants' performances across three different sensory domains. Building on Chapter 4's explanation, Fib strings allow the prediction of specific points by exploiting low-level statistical information: D points at Level 0 and D points at Level 1. In line with previous findings, which demonstrated low-level sequential statistical abilities in all three domains (cf. Section *3.1.4.*), we expected to find acquisition of D points at both Level 0 and 1 across the three sensory modalities via sequential implicit statistical learning. Specifically, since D points at Level 0 correspond to a fist-order transitional regularity (p (1|0) =1), while D points at Level 1 correspond to a second-order transitional regularity (p (0|11) =1), we predicted the former to be learned earlier than the latter. Based on the demonstrated auditory superiority in processing sequential statistical information over the visual domain (Saffran, 2002; Conway, Christiansen, 2005; 2009; cf. Section *3.1.2.*), we anticipated better auditory performance for these points than in the auditory domain. On the contrary, hypotheses for the tactile domain were open, since we found contrasting evidence concerning the comparison of this ability between the auditory and tactile domains (Conway, Christiansen, 2005; Pavlidou & Bogaerts, 2019; cf. Section 3.1.4.). Concerning the prediction of D points at Levels ≥ 2, our investigation centered on participants' ability to predict these points by exploiting the cognitive parsing strategy detailed in Section *4.2*. This strategy necessitates constructing recursive hierarchical representations, transitioning from the sequential to the hierarchical dimension. The goal of our investigation was to illuminate potentially domain-general and domain-specific aspects involved in this cognitive process. Aligned with prior studies (Martins et al., 2017; cf. Section *3.1.6.*), we expected to find evidence of this ability in the auditory domain. On the contrary, no specific expectations were set for the visual and tactile domains. Indeed, in the visual domain, previous studies focused on the investigation of the ability to represent recursion in *static* fractal images, with no exploration of the cognitive ability to form recursive hierarchical abstract representations from sequential stimuli (Martins, 2012; Martins et al., 2014; 2015; cf. Section *3.1.6.*). Similarly, the tactile domain lacked prior investigation into recursive (hierarchical) learning. Despite this, given the observed auditory advantage in processing sequential implicit statistical information over the visual sphere (cf. section *3.1.2.*) and considering that the ability to represent recursive hierarchical structures in our paradigm is intricately linked with the proficiency in sequential implicit statistical learning (cf. Section *4.2.*), we hypothesized to find auditory superiority over the visual domain in forming recursive hierarchical structures from sequential stimuli. However, we had no specific hypothesis concerning the tactile outcome. In summary, our study aimed to investigate the formation of recursive hierarchical structures from sequentially arranged stimuli in the auditory, tactile, and visual sensory domains. Specifically, we sought to unravel the relationship between sequential implicit statistical learning and the formation of recursive hierarchical representations, while also exploring potential domain-specific constraints in the process.

Overall, results indicated that our predictions have proven to be accurate: in line with our hypotheses, we found that the two low-level statistical regularities (D points at Level 0 and 1) were learned in all three modalities. Additionally, we observed domain-specific differences: as expected, the auditory domain proved to be superior to the visual domain, confirming findings from Saffran (2002), Conway, Christiansen (2005; 2009). Furthermore, an advantage of the auditory domain over the tactile domain was observed. Indeed, both at L0 and L1, the reaction times (RTs) for D points in the auditory domain decreased more steeply across blocks compared to the tactile and visual domains (i.e., the auditory learning curve was steeper). The comparison of D points between Block 1 and Block 3 within modalities at Level 0 and Level 1 revealed a more significant decrease in the auditory sphere, followed by the tactile sphere, and lastly, the visual spheres. This confirms a sequential statistical learning advantage for the auditory sensory domain over tactile and visual ones in acquiring both first- and second-order transitional regularities (D points at Level 0 and D points at Level 1, respectively). Additionally, it highlights a tactile domain advantage over the visual one. In addition to this, we also observed that D points at L1 were acquired later in the blocks in the visual modality (Block 2) as compared to the tactile and auditory modalities (Block 1). Our results align with findings from Saffran (2002), Conway, Christiansen (2005; 2009), demonstrating that the auditory domain is better at processing low-level sequential statistical information compared to the visual domain, which, in turn, is more suited for processing spatially arranged statistical information rather than temporally arranged (sequentially) information. As for the superiority of the tactile domain over the visual domain in processing first- and second-order transitional regularities in sequential input, our result is interesting, and novel compared to previous literature. In the face of contrasting results (Conway, Christiansen, 2005; Abrahamse, 2008; 2009; Pavlidou & Bogaerts, 2019), this is the first time we find a clear advantage of the tactile domain over the visual one. Regarding low-level statistical regularities (D points at Level 0 and 1), we also expected to observe that those at Level 0 would be learned before those at Level 1. In the auditory and tactile domains, we observed that both D points at Level 0 and Level 1 were learned within the first block. Therefore, from the conducted analysis, we cannot confirm whether D points at Level 0 were indeed learned before D points at Level 1. However, this is evident in the visual domain, where D points at L0 were learned in Block 1, while those at Level 1 were learned in Block 2.

Even regarding D points at levels ≥ 2, we have identified domain-specific learning differences. Specifically, we found similar learning performances between the auditory and tactile domains, while the visual domain demonstrated less proficiency in processing these regularities. Indeed, while we observed learning up to levels 4 and 5 in the auditory and tactile domains, respectively, in the visual domain learning stopped at Level 2. Additionally, at this level, we noted that learning occurred earlier in the auditory and tactile domains compared to the visual domain: in the second block in the auditory and tactile spheres, while in the visual domain only in the third block. In addition to this, the RTs curves of D points were steeper in the tactile and auditory modalities compared to the visual modality. As for D points at Level 3, we observed no learning effects in the visual modality, thus confirming that learning stopped at Level 2. On the contrary, in the auditory and tactile modalities, D points at Level 2 were learned in Block 2. D points at Level 3 were learned in Block 2 in the auditory modality, while in Block 3 in the tactile one. D points at Level 4 were learned in Block 3 both in the auditory and tactile modalities. At Level 5, we found no learning effects in the auditory modality, while in the tactile modality, D points were learned in Block 3. However, looking at the auditory RTs graph, we noted that there was a decrease in RTs from Block 2 to Block 3, although not significant. Thus, we do not rule out the possibility that the absence of learning effects at this level in the auditory modality was due to insufficient exposure to the string.

All in all, our results confirmed the ability to form recursive hierarchical abstract representations in the auditory domain, in line with what was found by Martins et al., 2017 (cf. Section *3.1.6.*). Crucially, our study also provided the first evidence that the visual and tactile domains are also able to form recursive hierarchical structures arising from sequentially presented input. Previous studies demonstrated the visual domain's ability to represent recursion (Martins, 2012; Martins et al., 2014; 2015), but they used paradigms that investigated this ability arising from *static* fractal images, rather than fading sequentially presented input. In other words, previously employed paradigms used fractal images, where recursive hierarchical structures unfolded in space, rather than in time. These studies showed that participants succeeded in tracking the interwoven hierarchical relationships between elements persisting in time, distributed in the spatial dimension, and to recursively apply these regularities across different hierarchical levels (cf. Section *3.1.6.*). In contrast, our study is the first to investigate the ability to form recursive hierarchical representation arising from sequentially presented input in the visual domain. Regarding the tactile domain, our study revealed pioneering results on the ability as well. Indeed, as we explained, to our knowledge, no study has ever investigated the ability to form recursive abstract representations in the tactile domain until now (cf. Section *3.2.*). Specifically, by designing a paradigm allowing direct comparison between the three sensory modalities, we were able to investigate possible domain-specific effects in the ability to form recursive hierarchical abstract structures from sequential stimuli, finding a clear advantage in the auditory and tactile domains over the visual domain. Overall, the obtained results corroborated our hypothesis that the auditory domain would outperform the visual domain in this particular skill. As we clarified earlier, the task at hand involves closely intertwining the ability to process recursive hierarchical structures with the proficiency in handling low-level transitional regularities. Hence, consistent with existing literature highlighting the auditory system's superior performance over vision in processing low-level transitional regularities within sequential input, we expected to find an advantage of the auditory over the visual domain also in forming recursive hierarchical structures from sequentially arranged stimuli. When it comes to the tactile domain, the absence of previous studies on this topic hindered us from forming specific predictions. However, aligning with the observed tactile learning advantage over the visual sphere in our study regarding the acquisition of first- and second-order transitional regularities, the tactile domain has proven to excel over the visual domain in the formation of recursive hierarchical structures as well.

Although, as explained, accuracy was not used to determine significance, the accuracy results still confirmed that there are learning differences between the three modalities. Despite the accuracy being high in this task, as expected, the interesting findings generally support the conclusions drawn from the RTs analysis. Just as RTs in the visual sphere were overall lower than those in the tactile and auditory spheres, we noted a specular trend in terms of accuracy rates as well: accuracy rates in the visual sphere were generally higher than those in the tactile and auditory spheres. Despite this general processing advantage of the visual sphere, going to observe the trend between blocks of accuracy rates on the two types of points (D and ND points), within the individual modalities, we found that it was the auditory sphere that performed best. In fact, in the tactile and visual spheres, apart from D points at L0 in the tactile sphere, in all other cases we observed that accuracy rates overall decreased along the task. This occurred on both D points and ND points. In contrast, in the auditory sphere, we observed an increase in accuracy rates on D points along the blocks at levels 0, 1, 2, and 3. The decrease in accuracy rates observed in the tactile and visual sphere might be linked to a fatigue effect due to the cognitive load required from the task, especially at higher levels. Crucially, however, despite the decrease, even in the tactile and visual spheres, the accuracy data have overall confirmed what was observed in the analysis of RTs. Specifically, in the tactile sphere, at levels 0, 1, 2, and 4, D points proved to be more accurate than ND points (at Level 0 and 1 from the first block, at Level 2 and 4 in the third block). The same pattern was observed in the visual sphere: At Level 0, D points were more accurate than ND points starting from the first block; at Level 1 in the first and second blocks; at Level 2 in the third block. Hence, overall, the results on accuracy rates generally supported the findings from the reaction time trends.

Regarding D points at levels ≥ 2, it is interesting to note that the observed learning effects align with the hypothesis that the human parser utilizes the Fib's cognitive parsing algorithm presented in Section *4.2.* As mentioned earlier, it is reasonably implausible to predict D points at level ≥ 2 in a SRT task using a flat statistical learning strategy (cf. Section *4.3.*). In order to predict points at different levels in a SRT task through a flat statistical learning strategy, the parser would need to simultaneously process multiple fading sequences, which overlap and progressively increase in length. This would impose a substantial workload, placing a considerable strain on human working memory resources and presenting a challenge to sustain effectively (cf. Section *4.3.*). On the contrary, the proposed Fib's cognitive parsing algorithm aims to be a more efficient cognitive strategy, reducing the workload on working memory. This strategy involves the formation of abstract hierarchical representations. Specifically, at levels ≥ 2, the mechanism by which the parser (human cognition) can incrementally predict points (i.e., D points) that would not have been predictable (i.e., ND points) at lower hierarchical levels is the recursive application of transitional regularities learned at levels 0 and

1, between increasingly larger embedded chunks (cf. Section *4.2.*). From the results obtained in our study, we have evidence that the parser applied the Fib's cognitive parsing algorithm. Indeed, we found confirmation that:

(i) The human parser processed D points differently than ND points at various levels. The former were generally learned earlier across the blocks than the latter and generally showed lower reaction times (RTs).

(ii) In every sensory modality, D points were processed differently across various levels, reflecting their computational complexity: D points at higher levels were learned later across the blocks compared to those at higher levels (or at most in the same block, but in any case, never earlier). Additionally, we noted that learning occurred incrementally, from lower to higher levels, with no cases of learning occurring at level n+1 in the absence of learning at level n. In other words, there were no learning jumps between levels. This result aligns with the Fib's cognitive parsing algorithm hypothesis. Indeed, we theorized that the prediction of D points at lower levels occurs before that of D points at higher levels - as it is computationally less complex – and, crucially, it is necessary for predicting D points at higher levels (cf. Section *4.2.*).

In conclusion, in our study, we found that all three sensory spheres can process both low-level sequential statistical information (i.e., D points at Level 0 and 1) and form abstract hierarchical representation to predict points that could not be predicted by exploiting a flat statistical learning strategy (i.e., D points at Level ≥ 2). Crucially, moreover, our results highlighted the presence of domain-specific differences. Specifically, participants showed much better learning performances in the tactile and auditory studies than in the visual one, in learning both low-level transitional regularities (D points at level 0 and 1) and higher-order transitional regularities which require the formation of recursive hierarchical representations (D points at levels ≥ 2). In particular, the auditory modality showed a major advantage over the tactile and visual spheres especially in tracking sequential low-level statistical information (i.e., D points at Level 0 and level 1). However, it was in the tactile sphere that we observed significant effects at the highest level, that is, up to Level

5. The result observed at L5 in the tactile study, however, should be interpreted with caution. In fact, as we have discussed, comparing the RTs graphs of the tactile and auditory studies at Level 5, we observe that in fact, RTs curves on D points in the two modalities show a very similar trend. Keeping in mind the fact that the higher we go with the levels, the fewer the points are, we do not rule out the possibility that the absence of learning found at specific levels might be related to the fact that subjects did not receive sufficient exposure to the regularities analyzed at those levels. It remains an open question as to whether a significant decrease in RTs could also be found at higher levels if participants were exposed to a longer Fib sequence.

The visual modality, instead, while showing overall significantly lower RTs and higher accuracy rates than the tactile and visual ones, turned out to be the least adept sensory modality at processing both sequential statistical information and abstract hierarchical representations arising from sequential stimuli. We attributed this result to a general processing advantage for the visual modality, independent from learning, over the auditory and tactile ones (cf. Abrahamse et al., 2009). This advantage is possibly linked to domain-internal factors such as more efficient communication channels connecting visual input processing and motor output, resulting in superior speed and accuracy. However, these are just speculations. Indeed, it is currently unclear to what this general processing advantage, independent of learning effects, can be attributed. One possibility might be related to our experimental design. In the visual task, the perceptual cues that participants could use to differentiate the two types of stimuli were twofold: color (blue square or red square) and spatial location (square presented on the right or square presented on the left). Specifically, the '0' of the grammar was presented as a red square that always appeared on the left side of the screen, while the '1' was a blue square that appeared on the right side of the screen. Thus, participants could rely on two visual perceptual cues to distinguish the stimuli: one related to color and the other to position on the screen. This contrasts with the tactile and auditory tasks, where only one cue was available to distinguish the stimuli. In the auditory task, the only cue was the different frequencies of the two stimuli ('0' of the grammar presented as a tone with a frequency of 333 Hz, while '1' was a tone of 286 Hz). There was no spatial cue, as both tones were presented to both ears. Similarly, in the tactile task, participants had only one perceptual cue to differentiate the stimuli. However, unlike the auditory task, the tactile task involved only a spatial discrimination. Participants felt the same vibration (120 Hz) either on the right thumb (for '1' of the grammar) or the left thumb (for '0' of the grammar). Thus, participants could only rely on a spatial perceptual cue (and not different intensities) for differentiation. In any case, participants could discriminate the stimuli based on a single cue in the tactile and auditory tasks, whereas they had two cues available in the visual task. Therefore, one possibility is that the presence of two simultaneous cues made the difference between the stimuli more salient, resulting in faster and more accurate overall performance. Future studies could focus on shedding more light on this point.

Another point we believe is worth highlighting concerns the results at L2 and L3 in the tactile experiment. In this case, we found that both at L2 and L3, D points decreased significantly across the blocks, confirming the occurrence of learning. However, at these levels, we did not find significant differences between D points and ND points. This partially contradicts our expectations. As we explained in Section *5.2*, if learning occurred using the cognitive parsing strategies proposed in Section *4.2.*, we would expect to see lower RTs or a more pronounced decrease, possibly occurring earlier across the blocks, for D points compared to ND points. Nevertheless, as we have discussed, finding a decrease in RTs for ND points is still consistent with the cognitive parsing strategies we proposed. Indeed, within the ND points at each level, we find the group of D points and ND points from level n+2 (cf. Section *4.2.*). Therefore, the fact that we found a decrease in RTs for ND points at both L2 and L3 is consistent with the fact that we found evidence of learning for D points at L4 and L5. Indeed, among the set of ND points at L2, there are the set of D points (plus the set of ND points) of L4. Similarly, among ND points of L3, we find the set of D points (and ND points) of level 5 (cf. Section *4.2.*).

Another interesting observation we believe is important to discuss is related to an effect found in the literature known as the *alternation advantage* (Bertelson, 1961; Fecteau et al., 2003; 2004; Gao et al., 2009; Williams, 1966). The alternation advantage is a cognitive phenomenon that can occur in Serial Reaction Time tasks with binary stimuli. It has been found that in these tasks, RTs might be influenced by cognitive biases unrelated to learning statistical regularities. For example, studies have shown that participants in a two-choice SRT task with randomized stimulus sequences tend to respond more quickly to alternating patterns (e.g., ABAB) compared to repeated patterns (e.g., AABB) (Soetens et al., 1985; Kirby, 1976). Interestingly, a recent study investigated the interaction between the alternation advantage and implicit statistical learning, shedding light on the cognitive sources underlying this phenomenon (Compostella et al., under review). This study provided evidence for the hypothesis that the alternation advantage can interfere with implicit statistical learning, further elucidating the cognitive sources of this effect. Specifically, it was proposed that shifts in (visuo)spatial attention play a role in the occurrence of the alternation advantage, and the perceptual dimension driving this mechanism is the spatial location of the stimulus. In other words, it has been found that the alternation advantage is related to the spatial characteristics of the stimuli that trigger shifts in (visuo)spatial attention before the stimulus onset. We propose that the alternation advantage may also have occurred in our study, specifically in the visual and tactile tasks. At first glance, our results seem to support this hypothesis. Indeed, in the tactile and visual studies, we observe that 0s have an advantage over 1s in terms of faster reaction times, in cases where the former corresponds to an alternation (i.e., 0 following 1; i.e., 1**0**) and in cases where the latter corresponds to a repetition (i.e., 1 following 1; i.e., 1**1**). This advantage appears to be independent of the learning of statistical regularities and the formation of recursive hierarchical representations, as it is present from the beginning of the task, being particularly pronounced in the first block, and then diminishing in subsequent blocks, where implicit statistical learning and the related formation of recursive abstract representations become more evident. Specifically, observing our data, we notice that in both the visual and tactile tasks, at L3, where D and ND points correspond to 0s following a preceding 1, i.e., an alternating stimulus (1**0**), the RTs are globally lower from the beginning of the task compared to the D points of L2 and L4, where D and ND points correspond to 1s following a previous 1, i.e. a repeating stimulus (1**1**). Crucially, this effect is observed in both the tactile and visual experiments, where the spatial dimension is an available perceptual feature that can be used to distinguish the two stimuli. Conversely, this effect is not present in the auditory task, where there is no spatial element to differentiate the two stimuli. Based on these observations, we believe it is important and interesting to further investigate this phenomenon in the future, conducting accurate and detailed statistical analyses to explore the phenomenon that, from initial observations, seems to align with and confirm the hypothesis put forward by Compostella et al. (under review). This hypothesis suggests that the alternation advantage arises from shifts in (visuo)spatial attention, triggered by the spatial arrangement of the two stimuli appearing in lateralized and opposite positions. Furthermore, the alternation advantage appears to be orthogonal implicit learning. Importantly, however, despite this cognitive bias, we observed that learning at various levels, following the proposed cognitive parsing algorithm, is confirmed. Summing up, on one hand, the tactile and visual tasks had spatial cues to distinguish between the two stimuli (for the tactile task, 0s in the grammar were presented on the left thumb, 1s on the right thumb; for the visual task, 0s were presented on the left side of the screen, 1s on the right side). In contrast, the auditory task lacked spatial cues, as both stimuli were presented to both ears. The presence or absence of a spatial perceptual cue seems to have influenced the overall RT results, giving an advantage in terms of shorter RTs to points corresponding to alternating stimuli (0 following 1) compared to stimuli featuring a repetition (1 following 1). Future studies could further explore this phenomenon. It would be interesting to replicate this study with experimental designs that remove the spatial dimension from the visual and tactile tasks, creating more comparable protocols across sensory modalities. This approach would provide a clearer measure of learning and comparison between the three sensory dimensions by eliminating the influence of the alternation advantage.

In conclusion, our study found results that strongly suggest participants processed the Fibonacci sequences and acquired regularities at different levels of complexity by using the cognitive parsing algorithm presented in Section *4.1.* As previously explained, it seems unlikely that the parser could have learned these regularities using a simple statistical learning strategy. The observed results align with the proposed patterns and hypotheses presented in our Fibonacci cognitive parsing algorithm (cf. Section *4.1.*). Importantly, the proposed cognitive mechanism has been shown to operate across various sensory domains, indicating it is a domain-general cognitive learning algorithm. However, as we have noted, we also identified domain-specific differences in how this cognitive algorithm is utilized.

An open problem, considered by Schmid (2023) but not analyzed in this thesis, concerns the nature of the hierarchical representations that the parser forms to predict points of increasing complexity at different levels. What happens to the parser's abstract hierarchical representation as it forms increasingly larger, recursively embedded chunks? Does the parser only retain the representation of the highest-level constituents while discarding the abstract hierarchical representation of the embedded sub-chunks? In other words, does the internal hierarchical structure of the constituents break down as hierarchical construction continues? Or, conversely, is the abstract representation of the sub-chunks maintained even when the parser embeds these chunks into larger chunks? As reported by Schmid (2023), several studies in the literature have assumed the hypothesis that during the abstract formation of chunks, the sequential steps taken to reach the chunks are erased, hence there is no record. This hypothesis is supported by several chunking models found in the literature (French et al., 2011; Goldwater et al., 2009; McCauley & Christiansen, 2014; Perruchet & Vinter, 1998; Robinet et al., 2011). Additionally, Schmid (2023) explains that according to the *subunit effect hypothesis*, once a chunk is learned, its subunits become less accessible. (Fiser & Aslin, 2005; Giroux & Rey, 2009; Orbán et al., 2008; Slone & Johnson, 2015; 2018). As Schmid correctly observes, an interesting analysis that could shed light on the nature of chunk representations would be to examine the trend of RTs within the hypothesized chunks, at different levels. Specifically, if the parser retains sub-chunks in memory, we should observe a deceleration in reaction times in correspondence to points immediately following the boundary of a chunk, as the sequence is processed from left to right (Schmid, 2023). This is undoubtedly an intriguing analysis that could be pursued in the future. Based on the results of our experiments, we are inclined to believe that the parser does retain sub-chunks in memory, even when creating embedded chunks. Indeed, in our analysis of D and ND points at various levels, we found that the RTs for D points at lower levels progressively decreased throughout the task, generally remaining lower than those for D points at higher levels. Importantly, the RTs for lower-level D points continued to decrease even after the higher-level D points had been learned. This might suggest that sub-chunks are maintained in memory. If this were not the case, we would expect to see that once higher-level D points are learned, the RTs for lower-level D points would progressively homogenize across different levels, eventually stabilizing at some point during the task. This hypothesis aligns with Schmid's observations (2023). However, it remains speculative, and future analyses could address this open issue.

In the next chapter, we will provide a broader context for our findings. We will explore the theoretical implications of uncovering a domain-general ability to form recursive hierarchical abstract representations from sequentially presented stimuli, and its close link to the ability to process low-level transitional regularities. Simultaneously, we will assess the potential causes of the observed domain-specific learning effects. Most importantly, we will conclude by discussing our findings within the framework of language acquisition, elucidating how our results contribute to understanding the fundamental mechanisms of language processing and acquisition.

6. Conclusion

In this thesis, we explored the ability to implicitly learn low-level statistical regularities and form recursive hierarchical abstract representations in sequentially arranged fading sequences of stimuli across three sensory domains: visual, auditory, and tactile. Our objective was twofold. Firstly, we aimed to elucidate the cognitive mechanisms involved in this process, with a specific focus on the transition from the linear to hierarchical dimension. Secondly, we sought to determine whether this ability is domain-general, present across all three sensory domains, and to explore any potential domain-specific differences. The choice to delve into this research topic stemmed from the observation of three crucial issues central to both theoretical and experimental linguistic discussions. The first concerns the role of recursion in the human language faculty. The second pertains to the role of abstract hierarchical representation and statistical learning in the acquisition and processing of human language, and their possible interplay. The third, closely linked, concerns the presence of domain-specific representational and learning constraints in language, alongside the role of domain-general learning abilities. Our journey into the exploration of this topic began with a comprehensive theoretical introduction to the linguistic debate, encompassing various theories and experimental findings in the current context. This literature review allowed us to clearly outline our research focus and develop a structured experimental design for investigating this research topic in the most effective manner.

In Chapter 1, we delved into the longstanding debate between nativist and usage-based approaches in the field of language acquisition. We examined arguments supporting both perspectives, with a particular emphasis on syntax acquisition. Then, we reviewed recent psycholinguistic studies that offered valuable insights into the role of implicit statistical learning in the acquisition and processing of syntactic phenomena. We also discussed findings from neural networks in the context of language acquisition, focusing particularly on recent deep neural networks that have achieved remarkable results in recent years. These results demonstrated the potential to create machines without innate language faculties that can learn linguistic abilities purely through exposure to linguistic data, by tracking statistical regularities in the data, with abilities almost on par with those of humans. Additionally, we presented studies providing compelling evidence of the richly structured and constrained nature of language, highlighting the existence of abstract structural representations during the complex processes of acquiring and processing syntax. Specifically, these abstract representations are hierarchical in nature. Crucially, these studies provided evidence for the fact that a purely sequential based model of learning fall short in capturing the core structure of human language syntax. Collectively, these findings suggested that language learners intricately rely on both surface-level statistical information and abstract representations of language structure. Consequently, we argued for the imperative acknowledgment, within contemporary language theories, of the pivotal role played by statistical learning, alongside the recognition of hierarchical boundaries and constraints. Importantly, we emphasized that there are various hierarchical phenomena within human syntax. Among these hierarchical phenomena, we explained that one of the most studied, debated, and yet controversial is recursion. Recursive embedding is thought to be a distinctive feature of human syntax, where a sentence can be embedded within another sentence, and a part of a structure can reflect the same organization as the entire structure. This capability allows for the creation of multi-level complex structures in which constituents are embedded within constituents of the same category, a remarkable feature of human syntax. As explained, this ability is considered by many scholars to be a unique aspect of human language syntax. Therefore, in this chapter, we outlined our research objective: to gain further insight into the role of recursion in human language syntax. Specifically, we aimed to investigate the mechanisms underlying this particular type of abstract hierarchical representation—namely, recursive hierarchical structures.

In Chapter 2, we focused on the concept of recursion. As discussed, despite the importance attributed to this phenomenon in linguistics, recursion was not clearly and universally defined for many years, leading to a proliferation of varied and sometimes conflicting definitions and causing significant terminological confusion. In this chapter, we aimed to provide a clear definition of recursion in cognitive science and linguistics. We defined recursion as the embedding of elements within other elements of the same type. We also made a clear distinction between iteration without embedding, iteration with embedding, and recursion. Specifically, we differentiated between types of recursion, including tail recursion and nested recursion. Importantly, we related each of these concepts back to linguistic phenomena, offering examples of different types of recursion as well as different types of iteration (both with and without embedding). Following our clarification of the concept of recursion, we critically examined the hypothesis by Hauser, Chomsky, and Fitch (2002) that recursion is a distinctive feature of human language, possibly absent in other cognitive domains and non-human species. This hypothesis asserts that recursion is a defining and universal trait of human language, setting it apart from other cognitive processes and non-human communication systems. However, our investigation uncovered several challenges to this view. While recursion is a key aspect of linguistic theory, we have seen that its prevalence in everyday language may not be as widespread as initially claimed (Karlsson, 2010; Verhagen, 2010). Indeed, studies suggest that complex recursive structures are uncommon in both spoken and written language. Additionally, the existence of languages like Pirahã, which convey complex ideas without recursion, questions the notion that recursion is essential to all linguistic systems. Moreover, other features of language, such as structure-dependence and duality of patterning, also contribute to the uniqueness of human language and can function independently of recursion (Kinsella, 2010). We also provided an overview of studies that explored recursion's role in non-linguistic cognitive domains. Recursion appears in processes such as numerical reasoning, navigation, and music, though its necessity is debated. For instance, while recursive strategies can be useful in navigation, they are not the only possible methods (Parker, 2006). Conversely, some non-linguistic domains, like music, visual perception, social cognition, and theory of mind, show clear instances of necessary recursion (Parker, 2006). Overall, our analysis suggested that while recursion is an important feature of human language, it is not uniquely linguistic and may stem from broader cognitive capacities. In summary, this chapter has shown that while recursion has traditionally been considered a unique and innate feature of human language, recent studies challenge this view, suggesting a more nuanced perspective. Building on these observations, this thesis aimed to investigate whether recursion is solely a linguistic trait or if it extends beyond language. Evidence indicates that human language learning involves cognitive biases related to the boundaries and constraints of the acquisition process. Various hypotheses propose that these biases may stem from either domain-general cognitive processes or be specific to language. Hence, we specified that our research seeks to clarify the extent to which recursion is domain-specific or domain-general by examining this ability across different sensory domains. After discussing the concept of recursion, we explored the intricate relationship between linear order, hierarchy, and their interaction in human language. We emphasized the importance of understanding the linear, temporal dimension to fully grasp the mechanisms of human language. We examined the historical debate within linguistic theory about the role of linear order, particularly focusing on Kayne's (1994) work, which highlights the close connection between linear order and hierarchical structure in syntax. Our discussion also addressed the broader cognitive implications of Kayne's theory, underscoring how crucial it is to consider both linear and hierarchical aspects to fully understand language processing and acquisition. By integrating insights on the importance of sequentiality with our study of recursion, we clarified the central focus of this thesis: the ability to process and generate recursive hierarchical abstract representations from sequential arrays of symbols. We posit that this cognitive ability is a key feature of human language syntax. In this context, we reviewed studies on cognitive mechanisms involved in processing sequential stimuli, from basic statistical computations to complex hierarchical representations, which explored the relationship between the different mechanisms at work in the process. We concluded by distinguishing between two types of recursion: one arising from temporally ordered sequences, typical of language and music, where the temporal dimension is primary, and another from spatially arranged stimuli, relevant to image processing, where the spatial dimension dominates. Therefore, a key focus of this thesis is to investigate how recursive hierarchical structures develop from temporally ordered sequences of stimuli.

In the second part of the chapter, we introduced the Artificial Grammar Learning (AGL) paradigm, a valuable tool for investigating implicit statistical learning. We found this tool particularly useful for exploring the research topics of this thesis and have thus employed it in our investigation, as detailed in Chapter 5.

Firstly, we explored various types of grammar within the Chomsky Hierarchy, which are commonly used in Artificial Grammar Learning (AGL) studies. We then reviewed studies on the ability to form recursive hierarchical abstract representations and identified a significant issue in the literature. In addition to considerable confusion about the concept of recursion—characterized by a lack of a clear, unified definition—we noted frequent misuse of artificial languages for studying recursive hierarchical structures. A notable example is the frequent use of the A^nB^n artificial language in recursion studies. However, as discussed, many of these studies only demonstrated supra-regular computational abilities without conclusively proving recursive capability. As a result, despite numerous attempts to test recursive abilities in AGL, there is a shortage of clear, irrefutable empirical evidence demonstrating this ability. We highlighted several crucial considerations to take into account for designing studies aimed at examining recursion. First, it is not enough for the tested language to be generated through a recursive process; participants may use non-recursive methods to process and learn the language. We emphasized the need to distinguish between algorithmic properties and the representational aspects of recursion, including what Martins (2012) refers to as distinctive signatures of recursion—such as depicting previously undefined dependency relationships or representing information within new hierarchical levels. Furthermore, we stressed the need for appropriate tools to study recursion in non-linguistic domains, acknowledging the challenge of the current shortage of suitable tools. We concluded the chapter by highlighting the potential of exploring recursion using grammars beyond the Chomsky Hierarchy. This introduction set the stage for our experimental study (Chapter 5), where we employed a grammar from the Lindenmayer Systems: the Fibonacci grammar.

In Chapter 3, we examined the relationship between the cognitive ability to form recursive hierarchical abstract representations and perception. Indeed, one of the key aims of this thesis is to determine whether the ability to generate recursive hierarchical abstract representations from sequential stimuli is domain-general or modality-dependent. Is the ability to form recursive hierarchical abstract representations from sequential stimuli a stimulus-dependent or a modality-based skill? Does it involve a single mechanism shared across domains, or are there modality-constrained mechanisms? Given that recursive hierarchical abstract representations might be formed from sequential stimuli of different sensory modalities—such as visual, auditory, or tactile—we aimed to investigate whether there are differences in the process across these three sensory domains. Could hearing excel over touch and vision in forming recursive hierarchical abstract representations arising from sequentially ordered stimuli, considering its crucial role in language and music processing, which are primarily conveyed through the auditory channel? Are we more proficient at learning and processing these structures in the auditory domain? Conversely, vision might demonstrate inferior abilities in handling these structures, having on the opposite greater proficiency forming recursive hierarchical abstract representation arising from static, spatially arranged stimuli, compared to sequential ones, since static hierarchical structures are primarily formed in the visual domain, as in the case of image processing. Indeed, when viewing an image, our visual system organizes static information hierarchically, allowing us to perceive the entire image composed of numerous hierarchically organized pixels, contributing to our comprehensive perception of the visual scene. And what about touch? Touch may excel in processing sequential rather than static hierarchical information. Indeed, detecting the shape of an object solely through simultaneous pressure on the skin is challenging, yet when the object is touched with a moving point or explored through tactile scanning, its shape becomes distinguishable (Lashley, 1951). As we explained, our phenomenological observations would lead us to hypothesize that the ability to form recursive hierarchical abstract representations from sequential (i.e., temporally ordered) stimuli might be more robust in the auditory or tactile domains, while the formation of recursive hierarchical abstract representations from static (i.e., spatially arranged) stimuli could be more robust in the visual domain. Alternatively, this ability might be stimulus-independent, allowing us to process these structures equally across visual, auditory, and tactile domains. However, in formulating our hypotheses, we did not solely rely on phenomenological discussions and speculations. We delved into the literature to explore whether previously conducted studies could provide further insights into the ability to form recursive hierarchical abstract representations in these three different sensory domains. This was our aim in Chapter 3. However, we found that no study has comprehensively explored this topic so far. Therefore, we dissected the phenomenon by conducting a review of studies that have investigated domain-specific spatiotemporal structure effects in implicit statistical learning of low-level transitional regularities in different sensory domains. Indeed, we believe that the ability to acquire low-level transitional regularities is a fundamental step in processing sequential stimuli, essential for subsequently creating recursive hierarchical abstract representations. Importantly, we did not find any studies that have investigated the presence of spatiotemporally domain-specific constraints in the tactile domain, while we found studies that have explored this question in the visual and auditory domains. This allowed us to verify whether previous studies had found evidence regarding the possible superiority of the visual domain over the auditory domain in the implicit statistical processing of spatially arranged stimuli and/or the superiority of the auditory domain over the visual domain for temporally, sequentially arranged stimuli. While no studies have examined the presence of spatiotemporal constraints in the tactile domain, we came across several recent and intriguing studies that delved into tactile sequential implicit statistical learning. These studies compared this ability with implicit statistical learning in the visual and/or auditory domains. Finally, we also examined studies exploring the ability to form recursive hierarchical abstract representations across different sensory domains. Crucially, we did not find any study investigating recursion in the tactile sensory domain. Conversely, we did find intriguing studies exploring this ability in the visual and auditory domains. Regarding low-level implicit statistical learning, research findings have indicated the presence of spatiotemporal domain-specific constraints: the auditory modality has been found to be superior to the visual modality in processing statistical information when stimuli were sequentially arranged (i.e., in the temporal dimension). Conversely, the visual modality demonstrated greater proficiency in learning when information was presented spatially rather than sequentially (Conway, Christiansen, 2005; 2009; Saffran, 2002). Hence, overall, results indicated that, concerning the processing of sequential statistical information, the auditory domain outperforms the visual domain. These results, as we have seen, contributed to the formulation of the Auditory Scaffolding Hypothesis (Conway et al., 2009). According to the theory, sound acts as a cognitive support or "scaffolding," aiding the development of general capacities for recalling, producing, and learning sequential information. Hence, Conway and colleagues' theory emphasizes the significant role of sound exposure in shaping cognitive abilities related to temporal and sequential patterns. The authors provided two sets of evidence supporting the theory: (i) congenitally deaf individuals show non-auditory sequencing abilities, and (ii) hearing populations exhibit modality-specific constraints, with better performance in sequencing tasks when the sense of hearing is involved rather than sight. However, recent studies, as discussed in Chapter 3, contradicted the first point, showing that deaf populations can successfully learn domain-general sequential information (Giustolisi et al., 2022; Giustolisi & Emmorey, 2018; Hall et al., 2018; von Koss Torkildsen et al., 2018 and Terhune-Cotter et al., 2021). Regarding the second point, we have pointed out that, while it has been demonstrated that hearing has an advantage in acquiring statistical sequential regularities compared to vision (Conway, Christiansen, 2005; 2009, Saffran, 2002), introducing a third variable changes the perspective on this advantage. Indeed, we have seen that contrasting results have emerged when comparing the auditory and tactile domains in processing sequential statistical information: Some studies suggested auditory superiority (Conway, Christiansen, 2005), while others suggested tactile superiority (Pavlidou, Bogaerts, 2019). Regarding the ability to form recursive hierarchical abstract representations, we found consistent evidence confirming this capacity in both the visual (Martins et al., 2014; 2015) and auditory domains (Martins et al., 2017). Overall, these findings suggest that recursion is a domain-general cognitive skill. However, upon examining these studies, we highlighted two critical considerations. Firstly, there is a distinction in the paradigms used to study recursion in the visual and auditory domains. Visual studies focused on static recursive structures presented spatially in fractal figures, while auditory studies centered on dynamic, sequential recursive structures heard over time. Despite evidence linking these abilities (Martins et al., 2017), it is crucial to acknowledge that they may involve different cognitive skills to some extent. Secondly, we underscored the significant gap in research on recursion in the tactile domain. This gap hampered our understanding of how recursion functions across various sensory modalities. In conclusion, we noted a shortage of studies that have developed methods to directly assess and compare the capacity to form recursive hierarchical abstract representations arising from sequentially presented input in the auditory, visual, and tactile sensory realms.

In Chapter 4, therefore, we introduced a grammar that is well-suited for developing a framework capable of directly comparing the three sensory domains in this ability: the Fibonacci grammar (Fib). Fib is a simple recursive rewrite system which is composed of only two symbols (0 and 1) and two rewriting rules (0→ 1; 1→01, i.e. 0 rewrites as 1; 1 rewrites as 01). By repeatedly applying these rewriting rules, we generate strings of 0s and 1s, potentially of infinite length. Fib binary sequences can be encoded onto different types of perceptual stimuli, allowing for the creation of directly comparable paradigms across different sensory modalities. Moreover, as we explained, the peculiar features of the Fibonacci sequence, such as self-similarity and aperiodicity, make it an optimal tool for studying how we form recursive hierarchical abstract representation arising from sequential stimuli, illuminating the entire process from sequence to hierarchy. To make the best use of these properties of Fib and accurately investigate recursion, we explained that it is crucial to choose an appropriate experimental paradigm and design. In our case, as we discussed, this is possible by adopting the Serial Reaction Time (SRT) task, which we introduced in Section *2.3.2*. Crucially, in SRT task where a sequence corresponding to a full generation of Fib is sequentially presented, we explained that participants can disambiguate (i.e., predict) specific points in the sequence by tracking low-level transitional regularities. However, for the reasons we have explained, there are points that we believe are implausible for the human parser to predict using a flat statistical strategy. Instead, these points require the creation of abstract hierarchical representations to be accurately predicted. In this context, we proposed a cognitive parsing algorithm specifically designed to process Fibonacci strings in an SRT task. This algorithm suggests how the human parser can learn points of varying cognitive complexity, starting from those hypothesized as simpler, which involve acquiring two low-level (first- and second-order) transitional regularities, to more complex ones, which involve forming increasingly larger embedded chunks and tracking conditional statistical information between these chunks. Importantly, we clarified how this could occur, namely through a recursive cognitive strategy, where the two low-level transitional regularities are applied across different hierarchical levels to incrementally form larger chunks and tracking transitional regularities between them. We also discussed why we think this algorithm may be the most compatible with human cognitive abilities when it comes to predicting points of increased complexity within Fibonacci sequences in a SRT task, discounting other potential mechanisms. In conclusion, we summarized the key findings from existing AGL studies with the Fibonacci grammar.

In Chapter 5 we presented the experimental design and results of our AGL study. Our study is the first to provide evidence and directly compare the ability to form recursive hierarchical abstract representations arising from sequentially presented input across the visual, tactile, and auditory sensory domains. Specifically, our dual objective was to (i) ascertain whether this ability is domain-general and shed light on any potential modality-specific learning differences; (ii) elucidate the computational mechanisms underlying the acquisition and processing of these structures, with particular attention to the relationship between sequential statistical learning and the formation of recursive hierarchical abstract representations. Three groups of adults participated in the study, each engaging in either a visual, auditory, or tactile experiment. In all three experiments, participants were exposed to the same sequence of stimuli, determined by the rules of the Fibonacci grammar. The grammar's symbols (0 and 1) were transmitted through different types of stimuli. In the auditory experiment, participants listened to two pure tones of equal amplitude but different frequencies via Bluetooth bone conduction headphones. In the tactile experiment, participants felt two gentle vibro-tactile impulses transmitted to their thumbs through the same headphones. In the visual experiment, participants observed sequential presentation of two colorful squares (blue or red) on a computer screen, appearing either to the right (for red squares) or left (for blue squares) of the screen. Participants underwent individual testing sessions. They were briefed that they would encounter a binary sequence of stimuli and were directed to respond to these stimuli by pressing designated keys on a computer keyboard swiftly and accurately. Not until the conclusion of the experiments were participants apprised that the stimulus sequence followed a non-random pattern. They were then queried about whether they had discerned any patterns during the task. In all three tasks, we measured participants' reaction times (RTs) and accuracy rates of responses on every point along the sequence. This included points we predicted could be anticipated through the proposed cognitive parsing algorithm, as well as those that could not be predicted, at each level of increasing hierarchical complexity as hypothesized by us. We then compared these data between predictable (referred to as *disambiguated*, *D Points*) and unpredictable (referred to as *non-disambiguated*, *ND Points*) points within each level. Our aim was to observe if there was an improvement in performance –specifically, a decrease in RTs, possibly accompanied by an increase in accuracy rates - by participants on predictable points throughout the task, or if they exhibited better performance compared to unpredictable points. This improvement would suggest learning of the predictable points.

Overall, our findings indicated that:

(i) The ability to form recursive hierarchical abstract representations arising from sequential stimuli is closely associated with the ability to grasp low-level transitional regularities.

(ii) Additionally, we showed that the cognitive capacity to form recursive hierarchical abstract representations from temporally ordered (i.e., sequential) stimuli is a domain-general ability. However, we also discovered domain-specific differences. While we found evidence of this ability across all three sensory domains, we observed a distinct advantage in the auditory and tactile domains compared to the visual domain.

In the next sections, we will thoroughly examine these two primary findings, striving to contextualize them within a broader perspective and discuss the theoretical implications of our results.

Processing and forming recursive hierarchical abstract representations from sequential fading stimuli. From linear order to hierarchical dimension: Statistical learning bootstraps hierarchical structure

The aim of our study was to uncover the cognitive foundations underlying human language faculty. Specifically, we sought to elucidate the mechanisms that enable the formation of recursive hierarchical abstract representations from sequentially arranged stimuli. The formation of such structures is believed to underpin complex cognitive phenomena in domains such as language (e.g., recursive syntactic phenomena) and music (e.g., key change modulation) (see Sections *2.1.2.*; *2.2.*; *3.2.*). Thus, we aimed to shed light on the cognitive mechanism behind this ability, observed across various cognitive domains. What are the foundations of the ability to build recursive hierarchy from sequential stimuli? Can we illuminate the mechanisms underlying the transition from the linear to the hierarchical dimension? Is this ability strictly linguistic or domain-general? By developing an AGL study with the Fibonacci grammar, we had the opportunity to investigate this ability in the absence of other characteristic features of language. Our paradigm excluded prosody, morphology, and semantics, allowing us to explore the deeper, possibly domain-general mechanisms of this cognitive ability. In our study, we demonstrated that humans possess the cognitive ability to create recursive hierarchical abstract representations from sequences of fading symbols, even in a non-linguistic context. Importantly, since this context was devoid of meaning, the ability to form recursive hierarchical abstract representations from sequential stimuli was shown to be independent of semantics and instead emerged purely from statistical learning phenomena and categorization. The fact that we achieved this result using the simplest type of temporal sequences, namely binary sequences, aligns with the findings of Planton et al. (2021) and supports Fitch's "dendrophilia hypothesis." This hypothesis posits that humans possess a multi-domain capacity and a natural inclination to infer tree structures from strings, even in the most straightforward scenarios (Fitch, 2014).

In our study, we have elucidated how sequential implicit statistical learning, and the formation of recursive hierarchical abstract representations arising from sequentially arranged stimuli are integral components of a unified cognitive process, representing a continuum from sequential to more abstract hierarchical processing, transitioning from the linear to the hierarchical dimension. Within our cognitive parsing algorithm (cf. Section *4.2.*), we posited that the cognitive mechanisms outlined in Dehaene et al.'s taxonomy (2015) are interlinked abilities in the formation of recursive hierarchical structures. Specifically, we proposed that sequential statistical learning, chunk formation, categorization, and the formation of abstract (recursive) hierarchical representation are computationally intertwined procedures, with each outcome serving as input for the next, progressing from simpler sequential mechanisms to more complex abstract hierarchical ones. Our experimental findings substantiated this (cf. Section *5.4.*), being in line with what has been found by Planton et al. (2021) and Radulescu et al. (2019). Indeed, we found evidence that the human parser processed the Fib sequence in this manner, predicting points of increasing complexity incrementally using our proposed cognitive parsing algorithm. Thus, in our study, we found that sequential statistical learning and the formation of recursive hierarchical representations can coexist; one does not preclude the other. On the contrary, our study sheds light on how sequential statistical learning (i.e., the acquisition of low-level transitional regularities) underpins the formation of recursive hierarchical abstract representations. In other words, sequential statistical learning serves as the foundation for the formation of recursive hierarchical abstract representations. Summing up, our experimental study revealed that statistical learning is essential for forming nested recursive structures. It is crucial for segmenting, chunking, and categorizing sequences, which then allows for the formation of recursive hierarchical structures of chunks. This leads us to suggest that these processes occur in language as well. We think that the human cognitive bias towards hierarchy and categorization observed in our study reflects a fundamental aspect of language. Our results illuminate the cognitive mechanisms underlying the domain-general, not exclusive to language ability to create recursive hierarchical abstract representations from sequentially arranged stimuli. Crucially, we posit this ability to be at work across various cognitive domains, including language and music.

However, it is important to highlight that language is inherently more complex than the sequences we used in our AGL study. In language, the formation of recursive hierarchical abstract representations is influenced by other language-specific factors such as semantics and prosody, which play major roles in the cognitive mechanism that transitions from sequence to hierarchy, such as chunking, categorization, and forming nested structures. Despite this, a fundamental aspect we believe to be language-independent, and shared between language and music, is the various mechanisms we demonstrated to be active in the transition from linearity to the formation of recursive hierarchical structures. We assert that, regardless of the cognitive domain (e.g., music or language), low-level statistical learning, as well as the formation of chunks and their categorization are closely linked and necessary steps for the formation of recursive hierarchical structures. Importantly, as we explained in Section *4.2.*, this process could not occur without categorization. Indeed, we believe categorization is an essential step for creating (recursive) hierarchical structures. In our experiment, categorization was based on perceptual attributes such as repetition, alternation, and distributive phenomena. Similarly, we think that in language, statistical distributive phenomena play a major role in chunking and categorizing, but these mechanisms rely on the presence of other factors, such as semantics, prosody, and morphology. Specifically, we think that categorization in language is inherently more complex due to its heavy reliance on semantics. Therefore, semantics plays a fundamental role in the formation of nested recursive structures in language. Semantics in language is undoubtedly a foundational element of recursive hierarchical structures. Crucially, we assert that without semantics, recursion in language might neither be necessary nor possible. The fundamental role played by semantics in the existence of recursive hierarchical structures in language has also been emphasized by Parker (2006). In this vein, Parker (2006) highlighted that while computer science acknowledges the iterative implementation of recursive algorithms, natural language processing does not function similarly. In natural language, semantics differentiates tail recursion from iteration, indicating a structural complexity not visible from the string alone. Tail recursion's strict ordering requirement, absent in iteration, underscores the importance of semantics in identifying the correct structure. An iterative description of sentence structure fails to capture the complex meanings they convey, demonstrating that semantics provides the necessary information to distinguish between iteration and recursion (cf. Section *2.1.1.*). However, differently from Parker (2006), which suggested that recursion in natural language might stem from the need to communicate recursive thought, we propose that the underlying cause is more aligned with efficiency and simplicity mechanism, driven by communicative needs, as we will discuss in more detail in the next section. Thus, in our hypothesis, recursion in language arises from a force driven by communicative needs, where a complex conceptual system must be channeled into the sequential and temporal nature of communication through the powerful yet finite cognitive capacities we possess as humans. Exploiting a recursive hierarchical mechanism, though seemingly complex, would be more efficient than using a flat, iterative algorithm, given the possibilities and limitations of human cognition. Regarding the question posed by Dehaene and colleagues (2015) concerning how the brain determines the optimal processing mechanism for a sequence, we believe that the parser employs increasingly complex and abstract mechanisms until it reaches the point where it finds the mechanism that allows for the elimination - or at least minimization- of prediction errors. In our case, the parser continued until it reached the final mechanism proposed by the taxonomy, namely the formation of recursive hierarchical representations, as this mechanism, given the properties of the Fib sequence, is the one that allows for the prediction of the greatest number of points and minimizes prediction errors. Overall, our hypotheses are fully aligned with the findings of Planton et al. (2021), who concluded that chunking and creating recursively embedded representations are essential for explaining human behavior when working memory capacity is exceeded and compression is most beneficial. They also align with Radulescu et al.'s hypothesis (2019), according to which the shift from low-level item-bound computations to rule induction and the formation of abstract categorization is an encoding mechanism gradually driven as an automatic response by the brain's sensitivity to input complexity (entropy) interacting with the limited encoding capacity of the human brain (channel capacity) (cf. Section *2.2.1.*).

The point raised by Radulescu and colleagues concerning the role of cognitive limitations (i.e. channel capacity) is particularly intriguing. We believe it would be both valuable and crucial for future research to focus on these aspects to gain further insights. Specifically, we think that exploring cognitive limitations and their relationship with the creation of recursive hierarchies in language is of significant interest. This exploration could also illuminate potential differences between the functioning of human language and the operations of modern large language models, as discussed in Section *1.2.3*. In light of Piantadosi's viewpoint, a key challenge for future research is to enhance models by incorporating architectural biases and principles that align more closely with human cognitive constraints. According to Piantadosi (2023) this might involve developing learning models that mimic the cognitive limitations observed in human learners. This approach is reflected in initiatives such as "The BabyLM Challenge" (Warstadt et al., 2023), which seeks to create models that can learn effectively from a developmentally realistic amount of data. As suggested by Piantadosi (2023), investigating the feasibility of efficient learning with limited resources and data, potentially through minor architectural adjustments, remains a compelling scientific question.

Another intriguing avenue for future research could be conducting experiments with Fibonacci sequences using modern large language models (cf. Section *1.2.3*.). Exposing modern large language models to Fibonacci grammar sequences could provide insights into the fundamental mechanisms behind these models. Specifically, if these models are designed to mimic human learning processes, their ability to learn from Fibonacci sequences should be comparable to human performance. This means that, while they might initially require exposure to longer sequences or more extensive training, they should ultimately achieve similar learning outcomes as humans. If the models do not perform comparably, it would suggest that their learning mechanisms differ from human cognition. Given the parallels between processing Fibonacci sequences and human syntax, such experiments could reveal whether large language models generate language through mechanisms distinct from those of human cognition. This could help us understand if these models truly replicate human-like learning or operate using different strategies.

To conclude, we think our findings can offer intriguing insights relevant to linguistic theories. Firstly, our results align with what we discussed in Sections *1.2.1.*; *1.2.2.*, namely that recent psycholinguistic experiments demonstrate both statistical learning and the formation of abstract hierarchical representations to be crucial in the acquisition of syntactic linguistic phenomena. Furthermore, our findings are consistent with Kayne's (1994) assertion that linear order and hierarchical dimension are closely intertwined. Therefore, our study aligns with and contributes to these perspectives by not only confirming the presence of these two abilities but also demonstrating their close connection and shedding light on the cognitive mechanisms involved in the transition and projection from linear to hierarchical dimensions. In essence, we believe our study provides interesting insights on the cognitive foundations of language. Regarding the AGL studies previously conducted with the Fibonacci grammar, our results are consistent with observations made by Schmid et al. (2023) and Vender et al. (2023). Specifically, our finding that statistical learning bootstraps the formation of hierarchical representations aligns with the proposals put forth by both studies, albeit these two studies proposed different underlying mechanisms for the projection from linear to hierarchical dimension (i.e., Bootstrapping Principle in Vender et al. 2023; merge of recursively deterministic transitions in Schmid et al. 2023; cf. Section *4.4.*). Moreover, our study serves as an intriguing follow-up to these previous studies. Indeed, we investigated how Fib sequences are processed in sensory domains beyond the visual domain (which was the sole domain investigated by Schmid et al., 2023, and Vender et al., 2023; cf. Section *4.4.*). Our investigation comprised the auditory and tactile modalities as well, where we observed learning at higher hierarchical levels than those found in Schmid et al. (2023) and Vender et al. (2023). By doing so, we shed light on domain-specific and domain-general aspects of learning, as discussed in further detail in the following section.

Processing and forming recursive hierarchical abstract representations from sequential fading stimuli: Domain-general ability with domain-specific learning differences

> "*On theoretical grounds we could expect complex systems to be hierarchies in a world in which complexity had to evolve from simplicity*."
>
> Simon Herbert – The Architecture of Complexity

Going into more detail, regarding (ii) (p.344), in our study, we found that the auditory domain displayed an advantage over both the visual and tactile domains concerning the acquisition of low-level sequential regularities. In turn, the tactile domain demonstrated superiority over the visual domain. However, when we moved to more complex and abstract levels of sequential processing requiring the formation of recursive hierarchical abstract representations, we found interesting results: the auditory sphere maintained a learning advantage over the visual sphere. Crucially, however, the tactile sphere turned out to be as efficient as the auditory sphere, showing similar learning trends across reaction times (RTs) data. As for the tactile sphere, we observe that it displayed a clear learning advantage over the visual sphere, both regarding the acquisition of low-level statistical regularities and the acquisition of more complex sequential regularities that require the formation of recursive hierarchical abstract representations. However, while the auditory and tactile domains demonstrated a clear learning advantage over the visual domain, we noted that the visual domain exhibited lower reaction times and higher accuracy rates overall. We ascribed this outcome to two possibilities. The first is a general processing advantage for the visual modality, independent of learning, over the auditory and tactile modalities. This advantage may stem from internal factors within the domain, such as more efficient communication pathways linking visual input processing and motor output, leading to superior speed and accuracy. The second possible cause of this effect could be related to our experimental design. Indeed, to discriminate between the two visual stimuli, two perceptual cues (different color and different location on the screen) were available. In contrast, for the auditory and tactile modalities, only one perceptual feature was available to distinguish between the stimuli (different frequency for the auditory domain and different spatial locations for the tactile domain). This might have made the visual stimuli more salient compared to the auditory or tactile stimuli, resulting in faster reaction times and higher accuracy rates in the visual task. For a detailed discussion, see Section *5.4*. Future studies are necessary to shed more light on this result.

Overall, we believe that our results do not entirely align with the assertions made by Conway et al. (2009). Indeed, we have observed that the tactile sphere has also proven to be particularly adept at processing sequential statistical information, showing a clear advantage over the visual sphere and furthermore demonstrating abilities comparable to those of hearing in acquiring sequential statistical information that necessitate the formation of recursive hierarchical abstract structural representations. In general, these findings do not entirely align with Conway's and colleagues' Auditory Scaffolding Hypothesis. The reason why touch, along with hearing, demonstrates such proficiency in this cognitive ability, and why they collectively outperform the visual sphere, remains unexplained. One possible explanation could be derived from evolutionary considerations. Touch may have developed a greater sensitivity to sequential information due to the evolutionary need to perceive the world in low-light conditions or absence of light. Sequential and temporal perception of an object's features through touch allows us to gather information about its three-dimensional shape. By exploring the surface of an object, for example, we can identify facets, detect subtle details, and form an abstract representation of its overall spatial structure. Sequential processing in the tactile domain might thus contribute to the construction of richer and more detailed mental representations of touched objects, with significant implications in object manipulation, navigation, and interaction with the surrounding environment. Similarly, the specialization of hearing in processing complex sequentially distributed information over time is evident. Hearing plays a fundamental role in processing complex sequential systems, such as language and music, contributing significantly to our ability to communicate, socialize, and interpret the surrounding world. In contrast, vision may have developed a stronger ability in processing spatially arranged stimuli during visual processing, at the expense of the ability to process sequentially presented stimuli. This is primarily what occurs during visual processing, where details, colors, and shapes are captured simultaneously, offering an immediate and comprehensive view of a scene. This ability is crucial for engaging with spatially distributed information, enabling a rapid and global visual exploration of the surrounding environment. Overall, our hypotheses are consistent with those formulated by Lashley (1951).

But now the question arises spontaneously: If our hypotheses were to be correct, is the human predisposition to process sensory information in specific ways the result of evolutionary adaptation or rather the acquisition of skills over the course of life? What can we expect in terms of the acquisition of tactile and visual sequential statistical information (at different levels of complexity and abstraction) in deaf individuals? From an evolutionary standpoint, we could consider that the advantage in sequential statistical processing observed in hearing and touch has been shaped over millennia to adapt to environmental and survival needs. On the other hand, it is possible that part of this predisposition is linked to individual development. Life experience and training could influence this ability in various ways across sensory domains, with individuals refining their skills based on the specific demands of the surrounding environment. If this were the case, deaf individuals might experience a heightened development of this ability in other sensory modalities, such as vision and touch, due to the absence of auditory stimuli. Brain plasticity could play a crucial role in adapting the available sensory modalities, allowing for greater specialization in response to individual and environmental needs. We believe that a future research direction to address this question could involve testing deaf people using our paradigm, comparing two groups of deaf individuals in our tactile and visual tasks, and then comparing the results with those obtained in our study with the typical population. This exploration could provide valuable insights into how sequential statistical learning – at different levels of abstraction and complexity, from low-level transitional regularities to recursive hierarchical abstract representations - operates in atypical populations. Deaf individuals, having deprived themselves of one of the primary senses, namely hearing, could serve as a particularly intriguing population to understand whether and how brain plasticity adapts to such conditions. Their capacity to form recursive hierarchical abstract representations could be analyzed and compared with that of the typical population to ascertain whether this ability is primarily innate, evolutionary, or influenced by individual experience. Delving into these dynamics could lead to new discoveries about the relationship between brain plasticity and implicit statistical learning. Furthermore, exploring how deaf individuals develop specific skills in sequential statistical processing could have important implications for education and rehabilitation. It might be possible to design targeted interventions to further enhance these skills adaptively, considering brain plasticity as a modifiable resource throughout life. Ultimately, exploring whether this predisposition is innate and evolutionary or influenced by experience offers interesting insights both to understand the complexity of the ability to form recursive hierarchical abstract representations across different sensory domains and to better comprehend the underlying mechanisms of statistical learning and brain plasticity. Furthermore, this could serve as a test for the Auditory Scaffolding Hypothesis. If Conway and colleagues' hypothesis holds true, we would expect to find limited evidence of learning among deaf individuals in our tactile and visual paradigms, or at least significantly lower learning effects compared to those observed in our study with the typical population.

The domain-specific differences we observed in our experimental study also tell us something interesting about the nature of implicit learning that took place. In the field of implicit statistical learning, two distinct perspectives exist regarding the nature of learning. The perceptual learning viewpoint suggests that individuals primarily gain knowledge of the stimulus sequence through forming associations between consecutive stimuli (known as stimulus-to-stimulus learning; see Remillard, 2003). Conversely, the motor learning perspective argues that learning predominantly occurs through associations between successive responses (referred to as response-to-response learning; see Nattkemper & Prinz, 1997). The results obtained in our study lead us to support the idea that the learning that took place was perceptual in nature, pertaining to specific stimuli rather than being entirely motor-based (see Abrahamse et al., 2008). In fact, if the learning were purely motor-based, we would not expect to find differences in learning across the three tasks (i.e. visual, tactile, and auditory) in our experimental study. This is because in all three tasks, participants were required to press the same keys - the *z* and *m* keys on the keyboard - in response to the perceived stimuli.

However, it is important to note that despite the domain-specific differences we found, all three sensory domains demonstrated the ability to represent sequential recursive nested structures, albeit with noticeable differences. In this thesis, we argued that the ability to form recursive hierarchical abstract representations from sequentially arranged stimuli is a key cognitive ability at work in human language, generating one of the various possible forms of hierarchical structures in language: syntactic recursive phenomena (cf. Sections *2.1*; *2.1.1.*). The discovery of the ability to process these structures across various sensory domains leads us to reject the hypothesis that the cognitive ability to process these structures are domain-specific to language (cf. Sections *1.1.1.*; *2.1.2.*), suggesting instead that they are domain-general in nature.

Taken together, our experimental findings suggest that the ability to form recursive hierarchical abstract representations from sequentially arranged stimuli is a domain-general phenomenon, in the sense that it is not a language-specific ability, but it can occur across different sensory domains. However, we did observe domain-specific differences. Indeed, different sensory domains exhibited varying levels of proficiency in dealing with the formation of these abstract structures. Consequently, there could be two plausible hypotheses regarding how this domain-general ability, which displays domain-specific differences, operates: either there are separate neural networks located in distinct cortical areas such as the visual, auditory, and somatosensory cortex that implement similar computational principles, or there exists a multi-modal region or partially-shared neural networks accessed by representations of stimulus inputs from specific modalities for further computation. These two hypotheses have been proposed by both Martins et al. (2017), regarding the ability to represent recursion, and Frost et al. (2015), concerning the ability of implicit statistical learning (cf. Sections *3.1.1.*; *3.1.6.*). In this vein, we think that another promising avenue for future research would be to shed light on the neural correlates of the ability to form recursive hierarchical abstract representations from sequential stimuli. Comparing the neural correlates activated in our paradigm across the three sensory domains could be particularly intriguing. This approach could reveal which common areas are involved in processing these structures across the three domains, while also identifying areas that are uniquely activated in specific sensory tasks. Crucially, it is important to consider, as we mentioned in Section *5.4.*, that differences in accuracy and learning rates across modalities do not necessarily imply distinct mechanisms. To delve deeper into this, future studies could explore several additional factors and methodologies. One such factor could be investigating transfer effects. For instance, creating a modified version of our serial reaction time task with Fib, where participants are exposed to visual stimuli in the first part and then switch to auditory or tactile stimuli in the second part, might reveal whether skills in recursive hierarchical representation transfer across domains. Another aspect worth examining is how the mode and speed of stimulus presentation impact each modality. By varying the speed and mode of presentation in the different modalities, we could determine if there are optimal conditions for each sensory domain, thereby refining our understanding of domain-specific processing capabilities (cf. Emberson et al., 2011). By pursuing these research directions, we could further unravel the complexities of how recursive hierarchical abstract representations are formed and processed across various sensory modalities, providing deeper insights into the underlying neural and cognitive mechanisms.

In any case, setting aside the open questions about the nature of the domain-specific differences observed, it is extremely interesting to note, for the research objectives and questions that motivated this thesis, that the ability to form recursive hierarchical abstract representations from sequential stimuli has been observed beyond language, across various sensory domains. This finding leads us to exclude it as a strongly domain-specific ability solely dedicated to language. In other words, it suggests that this characteristic did not emerge through the process of natural selection, driven by the demands of its linguistic purpose (Culbertson and Kirby, 2016). This hypothesis aligns with findings from computational modeling, which suggest that it is improbable for language learning to develop domain-specific hard constraints. This is mainly due to the fact that cultural evolution tends to magnify the influence of weak biases, as demonstrated by Kirby et al. (2007). Additionally, the rapid pace of language change, as noted by Chater et al. (2009), further undermines the likelihood of such constraints evolving (cf. Culbertson, Kirby, 2016). Instead, it seems more plausible that the capacity to form recursive hierarchical abstract representations evolved for purposes other than language. In this case, it would fall under what are termed *strong-domain general biases* (Culbertson and Kirby, 2016). Alternatively, this ability could be an outcome of broad architectural or computational principles governing how cognition operates, aligning with what Chomsky termed the *third factor of language design* (Chomsky, 2005). Crucially, however, these biases might still engage with language and its representations in domain-specific ways (Culbertson and Kirby, 2016). Indeed, according to Culbertson and Kirby (2016), domain-specificity in language can occur when domain-general biases interact with language in specific ways.

Regarding why we possess this domain-general tendency or cognitive ability to form recursive hierarchical abstract representations from sequential input, which is particularly prominent in language, one could speculate. In his work from 1962, Herbert sought to identify common properties among various types of complex systems. He noted that complexity often adopts the form of hierarchy. "[…] complexity frequently takes the form of hierarchy, and […] hierarchic systems have some common properties that are independent of their specific content. Hierarchy, I shall argue, is one of the central structural schemes that the architect of complexity uses." (Herbert, 1962, p.468). Herbert further explains that, on theoretical grounds, we could anticipate complex systems to exhibit hierarchical structures in a world where complexity evolves from simplicity. He argues that systems structured hierarchically possess evolutionary advantages. The time required for the evolution of a complex form, he contends, critically depends on the number and distribution of potential intermediate stable forms. Moreover, Herbert suggests that, at a cognitive level, hierarchical organization of information offers benefits. It facilitates efficient representation in memory and enhances information transmission by reducing the amount of information that, on the contrary, would be lost in the absence of hierarchical organization (Herbert, 1962). Similarly, we could argue that recursive hierarchical structures are present in various natural phenomena, both in organic forms in nature and in various cognitive phenomena.

For instance, consider the intricate patterns found in broccoli or the spiral arrangement of a sunflower, reflecting recursive organization at the biological level. In the realm of cognition, examples abound in music and in language, among the others (cf. Sections *2.1.2.*; *2.2.*). Hypothesizing about the evolutionary advantages of such structures, we might suggest that recursive hierarchical patterns allow for efficient utilization of resources and adaptation to environmental challenges. In biological systems, recursive structures could facilitate efficient resource allocation and energy conservation, contributing to the organism's survival and reproduction. Turning to cognitive phenomena, our hypothesis is that recursive hierarchical structures reduce computational load in terms of working memory. For example, in information transmission, it seems advantageous to group information into chunks, organize them hierarchically, and utilize algorithms that incorporate self-similarity. This approach would be more economical in terms of computational load. This hypothesis is in line with what has been proposed by both Planton and colleagues (2021) and Radulescu and colleagues (2019) (cf. Section *2.2.1.*). Overall, our hypotheses suggest that recursive hierarchical structures offer evolutionary and cognitive advantages by promoting efficiency and reducing computational costs. However, it is important to note that these are merely our speculations, and further insight could be gained through computational modeling studies. By simulating the dynamics of complex systems and analyzing the emergence and evolution of recursive hierarchical structures, modeling studies could shed light on the validity of these hypotheses.

References

Abrahamse, E. L., Lubbe, R. H. J. V. D., & Verwey, W. B. (2008). Asymmetrical learning between a tactile and visual serial RT task. *The Quarterly Journal of Experimental Psychology*, 61(2), 210–217.

Abrahamse, E. L., Lubbe, R. H. J. V. D., & Verwey, W. B. (2009). Sensory information in perceptual-motor sequence learning: Visual and/or tactile stimuli. *Experimental Brain Research*, 197, 175–183.

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford, England: Oxford University Press.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral cortex*, 17(7), 1493-1503.

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS biology*, 8(9), e1000489.

Aquinas, Thomas. (1951) Commentary on Aristotle's 'De Anima', Translated by Foster K., & Humphries S. New Haven: Yale University Press. Digital Edition.

Arnon, I. (2019). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science*, 11(3), 504–519.

Aslin, R. N, & Newport, E. L. (2012). Statistical learning: from acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistic by 8-month-old infants. *Psychological Science*, 9, 321–324.

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.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). Academic Press.

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., & Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain research*, 1298, 161-170.

Baker C.I., Olson, C.R, Behrmann M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15:460–466.

Bates, D., M. Mächler, B. Bolker & S. Walker (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1). 1–48.

Bates, E., & Elman, J. (1996). Learning Rediscovered. *Science*, 274(5294), 1849–1850.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, 106(5), 1672-1677.

Bernardy, J., & Lappin, S. (2017). Using deep neural networks to learn syntactic agreement. *Linguistic Issues in Language Technology*, 15, 1–15.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 13(2), 90–102.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the Stimulus Revisited. *Cognitive Science*, 35(7), 1207–1242.

Berwick, R.C. and Chomsky, N. (2017). *Why Only Us*. MIT Press.

Bickerton, D. (1996). An innate language faculty needs neither modularity nor localization. Open peer commentary to R.-A. Mueller. Innateness, autonomy, universality? Neurobiological approaches to language. *Behavioral and Brain Sciences,* 19, 631-632.

Bloom, P. (1994). Generativity within language and other cognitive domains. *Cognition, 51*, 177-189.

Boeckx, C., & Hornstein, N. (2004). The varying aims of linguistic theory. Unpublished manuscript, Harvard University, Department of Linguistics, and University of Maryland, Department of Linguistics, College Park.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European journal of neuroscience*, 23(10), 2791-2803.

Carnie, A. (2002). *Syntax: A generative introduction*. Oxford, England: Blackwell.

Chalmers, D.J. 1990. Syntactic transformations on distributed representations. *Connection Science*, 2, 53–62.

Chametzky, R. (2000). *Phrase structure: From GB to minimalism*. Oxford, England: Blackwell.

Charniak, E.,Santos, E. (1986). 'A Connectionist Context-free Parser which is not Context-free, but then it is not Really Connectionist Either', Department of Computer Science, Brown University.

Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1015–1020.

Chesi, C., Moro, A. (2014), Computational complexity in the brain. In: Newmeyer FJ, Preston LB, eds. *Measuring grammatical complexity*, Oxford: Oxford University Press.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., et al. (2014). Learning phrase representations using RNN Encoder–Decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734). Stroudsburg, PA: Association for Computational Linguistics.

Chomsky, N. (1956). Three Models for the Description of Language. *IRE Transactions on Information Theory* 2:113-124.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1959). On Certain Formal Properties of Grammars. I*nformation and Control*. Vol. 2, 137-167.

Chomsky, N. (1975). *The logical structure of linguistic theory*. Plenum Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* Praeger.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1993). *Lectures on Government and Binding: The Pisa Lectures.* DE GRUYTER MOUTON.

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

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22.

Chomsky, N. (2013). Problems of projection. *Lingua* 130, 33–49.

Chomsky, Noam. 2020. The UCLA lectures. https://ling.auf.net/lingbuzz/005485 (accessed 7 November 2023).

Chomsky, N. (2023, March 8). Noam Chomsky: The false promise of CHATGPT. *The New York Times.* Retrieved from https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Christiansen, M.H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure.* (Doctoral dissertation, University of Edinburgh).

Christiansen, M.H., Chater, N. (1999). Toward a Connectionist Model of Recursion in Human Linguistic Performance. *Cognitive Science*, 23: 157-205.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*, 13(2–3), 221–268.

Christiansen, M.H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, 11(3), 468–481.

Christiansen, M.H. & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics in perspective. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 19–75). Westport, CT: Ablex.

Christiansen, M.H., MacDonald, M.C. (2009). A Usage-Based Approach to Recursion in Sentence Processing. *Language Learning*, 59: 126-161.

Chun, M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360–365.

Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences*, 23(4), 305–317.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry*, 36(3), 315–332.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108, 804 – 809.

Compostella, A. (2019). [unpublished master thesis]. AGL with a modified Simon task. Disentangling sequence from hierarchical structure learning. University of Verona.

Compostella, A., Tagliani, M., Vender, M., Delfitto, D. (under review). On the interaction between implicit statistical learning and the alternation advantage: Evidence from manual and oculomotor serial reaction time tasks.

Conway, C. M. (2005). [Ph.D. thesis]. An Odyssey through Sight, Sound, and Touch: Toward a Perceptual Theory of Implicit Statistical Learning. Cornell University, Ithaca, NY.

Conway, C. M., Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, memory, and Cognition*, 31(1):24–39.

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, 21(4), 561–580.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The Importance of Sound for Cognitive Sequencing Abilities: The Auditory Scaffolding Hypothesis. *Current Directions in Psychological Science*, 18(5), 275–279.

Coolidge, F. L., Overmann, K. A., & Wynn, T. (2011). Recursion: What is it, who has it, and how did it evolve? *Wiley interdisciplinary reviews. Cognitive science*, 2(5), 547–554.

Coopmans, C. W., De Hoop, H., Kaushik, K., Hagoort, P., & Martin, A. E. (2022). Hierarchy in language interpretation: Evidence from behavioural experiments and computational modelling. *Language, Cognition and Neuroscience,* 37(4), 420–439.

Corballis, M. C. (2007). On phrase structure and brain responses: a comment on Bahlmann, Gunter, and Friederici (2006). *Journal of cognitive neuroscience*, 19(10), 1581–1583.

Cottrell, G. (1985). Connectionist Parsing. P*roceedings of the Seventh Annual Cognitive Science Society*, pp. 201-11.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.

Craig, J. C., & Rollman, G. B. (1999). Somethesis. *Annual Review of Psychology*, 50, 305–331.

Crain, S., & Nakayama, M. (1987). Structure Dependence in Grammar Formation. *Language*, 63(3), 522.

Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, 24, 139–186.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119 –1130.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16), 5842–5847.

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in psychology*, 6, 1964.

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: Deriving biases innoun phrase order from statistical properties of the world. *Language*, 96(3), 696-717.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3), 306–329.

Culicover, P. W. (2013). The role of linear order in the computation of referential dependencies. *Lingua*, 136, 125-144.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, 96, 233–262.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2, 133–142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.

De Vries, M.H., Christiansen, M.H., Petersson, K.M. (2011), Learning Recursion: Multiple Nested and Crossed Dependencies, *Biolinguistics* 5.1-2.

Dehaene, S., & Changeux, J. P. (1997). A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Science*s, 94(24), 13293-13298.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, 88(1).

Deroost, N., Coomans, D., & Soetens, E. (2009). Perceptual load improves the expression but not learning of relevant sequence information. *Experimental Psychology*, 56(2), 84–91.

Deroost, N., Zeischka, P., Coomans, D., Bouzza, S., Depessemier, P., & Soetens, E. (2010). Intact first- and second-order implicit sequence learning in secondary-school-aged-children with developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, 1, 1–12.

Descartes, R. (2003) [1637]. Discourse on method. In Elizabeth S. Haldane & George R. Thomson (trans.), *Discourse on Method and Meditations*, Mineola, NY: Dover, 1–52.

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, 107(3), 763–774.

Dienes, Z., Altmann, G., Kwan, L., Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology*, 21. 1322.

Dryer, M. S. (2018). On the order of demonstrative, numeral, adjective, and noun. *Language*, 94(4), 798–833.

Du, W., & Kelly, S. W. (2013). Implicit sequence learning in dyslexia: a within-sequence comparison of first- and higher-order information. *Annals of Dyslexia*, 63(2), 154–170.

Edelman, S., Hiles, B.P., Yang, H., & Intrator, N. (2002). Probabilistic principles in unsupervised learning of visual structure: Human data and a model. In T.G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, 14 (pp. 19–26). Cambridge, MA: MIT Press.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.

Elman, J.L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71-99.

Emberson, L. L., Conway, C.M., Christiansen, M.H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology* (Hove), 64, 1021–1040.

Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. Philadelphia, PA: Linguistic Society of America.

Everett, D. L. (1986). Pirahã. In D. Derbyshire & G. Pullum (Eds.), *Handbook of Amazonian languages* (Vol. 1, pp. 200-326). Berlin, Germany: Mouton de Gruyter.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology, 46*, 621-646.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York academy of science*s, 1169(1), 448-458.

Fanty, M. (1985). Context-Free Parsing in Connectionist Networks. Technical Report No. 174, Department of Computer Science, University of Rochester.

Fecteau, J. H., Au, C., Armstrong, I. T., & Munoz, D. P. (2004). Sensory biases produce alternation advantage found in sequential saccadic eye movement tasks. *Experimental Brain Research*. Advance online publication. https://doi.org/10.1007/s00221-004-2154-2.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews Neuroscience*, 4(6), 435–443.

Feigenson, L. (2011). Objects, sets, and ensembles. In S. Dehaene & E. Brannon (Eds.), *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Thought.* London: Elsevier.

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Science Advances, 6*(26), eaaz1002.

Fiser, J., & Aslin, R.N. (2001). Unsupervised learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134(4), 521–537.

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1598), 1933–1955.

Fitch WT, Hauser MD. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*. 303:377–80. PMID: 14726592 11.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1316(1), 87–104.

Fitch, W. T., Friederici, A. D., & Hagoort, P. (2012). Pattern perception and computational complexity: introduction to the special issue. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1598), 1925–1932.

Fitch, W. T. (2010). Three meanings of "recursion": key distinctions for biolinguistics. In R. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language: Biolinguistic Perspectives* (pp. 73–90). Cambridge, UK: Cambridge University Press.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, 11(3), 329–364.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computer*s, 35(1), 116–124.

Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, 81(5), 284–310.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, 118(4), 614–636.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction time task. *Psychological Research*, 61, 83–98.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 2458–2463. PMID: 16461904.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in cognitive sciences*, 19(3), 117–125.

Gao, J., Wong-Lin, K., Holmes, P., Simen, P., & Cohen, J. D. (2009). Sequential effects in two-choice reaction time tasks: Decomposition and synthesis of mechanisms. *Neural Computation*, 21(9), 2407–2436.

Geambaşu, A., Ravignani, A., & Levelt, C. C. (2016). Preliminary experiments on human sensitivity to rhythmic structure in a grammar with recursive self-similarity. *Frontiers in Neuroscience*, 10, 281.

Geambaşu, A., Toron, L., Ravignani, A., & Levelt, C. C. (2020). Rhythmic recursion? Human sensitivity to a Lindenmayer grammar with self-similar structure in a musical task. *Music & Science*, 3, 205920432094661.

Geldard, F. A. (1970). Vision, audition, and beyond. *Contributions to sensory physiology*, 4, 1–17.

Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204-1207.

Gerken, L. A., Wilson, R., & Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249–268.

Gibson, Edward (1998). Linguistic complexity: locality of syntactic dependencies, *Cognition*, 68: 1–76.

Giles, C., Miller, C., Chen, D., Chen, H., Sun, G., & Lee, Y. (1992). Learning and extracting finite state automata with second-order recurrent neural networks. *Neural Computation*, 4, 393–405.

Giles, C. & Omlin, C. (1993). Extraction, insertion and refinement of symbolic rules in dynamically driven recurrent neural networks. *Connection Science*, 5, 307–337.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, 33(2), 260–272.

Giustolisi, B., & Emmorey, K. (2018). Visual statistical learning with stimuli presented sequentially across space and time in deaf and hearing adults. *Cognitive Science*, 42(8), 3177–3190.

Giustolisi, B., Martin, J. S., Westphal-Fitch, G., Fitch, W. T., & Cecchetto, C. (2022). Performance of Deaf Participants in an Abstract Visual Grammar Learning Task at Multiple Formal Levels: Evaluating the Auditory Scaffolding Hypothesis. *Cognitive Science*, 46(2), e13114.

Goldberg, N. (2014). Imprints of Dyslexia: Implicit Learning and the Cerebellum. Utrecht, The Netherlands, LOT Publications.

Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, 112(1), 21–54.

Gomez, R. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70.109-135.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178–186.

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, 13(5), 431–436.

Goodfellow, L. D. (1934). An empirical comparison of audition, vision, and touch in the discrimination of short intervals of time. *American Journal of Psychology*, 46, 243-268.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: Genera*l, 121, 480 –506.

Greenberg, J. H. (1963). *Universals of Language*. Cambridge, Mass: MIT Press.

Grunwald, M. (Ed.). (2008). *Human haptic perception: Basics and applications* (1st ed.). Birkhäuser Basel.

Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. MIT Press.

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (pp. 1195–1205). Stroudsburg, PA: Assoc. Comput. Linguist.

Hall, M. L., Eigsti, I. M., Bortfeld, H., & Lillo-Martin, D. (2018). Auditory access, language access, and implicit sequence learning in deaf children. *Developmental Science*, 21(3), e12575.

Handel, S. (1988). Space is to time as vision is to audition: seductive but misleading. *Journal of experimental psychology. Human perception and performance*, 14(2), 315–317.

Hanson, S. and Kegl J. (1987). PARSNIP: A Connectionist Network That Learns Natural Language Grammar From Exposure to Natural Language Sentences, P*roceedings of the Ninth Annual Conference of the Cognitive Science Society*, 106-119. Hillsdale, NJ: Erlbaum.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56, 51–65.

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.

Herbert, A. S. (1962). The Architecture of Complexity, *Proceedings of the American Philosophical Society*, Dec. 12, 1962, Vol. 106, No. 6 (Dec. 12, 1962), pp. 467-482.

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of cognitive neuroscience*, 24(6), 1382-1397.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Hockett, C. (1960). The origin of speech. *Scientific American, 203*, 88-96.

Hoffmann, J., & Koch, I. (1997). Stimulus-response compatibility and sequential learning in the serial reaction time task. *Psychological Research*, 60, 87–97.

Hofstadter, D. R. (1980). *Gödel, Escher, Bach: An eternal golden braid*. London, England: Penguin.

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Process*, 15, 338 –346.

Hopcroft, J.E. & J.D. Ullman (1969). *Formal languages and their relation to automata.* Addison-Wesley Longman Publishing Co., Inc.

Horrocks, G. (1987). *Generative grammar*. London, England: Longman.

Howells, T. (1988). VITAL, a connectionist parser. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Huybregts, R. (1976). Overlapping dependencies in Dutch. In Utrecht Working Papers in Linguistics, number 1, p. 24–65.

Huybregts, R. (1984). The weak inadequacy of context-free phrase structure grammars. In de Haan, G. J., Trommelen, M., and Zonneveld, W., editors, *Van Periferie Naar Kern*, p. 81–99. Foris, Dordrecht.

Hunt, R. H. (2002). The induction of categories from distributionally defined contexts: Evidence from a serial reaction time task. Unpublished doctoral dissertation, University of Rochester, NY.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658–680.

Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system*. Oxford, England: Basil Blackwell.

Hurford, J. R. (2003). The language mosaic and its evolution. In M. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 38-57). Oxford, England: Oxford University Press.

Jäger G., Rogers J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 367. 1956- 70.

James, W. (1890). *The principles of psychology*. New York, NY: Holt.

Janata, P., & Parsons, L. M. (2013). Neural mechanisms of music, singing, and dancing. In M. A. Arbib (Ed.), *Language, Music, and the Brain: A Mysterious Relationship* (Vol. 10, pp. 307–328). Cambridge, Massachusetts: MIT Press.

Johansson, S. (2013). Biolinguistics or Physicolinguistics? Is the third factor helpful or harmful in explaining language? *Biolinguistics*, 7(October), 249-275.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548 –567.

Jordan, M.I. (1986). Serial order: A parallel distributed processing approach (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Science.

Joshi, A. (1985). How much context-sensitivity is necessary for characterizing structural descriptions? In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, p. 206–250. Cambridge University Press, New York.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automation perspective on the psycholinguistic results. *Language and Cognitive Processes*, 5(1), 1–27.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*, 29(1), 1–23.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217.

Karlsson, F. (2010). Syntactic recursion and iteration. In H. v. d. Hulst (Ed.), *Recursion and human language* (pp. 43–67). Berlin/New York: de Gruyter Mouton.

Karuza, E. A. (2014). Learning across space, time, and input modality: Towards an integrative, domain-general account of the neural substrates underlying visual and auditory statistical learning (Doctoral dissertation). University of Rochester, Department of Brain and Cognitive Sciences.

Kayne, R. (1994). *The Antisymmetry of Syntax*. MIT Press.

Kayne, R. (2022). Antisymmetry and Externalization. *Studies in Chinese Linguistics*, 43.1

Kidd, E. (2012). Implicit Statistical Learning Is Directly Associated With the Acquisition of Syntax. *Developmental Psychology,* Vol. 48, No. 1, 171–184.

Kinsella, A. R. (2010). Was recursion the key step in the evolution of the human language faculty? In H. van der Hulst (Ed.), *Recursion and human language* (pp. 179-191). New York, NY: De Gruyter Mouton.

Kirby, N. H. (1976). Sequential effects in two-choice reaction time: Automatic facilitation or subjective expectancy? *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 567–577.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173-203). Cambridge: Cambridge University Press.

Kirby, S., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 5241.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 169–181.

Koechlin, E., Jubault, T. (2006). Broca’s area and the hierarchical organization of human behavior. *Neuron* 50: 963974.

Koelsch, S. (2012). *Brain and Music*. London, UK: John Wiley & Sons.

Koelsch, S. (2013). Neural correlates of music perception. In M. A. Arbib (Ed.), *Language, Music, and the Brain: A Mysterious Relationship* (141-172). Cambridge, MA: MIT Press.

Koelsch, S., Maess, B., Friederici, A.D. (2000). Musical syntax is processed in the area of Broca: an MEG study. *Neuroimage*. 11: 56.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217-230.

Krivochen, D. & Saddy, D. (2018). Towards a classification of Lindenmayer systems. *ArXiv.* arXiv:1809.10542.

Krivochen, D., Phillips, B., & Saddy, J. (2018). Classifying points in Lindenmayer systems: Transition probabilities and structure reconstruction (v. 1.1). https://doi.org/10.13140/RG.2.2.25719.88484.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance,* 14(2), 318–320.

Kumaran, D., Melo, H.L. & Duzel, E. (2012). The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron,* 76: 653–666.

Kuznetsova, A., P.B. Brockhoff & R.H. Christensen (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13). 1–26.

Langendoen, D. T. (1975). Finite-State Parsing of Phrase-Structure Languages and the Status of Readjustment Rules in Grammar. *Linguistic Inquiry*, 6(4), 533–554.

Langendoen, D. T., & Postal, P. M. (1984). *The vastness of natural languages*. Blackwell Pub.

Langley, P. (1987). Machine learning and grammar induction. *Machine Learning*, 2, 5–8.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*; the Hixon Symposium (pp. 112–146). New York: Wiley.

Legate, J. A., & Yang, C. (2002). Empirical re-assessment of stimulus poverty arguments. *Linguistic Review*, 19, 151–162.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Li, G., Ning, N., Ramanathan, K., He, W., Pan, L., & Shi, L. (2013). Behind the magical number: hierarchical chunking and the human working memory capacity. *International Journal of Neural Systems*, 23(04), 1350019.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. Cognition, 89(3), 295–303.

Lindenmayer, A. (1968). Mathematical models for cellular interactions indevelopment I. Filaments with one-sided inputs. *Journal of Theoretical Biology.* 18: 280–299.

Linzen, T., & Baroni, M. (2020). Syntactic Structure from Deep Learning. *arXiv*:2004.10827. Retrieved from https://arxiv.org/abs/2004.10827.

Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4, 521–535.

Liu, Y., & Stoller, S. (1999). From recursion to iteration: What are the optimizations? *ACM Sigplan Notices, 34*, 73-82.

Lobeck, A. C. (2000). *Discovering grammar: An introduction to English sentence structure*. New York, NY: Oxford University Press.

Lobina, D. J. (2011). "A Running Back" and Forth: A Review of Recursion and Human Language. *Biolinguistics*, 5(1–2).

Loeper, H., ElGabali, M., & Neubert, P. (1996). Recursion and iteration in computer programming. *Kuwait Journal of Science and Engineering, 23*, 153-180.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail. *Journal of Acoustical Society of America*, 102, 1134 –1140.

Loudon, K. (1999). *Mastering algorithms with C*. Cambridge, MA: O'Reilly.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience*, 4(5), 540-545.

Mandelbrot, B. (1977). *The fractal geometry of nature.* Freeman.

Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 117(52), 30046–30054.

Marcus, G. (2006). Startling starlings. *Nature* 440, 1204-1207.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: a journal of general linguistics*, 5(1).

Martin, A., Ratitamkul, T., Abels, K., Adger, D., & Culbertson, J. (2019). Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard*, 5(1), Article 20180072.

Martins, M. J. D., Muršič, Z., Oh, J., & Fitch, W. T. (2015). Representing visual recursion does not require verbal or motor resources. *Cognitive Psychology*, 77, 20–41.

Martins, M. D., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of “musical fractals”: Processing hierarchy and recursion in the auditory domain. *Cognition*, 161, 31–45.

Martins, M. J. D., Krause, C., Neville, D. A., Pino, D., Villringer, A., & Obrig, H. (2019). Recursive hierarchical embedding in vision is impaired by posterior middle temporal gyrus lesions. *Brain,* 142(10), 3217–3229.

Martins, M. J., Fischmeister, F. P., Puig-Waldmüller, E., Oh, J., Geißler, A., Robinson, S., Fitch, W. T., & Beisteiner, R. (2014). Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage*, 96, 300–308.

Martins, M.D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598). 2055–2064.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, 122, 346–362.

Matusevych, Y., & Culbertson, J. (2022). Trees neural those: RNNs can learn the hierarchical structure of noun phrases. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44), 1848–1854.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11, 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, 9(3), 419–436.

McCoy, T., Frank, R., & Linzen, T. (2020). Does syntax need to grow on trees? Sources of hierarchical inductive bias in sequence-to-sequence networks. *Transactions of the Association for Computational Linguistics*, 8, 125–140.

Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C. E., & Friederici, A. D. (2005). Language processing within the human medial temporal lobe. *Hippocampus,* 15(4), 451-459.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology,* 56, 485–491.

Miller, G. A. (1967). Project Grammarama. In *The psychology of communication: Seven essays* (pp. 125–187). New York: Basic Books.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26, 393–425.

Mithun, M. (2010). The fluidity of recursion and its implications. In H. v. d. Hulst (Ed.), *Recursion and human language*, (pp. 17–41). Berlin/New York: de Gruyter Mouton.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 20, 67–85.

Moro, A. (2016). *Impossible languages.* The MIT Press.

Näätänen, R.,Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin,* 125(6): 826-859.

Nattkemper, D., & Prinz, W. (1997). Stimulus and response anticipation in a serial reaction task. *Psychological Research*, 60, 98 –112.

Newport, E.L., & Aslin, R.N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S.C. Howell, S.A. Fish, and T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Newport, E. L., & Aslin, R. N. (2004). Learning at distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology,* 48, 127– 162.

Niklasson, L. & van Gelder, T. (1994). On being systematically connectionist. *Mind and Language,* 9, 288–302.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19(1), 1–32.

O'Connor, N., & Hermelin, B. (1972). Seeing and hearing and space and space and time. *Perception & Psychophysics*, 11(1), 46–48.

O'Donnell, T. J., Hauser, M. D., & Fitch, W. T. (2005). Using mathematical models of language experimentally. *Trends in Cognitive Sciences*, 9(6), 284–289.

Opitz, B., Friederici, A.D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage*. 19: 1730–1737.

Opitz, B., & Friederici, A. D. (2007). Neural basis of processing sequential and hierarchical syntactic structures. *Human Brain Mapping*, 28, 585–592.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, 105(7), 2745–2750.

Öttl, B., Jäger, G., Kaup, B. (2015). Does Formal Complexity Reflect Cognitive Complexity? Investigating Aspects of the Chomsky Hierarchy in an Artificial Language Learning Study, *Plos One* 10 (4).

Parker, A. R. (2006). *Evolution as a constraint on theories of syntax: The case against minimalism* (Ph.D. thesis, University of Edinburgh).

Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology,* 22(7-8), 776-789.

Patel, A. D. (2013). Sharing and nonsharing of brain resources for language and music. In M. A. Arbib (Ed.), *Language, Music, and the Brain: A Mysterious Relationship* (pp. 329–355). Cambridge, Massachusetts: MIT Press.

Pavlidou, E. V., & Bogaerts, L. (2019). Implicit Statistical Learning Across Modalities and Its Relationship With Reading in Childhood. *Frontiers in psychology*, 10, 1834.

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. Topics in Cognitive Science, 11(3), 520–535.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246–263.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: Two approaches, one phenomenon. *Trends in Cognitive Sciences*, 10, 233–238.

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.

Phillips, B. (2017). Symmetry of Hierarchical Artificial Grammars and Its Effects on Implicit Learning. [Unpublished MSc dissertation]. University of Reading, UK.

Piantadosi, S. T. (2023). Modern language models refute Chomsky's approach to language. In E. Gibson & M. Poliak (Eds.), *From fieldwork to linguistic theory: A tribute to Dan Everett*. Language Science Press. Preprint available at https://ling.auf.net/lingbuzz/007180

Pinker, S. (1994). The language instinct. William Morrow & Co.

Pinker, S. (1999). Words and rules: The ingredients of language. Basic Books.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 16-37). Oxford, England: Oxford University Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition,* 95(2), 201–236.

Planton, S., Kerkoerle, T. V., Abbih, L., Maheu, M., Meyniel, F., Sigman, M., Wang, L., Figueira, S., Romano, S., & Dehaene, S. (2021). A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLOS Computational Biology*, 17(1), e1008598.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 1–49.

Pollack, J.B. (1988). Recursive auto-associative memory: Devising compositional distributed representations. *Proceedings of the Tenth Annual Meeting of the Cognitive Science Society*,33–39. Hillsdale, NJ: Lawrence Erlbaum.

Pollack, J.B. (1990). Recursive distributed representations. *Artificial Intelligence*, 46, 77–105.

Post, E. L. (1943). Formal reductions of the general combinatorial decision problem. *American Journal of Mathematics, 65*, 197-215.

Post, E. L. (1944). Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society, 50*, 284-316. (Reprinted in Davis, 1965, pp. 304-337).

Pothos, E. (2007). Theories of Artificial Grammar Learning. *Psychological Bulletin*, 133. 227- 244.

Premack, D. (2004). Is language the key to human intelligence? *Science* 303, 318-320.

Prusinkiewicz, P., Lindenmayer, A. (1990). *The algorithmic beauty of plants*. 2nd ed. New York: Springer- Verlag.

Radford, A. (1997). *Syntactic theory and the structure of English: A minimalist approach*. Cambridge, England: Cambridge University Press.

Radulescu, S., Wijnen, F., & Avrutin, S. (2019). Patterns Bit by Bit. An Entropy Model for Rule Induction. *Language Learning and Development*, 16(2), 109–140.

Rakison, D. H. (2004). Infants' sensitivity to correlations between static and dynamic features in a category context. *Journal of Experimental Child Psychology*, 89, 1–30.

Ramsey, W., & Stich, S. (1990). Connectionism and three levels of nativism. *Synthese*, 82(2), 177–205.

Reali, F., Christiansen, M. (2005). Uncovering the Richness of the Stimulus: Structure Dependence and Indirect Statistical Evidence. *Cognitive Science,* 29, 1007-1028.

Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, 6, 317–327.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115–119.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, 5, 333–361.

Reber, A.S., Kassin, S. M. (1980). On the Relationship Between Implicit and Explicit Modes in the Learning of a Complex Rule Structure. *Journal of Experimental Psychology: Human Learning and Memory*, Vol.6, No. 5, 492-502.

Rebuschat, P., & Monaghan, P. (2019). Aligning implicit learning and statistical learning: Two approaches, one phenomenon. Special Issue, *Topics in Cognitive Science*, 11(3), 455–586.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, 125, 123–138.

Redington, M., & Chater, N. (1998). Connectionist and Statistical Approaches to Language Acquisition: A Distributional Perspective. *Language and Cognitive Processes*, 13(2–3), 129–191.

Redington, M., Chater, N., Finch, S. (1998). Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425–469.

Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 518– 597.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085–1099.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review,* 77(6), 481.

Restle, F., & Brown, E. R. (1970). Serial pattern learning. *Journal of Experimental Psychology,* 83(1p1), 120.

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: A MDL-based cognitive model of inductive learning. *Cognitive Science*, 35(7), 1352–1389.

Roeper, T. (2009). The minimalist microscope: How and where interface principles guide acquisition. Paper presented at the Boston University Conference on Child Language.

Rumelhart, D. E., McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press.

Saffran J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language* 47, 172–196.

Saffran J. R. Johnson E. K. Aslin R. N. Newport E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology,* 39, 484 – 494.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science,* 274(5294), 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.

Saffran, J.R. and Wilson, D.P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy* 4, 273–284.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical cues in language acquisition: Word segmentation by infants. In Proceedings of the 18th annual Cognitive Science Society conference,376–380. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Sammler, D., Koelsch, S., Friederici, A.D. (2011). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* 47: 659–673.

Sampson, G. (1987). A Turning Point in Linguistics. *Times Literary Supplement*, June 12, p. 643.

Sauerland, U., & Trotzke, A. (2011). Biolinguistic Perspectives on Recursion: Introduction to the Special Issue. *Biolinguistics*, 5(1–2), 001–009.

Savin, H. B. (1967). On the successive perception of simultaneous stimuli. *Perception & Psychophysics*, 2, 479-482.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, 37(6), 1013-1025.

Schmid, S., Saddy, D., Franck, J. (2023). Finding hierarchical structure in binary sequences: evidence from Lindenmayer grammar learning. *Cognitive Science*, 47(1). e13242.

Schopenhauer, A. (1969). *The world as will and representation* (3rd ed., E. F. J. Payne, Trans.). New York, NY: Dover. (Original work published 1859).

Seidenberg, M., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.

Selman, B., Hirst G. (1985) A Rule-Based Connectionist Parsing System. Proceedings of the Seventh Annual Conference of the Cognitive Science Society.

Senghas, R. J. (2003). New ways to be deaf in Nicaragua: Changes in language, personhood, and community. In L. Monaghan, K. Nakamura, C. Schmaling, & G. H. Turner (Eds.), Many ways to be deaf: International, linguistic, and sociocultural variation (pp. 260–282). Washington: Gallaudet University Press.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science*, 305(5691), 1779–1782.

Servan-Schreiber, D., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592–608.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning*, 7, 161–193.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.

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

Shirley, E. J. (2014). Representing and remembering Lindenmayer-grammars. (Doctoral dissertation), University of Reading https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.658878.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language,* 81, 105–120.

Slone, L., & Johnson, S. P. (2015). Statistical and chunking processes in adults' visual sequence learning. Proceedings of the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA (pp. 2218–2223).

Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, 178, 92102.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 598–616.

Sopena, J.M. (1991). ERSP: A distributed connectionist parser that uses embedded sequences to represent structure (Tech. Rep. No. UB-PB-1-91). Departament de Psicologia Bàsica, Universitat de Barcelona, Spain.

St. John, M. F., & McClelland, J. L. (1988). Learning and applying contextual constraints in sentence comprehension. In Proceedings of the 10th Annual Cognitive Science Society Conference. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), Logical Aspects of Computational Linguistics. LACL 1996. (Lecture Notes in Computer Science, Vol. 1328). Springer, Berlin, Heidelberg.

Steedman, M. (1985). Dependency and coordination in the grammar of Dutch and English. *Language*, 61:523–568.

Stolcke, A. (1991). Syntactic category formation with vector space grammars. In *Proceedings from the Thirteenth Annual Conference of the Cognitive Science Society*, 908–912. Hillsdale, NJ: Lawrence Erlbaum.

Terhune-Cotter, B. P., Conway, C. M., & Dye, M. W. G. (2021). Visual sequence repetition learning is not impaired in signing DHH children. T*he Journal of Deaf Studies and Deaf Education*, 26(3), 322–335.

Tettamanti, M., Rotondi, I., Perani, D., Scotti, G., Fazio, F., Cappa, S. F., Moro, A. (2009). Syntax without language: Neurobiological evidence for cross-domain syntactic computations. *Cortex,* 45(7), 825-838.

Thiessen, E. D. (2017). What’s statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160056.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology,* 39, 706 –716.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants’ acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3, 73–100.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin,* 139(4), 792–814.

Thompson, S. P., & Newport, E. L. (2007). Statistical Learning of Syntax: The Role of Transitional Probability. *Language Learning and Development*, 3(1), 1–42.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Trask, R. L. (1993). *A dictionary of grammatical terms in linguistics*. London, England: Routledge.

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 25, 1322–1333.

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*, 36(6), 1078–1101.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J., H. and Koroshetz, W. J., Pinker, S. (1977). A Neural Dissociation within Language: Evidence that the Mental Dictionary Is Part of Declarative Memory, and that Grammatical Rules Are Processed by the Procedural System. *Journal of Cognitive Neuroscience* 9:2. 266-276.

Uriagereka, J., Reggia, J., & Wilson, G. (2013). A framework for the comparative study of language. *Evolutionary Psychology,* 11. 470-492.

van Der Hulst, H. (2010). *Recursion and Human Language*. Berlin, New York: De Gruyter Mouton.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). Attention is all you need. In U. von Luxburg (Ed.), *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (pp. 6000-6010). Red Hook, NY: Curran.

Vender, M., Compostella, A., Delfitto, D. (2023). Mapping precedence into containment: linear ordering in a bidimensional space. *Lingue e Linguaggio* 22(1), 49-88.

Vender M., Krivochen D.G., Compostella A., Phillips B., Delfitto D., Saddy D. (2020). Disentangling sequential from hierarchical learning in Artificial Grammar Learning: Evidence from a modified Simon Task. *PLoS ONE* 15(5): e0232687.

Vender M., Krivochen D.G., Phillips B., Saddy D., Delfitto D. (2019) Implicit Learning, Bilingualism, and Dyslexia: Insights From a Study Assessing AGL With a Modified Simon Task. *Frontiers in Psychology*, 2019;10: 1647.

Verhagen, A. (2010). What do you think is the proper place of recursion? Conceptual and empirical issues. In H. van der Hulst (Ed.), *Recursion and human language* (pp. 93-110). Berlin, Germany: De Gruyter Mouton.

von Humboldt, W. (1999). [1836]. The diversity of human language-structure and its influence on the mental development of mankind. In P. Heath (Trans.), Wilhelm von Humboldt: On Language (pp. 1–287). Cambridge: Cambridge University Press.

von Koss Torkildsen, J., Arciuli, J., Haukedal, C. L., & Wie, O. B. (2018). Does a lack of auditory experience affect sequential learning? *Cognition,* 170, 123–129.

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51), 20754-20759.

Waltz, D. L. and Pollack, J. B. (1985). Massively Parallel Parsing: A Strongly Interactive Model of Natural Interpretation. *Cognitive Science* 9, 51-74.

Warstadt, A., Parrish, A., Liu, H., Mohananey, A., Peng, W., et al. (2019). BLiMP: the benchmark of linguistic minimal pairs for English. *arXiv:1912.00582* [cs.CL]. Retrieved from https://arxiv.org/abs/1912.00582.

Watumull, J., Hauser, M. D., Roberts, I. G., & Hornstein, N. (2014). On recursion. *Frontiers in Psychology*, 4(JAN), 1–7.

Weckerly, J. & Elman, J. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society*, 414–419. Hillsdale, NJ: Lawrence Erlbaum.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103, 147–162

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.

Wilcox, E., Levy, R., Morita, T., & Futrell, R. (2018). What do RNN language models learn about filler–gap dependencies? In Linzen, T., Goldberg, Y., & Elhadad, M. (Eds.), *Proceedings of the Third Workshop on Analyzing and Interpreting Neural Networks for NLP* (pp. 211–221). Stroudsburg, PA: Assoc. Comput. Linguist.

Williams, J. A. (1966). Sequential effects in disjunctive reaction time: Implications for decision models. *Journal of Experimental Psycholog*y, 71(5), 665–672.

Yang, C.D. (2004). Universal Grammar, statistics or both? *TRENDS in Cognitive Sciences* Vol.8 No.10, 451-456.

Yerkes, R. M. (1943). *Chimpanzees: A laboratory colony*. New Haven, CT: Yale University Press.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, 57, 803– 815.

Zemel, R.S. (1993). A minimum description length framework for unsupervised learning. Unpublished doctoral dissertation, Department of Computer Science, University of Toronto, Canada.