Maria Vender, Arianna Compostella, Denis Delfitto

Mapping precedence into containment: linear ordering in a bidimensional space

(doi: 10.1418/107704)

Lingue e linguaggio (ISSN 1720-9331)
Fascicolo 1, gennaio-giugno 2023

Ente di afferenza:
Università di Verona (univr)

MAPPING PRECEDENCE INTO CONTAINMENT: LINEAR ORDERING IN A BIDIMENSIONAL SPACE

MARIA VENDER ARIANNA COMPOSTELLA
DENIS DELFITTO

ABSTRACT: There is an almost unanimous theoretical consensus according to which human languages are externalized as linear sequences of atomic units which are encoded according to specific hierarchical conditions. The nature of the interplay between the cognitive development of these hierarchical representations and their linearization on the string is however still not clear. In this paper, we aim to address this issue, exploring the relationship between precedence and containment by capitalizing on the results of a new experimental paradigm that has already provided interesting insights (Vender *et al.* 2019, 2020). More specifically, we report the results of two modified Simon Tasks in which the sequence of stimuli is determined by the rules of the Fibonacci grammar (Fib) or of its modifications Skip and Bif. All three grammars share the same transitional regularities, but they crucially differ in their structure: only Fib is characterized by the presence of so-called k-points, which provide, from a purely computational perspective, a potential bridge to full hierarchical reconstruction. We tested 64 adults' implicit learning skills, assessing learning of the statistical regularities in Fib, Skip and Bif, while also exploring the presence of hierarchical learning, in terms of the ability to predict k-points. Results provide evidence not only for the presence of statistically-based sequential learning, but also for hierarchical learning in Fib. We argue that the relations of precedence and containment are not antagonistic ways of processing a temporally ordered sequence of symbols; rather, they are strictly interdependent implementations of an abstract mathematical relation of linear ordering within a bidimensional computational space. We propose that the construction of this bidimensional space is primarily determined by labeling requirements, with the labeling algorithm emerging as the solution to the problem of mapping precedence into containment.

KEYWORDS: implicit learning, statistical learning, hierarchical learning, precedence vs. containment, Lindenmayer systems, chunking and labeling processes.

1. INTRODUCTION

In the current understanding in both theoretical linguistics and experimental psychology, precedence and containment (as well as the associated hierarchical concepts, such as c-command) are two essentially antagonistic ways of processing arrays of symbols (Christiansen & Chater 2016; Christiansen *et al.* 2012; Culbertson & Adger 2014; Jackendoff 2002; Martins 2012). More particularly, a conflict is deemed to exist between statistically-based processing of sequences of symbols and hierarchy-based processing of phrase-structures (Frank & Bod 2011; Frank *et al.* 2012; Frank & Christiansen 2018; Friederici 2017; Lobina 2011; Moro 2014). Cognitively, structural and linear mechanisms impose different computational requirements on the processing of external stimuli: intuitively, what can be representationally encoded by means of a node branching into two or more nodes cannot be encoded in linear terms, where an element in a sequence (for instance a spoken word) cannot be adjacent with more than two elements.

In formal approaches, linguistic conditions and dependencies are usually represented by reference to hierarchical relations expressed by recourse to trees, with a very limited space for conditions represented in terms of linear order (Chomsky 1995; Culicover 2013; Pinker 2000; Yang 2004). However, in the past three decades, statistical implicit learning models fostered the idea that either connectionist/associative views of cognition or Bayesian statistics, as implemented in the mind, could give rise to a set of powerful domain-general learning algorithms (Cornish *et al.* 2017; Griffiths & Kalish 2007; Tenenbaum & Griffiths 2001; Vapnik 1995). This stream of research strives towards modelling language acquisition and processing in terms of connectionist and dynamical systems or in terms of abstract probabilistic approaches relying on a Bayesian style of computation (Xu & Tenenbaum 2007). The issue arises to which extent these methods of analysis can account for the fact that properties of language are effectively processed and learned, including those properties that are standardly represented as structure dependent. The enigma of structure, as we view it, has two main components: (i) For which aspects is language processing based on sequential learning, and for which aspects is it based on hierarchical learning? (ii) Can we identify independent cognitive biases for sequential and hierarchical learning, and further investigate how these biases interact with each other?

1.1 Investigating linear and structural learning

For most linguists, syntactic operations are structure dependent. A large body of evidence in the acquisition literature converges to suggest that children abide by structure dependence as soon as they can be tested (Crain & Nakayama 1987; Guasti 2017). Moreover, compositional interpretation, in a large variety of domains, is based on hierarchical and not linear relations: this certainly holds for the interpretation of anaphoric relations, negative polarity items, scope and relative scope phenomena (Crain *et al.* 2017; Maclaran 1985). From an experimental perspective, the largest body of evidence in favor of the presence of hierarchical computing in language processing concerns the syntax of noun phrases, and more particularly the interaction between nouns and adjectives (Coopmans *et al.* 2022; Lidz *et al.* 2003; Culbertson & Adger 2014; Martin *et al.* 2020). These results suggest that the hierarchical bias must represent an inherent feature of the linguistic systems, as further confirmed by the observation that the gestural system developed by deaf children born to speaking parents (homesign) arguably involves hierarchy and structure-dependent operations (Coopmans *et al.* 2022; Goldin-Meadow 2005). Dehaene *et al.* (2015) suggest that this property might not be limited to language but also extend to other domains of cognition.

In this respect, the experimental paradigm used in the present contribution, crucially involving artificial languages generated by non-canonical grammars, whose output is significantly different from the output of finite-state or phrase-structure grammars, is intended to verify whether the human capacity to build structure also applies to sequences of symbols whose superficial properties are quite different from what we find in language, but which arguably permit to tease apart the predictions based on the statistical computation on the string from the predictions based on hierarchy-dependent calculations.

In the present contribution, we want to investigate the human capacity to build structure by testing the prediction skills of adult subjects when applied to sequences of symbols generated by a non-canonical grammar. We expect to find prediction skills that are arguably the result of at least some local hierarchy building procedure, under the hypothesis that humans are endowed with a strong cognitive bias to bootstrap sequences of symbols into hierarchical representations, the latter conceived in fact as a parallel linear ordering of the relevant symbols, based on the relation of dominance. Here are thus the research questions we would like to address in the present contribution:

(i) What are the cognitive foundations of the human sensitivity to hierarchy-based modes of language learning?

(ii) Are sequential learning and hierarchical learning two independent levels of language analysis or are these two modes of learning somehow interwoven?

(iii) Is there any way to define the algorithm by means of which humans bootstrap structure from linear order?

Problems already start with the proposed division of labor between sequential and hierarchical processing. The psycholinguistic and neurolinguistic evidence that has typically been offered in support for the claim that language processing is the computation of hierarchical relations and dependencies has constantly faced the objection that the same sets of data can be (re-)produced by essentially non-hierarchical statistical models of learning (Culbertson & Adger 2014; Ding *et al.* 2017; Frank & Christiansen 2018; Greenberg 1963).

On these conceptual and empirical grounds, weighing the relative success of statistically based and structure-based models of syntactic processing represents an increasingly unsatisfactory and problematic method. Besides delivering partial and controversial results, this method is based on the possibly wrong insight that sequential and hierarchical learning are radically antagonistic strategies of learning and processing. As we have already hinted at, the alternative possibility arises that these two strategies are in fact strictly interconnected, triggering the research question of establishing how this connection exactly works. Moreover, this strategy of investigation leaves the question concerning the cognitive foundations of the human capacity of hierarchy-based language processing and learning completely unanswered. However, the cognitive roots of the structure-building skills that humans manifest in their knowledge and use of language are certainly worth exploring, if only for the consequences that settling this issue might have for assessing the position of language within general cognition.

Our purpose is not simply that of arguing for the existence of hierarchy-driven computations in the processing of sequences of abstract symbols, to be understood as a cognitive bias that reflects in language. We have the more ambitious goal of finding the link between sequential and hierarchical computations, to be understood as a cognitive bias towards the creation of a multidimensional computational space, which arguably mirrors the interplay between linear and hierarchical organization as found in language, and explains why hierarchical computations constitute a sort of automatic effect of linearly based statistical computations.

In this paper, we present the results of two experimental studies that are based on a radical re-thinking of the relationship between linearly and hierarchically ordered arrays of symbols, and that further extend and improve on two previous studies exploiting the same framework of reference (Vender *et al.*

2019, 2020), to be introduced in some more detail below. The basic insight underlying these studies is very simple: we decided to modulate a formal set of symbols that constitute the object of our psycholinguistic inquiry in such a way that statistically-based accounts building on the calculation of transitional probabilities on the string and structure-based accounts building on a (possibly constrained) capacity for hierarchy-reconstruction make conflicting predictions about what should be learned and how it should be learned. In a nutshell, the new method we propose to address the enigma of structure capitalizes on using and re-adapting the Artificial Grammar Learning (AGL) research paradigm to define a set of artificial grammars that easily lend themselves to check the contrasting predictions made by statistically based learning strategies and hierarchy-based learning and parsing strategies. We use the output of methodologically sophisticated AGL research to inquire into the cognitive foundations of language that correspond to the capacity of chunking symbols together, categorizing chunks of symbols and constructing a linear order relation on a set X of symbols in a bi-dimensional space which exploits both the relation of precedence and the relation of containment. Crucially, from this perspective, we conceive of precedence and containment as potentially complementary ways of implementing an abstract mathematical notion of linear ordering and explore the formal and cognitive underpinnings of a position in which precedence and hierarchy may be interdependent, instead of representing alternative parsing strategies. This approach crucially enables us to address the bootstrapping problem (how we build containment from precedence) in a new original and revealing way, which paves the way for a more adequate understanding of the cognitive foundations of the human capacity for hierarchy-building (dendrophilia, in the sense of Fitch, 2014). The intriguing bunch of issues revolving around the relationship between the sequential ordering of sounds or gestures in language and the hierarchical arrangement of these very same symbols is of course per se not new in linguistics. It has been addressed in a series of contributions in theoretical linguistics across the years (Brody 2000; Kayne 1994; Moro 2000), in which the linearization problem was explicitly formulated: if knowledge of language involves a system of hierarchical representations but language has to be externalized as a temporal sequence of elements, how does hierarchy map into sequentially ordered arrays of terminals? In Kayne (1994)'s approach, for instance, precedence (the relation of linear ordering) is allowed only when it is paralleled by c-command (a hierarchical relation): x precedes y if and only if x c-commands y in the relevant hierarchical representation. On the other hand, the hierarchical relation of c-command is defined through the hierarchical relation of containment; for Kayne, the issue is thus how we map hierarchy into linear arrays of symbols, and not how we map linear order into hierarchy. In other words, hierarchy is simply assumed as given, and the issue is how we define its relationship with linear order; whereas from the perspective developed in this paper, the issue is how we create structure by starting with linear arrays of symbols. A large stream of research addresses the link between statistical computations and probabilistic expectations on one side and structure-building and hierarchy-based parsing on the other side. For instance, Chesi (2015;2021) propose a grammatical format in which structure-building, as driven by categorial top-down expectations, permits a reduction of the hiatus between parsing and generation, aligning hierarchy and linear order. In this model, adapted from Stabler's work on the formalization of Minimalist Grammars (cf. Stabler 2013), the order of structure-building parallels the order in which words are parsed left-to-right (cf. also Momma & Phillips 2018). A similar model of hierarchy-building is developed by Phillips (2003), where syntactic structure is built incrementally from left to right, by means of a top-down incremental process. Phillips (2003) shows that making hierarchical constructions depend on left-to-right linear processing involves an incremental process whereby constituents are progressively created and destroyed. In this way, it becomes possible to explain the different results of different syntactic diagnostics for constituency (coordination, movement, ellipsis, etc.), basically by relying on the differences in structure among the different derivational stages of structure-building on which each of these diagnostics is necessarily based. Another interesting class of theories of processing is the so-called 'surprisal theories' (cf. Levy 2008, Futrell & Levy 2017). Here, processing difficulty is related to the metrics of surprisal, which is defined, in turn, as a relationship between incremental probabilistic disambiguation of competing (structural) analyses and processing complexity. In other words, the surprisal effect is measured by the 'size' or degree of the update necessary to select the structural analysis that (most) correctly predicts the new word. In this way, the focus is on probabilistic grammatical models (in Levy's paper, the grammatical format of reference is a Probabilistic Context-Free Grammar). The interest of surprisal theory lies in the fact that whereas predictability is generally regarded as a function of the semantic context analysis, this theory links surprisal to a variety of language-internal sources, crucially including morphosyntax and phonology. On these premises, predictability gives rise to experimentally detectable effects in eye-tracking reading studies (in terms of reduced reading times and increased skipping probability for the predictable word) and in ERP-studies, in terms of a differential N400 effect. A competing theory of processing is Gibson's Dependency Locality Theory (Gibson 1998), in which the integration cost of a given word monotonically increases with the distance from the elements on which the words depend and on the number of intervening dependencies. These two theories often make conflicting predictions, as for head-final languages and the contrast between subject and object relatives. In a nutshell, the reason is that Surprisal Theory does not necessarily interpret the lexical material intervening between the point at which an expectation is originated and the point at which the expectation is satisfied as a trigger for processing difficulty, because in many cases this intervening material maximizes the expectation for a certain phrase (for instance, the head-final verb) with respect to competing words/phrases. Interestingly from the present perspective, surprisal theories can be distinguished from prediction-based connectionist models, since the latter are generally trained on raw corpora, i.e. strings where only linear order is relevant, whereas Levy's models (Levy 2008), for example, are trained on syntactically annotated corpora and the measure of expectation is calculated based on probabilistic grammars and hierarchical dependencies among words. Proponents of connectionist models, starting with Christiansen & Chater (1999), have often emphasized that it is exactly the relations that are difficult for a model (trained on raw corpora) to learn which are mostly difficult to process for the native-speakers of a language, like for instance nested dependencies in center-embedding. This might suggest that even exposition to non-annotated corpora, that is, models trained on essentially non-hierarchical parameters, are somehow able to bootstrap hierarchy from order in elaborating the expectation-based metrics that optimally increase the prediction power of the model. These considerations are especially plausible when considering the remarkable recent achievements, in terms of handling of (complex) syntactic dependencies, of Generative Pre-trained Transformers (GPT) in Natural Language Processing research, as these models are notoriously trained with raw (non-annotated) large amounts of text (Linzen & Baroni 2021). All in all, in surprisal theories hierarchy building and the relevant structural representations affect processing difficulty only through the mediation of probabilistic computations on word-expectancy, confirming, from the present perspective, that word-prediction remains the computational engine and the driving power for the development of increasingly efficient parsing strategies. However, the aim of the present paper does not consist only in establishing a firm link between the linear and hierarchical representations of words in natural language; rather, it consists in searching for a (possibly) domain-general cognitive bias, in humans, towards projecting sequentially-ordered arrays of symbols into graph-like structures. In our experimental studies, this bias already emerges at the binary level that is generally considered relevant, in formal syntax, for the application of the elementary structure-building operation 'Merge', in terms of the tendency to re-interpret precedence in deterministic bigrams of Fibonacci-sequences as a containment relation between the two relevant symbols.

This leads to the following research questions: assuming that language is processed from left to right as an ordered sequence of symbols (sounds or gestures), how can we account for the capacity of the parser to build hierarchy-based representations of these symbols? In other words, how do we shift from precedence to containment from a dynamic processing perspective? To address this questions, we propose a radical shift of perspective: what should be investigated is the human cognitive capacity of mapping sequences of objects into objects modelled as trees or graphs, instead of the capacity of linearizing trees/graphs, regarded as computational primitives. More particularly, the present study concentrates on the formal and cognitive underpinnings of mapping precedence into containment, leaving neurolinguistic extensions to future work. Our objective here is to propose an entirely new probe into the cognitive development of hierarchical representations and the way this development connects to statistical computations on the string.

1.2 Addressing the debate with a new experimental paradigm: Our first studies

In this section, we briefly summarize the experimental results that we previously obtained in the two published studies (Vender *et al.* 2019, 2020) in which we launched the original research paradigm sketched above, based on a new probe for disentangling learning effects at the sequential and hierarchical level. Our research program is based on the application of an AGL paradigm, in the form of a Serial Reaction Times (SRT) task, to a specific class of non-canonical grammars (Lindenmayer-systems, from now on L-systems), whose formal properties lend themselves to the inquiry into the relation between precedence and containment. In typical AGL paradigms, after a training phase in which participants are exposed to a sequence of visual or auditory stimuli generated by an artificial grammar and asked to memorize them, they are shown a new set of stimuli and asked for grammaticality judgments on the basis of what they had implicitly learned about the (regularities induced by the) grammar. Typically, people display an above chance performance, suggesting that implicit learning somehow took place, although this knowledge might concern either surface statistical properties of the input or abstract properties of the underlying generative system (Ettlinger *et al.* 2016; Pothos 2007; Reber 1967). Although it must be credited for a challenging and intriguing research output, the majority of AGL studies assessing implicit learning face two potential problems: (i) in terms of cognitive underpinnings, asking for grammaticality judgments undermines the implicitness of the task; (ii) in addressing language-related concerns, canonical grammars do not really exploit the potentials of AGL, since they are in fact not best suited to easily disentangle sequential and hierarchical learning (Saddy 2009, 2018).

The new research paradigm we have developed is crucially based on SRT tasks, which guarantee, with respect to (i) above, a considerably higher degree of implicitness of learning (Cleeremans & McClelland 1991; Nissen & Bullemer 1987). More particularly, we employ a modified Simon task (Simon 1969), in which the sequence of the stimuli is arranged following the rules of a noncanonical grammar, the Fibonacci grammar (Fib), which, as we will show, constitutes an optimal tool to assess both sequential and hierarchical learning (Krivochen *et al.* 2018; Lindenmayer 1968), thus addressing the challenge in (ii).

Fib is an asymmetric L-system (Krivochen *et al.* 2018; Krivochen & Saddy 2018), defined by the alphabet S = 0, 1 and the rewriting rules in (1):

(1) $0 \rightarrow 1$
$1 \rightarrow 01$

Applying these two rules produces sequences of symbols which can be represented as in Figure 1.

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 1: REPRESENTATION OF THE FIBONACCI GRAMMAR.

Fib is a simple deterministic recursive rewrite system, with some peculiar properties representing fundamental differences from the canonical grammars in the Chomsky Hierarchy. First, in classical L-systems there is no distinction between non-terminals and terminals, which means that every symbol can (and thus must) be rewritten. Second, all rules that may apply do so simultaneously: L-grammars feature no Traffic Convention (Greibach 1965; Hopcroft & Ullman 1969). Third, a major feature of L-systems is self-similarity: any structure-sensitive pattern found in the derivation can be mapped to an earlier stage of the derivation, at every scale. These features of L-systems are made visible in the Fib-representation in Figure 2.

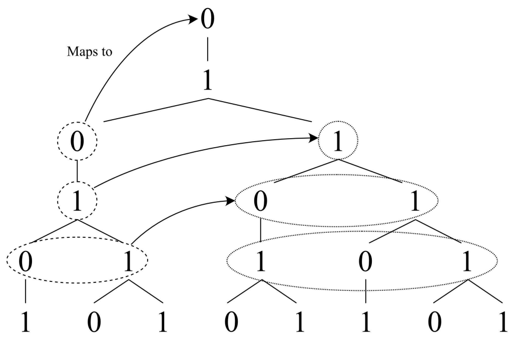

FIGURE 2: REPRESENTATION OF SELF-SIMILARITY IN THE FIBONACCI GRAMMAR.

Crucially for our present purposes, the properties of Fib that may be tied to statistical learning effects can be described as follows. Within each Fib-generation,[1] we find the following transitional regularities, which we will refer to as the Three Laws:

(a) First Law: A [0] is always followed by a [1], i.e. 00 is not produced by the grammar (henceforth, it is 'ungrammatical').

(b) Second Law: Two [1] are always followed by a [0], i.e. 111 is ungrammatical.

(c) Third Law: A single [1] can be followed by either a [0] or a [1], i.e., [1] does not provide enough information for a parser to determine what symbol will come next, since 010 and 011 are both possible continuations.

Taken together, the Three Laws entail that, by applying conditional statistics, the parser will be able to distinguish, within an arbitrary sequence of symbols corresponding to a Fib-generation, deterministic transitions from non-deterministic transitions (for instance, a 0 will always be followed by a 1, whereas a 1 may be followed by a 0 or by a 1). The reason why Fib lends itself to fruitfully investigate the distinction between sequential and hierarchical learning can be further clarified as follows. Consider the representation above of the first Fib generations: the sequence [01] is not only a natural constituent of the grammar, resulting from the rewriting rule $1 \rightarrow 01$, but it is also a sequence corresponding to a deterministic transition. This suggests that chunking the symbols 01 together as the 'constituent' [01] does not necessarily require that the parser figures out that [01] is the product of a specific rewriting

[1] Given the recursive format of the two generative rules ($0 \rightarrow 1$, $1 \rightarrow 01$) we define as a generation each of the sequences produced at a given recursive stage.

rule: it is sufficient that the succession between the two symbols is deterministic in terms of conditional statistics. The remaining 1s in the representation are dubbed k-points, and are highly relevant for our purposes as they permit to address the interplay between linear and hierarchical learning. As can be seen in Figure 3, k-points (circled with dotted line) are the 1s dominated by a 0 and dominating a 01.

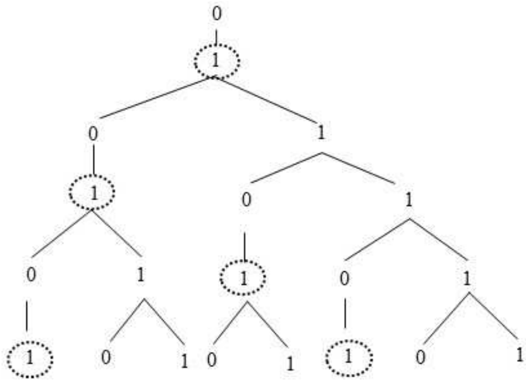

FIGURE 3: GRAPHICAL REPRESENTATION OF K-POINTS (01<u>1</u>, CIRCLED WITH DOTTED LINE).

Notice that, in linear terms, the second 1 within the sequence 01<u>1</u> is always a k-point. This is always, in strictly linear terms, an ambiguous point, since after a single 1 we may have either a 0 or a 1 (Third Law). Now, k-points are extremely interesting as a tool for inquiring into structure, for the following formal reason. If the parser were able to organize the linear array of symbols in progressively larger chunks (constituents), k-points might become progressively predictable (whereas they are not predictable, as just emphasized, as the result of linear statistical computations). Schmid *et al.* (2023) argue, for instance, that there is a gradual learning effects for k-points (i.e., the 1s that are not part of the deterministic bigram 01). They propose that the parser builds multiple levels of representation based on the possibility of treating 01, as a deterministic sub-sequence, as a unit which enters in turn some deterministic transitions at higher levels. So, at the first level of hierarchical computation, 01 deterministically follows 1, exactly as 1 deterministically follows 0 at the base-level; at the second level, the newly-constituted deterministic unit 101 deterministically follows 01, and so on at the successive levels with progressively larger chunks. Given this kind of hierarchical computations, 1s that were originally not predictable at the base-level become predictable as parts of higher-level deterministic transitions. However, this ‘structural’ strategy predicts an increasing predictability not only for the 1s, but also for the 0s that follow 01 (there would be no difference between the trigram 011 and the trigram 010). In our studies (Vender *et al.*, 2020, and the present one, as will be illustrated below), we found that RTs decrease faster in Fib for 1s, than for 0s, following 01. Here, we thus propose a different hypothesis about the role played by structural reconstruction, based on the possibility that the human parser applies a wired-in mapping principle ensuring bootstrapping from linear order into hierarchical representations. The basic insight is that if x precedes y, then x must be contained in y. The formal application of this principle ensures that the deterministic chunk 01 be categorized as 1 at the higher level, whereas the 1s following 01 (k-points) are categorized as 0s. As we will see below, in this way we are able to derive the progressive decreasing of RTs in $01\underline{1}$ as opposed to $01\underline{0}$, in line with our experimental findings for Fib.

Be it as it may, these considerations should elucidate why it is so interesting to inquire whether and under which conditions the parser is successful in predicting the occurrence of k-points (i.e., of 1s following a 01 chunk): essentially, these are points that cannot be predicted by only applying linear statistical calculations. Two other grammars, obtained by a modification of Fib, were used in our studies, Skip and Bif. Skip, which is obtained by manipulations of Fib, consists in replacing a generation n with a successive generation ($0\rightarrow$ 01, gen. 2; $1\rightarrow$01101, gen. 4). Crucially for our purposes, Skip features the same transitional regularities as Fib, the First and Second Law (*00; *111), while not preserving Fib's hierarchical structure, and most notably the structural properties of k-points. In other words, the 1s following 01 are still present in Skip, but their hierarchical properties are crucially different (that is, they are not immediately dominated by a 0 and dominating a 01, as in Fib). Interestingly, moreover, the frequency of 011 and 010 is different Fib and Skip: 011 is more frequent than 010 in Fib (ratio: 1.54), whereas the opposite holds in Skip, where 010 is more frequent than 011 (ratio: 1.67). In other words, the probability to have a 1 after 01 in Fib is 53.8%, whereas in Skip it decreases to 33%. The possibility thus arises that k-points learning in Fib be a pure effect of frequency statistics applied to the string, based on the higher frequency of 011.

Bif, instead, is a modification of Fib with a difference in the second generation rule: we have thus $0 \rightarrow 1$ (as in Fib) and $1 \rightarrow 10$ (Fib: $1 \rightarrow 01$). In Bif the Three Laws apply exactly as in Fib. The sequence 011 is present in the output of the grammar more frequently than its counterpart 010, as in Fib, although it cannot be considered a k-point, since the second 1 has a different hierarchical structure. The main features of the three grammars are summarized in Table 1, whereas their structure (the first five generations) is represented in Figure 4.

	Fib	Bif	Skip
Generation rules	0→1 1→01	0→1 1→10	0→01 1→01101
First law (*00)	valid	valid	valid
Second law (*111)	valid	valid	valid
011	Sequence present in the grammar's output, with specific hierarchical properties (the second 1 is a k-point, dominated by a 0 and dominating a 01)	Sequence present in the grammar's output, but devoid of hierarchical significance (the second 1 is not a k-point)	Sequence present in the grammar's output, but devoid of hierarchical significance (the second 1 is not a k-point)
Frequency of 011 and 010	011 is more frequent than 010 (ratio 1.54)	011 is more frequent than 010 (ratio 1.54)	010 is more frequent than 011 (ratio 1.67)

TABLE 1: MAIN FEATURES OF THE THREE GRAMMARS.

Fib	Bif	Skip
0	0	0
1	1	01
01	10	0101101
101	101	01011010101101011010101101
01101	10110	0101101010110101101010110101011010101101011010101101010110101011010101101010110101011010101101

FIGURE 4: REPRESENTATION OF THE FIRST GENERATIONS OF FIB, BIF AND SKIP.

In Vender *et al*. (2019), we administered the Modified Simon Task to four groups of 10-year-old monolingual and bilingual children, with and without a diagnosis of developmental dyslexia, with the aim of verifying whether the task was suitable for assessing implicit learning of typical and atypical populations, essentially focusing on statistical learning. The Simon Task (Simon, 1969) is a well-established experimental paradigm normally used in psycholinguistic studies to measure inhibition and conflict resolution; in a typical task, participants are presented with sequences of stimuli (e.g. a red square and a blue square) appearing on the left or on the right of the computer screen and are asked to respond to these visual stimuli based on their colour by making a leftward response to one stimulus (e.g. pressing a key on the left of the keyboard when they see a red square) and a rightward response to the other (e.g. pressing a key on the right of the keyboard when they see a blue square). The location of the stimuli on the display is irrelevant for the task, but it can influence the participants' reaction times (RTs) and accuracy in incongruent trials (when there is a mismatch between the location of the stimulus and the location of the key to be pressed) as compared to congruent trials (when the two locations match). The Simon effect represents the increase in RTs and accuracy observed in the incongruent conditions and is interpreted as a measure of inhibition and conflict resolution costs. While in traditional Simon tasks the sequence of the stimuli is random, in our modified Simon Task it was entirely determined by an artificial grammar, as those reported in Table 1 (Fib, Bif or Skip). In Vender *et al.* (2019) the version of the grammar that we employed was Bif. The attention was confined to learning the deterministic transitions corresponding to the two Laws, which was observed in four groups of school-aged children (monolingual and bilingual, with and without dyslexia), whereas the implicit learning of non-deterministic transitions (those corresponding to k-points) was not considered.

In Vender *et al.* (2020) we administered the same task to 22 9-year-old monolingual typically developing children with the aim of disentangling statistical and hierarchical learning effects. We found that the First Law was learned, whereas the Second Law was not; this is likely to depend on the fact that the task was shorter with respect to Vender *et al.* (2019) (330 vs. 432 trials), and thus confirms that the First Law (*00) is acquired earlier than the second Law (*111), as already observed in Vender *et al.* (2019). Interestingly, our results showed preliminary evidence for hierarchical learning, since subjects showed a sensitivity to k-points in Fib (RTs for the trigram 011 were faster in Fib than in Skip).

Although these results were far from conclusive and deserved a more thorough investigation, they were suggestive of the possibility that sensitivity to k-points in a sequence generated by a Fibonacci grammar depends on the bidimensional computation of some hierarchical properties associated with the sequence besides the computation of transitional regularities on the string. On these grounds, the two studies presented here were developed as a follow-up allowing the investigation of the role possibly played by a hierarchical mode of computation in the explanation of the learning effects found with k-points.

1.3 Research questions and predictions

Here is a brief synopsis of the import of the current study for the inquiry into the relationship between precedence-based and containment-based styles of computation, extending the results of Vender *et al.* (2019; 2020).

In Study 1, we employed 3 Skip blocks followed by 3 Fib blocks: if k-points learning is simply a matter of distributional statistics and thus determined by the higher frequency of 011 in Fib, we should expect similar learning effects for the sequence 010 in Skip, due to its higher frequency in this grammar. Conversely, any asymmetry between the learning effect for k-points (i.e. 011) in Fib and the learning effect for 0 in the sequence 010 in Skip, would strongly suggest that the parsing algorithm responsible for 'predicting' a k-point cannot be reduced to statistical sampling procedures applied to the string.

In Study 2, we aimed at comparing learning effects for k-points in the grammar Fib (where 1 is rewritten as 01) and in the grammar Bif (where 1 is rewritten as 10). Despite the superficial resemblance between these two grammars, k-points have a different status in these two grammars, and this enabled us to gain a better insight into the nature (global hierarchical computation vs. local hierarchical computation) and some further properties (hypothetical hierarchical reasoning as opposed to actual hierarchical reconstruction) of the parsing strategies that prompt the enhanced capacity to predict k-points as the SRT task progresses.

2. STUDY 1

2.1 Method

Participants

The experimental protocol was administered to 27 students of the University of Verona with no history of language or learning disorders (mean age: 23.7, SD = 3.8). None of the participants had any reports of brain damage, sensory impairments, or serious emotional, behavioural, or learning problems. Their vision was normal or corrected to normal.

Materials and procedures

Participants were administered a modified Simon Task, similar to those deployed in Vender *et al*. (2019, 2020). The task was run on an Asus 15.6' laptop using DMDX Automode version 6.0.0.4 software. Subjects were unaware of the real purposes of the study; they were simply asked to perform a traditional Simon Task, i.e. they were explained that they would see some blue or red squares appearing on the left or on the right side of the screen and that they had to press the number key 1 (on the left side of the keyboard) as they saw a red square, and the number key 0 (on the right side) as they saw a blue one, ignoring the position of the square on the screen. There were thus four conditions: two congruent conditions, when the red square appeared on the left side of the screen and the blue square on the right side, and two incongruent conditions, when the red square appeared on the right and the blue on the left. In these conditions, the task was complicated by the need to inhibit the tendency to press the key on the same side of the square. The presence of incongruent trials, occurring every sixth items, while being irrelevant for our research questions, was adopted as in our previous studies (Vender *et al.*, 2019, 2020) to keep the task engaging for the participants, reducing boredom, and was maintained also in the current research to ensure comparability with previous results. As in the previous studies, the stimuli were four blue or red squares (dimensions 1012x536 pixels, BMP files). Each trial started with a fixation cross which appeared in the middle of the screen and remained visible for 500 ms and which was followed by one of the four stimuli. Participants had 1000 milliseconds to press a key: if they did not provide an answer within this time window, the square disappeared, and the following item was shown (see Figure 5 for a representation of the experimental procedure). The timing started with the onset of the item and ended with the response of the subject. Both accuracy and RTs data were collected.

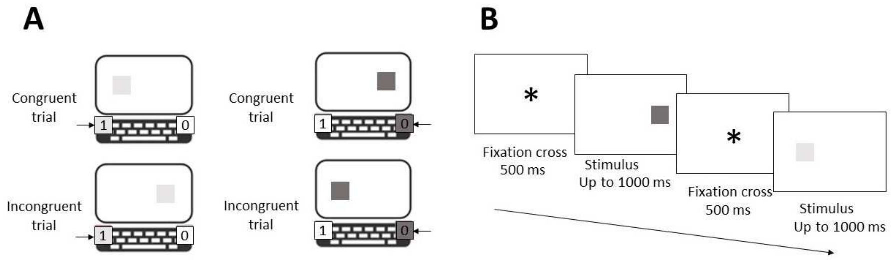

FIGURE 5: REPRESENTATION OF THE EXPERIMENTAL PROCEDURE (STUDY 1 AND STUDY 2). FIGURE A: REPRESENTATION OF CONGRUENT AND INCONGRUENT TRIALS OF THE MODIFIED SIMON TASK, WHERE PARTICIPANTS ARE ASKED TO PRESS 1 WHEN THEY SEE A RED SQUARE (IN LIGHT GREY HERE) AND 0 WHEN THEY SEE A BLUE SQUARE (IN DARK GREY HERE), INDEPENDENTLY ON THE POSITION OF THE STIMULUS ON THE SCREEN. FIGURE B: REPRESENTATION OF THE EXPERIMENTAL TASK.

The sequence of the stimuli was completely determined by the rule of an underlying grammar, Fib or Skip. As argued above Fib and Skip have the same statistical regularities, which translated into colours, where 0 corresponds to red and 1 corresponds to blue, are: (i) First Law: A red is always followed by a blue (i.e. the sequence red-red is ungrammatical) (ii) Second Law: A sequence of two blues is always followed by a red (i.e. the sequence blue-blue-blue is ungrammatical) (iii) Third Law: The sequence red-blue is ambiguous, as it can be followed either by a red, or by a blue (i.e. both red-blue-blue and red-blue-red are possible). The task was in this case considerably longer than in Vender *et al.* (2019, 2020): it was composed of 1092 trials, with three Skip blocks (97 trials each, in total 291) followed by three Fib blocks (267 trials each, in total 801).[2] There were 8 random practice trials in which subjects received feedback; after this short training, participants had the chance to ask questions before the experiment began. They were tested individually, in the presence of the experimenter, in a quiet room sitting in front of the computer screen. It took approximately 30 minutes to perform the task. The study was approved by the local Ethics Committee and conducted in accordance with the standards specified in the 2013 Declaration of Helsinki; moreover, written informed consent was obtained for all participants.

2.2 Results

Data were fitted to a series of linear mixed effects regression models using the statistical environment R (R Core Team 2020) and in particular the packages lme4 and lmerTest (Bates *et al.* 2015; Kuznetsova *et al.* 2017). For each model, we adopted a stepwise removal method starting from a fully specified model. We specified RTs as dependent variable, Point (Target points, vs. Ambiguous points) Congruency (Congruent trials vs. Incongruent trials) and Grammar (Skip vs. Fib) as independent variables with full interaction, with Subject as random intercept. To assess the presence of learning effects we selected the following target points (highlighted for clarity): $0\underline{1}$ for the First Law (Analysis 1), $11\underline{0}$ for the Second Law (Analysis 2) and $01\underline{1}$ for k-points (Analysis 3). In each analysis we compared the relevant points to the ambiguous points $01\underline{0}$.

Analysis 1: Learning of the First Law (01)

To assess learning of the First Law (a red is always followed by a blue, i.e. $0\underline{1}$, *00), we considered all 1s following a 0 across the Skip and Fib blocks and we compared them to the ambiguous points $01\underline{0}$, i.e. to all the 0s following the sequence 01. As shown in Table 2, RTs decrease significantly across blocks and are indeed considerably faster in Fib than in Skip for $0\underline{1}$ with respect to

[2] The difference in the number of trials between Skip and Fib in this Study is due to the fact that our principal aim was that of investigating structural learning in Fib, where there are k-points, which required a high number of trials to be learnt, as found in previous research (Vender *et al.*, 2020). Skip only acted as a control “foil” grammar to verify whether the learning of k-points was due to linear as opposed to hierarchical learning. Since there are no comparable structural points in Skip which we were interested in studying and Skip was only needed to verify whether the learning of k-points was due to linear effects, we decided not to further increase the total length of the experimental protocol by presenting a lower number of Skip stimuli. Notice further that in this case the Fib blocks follow the Skip blocks, whereas in Vender *et al*. (2020) Fib preceded Skip.

01<u>0</u>; accuracy is at ceiling in 0<u>1</u>, but it is lower in 01<u>0</u>, especially in incongruent trials.

	Skip Grammar	Fib Grammar	Block 1 (Skip)	Block 2 (Skip)	Block 3 (Skip)	Block 4 (Fib)	Block 5 (Fib)	Block 6 (Fib)
RTs Congruent	359.45 (84.16)	298.05 (62.03)	389.65 (87.04)	353.11 (84.36)	335.60 (81.07)	311.92 (59.84)	296.52 (52.40)	285.70 (73.84)
RTs Incongruent	491.85 (104.15)	429.80 (112.36)	496.48 (98.63)	495.21 (104.64)	483.85 (109.17)	458.94 (98.79)	427.93 (110.89)	402.53 (127.40)
Accuracy Congruent	0.99 (0.02)	0.99 (0.02)	0.99 (0.01)	0.99 (0.02)	0.99 (0.02)	0.99 (0.01)	0.99 (0.02)	0.98 (0.04)
Accuracy Incongruent	0.89 (0.16)	0.94 (0.08)	0.85 (0.15)	0.89 (0.13)	0.92 (0.20)	0.92 (0.10)	0.96 (0.06)	0.95 (0.07)

TABLE 2: MEAN (SDS) RTS AND ACCURACY RATES OF POINT 0<u>1</u> IN SKIP AND FIB GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS (ANALYSIS 1).

	Skip Grammar	Fib Grammar	Block 1 (Skip)	Block 2 (Skip)	Block 3 (Skip)	Block 4 (Fib)	Block 5 (Fib)	Block 6 (Fib)
010 RTs Congruent	394.61 (77.58)	382.25 (66.32)	400.98 (79,25)	398.23 (72.07)	384.63 (81.42)	379.16 55.13	385.99 (60.54)	381.59 (59.41)
010 RTs Incongruent	533.76 (127.54)	494.05 (131.95)	553.44 (111.63)	497.78 (121.19)	550.05 (149.79)	504.02 (125.57)	502.75 (139.72)	475.37 (130.55)
010 Accuracy Congruent	0.98 (0.03)	0.95 (0.09)	0.99 (0.02)	0.98 (0.05)	0.99 (0.02)	0.97 (0.05)	0.95 (0.08)	0.93 (0.13)
010 Accuracy Incongruent	0.80 (0.4)	0.73 (0.44)	0.77 (0.43)	0.87 (0.33)	0.75 (0.43)	0.77 (0.42)	0.75 (0.44)	0.69 (0.47)

TABLE 3: MEAN (SDS) RTS AND ACCURACY RATES OF THE AMBIGUOUS POINT 01<u>0</u> IN SKIP AND FIB GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS.

The best fitting model showed significant main effects of Point (β = 83.84, SE = 2.52, t = 33.30, p <.001), indicating that 0<u>1</u> were reacted to faster than 01<u>0</u>, Grammar (β = 61.67, SE = 2.43, t = 25.40, p <.001), with RTs being faster in Fib than in Skip, and Congruency (β = 130.27, SE = 3.31, t = 39.33, p <.001), with RTs being faster in congruent trials. The Grammar*Point interaction was significant as well (β = 49.22, SE = 4.14, t = 11.90, p <.001), showing that the decrease in RTs was significantly steeper for 0<u>1</u> than for 01<u>0</u>, and a significant Point*Congruency interaction (β = 24.07, SE = 6.97, t = 3.45, p <.001), indicating that the gap between 01 and 010 was sharper with congruent trials than with incongruent trials. The remaining Grammar*Congruency and Point*Grammar*Congruency were not significant. All in all, results indicate that the First Law was learnt, as shown by the fact that participants became faster in reacting to the deterministic 1s following a 0 (First Law) with respect to the ambiguous points 010.

Analysis 2: Learning of the Second Law (110)

To verify whether the second regularity (two blues are always followed by a red, i.e. 110, *111) was learnt, we considered each red trial following a sequence of two blues, i.e. 110, as compared to 010. As shown in Table 4, in both congruent and incongruent trials RTs decrease significantly across blocks with 110, being faster in Fib than in Skip, while accuracy remains high and stable.

	Skip Grammar	Fib Grammar	Block 1 (Skip)	Block 2 (Skip)	Block 3 (Skip)	Block 4 (Fib)	Block 5 (Fib)	Block 6 (Fib)
RTs Congruent	390.09 (68.90)	355.22 (58.06)	417.20 (69.99)	350.04 (67.73)	403.03 (68.97)	366.05 (53.88)	354.16 (56.73)	345.44 (63.56)
RTs Incongruent	499.00 (119.57)	445.16 (105.31)	528.59 (119.83)	468.25 (104.75)	500.16 (134.14)	452.35 (75.95)	453.85 (115.25)	429.29 (124.73)
Accuracy Congruent	0.98 (0.04)	0.96 (0.07)	0.98 (0.04)	1.00 (0.02)	0.96 (0.07)	0.97 (0.03)	0.95 (0.08)	0.95 (0.11)
Accuracy Incongruent	0.91 (0.22)	0.85 (0.21)	0.92 (0.18)	0.89 (0.23)	0.91 (0.24)	0.87 (0.19)	0.82 (0.22)	0.85 (0.21)

TABLE 4: MEAN (SDS) RTS AND ACCURACY RATES OF POINT 110 IN SKIP AND FIB GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS (ANALYSIS 2).

We found a significant main effect of Point (β = 30.01, SE = 2.80, t = 10.71, p <.001), Grammar (β = 12.16, SE = 3.43, t = 3.54, p <.001) and Congruency (β = 106.63, SE = 6.29, t = 16.94, p <.001), respectively showing that 110 points were reacted to faster than 010, that RTs were significantly lower in Fib than in Skip and in congruent than in incongruent trials. We also found a significant Grammar*Point interaction (β = 21.53, SE = 5.20, t = 4.14, p <.001), indicating that RTs decreased significantly faster in 110 than in 010, and a significant Grammar*Congruency interaction (β = 20.39, SE = 9.89, t = 2.06, p <.05), indicating that the gap between Fib and Skip was sharper with congruent trials than with incongruent trials. The remaining Point*Congruency and Point*Grammar*Congruency interactions were not significant. All in all, results indicate that the Second Law was learnt.

Analysis 3: Learning of k-points (011) vs. not-k points (010)

We analyzed RTs across the two grammars comparing k-points (each blue trial following a red-blue sequence, 011) with ambiguous not-k points (010). Mean accuracy and RTs across Blocks and grammars are reported in Table 5.

	Skip Grammar	Fib Grammar	Block 1 (Skip)	Block 2 (Skip)	Block 3 (Skip)	Block 4 (Fib)	Block 5 (Fib)	Block 6 (Fib)
k-points RTs Congruent	431.39 (79.89)	385.94 (67.05)	452.90 (86.42)	418.36 (82.27)	422.93 (70.99)	393.88 (64.25)	387.19 (65.96)	376.76 (70.94)
k-points RTs Incongruent	524.95 (146.86)	486.53 (139.30)	531.29 (140.29)	- -	518.62 (153.43)	489.25 (129.08)	491.23 (145.25)	479.13 (143.58)
k-points Accuracy Congruent	0.97 (0.06)	0.96 (0.07)	0.99 (0.04)	0.99 (0.04)	0.92 (0.12)	0.97 (0.06)	0.96 (0.08)	0.95 (0.07)
k-points Accuracy Incongruent	0.73 (0.43)	0.55 (0.50)	0.83 (0.38)	- -	0.64 (0.48)	0.57 (0.50)	0.54 (0.50)	0.54 (0.50)

TABLE 5: MEAN (SDS) RTS AND ACCURACY RATES OF K-POINTS (011) IN SKIP AND FIB GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS (ANALYSIS 3).

The fixed effects of the best fitting model are reported in Table 6. We found a significant main effect of Grammar ($p < .001$), with RTs being faster in Fib than in Skip, of Congruency ($p < .001$), with incongruent trials being faster than congruent ones, and of Point ($p < .001$), with not-k points (010) being faster than k-points (011). However, the significant Grammar*Point interaction indicated that k-points showed a steeper decrease in RTs passing from Skip to Fib, whereas this decrease was less sharp for not-k points, as visually displayed in Figure 6. Moreover, the significant Congruency*Point interaction showed that the difference between congruent and incongruent trials was more marked in not-k points that in k-points, both in Skip and in Fib. Finally, the significant Congruency*Grammar interaction showed that there was a wider difference between congruent and incongruent trials in Skip than in Fib. Taken together, these results indicate that, although not-k points are generally faster than k-points, only the latter showed a marked decrease in Fib. The fact that RTs decreased in both types of points can arguably be interpreted as a side-effect of the Simon Task, with adults simply becoming faster in performing the task: however, the fact that the slope was steeper for k-points, besides confirming that the two classes of points are different, suggests that the presence of a learning effect is significantly stronger for k-points.

Predictor	Estimate	Std.Error	t-value	p-value
Intercept	394.32	11.37	28.27	<.001
Grammar (Fib)	-12.45	3.29	-3.79	<.001
Point (k)	35.43	4.33	8.19	<.001
Congruency (incongruent)	126.84	7.31	17.36	<.001
Grammar (Fib)* Point (k)	-31.05	5.10	-6.09	<.001
Grammar (Fib) * Congruency (Incongruent)	-21.06	9.47	-2.22	<.05
Point (k) * Congruency (Incongruent)	-38.96	11.09	-3.51	<.001
Grammar (Fib) * Point (k) * Congruency (Incongruent)	23.38	13.66	1.71	ns (p = .09)

TABLE 6: SUMMARY OF THE FIXED EFFECTS OF THE MIXED EFFECTS MODELS OF ANALYSIS 3 (STUDY 2) – (N = 10579, LOG-LIKELIHOOD = -64446). RANDOM EFFECTS FOR SUBJECTS HAD SD OF 57.62.

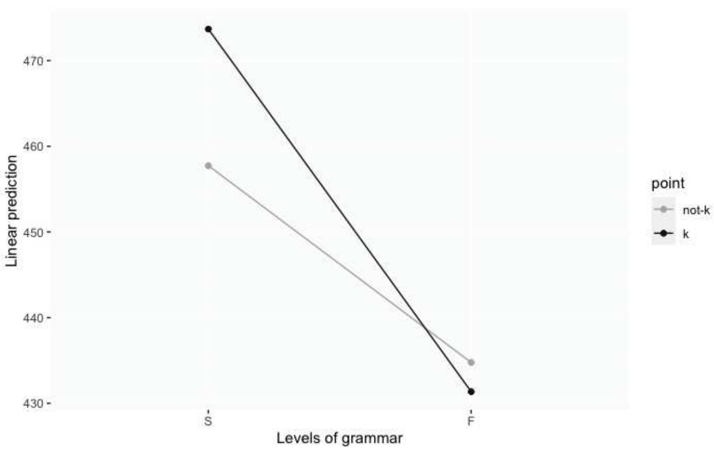

FIGURE 6: VISUAL REPRESENTATION OF THE GRAMMAR*POINT INTERACTION (STUDY 1, ANALYSIS 3). F = FIB, S = SKIP, K = K-POINTS (011), NOT-K = NOT-K, AMBIGUOUS POINTS (01<u>0</u>).

Furthermore, we ran an additional follow-up analysis comparing two distinct types of k-points in Fib: considering the longer sequence in which k-points are embedded, we can indeed distinguish between 01011 k-points and 11011 k-points. Since the former are in principle predictable in linear terms (there is always a 1 following two sequences of 01, i.e. *01010 is not attested in Fib), we called them predictable k-points, as opposed to 11011 which are unpredictable in linear terms (since 11010 are present in the grammar). This is particularly relevant for our purposes, as if participants predict k-points on linear grounds, we should find a difference between the two types of points. The best fitting model had Point (01011 vs 11011) and Congruency as fixed factors, while the interaction was excluded. We found only a main effect of Congruency (β = 91.60, SE = 4.72, t = 19.40, p <.001), but no effect of Point (β = .32, SE = 2.94, t = .11, p = .91), suggesting that there are no differences between these two types of points, and that the learning effects with k-points cannot be reduced to the result of statistical calculations on the sequence.

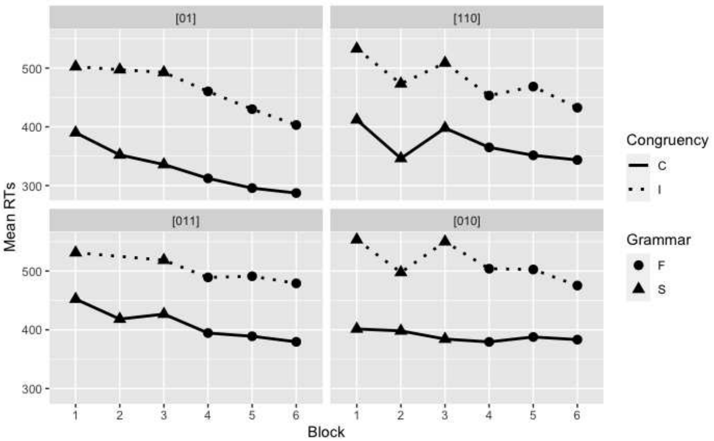

FIGURE 7: LEARNING CURVES OF MEAN RTS ACROSS BLOCKS IN 01, 110, 011 AND 010 (STUDY 1). C = CONGRUENT TRIALS; I = INCONGRUENT TRIALS; F = FIB; S = SKIP. 01 (FIRST LAW); 110 (SECOND LAW); 011 (K-POINTS); 010 (AMBIGUOUS POINTS).

All in all, results indicate that the First Law and the Second Law were learnt, as well as k-points. This is shown by the decrease in RTs from Skip to Fib, both for the deterministic points 01 and 110 and for k-points. In all cases, this decrease was sharper than that observed with the ambiguous points 010 (as also observable in Figure 7), indicating that the learning effects that we found in Analysis 1, 2 and 3 cannot be considered as the consequence of an increased practice with the task.

3. STUDY 2

3.1 Method

Participants

The experimental protocol was administered to 37 students of the University of Verona with no history of language or learning disorders (mean age: 24.9, SD = 3.1; age range: 21-37). None of the participants had any reports of brain damage, sensory impairments, or serious emotional, behavioural or learning problems. Their vision was normal or corrected to normal.

Materials and procedures

Participants were administered a modified Simon Task, which differed from the previous ones only in the structure, as it was composed of 4 Fib Blocks (89 items each, in total 356) and 4 Bif Blocks (89 items each, in total 356), for a total of 712 trials. Participants were tested individually in a quiet room. The whole task lasted approximately 20 minutes. Written informed consent was obtained for all participants.

3.2 Results

As in Study 1, data were fitted to a series of linear mixed effects regression models with RT as dependent variable, specifying Point (target, i.e. 0$\underline{1}$, 11$\underline{0}$ and 01$\underline{1}$ vs. ambiguous, i.e. 01$\underline{0}$) Congruency (Congruent trials vs. Incongruent trials) and Grammar (Fib vs. Bif) as independent variables and Subject as crossed-effects random intercept.

Analysis 1: Learning of the First Law (0$\underline{1}$)

Mean RTs and accuracy across blocks considering each blue trial following a red one (0$\underline{1}$) are reported in Table 7, whereas the data of the ambiguous sequence 01$\underline{0}$ is reported in Table 8.

We found a main effect of Point (β = 96.30, SE = 3.16, t = 30.50, p <.001), with 0$\underline{1}$ being faster than 01$\underline{0}$, of Grammar (β = 45.89, SE = 2.28, t = 20.15, p <.001), with Bif being faster than Fib, and of Congruency (β = 144.53, SE = 3.87, t = 37.35, p <.001), with congruent trials being faster than incongruent trials. We also found a significant Point*Grammar interaction (β = 36.23, SE = 4.37, t = 8.29, p <.001), showing that the RTs decrease from Fib to Bif was significantly sharper for 0$\underline{1}$ than for 01$\underline{0}$. There was also a significant

	Fib Grammar	Bif Grammar	Block (Fib)	1 Block (Fib)	2 Block (Fib)	3 Block (Fib)	4 Block (Bif)	5 Block (Bif)	6 Block (Bif)	7 Block 8 (Bif)
RTs Congruent	334.06 (71.27)	290.43 (57.42)	359.70 (88.19)	330.43 (70.99)	328.81 (67.91)	317.32 (58.00)	300.75 (58.66)	291.80 (53.48)	289.69 (55.38)	279.47 (62.17)
RTs Incongruent	488.72 (92.7)	436.375 (89.82)	502.06 (99.33)	480.05 (91.27)	487.93 (77.14)	484.86 (103.05)	454.79 (86.79)	443.80 (79.96)	441.80 (93.88)	405.11 (98.64)
Accuracy Congruent	0.99 (0.025)	0.99 (0.03)	0.99 (0.03)	0.99 (0.02)	0.99 (0.02)	0.99 (0.03)	1.00 (0.01)	0.98 (0.08)	0.99 (0.02)	0.99 (0.02)
Accuracy Incongruent	0.91 (0.14)	0.94 (0.11)	0.90 (0.15)	0.93 (0.12)	0.92 (0.13)	0.91 (0.16)	0.95 (0.10)	0.93 (0.14)	0.93 (0.11)	0.95 (0.09)

TABLE 7: MEAN (SDS) RTS AND ACCURACY RATES OF POINT 01 IN FIB AND BIF GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS (ANALYSIS 1).

	Fib Grammar	Bif Grammar	Block (Fib)	1 Block (Fib)	2 Block (Fib)	3 Block (Fib)	4 Block (Bif)	5 Block (Bif)	6 Block (Bif)	7 Block 8 (Bif)
010 RTs Congruent	396.10 (78.95)	386.26 (72.58)	401.14 (84.06)	396.13 (80.35)	389.33 (77.52)	397.81 (73.87)	395.03 (71.89)	386.79 (78.15)	386.70 (71.25)	376.51 (69.04)
010 RTs Incongruent	491.62 (114.17)	472.33 (124.19)	490.73 (111.47)	491.22 (119.45)	502.19 (116.06)	482.33 (109.70)	500.69 (133.98)	463.56 (112.93)	465.40 (116.40)	459.67 (133.43)
010 Accuracy Congruent	0.98 (0.06)	0.96 (0.07)	0.98 (0.06)	0.97 (0.05)	0.97 (0.05)	0.98 (0.07)	0.96 (0.05)	0.97 (0.06)	0.97 (0.06)	0.95 (0.09)
010 Accuracy Incongruent	0.80 (0.40)	0.72 (0.45)	0.85 (0.36)	0.82 (0.38)	0.76 (0.43)	0.76 (0.43)	0.73 (0.45)	0.80 (0.40)	0.70 (0.46)	0.67 (0.47)

TABLE 8: MEAN (SDS) RTS AND ACCURACY RATES OF THE AMBIGUOUS POINT 010 IN FIB AND BIF GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS.

Point*Congruency interaction (β = 63.91, SE = 7.92, t = 8.07, p <.001), showing that the difference between 01 and 010 was more marked with congruent trials. All in all, these results provide evidence for the learning of the First Law, as testified by shorter RTs in Bif than in Fib, in both congruent and incongruent trials (the former being faster).

Analysis 2: Learning of the Second Law (110)

We analyzed each red trial following a sequence of two blues; see Table 9 for RTs and accuracy data.

We found a significant main effect of Point (β = 30.01, SE = 2.80, t = 10.71, p <.001), Grammar (β = 12.16, SE = 3.43, t = 3.54, p <.001) and

	Fib Grammar	Bif Grammar	Block (Fib)	1 Block (Fib)	2 Block (Fib)	3 Block (Fib)	4 Block (Bif)	5 Block (Bif)	6 Block (Bif)	7 Block 8 (Bif)
RTs Congruent	365.61 (69.54)	343.04 (64.39)	383.06 (82.46)	385.74 (70.45)	376.55 (61.96)	317.10 (63.31)	353.94 (66.40)	350.87 (64.09)	334.89 (62.70)	332.45 (64.39)
RTs Incongruent	455.22 (76.05)	420.36 (81.81)	459.23 (83.48)	450.02 (70.52)	454.15 (74.61)	457.48 (75.58)	432.44 (68.99)	432.28 (88.03)	418.34 (85.49)	398.38 (84.72)
Accuracy Congruent	0.98 (0.04)	0.98 (0.03)	0.98 (0.05)	0.97 (0.05)	0.99 (0.03)	0.99 (0.02)	0.99 (0.02)	0.97 (0.05)	0.98 (0.03)	0.98 (0.03)
Accuracy Incongruent	0.87 (0.22)	0.84 (0.26)	0.90 (0.23)	0.92 (0.14)	0.86 (0.23)	0.82 (0.27)	0.85 (0.27)	0.85 (0.26)	0.84 (0.24)	0.84 (0.26)

TABLE 9: MEAN (SDS) RTS AND ACCURACY RATES OF POINT 110 IN FIB AND BIF GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS.

of Congruency (β = 106.63, SE = 6.29, t = 16.94, p <.001), indicating that RTs decreased in Bif, in both congruent and incongruent trials (the former being faster) and that 110 was faster than 010. We also found a significant Point*Grammar interaction (β = 21.53, SE = 5.20, t = 4.14, p <.001), indicating that RTs decrease from Fib to Bif was significantly steeper for 110 points than for 010, thus confirming that the Second Law was learnt.

Analysis 3: Learning of k-points (011) vs. not-k points (010)

We analyzed RTs across the two grammars comparing k-points (each blue trial following a red-blue sequence, 011) with not-k points (010). Mean accuracy and RTs across Blocks and grammars are reported in Table 10.

The results of the best fitting model are reported in Table 11. We found a significant main effect of Grammar (p <.01), with RTs being faster in Bif than in Fib, of Congruency (p <.001), with incongruent trials being faster than congruent ones, and of Point (p <.001), with k-points (011) being faster than not-k points (010). Crucially, the significant Grammar*Point interaction showed that k-points displayed a steeper decrease in RTs passing from Fib to Bif, whereas this decrease was less marked for not-k points, as visually displayed in Figure 8. Moreover, the significant Congruency*Point interaction revealed that k-points had faster RTs in congruent trials than not-k points. All in all, results show that k-points (011) are significantly faster than not-k points (010) and that the former decrease faster than the latter across the task, suggesting that subjects learnt to predict them more efficiently than with not-k points.

As a follow-up analysis we additionally compared the two types of k-points, the linearly predictable ones (01011) and the linearly unpredictable ones (11011), as in Study 1: the best fitting model had Point (01011 vs 11011),

	Fib Grammar	Bif Grammar	Block (Fib)	1 Block (Fib)	2 Block (Fib)	3 Block (Fib)	4 Block (Bif)	5 Block (Bif)	6 Block (Bif)	7 Block 8 (Bif)
k-points RTs Congruent	375.53 (70.80)	351.76 (59.02)	384.48 (78.09)	378.42 (70.85)	372.54 (64.81)	366.67 (69.46)	361.64 (56.11)	345.97 (62.48)	357.68 (64.05)	341.74 (53.45)
k-points RTs Incongruent	483.29 (129.65)	459.32 (113.44)	484. (120.97)	493.80 (146.09)	476.78 (116.04)	477.68 (135.51)	463.84 (105.42)	472.75 (129.51)	468.64 (105.37)	432.04 (113.46)
k-points Accuracy Congruent	0.98 (0.03)	0.97 (0.05)	0.98 (0.04)	0.98 (0.04)	0.98 (0.03)	0.98 (0.03)	0.98 (0.03)	0.96 (0.09)	0.97 (0.05)	0.98 (0.04)
k-points Accuracy Incongruent	0.75 (0.43)	0.68 (0.47)	0.81 (0.39)	0.7 (0.43)	0.70 (0.46)	0.72 (0.45)	0.71 (0.46)	0.64 (0.48)	0.74 (0.44)	0.64 (0.48)

TABLE 10: MEAN (SDS) RTS AND ACCURACY RATES OF K-POINTS (011) IN FIB AND BIF GRAMMARS, FOLLOWED BY MEAN (SDS) RTS AND ACCURACY RATES OF EACH BLOCK FOR BOTH CONGRUENT AND INCONGRUENT TRIALS (ANALYSIS 3).

Predictor	Estimate	Std.Error	t-value	p-value
Intercept	396.73	9.75	40.71	<.001
Grammar (Bif)	-10.18	3.62	-2.81	<.01
Point (k)	-24.13	3.22	7.50	<.001
Congruency (incongruent)	91.43	7.36	12.43	<.001
Grammar (Bif) *Point (k)	-10.63	4.58	-2.32	<.05
Grammar (Bif) * Congruency (Incongruent)	-11.88	9.93	-1.20	ns (.23)
Point (k) * Congruency (Incongruent)	18.53	8.99	2.06	<.05
Grammar (Fib) * Point (k) * Congruency (Incongruent)	3.72	12.74	.29	ns (.77)

TABLE 11: SUMMARY OF THE FIXED EFFECTS OF THE MIXED EFFECTS MODELS OF ANALYSIS 3 (STUDY 2) – (N = 9317, LOG-LIKELIHOOD = -56180). RANDOM EFFECTS FOR SUBJECTS HAD SD OF 57.29.

Grammar and Congruency as fixed factors with full interaction. We found only a main effect of Grammar (β = 19.65, SE = 4.27, t = 4.60, p <.001), of Congruency (β = 107.01, SE = 8.66, t = 12.36, p <.001), but no effect of Point (β = 1.59, SE = 3.71, t = .43, p = .67). Summarizing, as in Study 1 results provide evidence of learning of First Law, Second Law and k-points. This is shown by the decrease in RTs from Skip to Fib, both for the deterministic points 01 and 110 and for k-points. Again, this decrease was always sharper than that

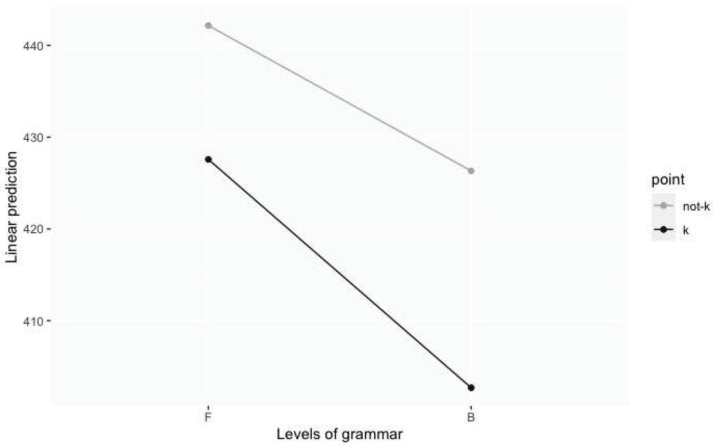

FIGURE 8: VISUAL REPRESENTATION OF THE GRAMMAR*POINT INTERACTION (STUDY 2, ANALYSIS 3). F = FIB, B = BIF. K-POINTS = 01<u>1</u>, NOT-K POINTS = 01<u>0</u>.

observed with the ambiguous points 01<u>0</u>, indicating that the learning effects that we found in Analysis 1, 2 and 3 cannot be seen as the consequence of an increased practice with the task, as represented in Figure 9.

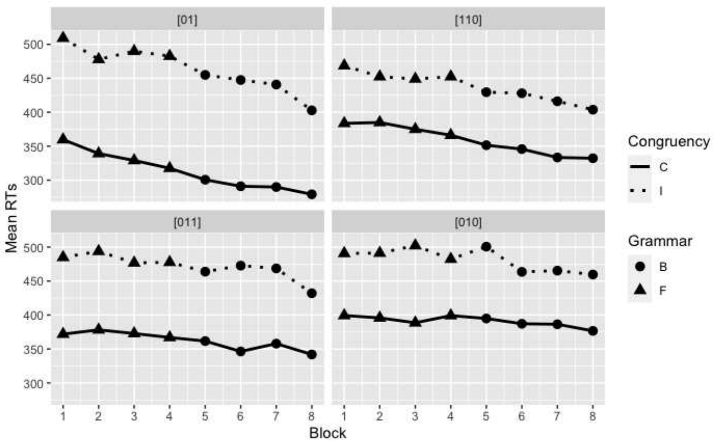

FIGURE 9: LEARNING CURVES OF MEAN RTS ACROSS BLOCKS IN 0<u>1</u>, 11<u>0</u>, 01<u>1</u> AND 01<u>0</u> (STUDY 2). F = FIB, B = BIF; C = CONGRUENT TRIALS, I = INCONGRUENT TRIALS.

4. GENERAL DISCUSSION

In this paper we aimed at disentangling the linguistically and cognitively relevant relationship between precedence and containment by reporting the results of two studies that address statistical and hierarchical learning using a SRT task (a modified Simon Task) with the Fibonacci grammar and its variants Skip and Bif. Besides confirming learning of the two Laws (i.e., learning of deterministic points 0<u>1</u> and 11<u>0</u>), both studies provided evidence for learning effects with non-deterministic points as well (the second 1 in the trigrams 011). In fact, we found that k-points (01<u>1</u>) learning is sensitive to the choice of the grammar (Fib vs. Skip); moreover, we found that both deterministic points and k-points are processed differently from the ambiguous trigram 010 in all the grammars considered, as shown by the significant Point*Grammar interactions effects that we found in both Study 1 and 2. Clearly, the relevant question to ask at this point is the following: What sort of computation leads the parser to develop the relatively early expectation that 01 is followed by 1, producing a decrease in RTs relatively to the trigram 011 that is not paralleled by the decrease relative to the trigram 010? The original hypothesis we want to put forward is that the parser is endowed with a possibly domain-general disposition to add a vertical axis of computation while processing sequences of symbols on a horizontal axis. It does so by re-interpreting the linear ordering based on precedence that is relevant for the horizontal axis in terms of a distinct instantiation of the very same abstract mathematical relation of linear ordering (reflexive, antisymmetric and transitive), that is, in terms of containment rather than of precedence, implementing this new relation on the newly added vertical axis. As we will see in more detail below, this state of affairs has two main consequences: (i) the mapping of precedence into containment represents the bootstrapping from linear order into hierarchy, it is thus arguably one of the cognitive sources of the bias towards hierarchy-based computations found in natural language; (ii) this mapping is triggered by the need the parser has to categorize the chunks it comes up with (as a result of the statistical computation on the string). For instance, in Fib 01 rapidly emerges as a chunk, based on the learning effects connected to the First Law, according to which 1 deterministically follows 0. Now, given the knowledge that 01 is a chunk (i.e., a single unit in the application of further computational processes), the question for the parser is whether this chunk should be categorized as 0 or as 1. How does the parser solve this problem? What could we infer based on our experimental results? Before addressing this issue, a preliminary point should be carefully elucidated. As we have seen, the parser does not distinguish between Fib and Bif (Study 2). In Bif, given that the generation rules have 1 rewritten as 10, it is the bigram 10 that represents a constituent. However, there is no indication that the parser develops any knowledge of this fact. On the contrary, the parser exhibits, while processing Bif generations, a learning effect related to Fib's k-points. Why should the parser deal with Bif generations as they were Fib generations? The answer is quite plausibly related to the fact that the deterministic transitions in Bif are not different from those in Fib: 0 is deterministically followed by 1 but it is never the case that 1 is deterministically followed by 0 (even in Bif, where 10 is the right constituent, grammatically speaking). For the parser, evidently, it is the existence of a deterministic transition that establishes when we have a chunk: it follows that for the parser the natural constituent, both in Bif and Fib generations, is 01, whilst 10 never qualifies as such. If this is correct, and we think the evidence to this effect is uncontroversial, these results clearly support the view, from the specific perspective introduced by our research paradigm, that knowledge of language is a dynamical process in which the formal properties of grammar play a very limited role (Öttl *et al.* 2015). Indeed, under certain conditions the parser could reconstruct a correct rule of grammar ($1 \rightarrow 01$), as is the case in Fib, but under certain other conditions it will be essentially inhibited from doing so, as is the case in Bif. The reason is that the parser decides which n-grams are constituents based on the analyses of the transitional probabilities on the string, and has, quite plausibly, no access to the rules of grammar unless the latter can be inferred by means of statistical generalizations on the sequence. On these grounds, let us go back to the parser's knowledge that 01 is a constituent. Given an arbitrary generation n of Fib, the 0s are always part of 01; the 1s are either part of 01 or independent constituents (k-points, which are highlighted below):

(n) $01\underline{1}01\underline{1}0101\underline{1}01$

Since the parser knows that 01 is a unit, the possibility arises of reducing the length of generation n across a series of progressively shorter layers, in each of which the bigram 01 is rewritten as the monogram 0 or as the monogram 1. We take this to be the primary cognitive trigger for the parser to project a vertical axis of computation, in which, proceeding bottom-up, each generation will be shorter than the preceding one, since the bigram 01 in generation n will have been rewritten as a single digit in generation n-1. The success of this style of computation clearly depends on one single factor: how does the parser decide for 0 or 1 as the label of the chunk [01]? As anticipated above, the core question is thus: What is the labelling algorithm? We propose that the parser adopts a labeling algorithm that is based on the re-interpretation of precedence (as a linear ordering relation) as containment (an alternative linear ordering relation) and that a strictly local application of this algorithm leads to the limited amount of hierarchical reconstruction that is compatible with our experimental results. There is in fact a formal property of Fib-generations that may act as a trigger for the association of precedence with containment. As seen in the Introduction, one of the most typical formal properties of the Fib-grammar is self-similarity: each generation n results from the combination of the two preceding generations (i.e. n-1 and n-2). For instance, given the generation n = 01101, the subsequence 01, which precedes the subsequence 101, is associated with generation n-2, whereas the subsequence 101 just corresponds to generation n-1 (see Figure 2). Interestingly, self-similarity also has a set-theoretic dimension. In fact, the three generations just considered are naturally analyzed as an ordinal set. In set-theory, a set x is an ordinal if it is transitive: $\forall y \in x, \forall z \in y.\ z \in x$. The three Fib-generations under scrutiny here clearly exemplify this property, as can be seen in (i), where each generation has been rendered in set-theoretic terms, based on the knowledge that the basic sets are 01 and 1 (which is how things are in Fib, as just seen):

(i) n-2 = $\langle 01 \rangle$;
n-1 = $\langle 1, \langle 01 \rangle\rangle$;
n = $\langle\langle 01 \rangle, \langle 1, \langle 01 \rangle\rangle\rangle$

In this way, indeed, precedence is automatically rendered as set-membership. For instance, the fact that the subsequence n-2 precedes the subsequence n-1 in generation n is reflected in the fact that n-2 belongs to n-1 (i.e. $\langle 01 \rangle$ is an element of $\langle 1, \langle 01 \rangle\rangle$). Derivation n is a transitive set because it is not only n-1 that belongs to n, but also n-2 (as seen in (i), $\langle 01 \rangle$ belongs to $\langle\langle 01 \rangle, \langle 1, \langle 01 \rangle\rangle\rangle$). Going on in the same vein, we have that n+1= 0101101, as a consequence of self-similarity (101 and 01101 are just the two preceding generations); in set-theoretic terms, n+1 = $\langle\langle 1, \langle 01 \rangle\rangle, \langle\langle 01 \rangle, \langle 1, \langle 01 \rangle\rangle\rangle\rangle$. Here, the fact that 101 precedes 01101 at generation n+1 is mirrored by the fact that n-1 (101) is an element of n (01101): $\langle 1, \langle 01 \rangle\rangle$ belongs to $\langle\langle 01 \rangle, \langle 1, \langle 01 \rangle\rangle\rangle$. What these formal remarks boil down to is that 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. More precisely, if we interpret ‘chunks’ as binary deterministic subsequences, the Bootstrapping Principle in (ii) holds:

(ii) If $x < y$ within a chunk, then $x \subseteq y$

This promptly provides a solution for the labelling algorithm: given (ii), it immediately follows that 01, as a chunk, should be labelled as 1 whenever considered as a unit for further computational operations. Indeed, if [01] is a 1, this categorization satisfies the Bootstrapping Principle in (ii): given the chunk [01], we have that 0 precedes 1 and that 0 is contained in 1 (based on the proposed labeling, 1 = 01, and since 0 is contained in 01, it follows that 0 is contained in 1). Conversely, (ii) is obviously not satisfied if the chunk 01 is categorized as 0. Consider now generation n above, reproduced below:

(iii) 0110110101101

The parser can proceed to build up shorter generations on the vertical axis, by rewriting 01 as 1. On the other hand, the parser is endowed with the knowledge that the shorter generations to be produced must still be sequences of 0s and 1s (as is the case for all Fib-generations). So, if the chunk 01 is reduced to 1 in the preceding generations, and, as we have seen, an arbitrary Fib-generation contains nothing else than 01s and 1s (1 a k-point), the obvious guess is that the k-points (the 1s that are not part of a 01 chunk) should be re-labeled as 0s at the previous generation: if this were not the case, the previous generation would be an illegitimate sequence of 1s. Endowed with this knowledge, which amounts to a solution for the labeling algorithm, the parser applies then the hypothetical structural reasoning that [01]s are labelled 1 and [1]s are labelled 0. For instance, given the generation 01101, the parser might engage in the limited amount of hypothetical hierarchical reconstruction shown in Figure 10 below, by activating a vertical axis within a multidimensional computational space.

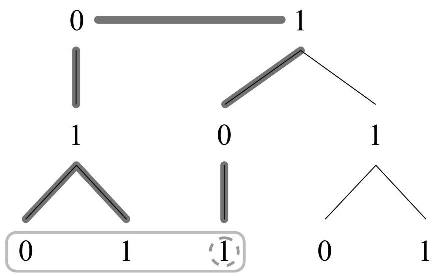

FIGURE 10: HYPOTHETICAL HIERARCHICAL REASONING BASED ON CONTEXT-INDEPENDENT LOCAL RECONSTRUCTION.

This accounts for the full array of experimental results that we have discussed above, and more particularly for the sensitivity to k-points, which is now simply a consequence of the process of hypothetical hierarchical reconstruction shown in Figure 10. Notice further that in our proposal the knowledge acquired on the horizontal axis (where linear ordering is implemented as precedence) is strictly interwoven with the knowledge acquired on the vertical axis (where linear ordering is implemented as containment, modulo the labeling algorithm). For instance, after the parser has projected [01] as [1] and [1] as 0 as shown in Figure 10, it proceeds horizontally by applying the First Law at generation n-2. Moreover, [01] is a chunk in need of labelling on the vertical axis only because it constitutes a deterministic transition on the horizontal axis. It is this mixed processing strategy - we submit - that further explains why there is no sensitivity to k-points in the foil-grammar Skip. Certainly, there are no k-points in Skip, but this is nothing more than a formal property of Skip, hardly accessible to the parser. In fact, since 01 is a natural constituent in Skip not less than in Fib or in Bif (i.e., 01 is a deterministic transition in Skip as well), the parser might engage in the very same kind of hypothetical structural reasoning that correctly derives our experimental results in Fib and Bif. If it does, however, it produces an expectation that conflicts with the results of distributional statistics on the horizontal axis: as discussed above, indeed, in Skip 011 is less frequent than 010, contrary to what happens in Fib, where 011 is more frequent than 010. The hierarchy-based expectation that 01 be followed by 1 is thus not supported by the statistical analysis of the string, contrary to what happens for Fib and Bif generations. Whenever this is the case - we submit - the parser abandons the vertical computations as a reliable source of information to enhance predictability on the original sequence and only concentrates on the generalizations derivable from linear computations. At this point, the question to be addressed is the following: how does hypothetical structural reasoning fare with respect to the competing hypotheses, that is, full hierarchical reconstruction and statistically based computations on the string? In order to address this question, notice first that the Fibonacci grammar is based on two simple transformation rules. Any system that learns to apply the rules could predict the structure for any position in the sequence. When confronted with a (long) generation of Fib, a human parser might have in principle recourse to two distinct procedures for predicting the next symbol in the sequence after exposure to the subsequence preceding that symbol. The first is full hierarchical reconstruction through the application of the two rewriting rules. As observed in the Introduction, this is essentially the line of analysis developed by Schmid *et al.* (2023). These authors propose that the parser builds multiple levels of representation based on the possibility of treating 01, as a deterministic sub-sequence, in terms of a unit which enters in turn some deterministic transitions at higher levels. In other words, if 01 is a deterministic chunk, it enters as such in higher level deterministic relations with other elements in the sequence. Intuitively, this holds for progressively larger chunks: at the second level of representation, for instance, we have that the chunk 101 deterministically follows the chunk 01, and so on for progressively larger chunks. Notice that this strategy is tantamount to a process of hierarchical reconstruction without 'explicit' labelling: the larger chunks that enter higher-order deterministic relations do so, in our terms, because they are labelled as 0s and 1s at the relevant higher level. We might say, in other words, that in Schmid *et al.*'s approach, labelling is 'implicit'. However, full hierarchical reconstruction is an unlikely processing procedure for predicting an arbitrary point within a given Fib-generation, due to its huge cognitive costs. As a matter of fact, Schmid *et al.* (2023) found that the predictability of non-deterministic points is enhanced only up to the first 2-3 levels of hierarchical reconstruction. Evidently, full-hierarchical reconstruction is, as it was to be expected, a resource-consuming strategy for enhancing predictability of the non-deterministic points on the string. More importantly, we did not find any evidence for full hierarchical reconstruction in our experimental studies. It predicts across-the-board learning effects. For instance, the two trigrams 011 and 010 are both unpredictable on linear grounds, since the transition from 01 to either 1 or 0 is not deterministic in a Fib-generation (as per the Third Law). If processing were based on full hierarchical reconstruction, we should expect learning effects to equally arise for both 1 and 0 after the bigram 01. Crucially, this is not what we found. What we found, in a nutshell, is a generalized learning effect for k-points, that is, the 1s following 01, whilst the interaction between 011 and 010 shows that the decrease in RTs is faster for k-points than for the 0s following 01. Full hierarchical reconstruction is thus not an option, because it cannot derive the asymmetry between the two indeterministic transitions 011 and 010 that constitutes one of our experimental findings. On the other hand, linear order alone is also unlikely to derive the strong learning effect that we found with k-points. Here is why. In principle, linear predictability increases with the size of n-grams. For instance, if we only consider bigrams, we can exclude 00 (i.e. *00). For trigrams, we can additionally exclude 111 (i.e. *111). For 5-grams, we also learn that 01010 is ungrammatical (i.e. *01010). In fact, it can be shown that predictability increases with the size of n-grams, but only if n is a member of the Fibonacci-sequence (1, 2, 3, 5, 8, 13, 21, 34, etc.); for instance, there is no predictability growth if we shift from trigrams to 4-grams (all we know is still that *00 and *111). Moreover, the relationship between n-gram size and prediction power is not linear, that is, the increase in prediction power becomes smaller at each step in the growth of n-grams. So, what does this entail for k-point learnability? Essentially, it entails that increasing the size of the n-gram preceding the last symbol in 011 and 010 may increase our power to predict 1 or 0. For instance, whereas a 01 may be followed by either 1 or 0, the larger n-gram 0101 is obligatorily followed by a 1, and the same holds for the even larger 101 01101 01101; similarly, the n-gram 01 011 01101 is obligatorily followed by a 0, and the same holds for the even larger 1010110101101101101101. In other words, the parser might increase its prediction power by processing progressively larger sequences than 01. In this case, k-point predictability would be an exclusive function of statistically grounded linear computations on arbitrary subsequences of a (long) Fib-generation. Is this in fact what the human parser does? No, since this large-scale linear style of computation meets essentially the same objections that apply to full hierarchical reconstruction, as discussed above. First, this linear strategy (as was the case for hierarchical reconstruction) requires a tremendous use of cognitive resources. But even putting this aside, what the linear strategy should produce is a generalized increase of prediction power (in terms of the learning effects manifested by RT-decrease) not only for $01\underline{1}$ but also for $01\underline{0}$, since, as we have seen, considering larger chunks leads to an across-the-board growth of prediction power. Again, this would leave us without any explanation for the asymmetry between the trigrams 011 and 010 that consistently emerged from our experimental results. In fact, another experimental result is worth considering. If the 5-gram 01011 is predictable on linear grounds, we should find a difference between the learning effects for linearly predictable k-points (01011) and for linearly unpredictable ones (11011). However, we found no statistically significant difference in RTs between these two classes of k-points, as shown by the relevant follow-up analyses conducted in both studies. Given these considerations, the best strategy to derive the asymmetric learning effect with k-points consists in proposing a processing strategy based on a strict interplay between linear and hierarchical processing. This mixed style of computation, whereby linear statistical computation interacts with locally restricted structural hypothetical reasoning, has two main advantages: (i) it effectively derives the asymmetrical learning effect with indeterministic k-points ($01\underline{1}$), in relation to the indeterministic points ($01\underline{0}$); (ii) it does not require an unreasonable amount of cognitive resources, since both the linear and the hierarchical computation are strongly constrained, given the gradual and successive learning of the two laws in the case of the linear computation, and the strictly local and hypothetical nature of hierarchical reconstruction, which is in fact promptly abandoned (as arguably happens in Skip) whenever in patent conflict with the results of the linear computation.

5. CONCLUSIONS

In this contribution, we have argued that the relations of precedence and dominance/containment are not antagonistic ways of processing a temporally ordered sequence of symbols; rather, they should be seen as strictly interconnected relations whose application leads to the construction of a bidimensional computational space. From this perspective, hierarchy is simply an additional interpretation of the very same abstract mathematical relation of linear ordering (a reflexive, antisymmetric and transitive relation, as are both precedence and containment), originally instantiated as precedence in the analysis of some finite array of symbols. While processing the output of our set of asymmetric L-systems (the grammars Fib, Bif and Skip), the participants in our studies arguably interpreted "x precedes y" in terms of an additional linear ordering based on containment (x is contained in y). As we have shown, this re-interpretation of precedence as containment is inherently linked to the operation of labelling chunks emerging from deterministic transitions on the sequence, under the constraint that no new symbols be added. According to the emerging picture, the linear ordering instantiated by precedence is hardly amenable to mere externalization requirements; rather, hierarchy is almost literally projected from linear order. We believe these findings have important consequences for properly addressing some core properties of natural language syntax, while leaving these matters to future work. At the same time, we have indicated that theoretical interpretation is dependent on the emerging experimental results (as, for instance, the asymmetry between the indeterministic points 01$\underline{1}$ and 01$\underline{0}$). In this vein, it is important to refine and broaden our experimental design, in search of fine-grained results able to disentangle all the variables at stake (such as habituation, independent cognitive biases like repetition or alternation of the same symbol, etc.). One possibility that we are currently pursuing (Compostella *et al.* under review) consists in simplifying the Simon Task as administered so far, by (i) eliminating the incongruent trials in order to avoid potentially disrupting processing costs related to inhibition and conflict resolution, and (ii) adopting the eye-tracking methodology to investigate the participants' anticipation skills (hence their ability to predict upcoming stimuli) while disentangling perceptual from motor aspects of learning.

REFERENCES

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

Brody, M. (2000). Mirror theory: Syntactic representation in perfect syntax. *Linguistic inquiry*, 31(1). 29–56. https://doi.org/10.1162/002438900554280.

Chesi, C. (2015). On directionality of phrase structure building. *Journal of psycholinguistic research*, 44. 65–89. https://doi.org/10.1007/s10936-014-9330-6.

Chesi, C. (2021). Expectation-based minimalist grammars. *arXiv preprint*. https://doi.org/10.48550/arXiv.2109.13871.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA, USA: MIT Press.

Christiansen, M.H. & N. Chater (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2). 157–205. https://psycnet.apa.org/doi/10.1016/S0364-0213(99)00003-8.

Christiansen, M.H. & N. Chater (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39. https://doi.org/10.1017/S0140525X1500031X.

Christiansen, M.H., C.M. Conway & L. Onnis (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, 27(2). 231–256. https://doi.org/10.1080/01690965.2011.606666.

Cleeremans, A. & J.L. McClelland (1991). Learning the structure of event sequences. *Journal of experimental psychology: general*, 120(3). 235. https://doi.org/10.1037/0096-3445.120.3.235.

Compostella, A., M. Tagliani, M. Vender & D. Delfitto (under review). On the interaction between implicit statistical learning and the alternation advantage: Evidence from manual and oculomotor serial reaction time tasks. *Submitted to Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Coopmans, C.W., H. De Hoop, K. Kaushik, P. Hagoort & A.E. Martin (2022). Hierarchy in language interpretation: evidence from behavioural experiments and computational modelling. *Language, Cognition and Neuroscience*, 37(4). 420–439. https://doi.org/10.1080/23273798.2021.1980595.

Cornish, H., R. Dale, S. Kirby & M.H. Christiansen (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PLoS ONE*, 12(1). e0168532. https://doi.org/10.1371/journal.pone.0168532.

Crain, S. & M. Nakayama (1987). Structure dependency in grammar formation. *Language*, 63(3). 522–543. https://doi.org/10.2307/415004.

Crain, S., L. Koring & R. Thornton (2017). Language acquisition from a biolinguistic perspective. *Neuroscience & Biobehavioral Reviews*, 81. 120–149. https://doi.org/10.1016/j.neubiorev.2016.09.004.

Culbertson, J. & D. Adger (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16). 5842–5847. https://doi.org/10.1073/pnas.1320525111.

Culicover, P.W. (2013). The role of linear order in the computation of referential dependencies. *Lingua*, 136. 125–144. https://doi.org/10.1016/j.lingua.2013.07.013.

Dehaene, S., F. Meyniel, C. Wacongne, L. Wang & C. Pallier (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1). 2–19. https://doi.org/10.1016/j.neuron.2015.09.019.

Ding, N., L. Melloni, A. Yang, Y. Wang, W. Zhang & D. Poeppel (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience*, 11. 481. https://doi.org/10.3389/fnhum.2017.00481.

Ettlinger, M., K. Morgan-Short, M. Faretta-Stutenberg & P.C. Wong (2016). The relationship between artificial and second language learning. *Cognitive science*, 40(4). 822–847. https://doi.org/10.1111/cogs.12257.

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. https://doi.org/10.1016/j.plrev.2014.04.005.

Frank, S.L. & R. Bod (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological science*, 22(6). 829–834. https://doi.org/10.1177/0956797611409589.

Frank, S.L. & M.H. Christiansen (2018). Hierarchical and sequential processing of language: A response to: Ding, Melloni, Tian, and Poeppel (2017). Rule-based and word-level statistics-based processing of language: insights from neuroscience. *Language, Cognition and Neuroscience*, 33(9). 1213–1218. https://doi.org/10.1080/23273798.2018.1424347.

Frank, S.L., R. Bod & M.H. Christiansen (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, 279(1747). 4522–4531. https://doi.org/10.1098/rspb.2012.1741.

Friederici, A.D. (2017). *Language in our brain: The origins of a uniquely human capacity*. MIT Press.

Futrell, R. & R. Levy (2017). Noisy-context surprisal as a human sentence processing cost model. *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 1, long papers. Valencia, Spain: Association for Computational Linguistics*. 688–698. https://aclanthology.org/E17-1065.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1). 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press. https://doi.org/10.2307/j.ctv1w9m9ds.

Greenberg, J.H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Human Language*, 2. 73–113.

Greibach, S.A. (1965). A new normal-form theorem for context-free phrase structure grammars. *Journal of the ACM (JACM)*, 12(1). 42–52. https://doi.org/10.1145/321250.321254.

Griffiths, T.L. & M.L. Kalish (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, 31(3). 441–480. https://doi.org/10.1080/15326900701326576.

Guasti, M.T. (2017). *Language acquisition: The growth of grammar*. MIT press.

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

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Kayne, R.S. (1994). *The antisymmetry of syntax*, volume 25. Cambridge, MA, USA: MIT press.

Krivochen, D. & D. Saddy (2018). Towards a classification of Lindenmayer systems. *ArXiv:1809.10542 [Cs]*. http://arxiv.org/abs/1809.10542.

Krivochen, D., B. Phillips & D. Saddy (2018). Classifying points in Lindenmayer systems: Transition probabilities and structure reconstruction. https://doi.org/10.13140/RG.2.2.25719.88484.

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.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3). 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Lidz, J., S. Waxman & J. Freedman (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition*, 89(3). 295–303. https://doi.org/10.1016/S0010-0277(03)00116-1.

Lindenmayer, A. (1968). Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. *Journal of theoretical biology*, 18(3). 280–299. https://doi.org/10.1016/0022-5193(68)90079-9.

Linzen, T. & M. Baroni (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, 7. 195–212. https://doi.org/10.1146/annurev-linguistics-032020-051035.

Lobina, D.J. (2011). Recursion and the competence/performance distinction in AGL tasks. *Language and Cognitive Processes*, 26(10). 1563–1586. https://doi.org/10.1080/01690965.2011.560006.

Maclaran, R. (1985). Tanya Reinhart, anaphora and semantic interpretation. *Journal of Linguistics*, 21(1). 221–226. https://doi.org/10.1017/S0022226700010124.

Martin, A., A. Holtz, K. Abels, D. Adger & J. Culbertson (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). https://doi.org/10.5334/gjgl.1085.

Martins, M.D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598). 2055–2064.

Momma, S. & C. Phillips (2018). The relationship between parsing and generation. *Annual Review of Linguistics*, 4. 233–254. https://doi.org/10.1146/annurev-linguistics-011817-045719.

Moro, A. (2000). *Dynamic antisymmetry*. 38, Cambridge, MA, USA: MIT press.

Moro, A. (2014). On the similarity between syntax and actions. *Trends in Cognitive Sciences*, 18(3). 109–110. https://doi.org/10.1016/j.tics.2013.11.006.

Nissen, M.J. & P. Bullemer (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, 19(1). 1–32. https://doi.org/10.1016/0010-0285(87)90002-8.

Öttl, B., G. Jäger & B. Kaup (2015). Does formal complexity reflect cognitive complexity? Investigating aspects of the Chomsky hierarchy in an artificial language learning study. *PLoS ONE*, 10(4). e0123059. https://doi.org/10.1371/journal.pone.0123059.

Phillips, C. (2003). Linear order and constituency. *Linguistic inquiry*, 34(1). 37–90. https://doi.org/10.1162/002438903763255922.

Pinker, S. (2000). *The language instinct: How the mind creates language*. New York: Harperperennial.

Pothos, E.M. (2007). Theories of artificial grammar learning. *Psychological bulletin*, 133(2). 227. https://doi.org/10.1037/0033-2909.133.2.227.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, 6(6). 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X.

Saddy, D. (2009). Perceiving and processing recursion in formal grammars. *Paper Presented at Recursion: Structural Complexity in Language and Cognition Conference at the University of Massachusetts. Amherst, MA: University of Massachusetts*.

Saddy, D. (2018). Structure, constituency, and labelling. *North Ryde, AU: Macquarie University*.

Schmid, S., D. Saddy & J. Franck (2023). Finding hierarchical structure in binary sequences: evidence from Lindenmayer grammar learning. *Cognitive Science*, 47(1). e13242. https://doi.org/10.1111/cogs.13242.

Simon, J.R. (1969). Reactions toward the source of stimulation. *Journal of experimental psychology*, 81(1). 174. https://doi.org/10.1037/h0027448.

Stabler, E.P. (2013). Two models of minimalist, incremental syntactic analysis. *Topics in cognitive science*, 5(3). 611–633. https://doi.org/10.1111/tops.12031.

Tenenbaum, J.B. & T.L. Griffiths (2001). Generalization, similarity, and Bayesian inference. *Behav Brain Sci*, 24. 629–640. https://doi.org/10.1017/S0140525X01000061.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-2440-0.

Vender, M., D.G. Krivochen, B. Phillips, D. Saddy & D. Delfitto (2019). Implicit learning, bilingualism, and dyslexia: Insights from a study assessing AGL with a modified Simon Task. *Frontiers in Psychology*, 10. 1647. https://doi.org/10.3389/fpsyg.2019.01647.

Vender, M., D.G. Krivochen, A. Compostella, B. Phillips, D. Delfitto & D. Saddy (2020). Disentangling sequential from hierarchical learning in Artificial Grammar Learning: Evidence from a modified Simon Task. *PLoS ONE*, 15(5). e0232687. https://doi.org/10.1371/journal.pone.0232687.

Xu, F. & J.B. Tenenbaum (2007). Word learning as Bayesian inference. *Psychological review*, 114(2). 245–272. https://doi.org/10.1037/0033-295X.114.2.245.

Yang, C.D. (2004). Universal Grammar, statistics or both? *Trends in cognitive sciences*, 8(10). 451–456. https://doi.org/10.1016/j.tics.2004.08.006.

Maria Vender
University of Verona, Department of Cultures and Civilizations
Via dell'Università 4, 37129 Verona
Italy
e-mail: maria.vender@univr.it
https://orcid.org/0000-0003-1985-9184

Arianna Compostella
University of Verona, Department of Foreign Languages and Literatures
Lungadige Porta Vittoria 41, 37129 Verona
Italy
e-mail: arianna.compostella@univr.it
https://orcid.org/0000-0002-4766-4168

Denis Delfitto
University of Verona, Department of Cultures and Civilizations
Via dell'Università 4, 37129 Verona
Italy
e-mail: denis.delfitto@univr.it
https://orcid.org/0000-0003-0468-1248