Structure and Learning in Natural Language

A Dissertation Presented

by

Jonathan Rawski

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Linguistics

Stony Brook University

2021

Stony Brook University
The Graduate School

Jonathan Rawski

We, the dissertation committee of the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Jeffrey Heinz – Dissertation Advisor
Professor, Department of Linguistics

Mark Aronoff – Chairperson of Defense
Professor, Department of Linguistics

Thomas Graf
Associate Professor, Department of Linguistics

Jane Chandlee
Assistant Professor, Department of Linguistics, Haverford College

This dissertation is accepted by the Graduate School

Eric Wertheimer
Dean of the Graduate School

Abstract of the Dissertation

Structure and Learning in Natural Language

by

Jonathan Rawski

Doctor of Philosophy

in

Linguistics

Stony Brook University

2021

Human language is an incredibly rich yet incredibly constrained system. Learning and generalizing these systematic constraints from small, sparse, and underspecified data presents a fundamental inference problem. The rapidity and ease by which humans learn these constraints has made this a foundational study in cognitive science, linguistics, and artificial intelligence. Traditional approaches treat this problem as grammar induction, positing structured mental representations where statistical learning strategies form inductive biases for heuristically privileging some types of constraints over others. This dissertation shows how structural properties of the space of possible grammars themselves enable learning, revealing that the role of statistical heuristics is overrated for a variety of linguistically relevant learning problems. The representational primitives of a grammar themselves – whatever they may be – form a partial order, and the dissertation presents a learning algorithm which traverses this space to select a grammar. Since the algorithm is agnostic to the type of representations, the dissertation provides a computational separation between the mental structures learners extract from data, and the learning strategy they use to generalize.

The dissertation then demonstrates the effectiveness of the algorithm on several well-understood phonological patterns governing the distribution of sounds into words. While the learning algorithm succeeds for typical representations advocated by phonologists, it reveals that the constraint space is not only large but also redundant, and the algorithm is guaranteed to find all surface-true grammars. For this reason, induction alone is insufficient for successful learning, and the dissertation describes additional non-statistical abductive principles for selecting particular grammars over others. Finally, while the representations considered by the algorithm are discrete, the dissertation shows how to translate these structures and constraints into the distributed representations characteristic of neural learning systems via tensor algebra. In this way, the thesis addresses fundamental questions about structure and learning.

Overall, these results clarify the role of induction and abduction in grammatical inference. from induction to abduction, help us understand the role structure and statistics play in these processes, and provide an analytical link between the cognitive issues of structure, learning, and bias in natural language.

Acknowledgements

This dissertation is, hopefully, a expression of gratitude for the encouragement, dedication, and support that have led to me being a professional linguist. It is entirely the product of the wonderful ecosystems I've found myself in at Stony Brook and beyond, whose contributions are uncountably infinite. I hope that I've been as much a source of support and inspiration for them as they've been for me. The title is an homage to Charles Yang's dissertation on learning, which I hoped to someday emulate when I first read it as an undergraduate. Anja Łubowicz convinced me in her undergraduate phonology courses that such learning problems are frustrating and interesting. Jeffrey Heinz, my graduate advisor, does so every day, and whatever beauty there is to be found in this work is as much a product of his insight as of mine. Dakotah Lambert was an invaluable help in both the implementation and discussion of the work. Jane Chandlee and Rèmi Eyraud helped with the original BUFIA algorithm, and Jane helped me better articulate my conceptions of what structure means. Thomas Graf helped me make sure the details actually worked out. Mark Aronoff helped me make sure those details are actually linguistic. Jim Rogers helped plant the seeds for much of the ideas of structure. My partner Lucie Baumont helped by encouraging and entertaining a sea of nonsense and procrastination. My mother gave 29 years of constant love and support to get me to this point, and I hope whoever reads this work is as proud as she is when they finish.

No one can reveal to you anything but that which already lies half asleep in the dawning of your knowledge.

— Kahlil Gibran

Contents

List of Figures

List of Tables

Chapter 1

Grammatical Inference and Linguistic Theory

This dissertation concerns the relationships between the structure of language and the learning of language. Natural language is an incredibly rich yet incredibly constrained system. The remarkable capacity both of the cognitive faculties required for natural language, and the learning mechanism which enables the growth of natural language in the maturing individual, have been a core subject of human inquiry dating back at least to the ancient Indian logicians and grammarians. They developed a theory of the "spiritual quality of speech", the invariant mental structure imposed upon the sensory experience that over time organized into linguistic knowledge (Matilal, 1990). Their line of inquiry continued through to thinkers in the seventeenth century, who were astonished at

> this marvelous invention of composing from twenty-five or thirty sounds an infinite variety of expressions, which although not having any resemblance in themselves to that which passes through our minds, nevertheless do not fail to reveal to others all of the secrets of the mind, and to make intelligible to others who cannot penetrate into the mind all that we conceive and all of the diverse movements of the soul.
>
> (Arnauld and Lancelot, 1660)

Modern developments in linguistic theory have joined their puzzlement about what seems obvious on the surface of linguistic ability. The ability of humans to quickly acquire the constraints and richness of linguistic structure represents a fundamental inference problem, and is now key to linguistics, cognitive science, and artificial intelligence. The cognitive approach to linguistic theory attempts to ascertain the underlying factors which give rise to this "marvelous invention" or cognitive faculty possessed by humans. Smolensky and Legendre (2006a) summarize this in the following way:

> "A complete cognitive architecture must provide a formal foundation for developing, within a single unified framework, theories that promise to
>
> 1. formally explain central aspects of higher cognition, such as grammatical universals and the productivity of cognition
> 2. provide a formal framework for theories of cognitive processes that explain human behavior and show explicitly how cognition can be reduced to basic computational operations
> 3. reduce mental computation to neural computation, showing how the basic computational operations can be, and are, physically realized in the brain"
>
> (Smolensky and Legendre, 2006b, pg. 556)

This dissertation takes a strong cognitivist position, and adapts these central desiderata of a cognitive architecture to the central problem of relating structure and learning. The key areas of inquiry of this dissertation are

- What are the various notions of "linguistic structure" such that they satisfy (1)?
- What do these notions of structure explain in the context of learning, and can we algorithmize them to satisfy (2)?
- How is this structure cognitively represented, such that it satisfies (3)?

Each of these questions will be given a rigorous mathematical treatment, in accordance with the formal and computational requirements of cognitive theory. This dissertation is particularly concerned with a general problem of selecting the most general grammar from a set of grammars which describe linguistic data equally well. The dissertation will consider this question both analytically and with simulations on data. This puts the dissertation in a line of work which is often termed *Grammatical Inference* (de la Higuera, 2010a; Heinz et al., 2015a; Niyogi, 2006). The study of grammatical inference provides non-trivial conditions of explanatory adequacy on theories of natural language. It refines each of the three questions above by considering the ingredients of successful and unsuccessful learning, constraints on how these ingredients may be mixed, and the results of these mixtures in the form of analytical proofs and algorithmic guarantees. Osherson and Weinstein (1983) explain:

> For a class of languages to be the natural languages, the class must be learnable by children on the basis of the kind of linguistic exposure typically afforded the young. Call this the learnability condition on the class of natural languages. Formal learning theory is an attempt to deploy precise versions of the learnability condition in the evaluation of theories of natural language. In the present context, such a theory will specify (a) the kind of linguistic input available to children, (b) the process by which children convert that experience into successive hypotheses about the input language, and (c) the criteria for "internalization of a language" to which children ultimately conform. From (a)-(c) it should be possible to deduce (d) the class of languages $\mathcal{L}$ that can be internalized in the sense of (c) by the learning mechanism specified in (b) operating on linguistic input of the kind characterized in (a).
>
> (Osherson and Weinstein, 1983, pg. 37)

What makes a class of languages $\mathcal{L}$ learnable? Results in this area have permeated almost every theory of language, and consequently every problem of learning related to structure. The scope of these results is immense, so readers wishing a full treatment may study Niyogi (2006), as well as Heinz (2016) for connections to psycholinguistics and Heinz and Rawski (forthcoming) for a history of learnability in phonology. This chapter motivates the study of grammatical inference in linguistic theory, by placing each of the questions above into linguistic context.

1.1 Structure in Learnability and Computability

This section describes how the broad question of learning the grammar of a language necessarily requires various sorts of structure. The important question then turns to the nature of that structure.

1.1.1 Membership Problems

Let us consider more precisely the problem a cognitive system, such as a child, faces when acquiring their language. We can view any particular language as some infinite set of expressions S, whose members may be viewed as sequences of symbols from some finite set (or alphabet) Σ. Note that use of such sequences is not a claim that the literal content of linguistic experience is a string. The particular nature of this set S is unknown to the child before acquisition begins. According to the computability thesis (Turing, 1937), every set S can equivalently be thought of in terms of an indicator function, or grammar, so that the extension, or language, of the grammar $L(G) = S$. This means the properties of the language are fully encoded in the function, or grammar, which can generate it (Chomsky, 1959; Post, 1944). This notion of generation presupposes a mapping between the strings belonging to a language and specific structural descriptions with respect to a grammar.

> For a class of grammars to have linguistic interest, there must be a procedure that assigns to any pair (σ, G), where σ is a string and G a grammar of this class, a satisfactory structural description of the string σ with respect to the grammar G. In particular, the structural description should indicate that the string σ is a well-formed sentence of the language $L(G)$ generated by G, where this is the case. If it is, the structural description should contain grammatical information that provides the basis for explaining how σ is understood by the speakers who have internalized the grammar; if it is not, the structural description might indicate in what respects σ deviates from wellformedness.
>
> (Chomsky and Schützenberger, 1959, pg. 119)

Since every grammar is associated to a function, the issue turns to the nature or type of that function. Such a function may map strings to binary values such as 0 or 1. One could also consider other functions which take strings as input and return various values, depending on the desired properties of the grammar. For example, a grammar can be made stochastic by replacing the binary mapping with a mapping to computable real values between 0 and 1, or to represent linguistic transformations by changing one representation into another. Some example functions are listed in Table 1.1 below, along with the linguistic intuition they encode.

Function	Description	Linguistic Correlate
$f : \Sigma^* \to \{0,1\}$	Binary classification	(well-formedness)
$f : \Sigma^* \to \mathbb{N}$	Maps strings to numbers	(well-formedness)
$f : \Sigma^* \to [0,1]$	Maps strings to real values	(gradient classification)
$f : \Sigma^* \to \Delta^*$	Maps strings to strings	(single-valued transformation)
$f : \Sigma^* \to \mathcal{P}(\Delta^*)$	Maps strings to stringsets	(multi-valued transformation)

Table 1.1: Grammars as functions. Adapted from Rawski and Heinz (2019)

Crucially, children are not exposed directly to the grammar — they get exposure to the expressions of their language, along with extrammatical and nonlinguistic cues, yet they reliably acquire the grammar that provides a compact encoding of the ambient language they are exposed to in a particular linguistic community. We will call this basic question of whether some string in a language is well-formed according to a given grammar a *membership problem.*

The statement of the membership problem allows us to clearly formulate a definition of learning. In particular, is there a learning algorithm A which takes as input any finite subset of data D of the possible forms of a language S, and returns a grammar G which solves the membership problem for the target language?

1.1.2 Enumeration and Universal Grammar

Does every possible language have a solution to a membership problem? The answer is no. One consequence of the computability thesis is that most languages have no solution to the membership problem. Why can a learner not just entertain every possible grammar, moving one by one until a grammar is selected which works?

This problem may be formed concretely by understanding enumeration. A set S is enumerable, or countable, if its members can be arranged in an ordered list where each member will eventually be encountered. Formally, there is a function f which maps positive integers to members of S such that f is *onto*, i.e. for every element s in the co-domain there is some x in its domain such that $f(x) = s$. For example, consider this list of natural numbers:

$$1, 2, 3, 4, 5, 6, \ldots$$

Every positive natural number in the list will appear in some finite amount of time. Every element can be enumerated — a number can be assigned to it by a function which gives the value n to each positive nth integer. Now consider a different list of natural numbers:

$$1, 3, 5, \ldots, 2, 4, 6, \ldots$$

For this list, it is impossible to assign an index to any even numbers: they will not all be encountered in a finite amount of time. In an acceptable list, each item must appear sooner or later as the nth entry, for some finite n.

The infinite set Σ^* of all strings (sequences) over a finite set (or alphabet) of symbols Σ is enumerable. The usual way to enumerate strings in Σ^* is to order them by length and then alphabetically within strings of the same length, as shown below over the alphabet $\Sigma = \{a, b, c\}$.

$$0 \rightarrow \varepsilon$$
$$1 \rightarrow a$$
$$2 \rightarrow b$$
$$3 \rightarrow c$$
$$4 \rightarrow aa$$
$$5 \rightarrow ab$$
$$6 \rightarrow ac$$
$$\ldots$$

However, what about languages or stringsets, the subsets of the powerset $\mathcal{P}(\Sigma^*)$? Recall that a powerset is the set of all subsets of a set S, or the set of all languages in the case of Σ^*. An argument from Georg Cantor (1892) demonstrates that the powerset of any countable set is uncountable, meaning there is no way to properly enumerate the members of that set. This has immediate and far-reaching consequences for learnability. Any program or grammar that solves the membership problem is of finite length. This means it can be written as a finite string, and is an element of Σ^*. Consequently, there are at most countably many languages S which have programs which solve the membership problem of S. But there are uncountably many languages (elements of $\mathcal{P}(\Sigma^*)$), so most languages have no solution to the membership problem. Consequently, any possible learning framework whose target language S is non-enumerable, or uncomputable, will never have a solution. The learning algorithm cannot ultimately return a grammar which solves the membership problem for S.

It is now apparent that learnable languages, must come from a well-defined class and that pure tabula rasa learning is impossible. This forms what in the cognitive study of language is often called the "logical problem of language acquisition" (Baker and McCarthy, 1981; Hornstein and Lightfoot, 1981). This is the observation that "there is no free lunch — no way to generalize beyond the specific training examples, unless the learner commits to some additional assumptions" (Mitchell, 2017). Given a finite set of observations, there is an infinite number of possible generalizations consistent with them. This is true for every problem that can be abstracted as inducing generalizations from data.

Thus, in order to arrive at a single, "correct" generalization, learners must be constrained in specific ways. Such constraints must exist a priori on the hypothesis space of the learner, and are thus separate from the observed data. Note that this is a logical truth of learning theory. The controversy is in where such priors must reside. A connectionist would assume that the biases correspond to the features defining the topology of a network, while a behaviorist might assume constraints on the way generalizations are made, and a nativist would say they reside in innate categories, operations, and principles.

This perspective had long been noted in the field of ethology which, after the behaviorist turn in the early 1900s, gained a new strength thanks to the newfound appreciation for the limits of reasoning and abilities. Eric Lenneberg, a crucial influence on the biological study of language, championed the idea of structural limits in learning since the early 1950s. In his "*Biological Foundations of Language*", Lenneberg noted that "there is no possible way in which we could think of a device, natural or artificial, that is freed from all structural information" (Lenneberg, 1967, pg. 394). In the context of neuroscience, Zador (2019) elaborates on this point, noting

> the role of innate mechanisms goes beyond simply establishing responses to sensory representations. Indeed, most of the behavioral repertoire of insects and other short-lived animals is innate. There are also many examples of complex innate behaviors in vertebrates, for example in courtship rituals. A striking example of a

> complex innate behavior in mammals is burrowing: Closely related species of deer mice differ dramatically in the burrows they build with respect to the length and complexity of the tunnels. These innate tendencies are independent of parenting: Mice of one species reared by foster mothers of the other species build burrows like those of their biological parents.
>
> (Zador, 2019, pg. 1)

Ethologists like Lenneberg understood the importance of hypothesis *classes* for learning, noting that "within the limits set, however, there are infinitely many variations possible. Thus the outer form of languages may vary with relatively great freedom, whereas the underlying type remains constant" (Lenneberg, 1967, pg. 374). Compare this to a later statement by the language acquisition researcher Lila Gleitman (1990), that "the trouble is that an observer who notices *everything* can learn *nothing*, for there is no end of categories known and constructible to describe a situation".

Thus, any debate that exists regards the *nature* of the innate constraints required for learning, not the *existence* of such innate constraints themselves. This pint is acnowledged even by those who paint themselves as empiricists (Clark and Lappin, 2011a; Chater et al., 2015). It is definitional, and therefore trivial, that there must be some type of constraint or bias on the structure of the hypothesis space and of the structure of the learning algorithm. This fact takes the form of the linguistic notion of a Universal Grammar, though the universality is merely a cover term for this mathematical result. There is no such thing as bias-free learning.

Learning theory is also concerned with "the circumstances under which these hypotheses stabilize to an accurate representation of the environment from which the evidence is drawn. Such stability and accuracy are conceived as the hallmarks of learning" (Osherson et al., 1986). How difficult is it for a learner to arrive at a generalization, and what role does the evidence a learner receives play? There are several general statements one can make. Learners exposed only to *positive evidence* (only elements of the language, like the one specified above) have a harder time than those given both positive and *negative* evidence (both elements in and outside of the prperly identified language). Learners' evidence, which may be mislabeled or "noisy," have a more difficult task than those given accurate evidence. Learners who may ask questions about the acceptability of a form have access to more information than those who cannot. Requiring that a learner selects the exact target grammar is more strict than requiring them to select an approximately correct target grammar. For further definitions and discussions of learning paradigms and their consequences for theories of cognition, see Heinz (2016).

Following the computability thesis of Church and Turing, we will consider only computably enumerable languages. A computably enumerable language has a potentially infinite number of sentences and has a finite representation in terms of a Turing machine, the Lambda Calculus, Post rewrite systems, abstract state machines, and other equivalent definitions. Thus languages, machines, and grammars are formally equivalent ways of specifying the same set. The computably enumerable languages constitute an enumerable set and effective procedures exist to enumerate the grammars (machines) that generate them (Hopcroft et al., 1979). Since a grammar is a finite representation of the language, it is unavoidable that linguistic cognition works with these finite representations rather than the infinite languages themselves.

Also, since many different grammars may be compatible with the same language, this raises the question of intensional versus extensional learning, known in linguistics as the I-language and E-language distinction (Chomsky, 1995). This distinction is a more fleshed-out concept of Universal Grammar, known in linguistics since at least the 18th century. James Beattie, in particular, notes that

> Languages, therefore, resemble men in this respect, that, though each has peculiarities, whereby it is distinguished from every other, yet all have certain qualities in common. The peculiarities of individual tongues are explained in their respective grammars and dictionaries. Those things, that all languages have in common, or that are necessary to every language, are treated of in a science, which some have called Universal or Philosophical grammar.
>
> (Beattie, 1788, pg. 766)

As Niyogi (2006) notes, this means for a membership problem there is really a collection $\mathcal{G}$ of possible target grammars and the class of languages $\mathcal{L}$ is then defined as $\mathcal{L} = \{L_G \mid g \in \mathcal{G}\}$. Given any computably enumerable language L there are an infinite number of G_i 's such that the language of said grammar $L_{G_i} = L$. Then any collection of grammars may be defined by specifying their indices in some acceptable enumeration.

However, while the debate on the relevance of computability to cognitive science and theory of mind is vast (McCulloch and Pitts, 1990; Piccinini and Bahar, 2013; Pylyshyn, 1984; Putnam, 1967; Sprevak, 2010, a.o.), the restriction of a hypothesis space to just computably enumerable languages seems to be of little interest from the perspective of establishing a cognitive theory relating structure and learning, as all languages that can be defined by some formal grammar fit into it. Given this generality, computably enumerable languages offer the weakest condition on the complexity allowed for possible natural languages, and thus lack a precise specification of essential computational requirements. Noam Chomsky argues that

> [This] condition, on the other hand, has no interest. We learn nothing about a natural language from the fact that its sentences can be effectively displayed,i.e., that they constitute a recursively enumerable set. The reason for this is clear. Along with a specification of the class F of grammars, a theory of language must also indicate how, in general, relevant structural information can be obtained for a particular sentence generated by a particular grammar.
>
> (Chomsky, 1959, pg. 138)

Perhaps natural languages are the result of stronger restrictions on the hypothesis class of grammars that will enable learning to succeed. This is a core scientific task of the linguist, to determine the restrictions on the class of grammars that may plausibly constitute linguistic ability and generate the patterns seen in the worlds languages. In contrast, a developmental psychologist focuses on the nature of and restrictions on the learning algorithm itself. This dissertation will consider both: further structured or restricted grammars such that they are

1. linguistically relevant in the above sense
2. learnable in the above sense
3. computable

These restrictions are what the various common notions of Universal Grammar are taken to be. Chomsky puts this succinctly.

> "The real problem is that of developing a hypothesis about initial structure that is sufficiently rich to account for acquisition of language, yet not so rich as to be inconsistent with the known diversity of language."
>
> Chomsky (1965, p.58)

There is thus a tension between a theory of language which is descriptively rich and one that can be learned both in principle and effectively by humans. Both of these notions inevitably revolve around considerations of structure, as shown earlier. Computability is a weak but informative structural restriction on what it means to be a learning problem such that it even has a solution or set of solutions. The chapters that follow will consider further restrictions on the space of grammars that form solutions to the learning problem. As always, the tension will remain. The point is that this first notion of structure, restrictions on a problem such that it forms a learning problem, is of central importance.

1.2 Structure in Inductive and Abductive Inference

The preceding section described how the broad question of learning the grammar of a language necessarily requires various sorts of structure. The following section will refine this concept of structure: in the particular way the learning problem is formulated, and in the class of solutions that said formulation entertains.

1.2.1 Inductive Grammatical Inference

The first study of this problem of identifying languages or stringsets came from the pioneering work of Gold (1967), which inspired much work, including work by Feldman (1972); Blum and Blum (1975); Angluin (1980a); Angluin and Laird (1988); Osherson and Weinstein (1983); Pitt (1989); Jain et al. (1999) and many others. Niyogi (2006) provides an excellent introduction to this literature in the context of language acquisition. This section overviews a common inductive criterion for grammatical inference problems, *Identification in the Limit* (Gold, 1967). There are of course many other frameworks inspired by Gold, and readers are referred to the overviews mentioned earlier.

Gold (1967), inspired directly by mathematical linguists' application of the theory of computation, introduced the first learnability results for a type of learner over classes of formal languages. In his framework, learning is a continuous process unfolding in time with no end, an idea which persists today under the rubric of continual learning. Evidence comes incrementally, and the learning algorithm incrementally outputs grammars based on its experience thus far. As time goes on, the grammars outputted must be identical (convergence) and must solve the membership problem for the target language in order for learning to succeed (correctness).

More precisely, according to the Identification in the Limit framework there are no limits on the learner's computational resources or time, and each input is assumed to be a finite initial portion of an infinitely long data stream drawn from the target language S. Learners map these finite pieces of the data stream to grammars. A particular piece of evidence, a *positive presentation* of the target language S is a function $f : \mathbb{N} \rightarrow S$ such that f is onto, meaning for every string $s \in S$, there is some number $n \in \mathbb{N}$ such that $f(n) = s$.

Gold did not limit the discussion solely to positive evidence, as is commonly misunderstood. He also considered evidence of different forms: positive evidence, positive and negative evidence, and arbitrary/primitive recursive evidence. Positive evidence (often called positive data) for a formal language S is such that every element of S can be observed at least once. Positive and negative evidence includes every logically possible string in Σ^* at least once, along with a label indicating whether it belongs to the target language S or not, similar to what is today called *supervised learning*. Primitive recursive evidence requires that the finite data presentations are only those generable by primitive recursive functions, a subset of the computably enumerable functions. See Heinz (2016) for more discussion on the merits of these.

As the learner encounters successive data points from this stream, it generates a corresponding stream of hypothesis grammars. In this framework, a learner *converges* to a grammar G if at some finite point, every future hypothesis it generates from new data is exactly G. The learner is said to identify a language in the limit if G generates the target language for *any* such presentation of data from the target language. The learner is said to identify a *class* of languages if it identifies in the limit every member language of the class. Readers are referred to Niyogi (2006) for a description of Probably Approximately Correct (PAC) learning in this setting.

There are two very important constraints on the learner given by this framework. First, success for the learning algorithm means that it converges over time to a correct generalization. In particular, at some time point n, the algorithm must output the same program and this program must solve the membership problem for S. The second point, which follows naturally, is that the algorithm can make only finitely many mistakes in generalizing.

Gold (1967)'s original paper derives several important results from these constraints, both positive and negative. A learner exists which identifies the class of finite languages in the limit from arbitrary positive evidence. However, this is not the case for *super*-finite classes of languages, which include all finite languages and *at least one* infinite language. There is no learner which can identify any super-finite class in the limit from arbitrary positive evidence. The major consequence of this result is that no traditional type of the computable languages (regular, context-free, context-sensitive) as discussed before is identifiable in the limit from positive data within this paradigm (see Niyogi (2006) for more discussion).

However, learners which have access to positive and negative data are able to learn any of the computable languages. Additionally, it is the case that learners which work over subclasses of these major language classes can and do succeed, often with extremely efficient learning results (de la Higuera, 2010b; Heinz et al., 2015b; Clark and Lappin, 2011b; Lambert et al., 2021). The central goal of this dissertation is to understand further the notions of structure that enable the successful learning of grammatical patterns through restricted classes of grammars, and a precise but flexible definition of linguistic representation. For extensive discussion of these results and their influence and importance in cognitive science, see Heinz (2016).

Building on the work of Gold, Heinz (2010c) considered a class of hypothesis grammars are a class known as *string extension grammars*, are finite subsets of some set A. The class of languages they generate are determined by a function f which maps strings to finite subsets of A (chunks of grammars). Since the size of the canonical grammars is finite, a learner which develops a grammar on the basis of the observed forms and the function f identifies this class in the limit from positive data. These learners are called String Extension Learners because each string in the language can be mapped or "extended" to an element of the grammar, which in every case, is conceived as a finite set of elements. Later, Heinz et al. (2012) generalized the finite subsets of the set A to be elements in a lattice, and showed that such “lattice-class” learners can be identified in the limit, and are incremental, globally consistent, locally conservative, and set-driven, and strongly monotonic, making them also learnable in the Probably approimately Correct sense (Valiant, 1984)

1.2.2 Constraint-Satisfaction and Abductive Grammatical Inference

Identification in the Limit and its variants are an illustrative case for which classes of grammars can an instance of the general problem of induction can be overcome. In particular, successful inference, defined as identification in the limit, succeeds under a certain combination of constraints on the learning problem, as well as constraints on the nature of the evidence and the type of the hypothesis grammars.

The idea of problems as constraints has been taken from the problem solving literature in cognitive psychology (Simon, 1977), and is distinct from the notion of “constraint” used in linguistic theory, which refers to a ban on ill-formed representations under some criterion. Briefly, the constraint-inclusion view depicts an inference problem as consisting of all the constraints on its solutions, plus requiring that a solution can be found (Nickles, 1981). The constraints form the problem itself by characterizing the problem and give it structure. The explicit demand that the solution be found is prompted by a consideration of the aims of the inference, the pursuit of which is intended to fill the outstanding gaps in the problem’s structure (Haig, 2018).

Each constraint contributes to a characterization of the problem by eliminating some potential solutions as disallowed. However, at any one time, only a manageable subset of the problem’s constraints will be relevant to the specific inference task at hand (Haig, 2018). Also, by including all the constraints in the problem’s articulation, the problem enables the learner to direct inference effectively by pointing the way to its own solution. As Haig (1987) puts it, “in a very real sense, stating the problem is half the solution!”

Nickles (1981) notes that the constraint-inclusion learning stresses the fact that hypotheses typically evolve from an ill-structured state and eventually attain a degree of well-formedness so that a solution becomes possible. From the constraint-inclusion perspective, a hypothesis is ill-structured to the extent that it lacks the constraints required for its solution. The learner deals with such problems all the time; problems are generated, selected for consideration, developed, and modified during a particular presentation of data.

Under this view, successful grammatical inference is a product of the conditions placed on it. In a stronger sense, the class $\mathcal{L}$ of languages we are considering is the outcome of a learning process constrained in a particular way such that they are the natural outcomes. One consequence of this is that successful language learning is the product of modular influences, what learning theorists call parallel learning (Case and Moelius, 2007) and what cognitive scientists call modular learning (Gallistel and King, 2009). A concrete proposal of this is Heinz (2010a). More concretely, the underdetermination of theory by data, combined with a view of learning pursuing multiple goals as stated earlier, leads to a multi-criterial view of theory appraisal. Applying this same concept to the biological substrate of a learning system, Gallistel (1999) notes that

> Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms [...] From a biological perspective, the idea of a general learning mechanism is equivalent to assuming that there is a general purpose sensory organ, which solves the problem of sensing.
>
> (Gallistel, 1999, page 1179)

In the weakest sense, we may say that a successful learner is a structured learner. In the strongest sense, we may say that general-purpose learning is incoherent in the context of language acquisition, because it shifts the explanatory burden to the data. An empiricist or inductive reconstruction of learning problems as constraints would normally take such problems to comprise those constraints that regulate the testing of theories for their adequacy with respect to the data at hand. Because of this concern with evidence, grammatical inference has been presented as essentially inductive or data-driven in nature.

This dissertation shifts the understanding of grammatical inference from induction to abduction. In contrast to induction, which infers a general case from instances, abduction, or retroduction, is explanatory inference from puzzling data to a conception of one or more causal mechanisms which, because of their prima facie plausibility, deserve to be further investigated. With the underdetermination of theories by data occurring in both the context of the class of hypotheses entertained and the way they are selected from evidence, the learner must necessarily appeal to conceptual as well as data-driven constraints. Thus the structure of the class of natural languages is the direct outcome of a structured learning problem, instantiated by a specific constraints. The very real consequence is that the notion of structure and learning indispensable.

One popular framing of abductive inference is as *Inference to the Best Explanation* (Haig, 2018). This framework recognizes that a learner's hypotheses are based on considerations of explanatory worth. Hypotheses are accepted when they when they are judged to provide a better explanation of the evidence than its rivals do. In science, inference to the best explanation is often used to adjudicate between well-developed, competing theories (Thagard, 1988; Day and Kincaid, 1994; Lipton, 2004; Thagard, 1992; van Rooij and Baggio, 2021). Lipton proposed that inference to the best explanation is not an inference to the likeliest explanation, but to the "loveliest" explanation, comprising explanatory virtues such as theoretical elegance, simplicity, and coherence. For Lipton and others, explanatory virtues provide the guide to inference about causes. Thagard (1992)'s formulation of inference to the best explanation identifies, and systematically uses, a number of evaluative criteria in a way that has been shown to produce reliable judgments of best explanation in science.

The abductive view of grammatical inference has a number of advantages, as will be seen in the specific cases we will consider. Since the features of the learning problem are entirely determined by the constraints imposed on them, we may examine the effect of a particular constraint on the overall inference problem, and consider others, without compromising the (perhaps numerous) other constraints already put. The definition and extensions of String Extension Grammars from partitions to elements of a lattice described in the previous section form an example of this. One may factor a learning problem into its parts, examine their respective effects on the learning problem as a whole, and consider other ones which behave similarly.

1.3 Structure in Representations

So far we have considered two notions of structure as they relate to the problem of language learning: structure in the definition of a problem as a solvable learning problem, and structure in the constraints placed upon a learner such that the problem of underdetermined data may be overcome and result in a restricted class of candidate grammars which the learner converges to based on evidence.

This dissertation also distinguishes a third type of structure, the structural content present in the individual examples the learner extracts as evidence. It is common in inductive learning frameworks to consider the data as a sort of oracle, which has outsized explanatory power in the context of the algorithm. However, in a very real sense, data is not as important to a learner as the structure the learner extracts from it. This is because all data is "cooked" rather than "raw", meaning there is some process that extracts structural information from it (Hammarberg, 1981).

Angluin (1980a) defined a benchmark for necessary and sufficient structure in a class of formal languages. If every language L in a class contains a finite set S, where no other language L' in the class is simultaneously a superset of S and proper subset of L, then this hypothesis space is sufficiently structured such that identification in the limit from positive data can succeed. She calls such a finite set S a *tell-tale set*, and the above property of hypothesis spaces is the *tell-tale property*.

The tell-tale property is necessary and sufficient for learning, because a learner who guesses L after exposure to its tell-tale set is guaranteed to have hypothesized *the smallest language* in the class consistent with the data sample. In other words, they are guaranteed to never overgeneralize in the way described above. Conversely, if a learner always guesses the smallest language in the class consistent with the positive data sample, with a large enough sample the learner will never converge to an incorrect target grammar. Characterizing the tell-tale sets of a hypothesis space, and more generally, the nature of the finite experience a learner needs to generalize correctly to the patterns in a hypothesis space, is one of the important lessons of learning theory for linguistics (see Heinz and Rawski (forthcoming) for further detail).

The tell-tale set is an important restriction on structural information given by evidence, particularly for constraints that learners should be error-correcting. If the observed sample of data is not consistent with a current hypothesis H, it is abandoned in favor a hypothesis that is consistent with the data. If the grammar generating a smallest language consistent with the data can be identified then this is a natural choice. However, it is not always the case that there is one smallest language in the class consistent with the data seen so far, nor is it always the case that a given positive sample of data includes a tell-tale set. This typically occurs when the grammatical space has been determined by descriptive linguistic considerations to the exception of learnability ones. In these cases, there is a puzzle: what hypothesis should H be replaced with?

The credit (or blame) problem (Clark, 1989) exacerbates this puzzle because covert structure prevents the learner from directly observing the source of this error in H (Dresher, 1999). When learning phonological mappings to underlying forms, errors can arise from hypothesizing the wrong lexical representation or the wrong phonological mapping, but the learner must still somehow determine which is the source of the error. For example, metrical footing is not present in the signal, so if a learner notices an error, which aspects of the grammar must it assign credit to? It is not immediately clear.

This is not unlike the problem determining which, of several possible sources account for given observations. Here is a concrete case. Say a learner observes a trisyllabic word with stress on the medial syllable. Should the learner hypothesize left-aligned iambs or right-aligned trochees?

Several learning proposals in phonology explicitly engage the credit problem, notably Dresher and Kaye (1990); Tesar and Smolensky (2000); Jarosz (2006, 2013) and (Tesar, 2014). Jarosz (2019) makes the argument that the nature of statistical inference enables inferences about sources of error/blame (see also Nazarov and Jarosz (2017) on learning interdependent parameters for stress).

Where does this structure come from? It is mistakenly, yet commonly, thought in linguistic and cognitive contexts that experience is causal on sensory experience. This is backwards. To the extent that data is meaningful, it is because the learner imposes structure on it. This is the real notion of "representation" meant in cognitive science: organisms must parse the world according to their innate characteristics. Thus Zador (2019) writes "the answer is that much of our sensory representations and behavior are largely innate."

Zador describes many such innate sensory representations in biology. For example, many olfactory stimuli are innately attractive or appetitive (blood for a shark) or aversive (fox urine for a rat). Responses to visual stimuli can also be innate. Mice respond defensively to looming stimuli, which may allow for the rapid detection and avoidance of aerial predators. Zador concludes that for biological systems, "it appears that a large component of an animal's behavioral repertoire is not the result of clever learning algorithms—supervised or unsupervised—but rather of behavior programs already present at birth."

This finding is a natural outcome of the abductive view of inference. Until a learner has collected and fully parsed data, it will not really know what the particular grammatical inference problem is, even though some of the constraints for its satisfaction will be available to it. As Haig (1987) notes, "The point is that it is the puzzle, or puzzles, thrown up by our data analyses that prompt the generation of new explanatory theories." The point is that the learning problem is necessarily dependent on a particular parsing strategy from experience to evidence, but the parsing strategy and the learning strategy are necessarily separate.

This point is central to the treatment of linguistic structures in this dissertation. The next chapter develops a mathematical notion of linguistic structure which is highly general, such that a learner always expresses its representations in that form, but where the content of the representations is agnostic. This is linguistically advantageous, as linguistic theory has developed a rich and diverse repertoire of data structures across various subfields, consisting of strings, trees, graphs, and others.

The main consequence of this view of representations for this dissertation is that the learning algorithms are highly factored. We are more interested in the behavior of a learner for some particular representation, than making a claim about the correctness of one or more representations for learning. Since as of yet we cannot directly observe linguistic representations beyond extremely shallow measurement, it is more coherent to explore the behavior of a learner using a variety of particular representations parsed from some data.

One major finding of the dissertation is that for the learning problems we consider, there are many possible extensionally equivalent grammars. That is, the nature of representations considered by the learner naturally leads to multiple characterizations of the same set. The solution to this, as we will see, is to consider various constraints for narrowing the scope of these possible solutions.

In this way, the dissertation shows a dependence between learning and the representations learning works over

1.4 Layout of the Dissertation

This dissertation synthesizes the three increasingly specific notions of structure described above to form a coherent abductive theory of grammatical inference. In what follows, we will discuss each of these starting from the last one. We will first discuss structure in the representation which I've us a structured hypothesis, which give us additional constraints on the learning problem, and so on. First, Chapter 2 mathematically describes a unified model-theoretic notion of structural representational information, uses it to define several linguistically relevant structural representations, and defines a notion of grammar whose components are sub-structures of these representations. The chapter then shows how the inherent model-theoretic properties of these representations structure the space of possible components of a grammar into a partial order.

Chapter 3 considers the question of learning from a constraint-based perspective. The partial order structure that the components of the grammar possess directly leads to a property called "grammatical entailment" which allows a learner to prune out vast swathes of possible components of grammars given evidence. The Chapter develops a provably correct abductive algorithm which traverses this partial order to select the most general grammar. The chapter shows that the entertained grammars are highly redundant and prohibitively large, and introduces a variety of additional abductive principles to constrain the candidate solutions further.

Chapter 4 will demonstrate the effectiveness of these algorithmic variants on a variety of well-understood phonotactic data. Phonotactics, the problem of classifying phonological surface structures as well-formed or ill-formed, is a natural test case for these algorithms. Each of these case studies allows the different behavior of the learning algorithm to come out, and to be compared to the UCLA Maximum Entropy Phonotactic Learner (Hayes and Wilson, 2008), a statistical inductive inference learning algorithm. In the cases studied, the main result is that for a variety of linguistically relevant learning problems, statistical generalization is unnecessary, that the structure-based learning algorithm correctly finds the grammars from data, and that the the success of the UCLA learner is largely due to *structural* choices.

Finally, Chapter 5 considers the cognitive instantiation of the representations discussed in Chapter 2. In particular, it discusses how to embed the structured discrete symbolic representations into vector spaces via tensors, and defines several constrained classes of grammars as operations using tensor calculus. In this way, the unified notion of structure gains another interpretation, one closer to the notions of distributed computation that emerged from the connectionist literature over the past decades.

Chapter 2

Structures and Grammars

This chapter defines the central ideas of model-theoretic representations and considers various representations of linguistic structure. This involves deciding what kind of objects we are reasoning about and what relationships between them we are reasoning with. The presentation in Sec. 2.2–2.3 will largely follow Lambert et al. (2021). We will first discuss a general notion of structural information, and use it to derive a notion of substructures. In contrast to previous approaches, this will allow us to describe several distinct representations of words in a uniform way. Structural information is defined relationally in terms of model theory. Finite model theory provides a unified ontology and a vocabulary for representing many kinds of objects, by considering them as relational structures (see Libkin 2004 for a thorough introduction). This allows flexible but precise definitions of the structural information in an object, by explicitly defining its parts and the relations between them. This makes model-theoretic representations a powerful tool for analyzing the information characterizing a certain structure.

It is important to note that this application of finite model theory has a rich history in linguistics. Much of this work has occurred in the domain of syntactic theory (Rogers, 1997, 1998; Pullum and Scholz, 2001; Rogers, 2003a; Morawietz, 2003; Pullum, 2007; ter Meulen, 2012; Graf, 2013). Building on earlier insights from Declarative Phonology (Bird, 1995; Coleman, 1998), in recent years, there has been a rise of model-theoretic approaches to phonology (Potts and Pullum, 2002; Graf, 2010; Vu et al., 2018; Strother-Garcia, 2019; Chandlee and Jardine, 2019; Danis and Jardine, 2019; Oakden, 2020; Rogers and Lambert, 2019b,a; Rogers et al., 2013; Dolatian, 2020).

2.1 Preliminaries

A n-ary relation R over a certain non-empty set X is $R \subseteq X^n$. When$\langle x, y \rangle \in R$, we also write $R(x, y)$; when $\langle x, y \rangle \notin R$, we also write $\neg R(x, y)$.

A relation R is

- *transitive* iff $\forall x \forall y \forall z (R(x, y) \wedge R(y, z) \rightarrow R(x, z))$.
- *reflexive* iff $\forall x R(x, x)$.
- *irreflexive* iff $\forall x (\neg R(x, x))$.
- *symmetric* iff $\forall x \forall y (R(x, y) \leftrightarrow R(y, x))$.
- *asymmetric* iff $\forall x \forall y (R(x, y) \rightarrow \neg R(y, x))$.
- *anti-symmetric* iff $\forall x \forall y (R(x, y) \wedge R(y, x) \rightarrow x = y)$.
- a *strict partial order* iff it is irreflexive, transitive, and asymmetric.
- a *partial order* iff it is reflexive, transitive, and anti-symmetric.
- a *strict total order* if it is a strict partial order and $\forall xy (x \neq y \rightarrow (R(x, y) \vee R(y, x))$.
- a *total order* if it is a partial order and $\forall xy (R(x, y) \vee R(y, x))$.

If R is a partial ordering of a set X, then $R - \{\langle x, x \rangle : x \in X\}$ is a strict partial ordering of X. We often use $\leq$ to denote a partial order R. When $(x, y) \in R$, we write $x \leq y$ or $y \geq x$. Note that while $\leq$ is often used to denote the numerical less than or equal to relation over numbers, we use $\leq$ to denote other partial orders as well. A *poset* (partially ordered set) $\langle X, \leq \rangle$ is a set X where the binary relation $\leq$ is reflexive, transitive and antisymmetric.

A finite, non-empty set Σ is called an *(unranked) alphabet*. The *Kleene closure* of Σ is $\Sigma^* := \bigcup_{i \geq 0} \Sigma^i$. A Σ-*string* is some element of Σ^*. It is of *length* n, written $|s| = n$, iff it is a member of Σ^n. The special string s with $|s| = 0$ is called the *empty string* ε. When the choice of alphabet is clear from context, we simply speak of strings rather than Σ -strings. Since strings are sequences over Σ the same operations can be applied to them. Strings can be concatenated: given strings $s := \langle \sigma_1, \ldots, \sigma_i \rangle$ and $t := \langle \sigma_{i+1}, \ldots, \sigma_n \rangle$, $s \cdot t := \langle \sigma_1, \ldots, \sigma_i, \sigma_{i+1}, \ldots, \sigma_n \rangle$ is their *string concatenation*, also written st. Note that $s \cdot \varepsilon = \varepsilon \cdot s = s$.

Given two Σ -strings $s := \sigma_1 \ldots \sigma_n$ and t, s is a *subsequence* of t iff there are Σ-strings $u_0, \ldots, u_n$ such that $u_0 \cdot \sigma_1 \cdot u_1 \cdot \sigma_2 \ldots u_{n-1} \cdot \sigma_n \cdot u_n = t$. If $u_i = \varepsilon$ for all $1 \leq i < n, s$ is a *substring* of t. If furthermore $u_0 = \varepsilon$, then s is a *prefix*, and if $u_n = \varepsilon$, then s is a *suffix*. s might be both a prefix and a suffix of t yet $s \neq t$ (such as when $s := a$ and $t := aa$).

2.2 Model-Theoretic Representations of Strings

A relational structure in general is a set of domain elements, dom, which is augmented with a set of relations of arbitrary arity, $R_i \subseteq \mathrm{dom}^{n_i}$. The relations provide information about the domain elements. The *model signature* $\mathcal{M} = \langle \mathrm{dom}; R_i \rangle$ collects these parts and defines the nature of the structure in terms of the information in the model. Let w be a string over some alphabet Σ. Then a model for a word w is a structure:

$$\mathcal{M}_{\Sigma}^{R_i}(w) := \left\langle \mathrm{dom}_w; R_i, \sigma_w \right\rangle_{\sigma \in \Sigma}$$

where dom_w is isomorphic to an initial segment $\langle 1, \ldots, |w| \rangle$ of the nonzero natural numbers and represents the positions in w, and each σ_w is a unary relation that holds for all and only those positions at which σ occurs. Note that it is assumed that the set $\{\sigma_w\}_{\sigma \in \Sigma}$ is a partition of dom_w.[1] Consider an alphabet $\Sigma = \{\text{s}, \text{ʃ}, \text{á}, \text{à}\}$, which represent two types of sibilants and a vowel with either low or high tone. Strings are combinations of these symbols at certain events, like the word ‘sásàʃá’.

The remaining R_i are the other salient relations, which are used to define order in a particular structure. One model signature for strings, called the *precedence model*, is given as

$$\mathcal{M}^{<}(w) = \langle \mathrm{dom}_w; <_w, \text{s}_w, \text{ʃ}_w, \text{á}_w, \text{à}_w \rangle.$$

[1] One can convert a model in which multiple unary relations may apply to a given domain element into a partitioned normal form by simply replacing these unary relations with their powerset.

This model says that for every symbol σ in alphabet Σ, there is a unary relation R_σ in $\mathcal{R}$ that can be thought of as a labelling relation for that symbol. For our set $\Sigma = \{\text{s}, \text{ʃ}, \text{á}, \text{à}\}$, $\mathcal{R}$ includes the unary relations R_{s}, $R_{\text{ʃ}}$, $R_{\text{à}}$, and $R_{\text{á}}$. It also defines a binary relation $x < y$, the general precedence relation on the domain dom, as follows:

$$x < y := \{(i, j) \subseteq D \times D | i < j\}$$

A visual of the word model for ‘sásàʃá’ under this signature is given in Figure 2.1. This and the following figures also visually depict a particular *connected substructure* of the word model, called a *factor*, which we will later define more rigorously (see also Rogers and Lambert (2019b,a).

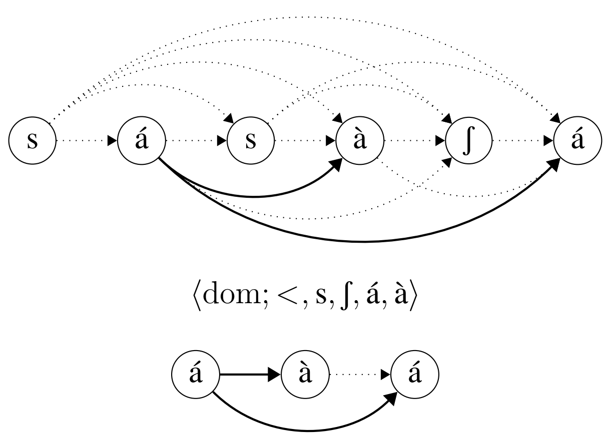


Figure 2.1: The general precedence model of ‘sásàʃá’ (top), along with the 3-factor ‘áàá’ (bottom).

The general precedence relation describes a notion of structural information purely in terms of whether a node precedes another one. While the information that, say, the last element in a string comes after the first is immediately accessible from the model, this distinction collapses the notions of immediate and general structural adjacency. Building on this precedence relation we can derive different types of relational structure.

These refine the model of a word to describe immediate, relativized, or multiply-relativized adjacency.

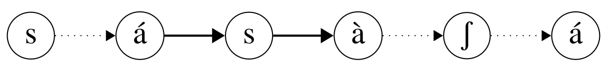
Perhaps we would like to consider only immediately adjacent elements. Rather than a general precedence relation $<$, we may consider an immediate precedence, or successor, relation $\lhd$. The standard successor relation is the transitive reduction of the precedence relation and is first-order definable from the latter as follows:

$$x \lhd y \;:=\; x < y \;\wedge\; (\forall z)[x < z \;\Rightarrow\; y \leq z].$$

This relation gives a different word model, where elements are arranged according to immediate adjacency, commonly called the *successor model*. The signature for this model is given as

$$\mathcal{M}^{\lhd}(w) = \langle \mathrm{dom}_w; \lhd_w, \mathrm{s}_w, \int_w, \text{á}_w, \text{à}_w \rangle.$$

A visual of the successor word model for the word ‘sásàʃá’ is given in Figure 2.2.



$$\langle \mathrm{dom}; \lhd, \mathrm{s}, \int, \text{á}, \text{à} \rangle$$

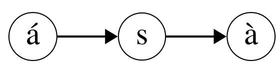


Figure 2.2: The immediate successor model of ‘sásàʃá’ (top), and a particular factor ‘ásà’ (bottom)

The general precedence relation can alternatively be refined to discuss a form of immediate adjacency relativized to certain unary relations in the signature. In particular, we can form relations between subsets of the alphabet, commonly called a *tier-alphabet*. For example, we may want to discuss the relations between only the sibilant elements present in a word, to the exclusion of all others. Similarly to how the successor relation is derived, we can restrict the precedence relation to the intended tier-alphabet τ and first-order define a similar tier-successor relation $\lhd^{\tau}$:

$$x \lhd^{\tau} y \;:=\; \tau(x) \wedge \tau(y) \;\wedge\; x < y \;\wedge\; \big(\forall z\big)\big[\big(\tau(z) \;\wedge\; x < z\big) \;\Rightarrow\; y \leq z\big].$$

Adjusting the model signature appropriately, shown below, we get a tier-based notion of structure, depicted in Figure 2.3.

$$\mathcal{M}^{\lhd^{\{\mathrm{s},\int\}}}(w) = \langle \mathrm{dom}_w; \lhd_w^{\{\mathrm{s},\int\}}, \mathrm{s}_w, \int_w, \text{á}_w, \text{à}_w \rangle.$$

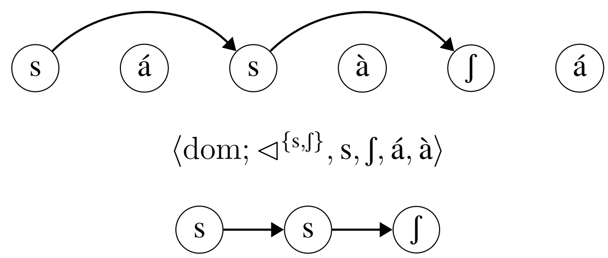


Figure 2.3: The tier-successor model of 'sásàʃá' where the tier $\tau = \{\text{s}, \text{ʃ}\}$ (top), and a factor 'sʃs' (bottom)

Because the unary relations partition the domain elements, we can create a tier-adjacency relation for each element of the powerset of these relations. This merely amounts to adding tier-adjacency relations to the model signature to create a multi-tier signature. A model of the multi-tier relations is shown in Figure 2.4.

$$\mathcal{M}^{\triangleleft^{\{\text{s},\text{ʃ}\}},\triangleleft^{\{\text{à},\text{á}\}}}(w) = \langle \text{dom}_w; \triangleleft_w^{\{\text{s},\text{ʃ}\}}, \triangleleft_w^{\{\text{à},\text{á}\}}, \text{s}_w, \text{ʃ}_w, \text{á}_w, \text{à}_w \rangle.$$

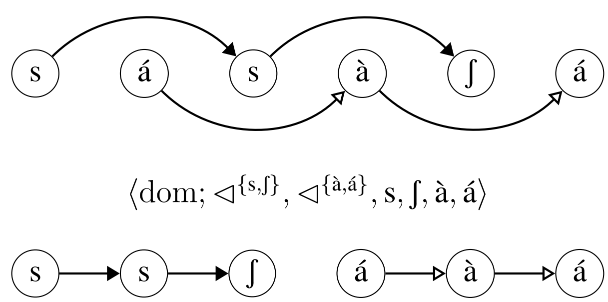


Figure 2.4: The multi-tier-successor model of 'sásàʃá' with tiers $\{\text{s}, \text{ʃ}\}$ and $\{\text{à}, \text{á}\}$ (top), and the factors 'sʃs' and 'áàá' (bottom)

Figure 2.5 depicts the relationships among these ordering relations.

These four conventional model signatures of strings are by no means the only relational word models that may be considered. What makes them conventional is the unary relations which essentially label each domain element with a single, mutually exclusive, property: the property of being some $\sigma \in \Sigma$. In contrast, unconventional models for strings recognize that distinct alphabetic symbols may share properties, and the model signature includes these properties as unary relations (Strother-Garcia et al., 2016; Vu et al., 2018). A conventional model of the symbol

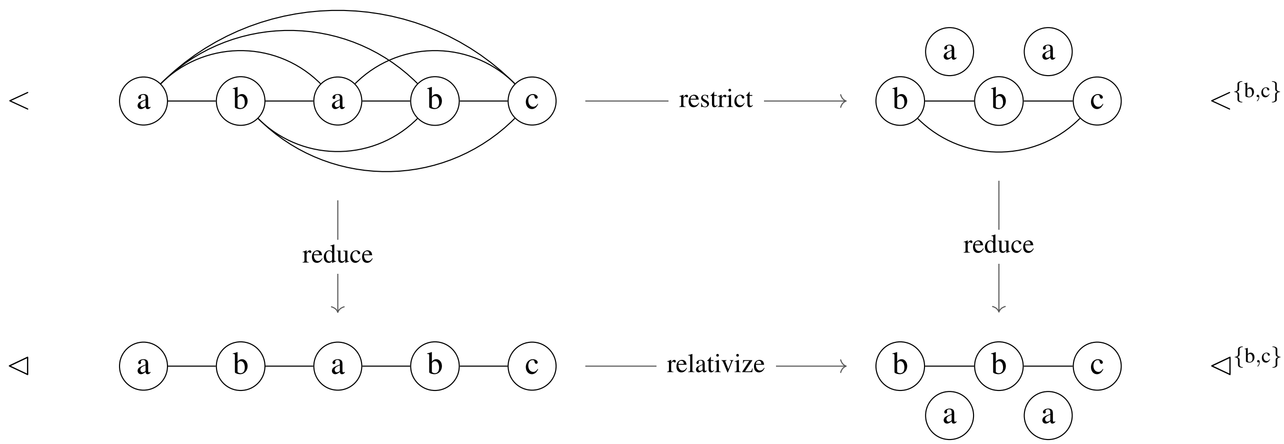


Figure 2.5: Word models for "ababc" using general precedence, immediate successor, and relativized variants of each, showing the relationships among these relations. Domain elements that are not ordered are pulled aside from the structure. In these examples, the alphabet is $\Sigma = \{\text{a}, \text{b}, \text{c}\}$, and the salient symbols for the relativized relations are $\tau = \{\text{b}, \text{c}\}$. Adapted from Lambert 2021

alphabet $\Sigma = \{\texttt{a}, \ldots, \texttt{z}, \texttt{A}, \ldots, \texttt{Z}\}$ would include 52 unary relations, one for each lowercase and capital letter. On the other hand, an unconventional model, might only include 27: 26 for the letters, and one unary relation indicating the property `Capital`.

As another example, a core tenet of modern linguistics states that perceived entities are composed of multiple atomic, shared properties, termed features (Jakobson et al., 1952). The centrality of linguistic features is at this point uncontroversial. Features have been described as "the most fundamental insight gained during the last century" (Ladefoged, 2000). The reason for this is that they allow for linguistic generalizations to be factored into the interaction of smaller, more atomic parts. The targets of phonological rules are thus the identifying information of a segment rather than segments themselves, a powerful cognitive hypothesis that learning and knowledge of language is organized around mind-internal properties of phonological information.

If the atomic properties of the sequential elements are taken to be a set of phonological features, one may use multiple unary relations to describe the same domain element. Properties like anteriority ($\pm$`ant` — whether it occurs in the anterior of the vocal tract), stridency ($\pm$`str` — whether it produces a high-intensity fricative noise), or vocalicity ($\pm$`voc` — whether it is produced by vibrating the vocal chords), whether it has a high or low tone ($\pm$`H`, $\pm$`L`) among others (Hayes, 2009). Each sound at some position x is represented as satisfying relations

$R \in \{\pm\texttt{syll}(x), \pm\texttt{str}(x), \ldots, \pm\texttt{ant}(x)\}$. Under this signature, a word model for the example 'sásàʃá' is given in Figure 2.6.

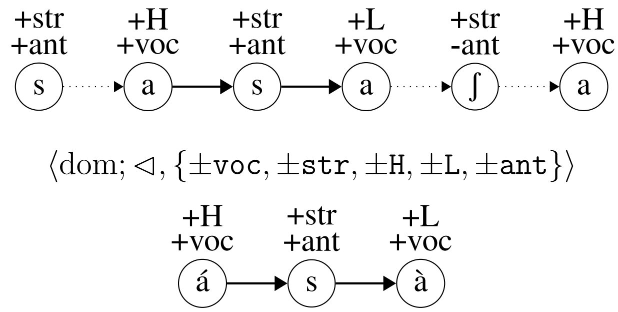


Figure 2.6: The Unconventional successor model of 'sásàʃá', along with a particular 3-factor

Generally, this reduction can be exponential: an alphabet of size 2^n can be represented with n unary relations in the model signature.

2.3 Model-theoretic representations of nonlinear structures

The previous model signatures characterized information based on a string data structure, as in the work of Büchi (1960); McNaughton and Papert (1971); Rogers et al. (2013); Thomas (1982), and others. However, the model-theoretic framework is not restricted to simple string models, but in fact applies to any structure that can be characterized as a graph. In this sense strings are a fundamental special case, for which there are many variants which characterize many different types of structures used in linguistic theory.

2.3.1 Autosegmental graphs

An example of a nonlinear structure where the graph perspective is clearly relevant to linguistic research concerns autosegmental representations in phonology. Graphs were proposed to handle a variety of prosodic phenomena for which the string-based perspective was inadequate. Phonological processes affecting domains larger than two adjacent segments, such as tonal alternations in tonal languages, have temporal properties that do not always map consistently onto discrete vowel segments in one-to-one fashion (Goldsmith, 1976; Williams, 1976). Goldsmith introduced a model of the phonological word where tonal features formed an independent string from the segmental string, called a tier. Segments on the two strings are linked via many-to-one relations, turning the structure into a graph.

In practice, encoding these adjustments into a word model involves adding more relational structure. Jardine (2017) uses a binary relation $\alpha(x, y)$ to encode the association relation between autosegmental tiers. Augmenting the successor model signature used throughout this paper gives a signature as

$$\mathcal{M}^{\alpha,\lhd}(w) = \langle \mathrm{dom}_w; \alpha_w, \lhd_w, \mathrm{s}_w, \int_w, \mathrm{a}_w, \mathrm{H}_w, \mathrm{L}_w \rangle.$$

Here the domain is increased to accommodate the new autosegments, and the successor relation holds between elements on both tiers. The unary relations encoding vowels with tonal features have been split, into a relation ‘a’ for vowel information, and distinct ‘H’ and ‘L’ relations for tonal information. Under this signature, a word model for the example ‘sásàʃá’ is given in Figure 2.7.

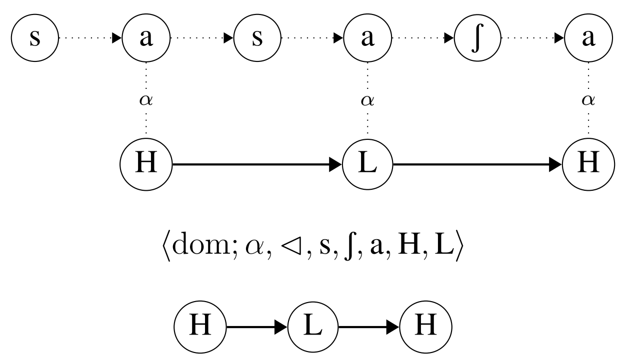


Figure 2.7: The autosegmental successor model of ‘sásàʃá’, along with its 3-factor ‘HLH’. The α relation is shown without tips because it is symmetric

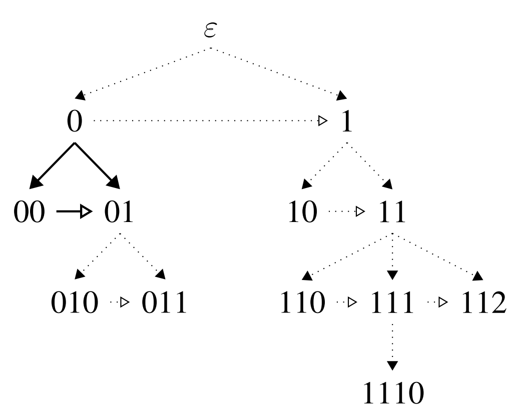
2.3.2 Tree models

The model-theoretic framework also allows describing tree structures, and opens the door to study parallels between phonological and syntactic structures (Graf, 2014). Rogers (2003b) describes a model-theoretic characterization of trees of arbitrary dimensionality. Each node in a tree can be associated with a unique numeric address, represented as a tuple. The root node has address $\langle\rangle$ (also written ε), and the daughter of a node with address α and i -many left siblings is assigned

$\alpha \cdot \langle i \rangle$. The node addressing system allows for a very elegant definition of trees, by specifying the domain dom as a Gorn tree domain (Gorn, 1967). This is a hereditarily prefix closed set D of node addresses, that is to say, for every $d \in D$ with $d = \alpha i$, it holds that $\alpha \in D$, and if $i \neq 0$ then $\alpha(i-1) \in D$ as well.

A Σ-labeled tree is a pair $T_\Sigma := \langle D, \ell \rangle$ such that D is a tree domain and $\ell : D \to \Sigma$ labels every node with some symbol drawn from Σ. In this view, a string may be called a one-dimensional or unary-branching tree, since it has one axis along which its nodes are ordered. In a standard tree, on the other hand, the set of nodes is ordered as above by two relations, "dominance" and "immediate left-of". We may explicitly write them out model-theoretically so that a signature for a Σ-labelled two-dimensional tree T is $\mathcal{M}^{\triangleleft^{\downarrow}, \triangleleft^{\rightarrow}} = \langle D; \triangleleft^{\downarrow}, \triangleleft^{\rightarrow}, R_\sigma \rangle_{\sigma \in \Sigma}$ where $\triangleleft^{\downarrow}$ is the immediate dominance relation and $\triangleleft^{\rightarrow}$ is the immediate right-of relation. Model signatures that include the transitive closures of each of these relations have also been studied.

Note that in contrast to other definitions, we do not need to say anything about dominance and precedence as they are already implicit in the fact that D is a tree domain. Hence x is the mother of $y (x \triangleleft y)$ iff $y = xi$ for some digit i, and x is the left sibling of y iff $x = \alpha i$ and $y = \alpha(i+1)$. Suppose s is the mother of two nodes t and u in some standard tree, and also assume that t precedes u. Then we might say that s dominates the string tu.



$\langle D; \triangleleft^{\downarrow}, \triangleleft^{\rightarrow}, R_\sigma \rangle_{\sigma \in \Sigma}$

Figure 2.8: A tree model. Nodes are organised by immediate dominance (black tip) and immediate right-of (white tip) relations. Labelling relations are omitted to show Gorn addresses. A particular factor is noted with solid thick lines.

2.4 Factors, Superfactors, Ideals and Filters

We referred to the preceding models of strings, trees, and graphs w as a *structure* (Rogers and Lambert, 2019b; Chandlee et al., 2019). However, structures are more general in that they correspond to any mathematical structure conforming to the model signature. As such, while a model of a string w will always be a structure, a structure will not always be a model of a string w. A structure in M is any $S = \langle D, \{R_1, R_2, \ldots, R_n\}\rangle$ where D is finite and each R_i is an a_n-ary relation over D. Since D is finite, its elements are standardly given as elements of $\mathbb{N} : D = \{1, \ldots, k\}$ for some $k \in \mathbb{N}$. The *size* of a structure S, denoted $|S|$, coincides with the cardinality of its domain.

We next wish to introduce a partial ordering over structures. To do so, we must define the terms *connected, restriction*, and *factor*. For each structure $S = \langle D; \lhd, R_1, \ldots R_n\rangle$ let the binary *connectedness* relation C be defined as follows.

$$C \stackrel{def}{=} \left\{(x, y) \in D \times D \mid \exists i \in \{1 \ldots n\}, \exists (x_1 \ldots x_m) \in R_i, \exists s, t \in \{1 \ldots m\}, x = x_s, y = x_t\right\}$$

Informally, domain elements x and y belong to C provided they belong to some non-unary relation. Let C^* denote the symmetric transitive closure of C.

Definition 1 (Connected structure). *A structure* $S = \langle D; \lhd, R_1, R_2, \ldots, R_n\rangle$ *is* connected *iff for all* $x, y \in D$, $(x, y) \in C^*$.

For example, consider a string like *abba*, under the successor model signature $M^{\lhd}$. $M^{\lhd}(abba)$ is a connected structure, depicted as

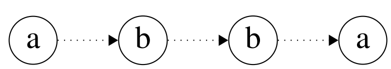


Figure 2.9: A Connected Structure

However, the structure $S_{ab,\ ba}$ depicted below which is identical to $M^{\lhd}(abba)$ except it omits the pair (2,3) from the order relation is not connected since none of (1,3),(1,4), (2,3) nor (2,4) belong to C^*. $S_{ab,\ ba} = \left\langle D = \{1, 2, 3, 4\}; \lhd = \{(1, 2), (3, 4)\}, R_a = \{1, 4\}, R_b = \{2, 3\}, R_c = \varnothing\right\rangle$

Note that no string in Σ^* has structure $S_{ab,\ ba}$ as its model.

Definition 2. $A = \left\langle D^A; \lhd, R_1^A, \ldots, R_n^A\right\rangle$ is a restriction of $B = \left\langle D^B; \lhd, R_1^B, \ldots, R_n^B\right\rangle$ *iff* $D^A \subseteq D^B$ *and for each* m*-ary relation* R_i*, we have* $R_i^A = \{(x_1 \ldots x_m) \in R_i^B \mid x_1, \ldots, x_m \in D^A\}$.

Figure 2.10: A Disconnected Structure

Informally, one identifies a subset A of the domain of B and strips B of all elements and relations which are not wholly within A. What is left is a restriction of B to A.

Definition 3. *Structure A is a* factor *of structure B ($A \sqsubseteq B$) if A is connected, there exists a restriction of B denoted B', and there exists $h : A \to B'$ such that for all $a_1, \ldots, a_m \in A$ and for all R_i in the model signature: if $h(a_1), \ldots h(a_m) \in B'$ and $R_i(a_1, \ldots, a_m)$ holds in A then $R_i(h(a_1), \ldots, h(a_m))$ holds in B'. If $A \sqsubseteq B$ we also say that B is a* superfactor *of A.*

In other words, properties that hold of the connected structure A also hold in a related way within B.

If $A \sqsubseteq B$ and $|A| = k$ then we say A is a k-factor of B. For all $w \in \Sigma^*$, and for any model M of Σ^*, let the factors of w be $\texttt{Fact}(M, w) = \{A \mid A \sqsubseteq M(w)\}$ and the k-factors of w be $\texttt{Fact}_k(M, w) = \{A \mid A \sqsubseteq M(w),\ |A| \leq k\}$. We also define $\texttt{Fact}(M, \Sigma^*)$ to be $\bigcup_{w \in \Sigma^*} \texttt{Fact}(M, w)$ and $\texttt{Fact}_k(M, \Sigma^*)$ to be $\bigcup_{w \in \Sigma^*} \texttt{Fact}_k(M, w)$. When M is understood from context, we write $\texttt{Fact}(w)$ instead of $\texttt{Fact}(M, w)$. We define the sets of superfactors $\texttt{Supfact}(M, w)$ and $\texttt{Supfact}(M, \Sigma^*)$ similarly.

The next two lemmas show how this representational perspective unifies the treatment of substrings, subsequences, their tier-based variants, as well as more general objects like graphs and trees. They are factors under the successor and precedence models, respectively. A string $x = x_1, \ldots, x_n$ is a substring of y iff there exists l, r such that $y = lxr$. String x is a subsequence of y iff there exists $v_0, v_1, \ldots, v_n$ such that $w = v_0 x_1 v_1, \ldots, x_n v_n$.

Lemma 1 (Substrings are factors under $M^{\lhd}$)**.** *For all strings $x, y \in \Sigma^*$, x is a substring of y iff $M^{\lhd}(x) \sqsubseteq M^{\lhd}(y)$.*

Proof. Note that the result trivially holds for $x = \lambda$: we restrict ourselves to the case $x \neq \lambda$. Let $M^{\lhd}(x) = \langle D^x; \lhd, [R^x_\sigma] \rangle$ and $M^{\lhd}(y) = \langle D^y; \lhd, [R^y_\sigma] \rangle$

($\Rightarrow$). Suppose x is a substring of y: it exists l, r such that $y = lxr = \sigma_1, \ldots, \sigma_{|l|}\sigma_{|l|+1}, \ldots, \sigma_{|l|+|x|}\sigma_{|l|+|x|+1}, \ldots, \sigma_{|l|+|x|+|r|}$. This implies that, for all i, $1 \leq i \leq |x|$,

$d \in R^y_{\sigma_{|l|+i}}$ iff $d \in R^x_{\sigma_i}$. Thus, if we set the isomorphism φ to be such that $\varphi(i) = |l| + i$ for $1 \leq i \leq |x|$, we have $\varphi(M^{\lhd}(x))$ that is a restriction of $M^{\lhd}(y)$, and therefore $M^{\lhd}(x) \sqsubseteq M^{\lhd}(y)$ by definition.

($\Leftarrow$). Let y be the sequence of letters $\sigma_1 \dots \sigma_{|y|}$ and suppose $M^{\lhd}(x) \sqsubseteq M^{\lhd}(y)$: there exists a isomorphism $\varphi : \{1, \dots, |x|\} \to \{1, \dots, |y|\}$ such that $\varphi(M^{\lhd}(x))$ is a restriction of $M^{\lhd}(y)$. This means that $\varphi(D^x) \subseteq D^y$ and for all σ: $\varphi(R^x_\sigma) = \{\varphi(i) \in R^y_\sigma \mid \varphi(i) \in \varphi(D^x)\}$ (Definition 2). This implies that $x = \sigma_{\varphi(1)} \dots \sigma_{\varphi(|x|)}$. Given that $\lhd = \{(i, i+1) \in D \times D\}$, we have $\varphi(i+1) = \varphi(i)+1$ and thus there exist l and r in Σ^* such that $y = l\sigma_{\varphi(1)}, \dots, \sigma_{\varphi(|x|)}r = lxr$. □

The following proofs are similar to the first and are left to the reader.

Lemma 2 (Subsequences are factors under $M^{<}$)**.** *For all strings $x, y \in \Sigma^*$, x is a subsequence of y iff $M^{<}(x) \sqsubseteq M^{<}(y)$.*

Lemma 3 (Tier-substrings are factors under $\mathcal{M}^{\lhd^T}$)**.** *For all strings $x, y \in \Sigma^*$, x is a tier-substring of y iff $\mathcal{M}^{\lhd^T}(x) \sqsubseteq \mathcal{M}^{\lhd^T}(y)$.*

Lemma 4 (Subgraphs are factors under $\mathcal{M}^{\alpha,\lhd}$)**.** *For all autosegmental graphs $x, y \in \Gamma^*$, x is a subgraph of y iff $\mathcal{M}^{\alpha,\lhd}(x) \sqsubseteq \mathcal{M}^{\alpha,\lhd}(y)$.*

Lemma 5 (Subtrees are factors under $\mathcal{M}^{\lhd^\downarrow,\lhd^\rightarrow}$)**.** *For all trees $x, y \in T_\Sigma$, x is a subtree of y iff $\mathcal{M}^{\lhd^\downarrow,\lhd^\rightarrow}(x) \sqsubseteq \mathcal{M}^{\lhd^\downarrow,\lhd^\rightarrow}(y)$.*

Observe that $(\texttt{Fact}(M, w), \sqsubseteq)$ is a partially ordered set (poset). The next definition and lemma establishes that models of strings are principal elements of ideals and filters.

Given a poset $\langle X, \leq \rangle$, and $x_1, x_2, y \in X$, y is an *upper bound* of x_1, x_2 if $x_1 \leq y$ and $x_2 \leq y$; y is the *least upper bound (lub)* of x_1, x_2 if y is an upper bound of x_1, x_2 and for every z that is an upper bound of x_1, x_2, $y \leq z$. The lub of x_1, x_2 is often denoted $x_1 \vee x_2$, and is also known as the *join* of x_1, x_1 and the *supremum (sup)* of x_1, x_2. (Note that we overload $\vee$ to denote both logical or and the join operator; its meaning should be clear from context.)

Definition 4 (Ideals)**.** *A subset I of a poset is an Ideal if*

- *I is non-empty,*

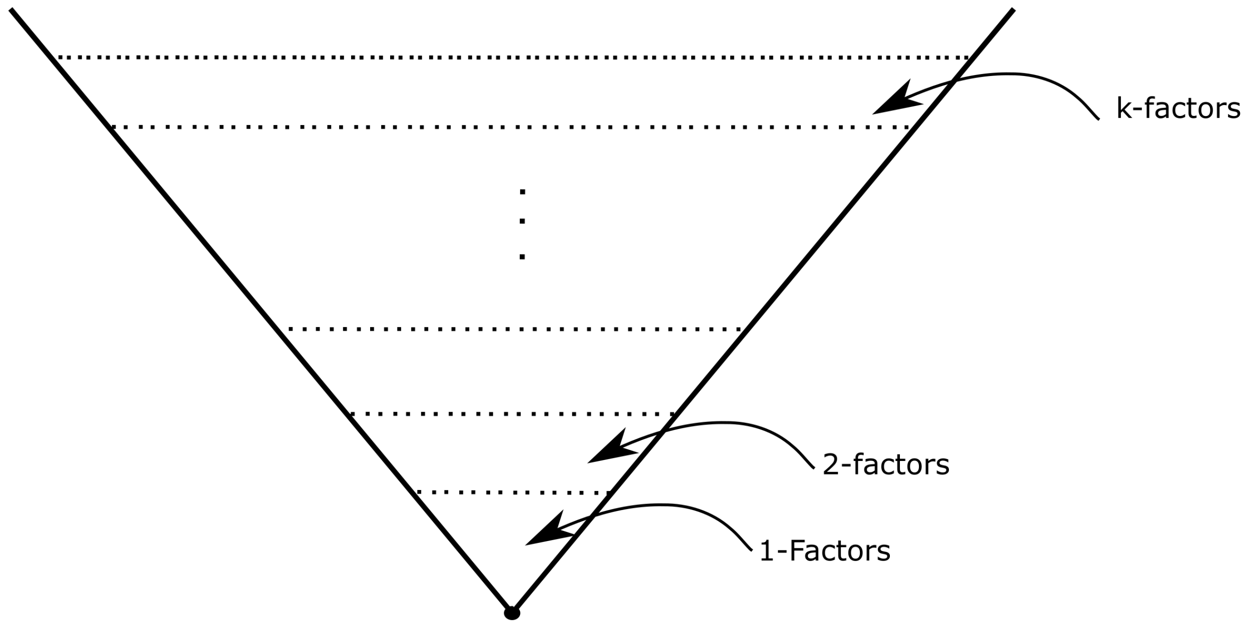


Figure 2.11: The partially ordered space of factors

- *for every x in I, $y \leq x$ implies that y is in I, and*
- *for every x, y in I, there exist some element z in I, such that $x \leq z$ and $y \leq z$.*

The dual of an ideal is a filter.

Definition 5 (Filters)**.** *A subset F of a poset is a* filter *iff*

- *F is non-empty,*
- *for every x in F, $x \leq y$ implies that y is in F, and*
- *for every x, y in F, there exist some element z in F, such that $z \leq x$ and $z \leq y$.*

Definition 6 (Principal Ideals, Filters and Elements)**.** *For any poset $\langle X, \leq \rangle$, the smallest filter containing $x \in X$ is a* principal filter *and x is the* principal element *of this filter. Similarly, the smallest ideal containing $x \in X$ is a* principal ideal *and x is the* principal element *of this ideal.*

Lemma 6. *Given a model M of Σ^* and $k > 0$, $\texttt{Fact}_k(M, w)$ is a principal ideal in $\texttt{Fact}(M, \Sigma^*)$ whose principal element is $M(w)$. $\texttt{Supfact}_k(M, w)$ is a principal filter in $\texttt{Supfact}(M, \Sigma^*)$ whose principal element is $M(w)$.*

Remark 1 (The empty structure)**.** *Given a model M, the empty structure $\langle\varnothing;\varnothing,\dots\varnothing\rangle$ is a factor of every structure in* $\texttt{Fact}(M,\Sigma^*)$.

Remark 2 (Factors and ideals)**.** $\texttt{Fact}_k(M,w)$ *is a principal ideal in* $\texttt{Fact}(M,\Sigma^*)$ *whose principal element is $M(w)$.*

Remark 3 (Superfactors and filters)**.** $\texttt{Supfact}_k(M,w)$ *is a principal filter in* $\texttt{Supfact}(M,\Sigma^*)$ *whose principal element is $M(w)$.*

Consider the following example, building on the unconventional string models discussed earlier. Consider the set of distinctive features +N (nasal), +V (voice), and +C (coronal), each of which has a corresponding negative feature (−N, −V, −C). Each of these properties may be true of a particular domain element in a model. Because of this, featural descriptions may be structured into an ideal according to generality. An example with the aforementioned features is given below. The featural description [−N,+V,+C] is a superfactor of [−N,+C] and [−N,+V], which in turn are superfactors of [−N].

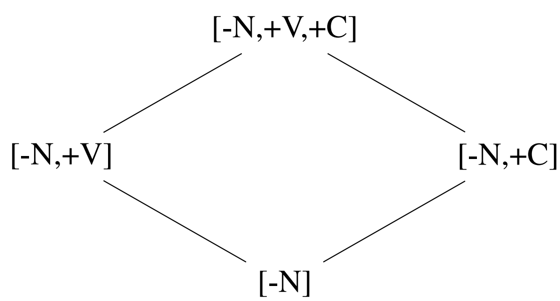


Figure 2.12: Factor ideals and filters for a unary domain element labeled with feature predicates

2.5 Grammars, Languages, and Language Classes

Factors can define grammars, formal languages, and classes of formal languages. Usually a model signature provides the vocabulary for some logical language. Sentences in this logical language define sets of strings as follows. The language of a sentence φ is all and only those strings whose models satisfy φ. Within the regular languages, many well-known subregular classes can be characterized logically in this way (McNaughton and Papert, 1971; Rogers and Pullum, 2011; Rogers et al., 2013; Thomas, 1997; Lambert et al., 2021).

Intuitively, the grammars that will be considered here consist of a finite list of *forbidden* factors, whose largest size is bounded by k. Permissible forms in the language of this grammar are those which do not contain any forbidden factors. In this way these grammars are like logical expressions which are "conjunctions of negative literals" (Rogers et al., 2013) where the negative literals are played by the forbidden factors.

Each forbidden factor is a principal element of a filter and the language is all strings whose models are not in any of these filters. For each k, there is a class of languages including all and only those languages that can be defined in this way. For example, the Strictly k-Local (SL_k), Strictly k-Piecewise, k-Tier-Based Strictly Local, and k-Multi-Tier-Based Strictly Local languages can be defined in this way. These classes contain languages which forbid finitely many substrings, subsequences, tier substrings, or multiple tier substrings respectively, according to the ordering relation which characterizes the particular k-factors for that particular model signature (successor, precedence, tier successor, multi-tier successor, respectively) (Garcia et al., 1990; Rogers et al., 2010; Lambert et al., 2021). Formally:

Definition 7. *Let k be some positive integer, and M a model of Σ^* with signature Γ. A* grammar *G is a subset of $\texttt{Fact}_k(M, \Sigma^*)$. The* language *of G is $L(G) = \{w \in \Sigma^* \mid \texttt{Fact}_k(M, w) \cap G = \varnothing\}$. The* class of languages *$\mathcal{L}(M, k) = \{L \mid \exists G \subseteq \texttt{Fact}_k(M, \Sigma^*), L(G) = L\}$.*

The elements of G are principal elements of filters, and are called forbidden factors.

As an example, let $\Sigma = \{a, b, c\}$ and consider $G = \{M^{\lhd}(aa), M^{\lhd}(bb), M^{\lhd}(c)\}$. $L(G)$ includes the strings $(ab)^+$ and $(ba)^+$ and no other strings, because the substrings aa, bb, and c are all forbidden. This language belongs to the class of languages $\mathcal{L}(M^{\lhd}, 2)$.

Lemma 7. *For each $w \in L(G)$ and each $g \in G$, $\texttt{Fact}(M, w)$ has a zero intersection with $\texttt{Supfact}(g)$.*

Proof. Suppose it is the case: it exists $A \in \texttt{Fact}_k(\Sigma^*)$ such that $A \sqsubseteq M(w)$ and $g \sqsubseteq A$. This implies that $g \sqsubseteq M(w)$ and thus that $\texttt{Fact}_k(M, w) \cap G \neq \varnothing$, which contradicts Definition 7. $\square$

In other words, the principal ideal of $M(w)$ is disjoint from the principal filters of the elements of G.

To conclude, this chapter defined a rigorous yet flexible notion of linguistic structure using the features of finite model theory. Linguistic structure is characterized as the signature of some model, for which individual linguistic structures are models. The chapter described various notions of linguistic structure this way, and showed how the model-theoretic notion of structure leads to a notion of a sub-structure, called a "factor". Sets of these factors form the components of a grammar, assigning a factor some measure of well-formedness. The factors themselves were shown to be organized into a partial order. The next chapter will use these notions to develop a concept of learning with these factors.

Chapter 3

Learning with Partially Ordered Representations

3.1 The Learning Problem[1]

The previous chapter discussed how the notion of a linguistic structure can be given a unified description using the language of model theory, and showed how logical statements over such model-theoretic structures function as grammars by solving membership problems for sets of structures. It further showed that the nature of these representations structures the space of possible "forbidden" structures characterizing a class of grammars into a particular structure itself—a partial order. It was further shown that the individual elements of this space form grammatical ideals, and thus that the sets of "forbidden factors" which characterize grammars are collections of ideals.

For some M, k, is $\mathcal{L}(M, k)$ learnable from positive data? The inductive answer appears to be Yes (Heinz, 2010c; Heinz et al., 2012). The solution presented in these studies uses the function $\texttt{Fact}_k(M, w)$ to identify permissible k-factors in words w in the positive data. The k-factors that are not permissible are forbidden. With sufficient positive data, such a learning algorithm will converge to a grammar that generates any target language in the class. Since there is a finitely sized

[1]This chapter, particularly Sections 3.1-3.4, expands on prior collaborative work, see Chandlee et al. (2019)

grammar of k-factors that will identify the language, there is some point at which the grammar will not add any new factors. Thus, it is guaranteed to identify in the limit the target language. This situation holds for any grammar that can be thought of as a set of k-factors over a particular model signature.

However, the problem of induction established in Chapter 1—that for any given pattern there is a prohibitively large, often infinite, number of candidate hypotheses or explanations for that pattern—remains. The issue is that for some word models, there are many factors of that model which work equally well. When factors are interpreted as well-formedness constraints for a grammar, the issue is that there are many constraints which will describe a particular pattern equally well.

The issue is quite easy to see in the domain of phonological representations and grammars using non-canonical string models. As mentioned in the previous chapter, phonological representations often eschew unary segmental properties for more fundamental distinctive features. A particular representational element may have multiple of these features, which in concert pick out a particular segment. These multiple shared properties pose a crucial problem for inductive inference. In a paper arguing for the need for statistical generalization to overcome gappy data, Wilson and Gallagher (2018) describe the problem succinctly:

> What about[...]a nonstatistical model[...]that learns by memorizing feature sequences? The immediate problem confronting such a model is that any given segment sequence has multiple different featural representations.
>
> (Wilson and Gallagher, 2018, pg. 616)

Here is an example. Imagine the substring nt is not present in a corpus. There are many possible equivalent forbidden substructures: *nt, *[+nasal][+coronal], *[+consonant][+coronal,-continuant], *[+sonorant][-sonorant], etc. How can a learner decide which of these constraints is responsible for the absence of nt? Some may misclassify positive data. Or, to take another example, Wilson and Gallagher (2018) describe a segmental co-occurrence pattern from Quechua, which will be dealt with in more detail in the next chapter. They describe a surface attested dorsal-tier trigram substring [oqa]. As they describe, one could come up with many equivalent substructures using distinctive phonological features which would rule out the pattern. The substring [oqa] could be represented in the following ways:

- with factors using very general feature classes
  - [+syll][syll][+syll] = VCV
- with factors using maximally specific feature classes
  - [+syll,high,low, +back][cont,son,+dorsal,high,cg][+syll,high, +low] = [oqa]
- with factors using features at intermediate levels of granularity
  - [+syll,high,low][cont,son,+dorsal,high][+syll,high, +low] = EQA

The issue here is exactly that noted in the introduction: the fundamental problem of inductive inference. For any given pattern, there are many parameters of a grammar which will account for the pattern. Unless there is some predetermined method for deciding which hypotheses are to be preferred over others, there is no hope of ensuring that the best solution is found. In linguistic terms, this is the problem of poverty of the stimulus. The primary data does not in itself give any hint on how to accurately select between grammars which will equivalently explain the phenomena at hand. In this case the parameters are the factors, and the presence of multiple shared properties in the form of the unary relational predicates means that there are many equivalent ways of describing a particular structure. In the terms of the previous chapter, there are many equivalent factors for deciding well-formedness. Wilson and Gallagher (2018) state the problem in this way:

> If a hypothetical [learner] judged a substring to be legal as long as it satisfied any attested featural description, it would tolerate (among other structures) every VCV trigram and thus massively overgeneralize. If the model instead required all feature representations of a substring to be attested, it would be equivalent to [memorizing segmental trigrams] [. . .] Lacking a method for deciding which representations are relevant for assessing well-formedness—precisely the role played by statistics [in inductive linguistic models]—learning . . . is doomed.
>
> (Wilson and Gallagher, 2018, pg. 617)

Under what criteria does a learner select between grammatical hypotheses that are otherwise equivalent? Schmidhuber et al. (2002) asks a similar question. “We should somehow specify additional constraints on the generating process. But which constraints are plausible? Which reflect the “natural” concept of simplicity?” Here Schmidhuber brings in a notion of “simplicity” of a hypothesis which is also not new in grammatical inference. It is often claimed, under many names, that particular inferences should preferably depend on the simplicity of their parameters. In this sense, the grammar a learner entertains should be “simple” or “most general” under some definition of simplicity or generality.

The following sections will describe how the properties of factors and superfactors themselves enable a property that overcomes this problem to a significant degree, enabling successful inference. In particular, it shows how the partial order that the factor space has enables certain entailments, which the learner may use to successfully prunce the factor space and select the most general grammars.

3.2 Grammatical Entailment

The structure of the space of k-factors of a model is, as mentioned in the prevous chapter, a partial order. Additionally, the factors themselves form principal ideals and filters within that space of factors such that factors and superfactors are ordered in a highly structured way.

What kind of inferential abilities does this structure allow? One particularly salient inferential property is that information about other factors entails information about other factors.

Consider the problem from the domain of segments labeled with distinctive phonological features. Consider the example given in the previous chapter, of the set of distinctive features +N (nasal), +V (voice), and +C (coronal), each of which has a corresponding negative feature (-N, -V, -C). A segment may be specified for one or more of these features. Because of this, featural descriptions may be structured into an ideal according to generality. An example with the aforementioned features is given below. The featural description [-N,+V,+C] is more specific than [-N,+C] and [-N,+V], which in turn are more specific than [-N]. If each of these are taken to be possible constraints, the more general ones cover more possible forms. Banning all non-nasals is a more general statement and bans more forms than banning voiced non-nasals, which is more general than banning coronal voiced non-nasals.

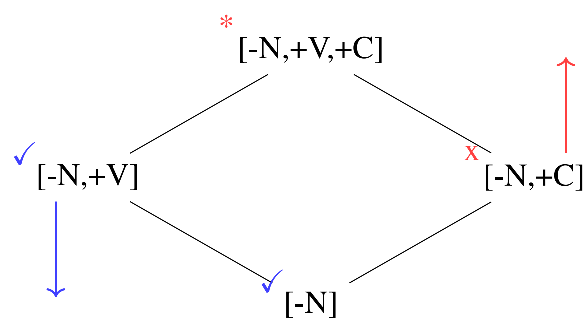


Figure 3.1: Feature-based factor ideals and filters. Red indicates ill-formedness, and blue indicates well-fomedness of a factor with respect to some grammar

Consider the feature ideal described in Figure 3.1. The learner may directly draw two inferences about the nature of these hypotheses. If a learner has evidence that voiced non-nasals are allowed, meaning the structure [-N,+V], then the learner may directly infer that non-nasals are allowed. This means that the learner does not have to consider this particular constraint, it gets it for free. Similarly, if the learner has evidence that non-nasal coronals ([-N,+C]) are banned, then the learner can directly infer that non-nasal voiced coronals ([-N,+V,+C]) are also banned. This means the learner does not have to consider this constraint. Since each node in the structure may be a principle ideal or filter, information about the grammaticality of one node entails information about the grammaticality of other nodes, in either direction.

In this way, the ideals and filters within a a particular model noted above give rise to these entailment properties of grammaticality with respect to the hypothesis space. If the learner constructs filters, then the grammar G will allow structures such that language membership is downward entailing with respect to the grammar G, and language non-membership is upward entailing with respect to the grammar G.

The advantage of this may be seen more clearly when considering the larger hypothesis space of featural descriptions,capped at three descriptions per segment. This is shown below in Figure 3.2. The lowest element in the space is a segment with total feature underspecification. Going up one layer in the structure adds a particular valued feature at a time. The ordering relations are drawn in between.

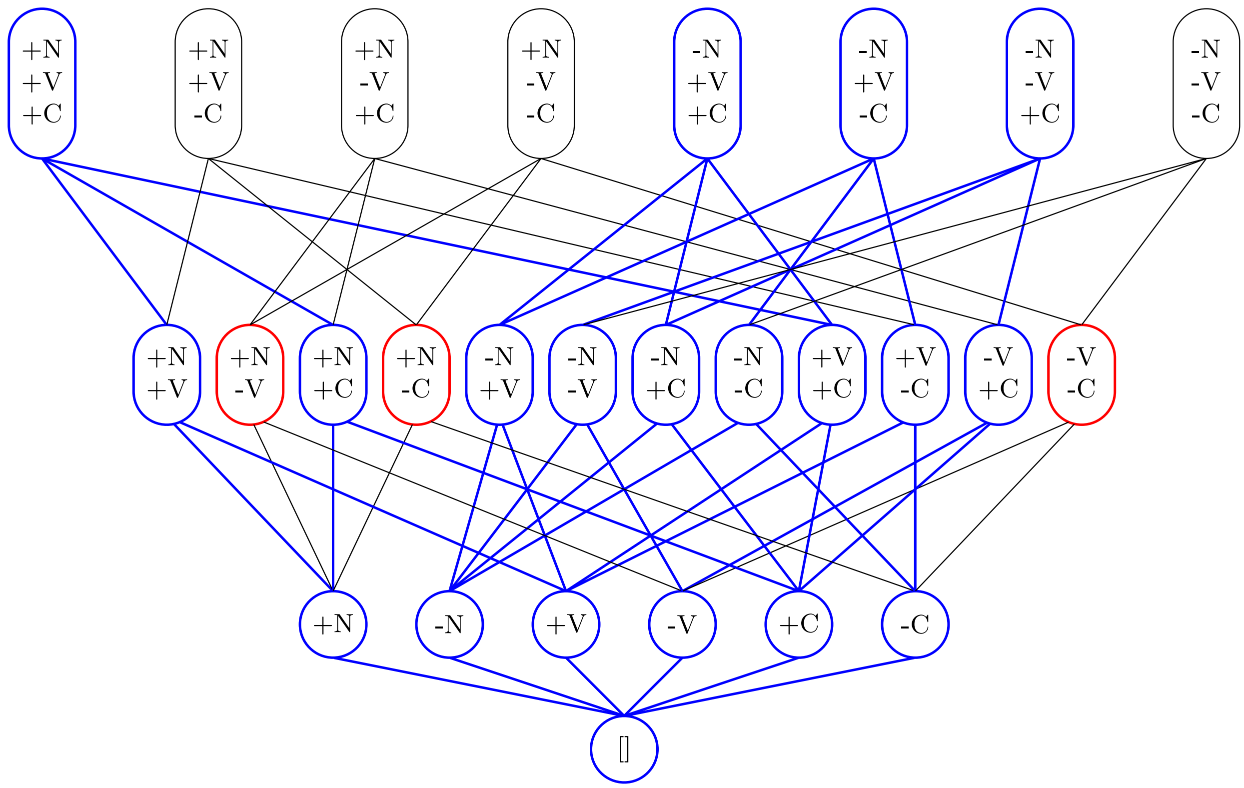


Figure 3.2: Featural hypothesis space

Given a grammar G, we call a factor s in $\texttt{Fact}(\Sigma^*)$ *ungrammatical* if it belongs to a principal filter of any element of G. Factors that are not ungrammatical are called *grammatical*. Lemma 6 (pg. 35) ensures that grammaticality is downward entailing, in the sense that if a model of the word $M(w)$ is not contained in the principal filters of the elements of the grammar, then neither are the factors of $M(w)$. But it also ensures that ungrammaticality is upward entailing: if a model of the word $M(w)$ belongs to the principal filters of the elements of the grammar, then all of the superfactors of $M(w)$ in that filter are likewise contained.

3.2.1 Example: Orthographic Capitalization

As an example, consider capitalized letters as discussed above. In an unconventional word model, each capital letter at some position x is represented as satisfying one of the relations

$R \in \{\mathtt{a}(x), \mathtt{b}(x), \ldots, \mathtt{z}(x)\}$ as well as the unary relation $\mathtt{capital}(x)$. Thus the relation $\mathtt{a}(x)$ is true of both lowercase a and uppercase A, but $\mathtt{a}(x) \wedge \mathtt{capital}(x)$ is only true of uppercase A. Note also that in this model no position x of a structure can satisfy both predicates $\mathtt{a}(x)$ and $\mathtt{b}(x)$.

To ease the exposition here, we will use square brackets to delimit the domain elements and write the unary relations within them instead of specifying the model in mathematical detail. For example, for the string Bab, we will write $[\mathtt{b}, \mathtt{capital}][\mathtt{a}][\mathtt{b}]$.

Figure 3.3 showcases the relationship among these structures under a model M. The structure for A, $[\mathtt{capital}, \mathtt{a}]$, contains as factors $[\mathtt{capital}]$, $[\mathtt{a}]$, [], and the empty structure (not shown). The empty structure is a factor of [], and [] in turn is a factor of $[\mathtt{capital}]$ and $[\mathtt{a}]$. The factor $[\mathtt{a}]$ contains the factor [], the domain element with no relations, but has superfactors [capital, a], which has one domain element and two relations, and [a][], which has two domain elements, and the first satisfying the property a. Factors and superfactors are listed above and below each other, respectively, with lines between them. Members of one ideal are noted with a blue checkmark, and members of a filter are noted by a red asterisk.

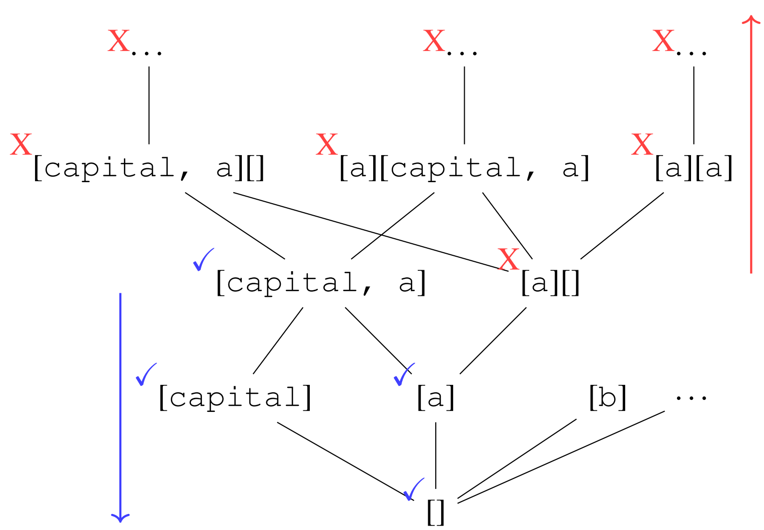
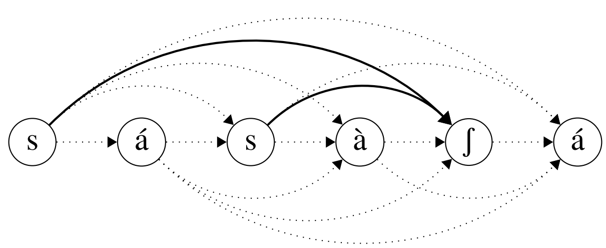


Figure 3.3: The Structure ideals(blue) and filters(red) for a capitalized letter model.

Applying this to the example in Figure 3.3, if the structure $[\mathtt{capital}, \mathtt{a}]$ is grammatical, then all of its factors, such as [capital] and [a], and [] are grammatical. Since those are grammatical, each of their factors is also grammatical, which in this case is just [∅], shown in blue in Figure 3.3. Conversely, if the structure [a][] is known to be ungrammatical, then any structure which has it as a factor is also ungrammatical (in this example, [`capital,a`][], shown in Red in Figure 3.3. To see the importance, consider a string with only lowercase letters. In a connected model, the grammar would ban 26 forbidden factors (A,B,C,...), but the “capital” model bans just one, [`capital`].

3.2.2 Example: Long Distance Phonological Dependencies

In many languages, the presence of certain segments is dependent on the presence of another segment. In Samala, subsequences like `s...s` are allowed but `s...ʃ` are not (Hansson, 2010). Consider the hypothetical word modeled in the previous chapter, [sasaʃa], with its model under the precedence signature depicted again below. This word would be ruled as ungrammatical under this linguistic constraint, simply because of the presence of the forbidden subsequence, in this case twice (between the first and fifth element, and again between the third and fifth element). In contrast a word like sasasa would be allowed, because the subsequence under consideration is not present in the word.



$\langle$dom; $<$, s, ʃ, á, à$\rangle$

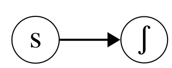
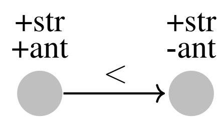


Figure 3.4: Precedence model of ‘sásàʃá’, along with the forbidden 2-factor ‘sʃ’, which is contained twice

Note also that in this model no position x of a structure can satisfy both predicates $+\mathtt{str}(x)$ and $-\mathtt{str}(x)$. We again use square brackets to delimit the domain elements and write the unary features within them, so a model representation like $\begin{bmatrix}\texttt{+str}\\ \texttt{+ant}\end{bmatrix}\begin{bmatrix}\texttt{+str}\\ \texttt{-ant}\end{bmatrix}$ has the following visual representation:



To ease the exposition, we will use square brackets to delimit the domain elements and write the unary relations within them instead of specifying the model in mathematical detail. In an unconventional subsequence word model, then, one possible structure of the subsequence $\texttt{s...ʃ}$ is written $\begin{bmatrix}\text{+str}\\ \text{+ant}\end{bmatrix}\begin{bmatrix}\text{+str}\\ \text{-ant}\end{bmatrix}$.

In an unconventional model, banning structures of the form $\texttt{[+str][+str]}$ is insufficient, since all these segments share that stridency property, while a structure like $\begin{bmatrix}\texttt{+str}\\ \texttt{+ant}\end{bmatrix}\begin{bmatrix}\texttt{+str}\\ \texttt{-ant}\end{bmatrix}$ will distinguish them, since they disallow stridents which disagree on the $\pm\texttt{ant}(x)$ relations. The structure $\texttt{[+ant][-ant]}$ however, is insufficient, since consonants like $\texttt{p,b,m}$ have that feature, and would incorrectly ban acceptable strings. To see the importance, a conventional string model mus ban multiple sibilant factors sʃ,zʃ,sʒ,zʒ, while an unconventional model can just ban one, $\begin{bmatrix}\texttt{+str}\\ \texttt{+ant}\end{bmatrix}\begin{bmatrix}\texttt{+str}\\ \texttt{-ant}\end{bmatrix}$.

Figure 3.5 showcases the relationship among these structures under a precedence model $M^<$.

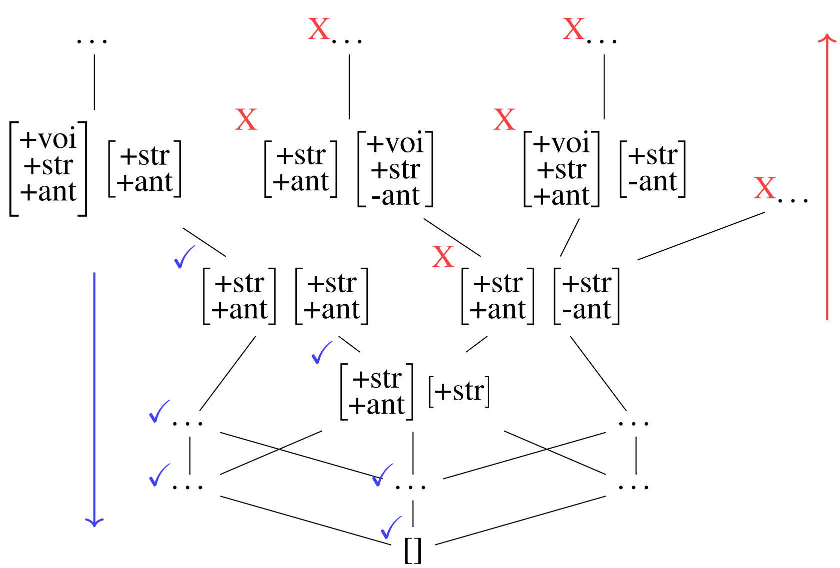


Figure 3.5: Structure ideals(blue) and filters(red) for a phonological precedence model.

The structure for $\begin{bmatrix}\text{+str}\\ \text{+ant}\end{bmatrix}$ [+str] contains as factors (among others) [+str] [+str], [+str] [] , [], and the empty structure (not shown). The empty structure is a factor of [], and [] in turn is a factor of

$[+\texttt{ant}]$ and $[-\texttt{str}]$, and so on. The factor $[+\texttt{ant}]$ contains the factor [], the domain element with no relations, but has superfactors [+ant, +str], which has one domain element and two relations, and $[+\texttt{ant}][]$, which has two domain elements, and the first satisfying the property +ant. factors and superfactors are listed above and below each other, respectively, with lines between them. Members of one ideal are noted with a blue checkmark, and members of a filter are noted by a red X. If the structure $\begin{bmatrix}\text{+str}\\ \text{+ant}\end{bmatrix}\begin{bmatrix}\text{+str}\\ \text{+ant}\end{bmatrix}$ is grammatical, then all of its factors, are grammatical, and so are their factors, in turn. Conversely, if the structure $\begin{bmatrix}\text{+str}\\ \text{+ant}\end{bmatrix}\begin{bmatrix}\text{+str}\\ \text{-ant}\end{bmatrix}$ is known to be ungrammatical, then any structure which has it as a factor is also ungrammatical (for example, $\begin{bmatrix}\text{+voi}\\ \text{+str}\\ \text{+ant}\end{bmatrix}\begin{bmatrix}\text{+str}\\ \text{-ant}\end{bmatrix}$, where the first segment is also voiced +voi), shown in Red in Figure 3.5.

Generally, these reductions can be exponential: an alphabet of size 2^n can be represented with n unary relations in the model signature. However, this exponential reduction does not necessarily make learning any easier. The reason for this is that the size of $\texttt{Fact}_k(M, \Sigma^*)$ equals $\sum_{i=1}^{k}(2^n)^i$ where n is the number of unary relations. Since a grammar is defined as a subset of $\texttt{Fact}_k(M, \Sigma^*)$, the number of considered grammars is thus very large. Therefore, the problem of how to search this space effectively is paramount.

3.3 The Learning Problem, Revisited

The structural filters give the learner an advantage when confronting hypothesis spaces under a particular model. In particular, it allows the learner to prune vast swathes of the hypothesis space as it reaches for principal elements of features. If a learner identifies one structure as being grammatical, the learner may infer that all of its factors are also grammatical and not have to consider them. Alternatively, if the learner knows a structure is ungrammatical, it may infer that the ideals above it are also ungrammatical.

Consider a learner working over the space in Figure 3.2 (pg. 44) who has observed the following input structures: nasal voiced coronals, non-nasal voiced coronals, non-nasal voiced non-coronals, and non-nasal voiceless coronals. The maximally specified feature descriptions of these are [-N,+V,+C] [-N,+V,+C] [-N,+V,+C] [-N,+V,+C], and are circled in blue in the top of the figure. What constraints is the learner to posit? One strategy would simply be to ban the four most specified forms not present in the data. However, there is a more general strategy. If the learner takes advantage of the grammatical entailments, it may conclude that any more general variant of the accepted forms is licit, and thus not able to be considered as a possible constraint, noted in blue in the figure. The most general constraints left are guaranteed to ban exactly the forms that are not seen in the data. In addition, the learner only has to store three constraints, rather than four, and they are more general. This is a crucial generalization step.

As overviewed in Chapter 1, this section uses the concept of grammatical entailments to state this inference problem using a constraint-inclusion view of inference problems (Haig, 1987; Nickles, 1981). As a first attempt in this spirit, Chandlee et al. (2019) state the learning problem not in terms of converging to a correct grammar in the limit as previously studied, but instead as returning an 'adequate' grammar given a finite positive sample. Determining what counts as an adequate grammar is what (Raedt, 2008) calls a Quality Criterion.

Definition 8 (The Learning Problem). *Fix Σ, model M, and positive integer k. For any language $L \in \mathcal{L}(M, k)$ and for any finite $D \subseteq L$, return a grammar G such that*

1. G is consistent, that is, it covers the data: $D \subseteq L(G)$;
2. $L(G)$ is a smallest language in $\mathcal{L}$ which covers the data: so for all $L \in \mathcal{L}$ where $D \subseteq L$, we have $L(G) \subseteq L$; and
3. G includes structures S that are restrictions of structures S' included in other grammars G' that also satisfy (1) and (2): for all G' satisfying the first two criteria for all $S' \in G'$, there exists $S \in G$ such that $S \sqsubseteq S'$.

The first criterion is self-explanatory. It enforces the constraint that the grammar explains or generates the data. The second criterion is motivated by Angluin's (1980) analysis of identification in the limit. It says that for all languages L for which our data D is a subset, the language of the grammar is a subset of L. Both of these criteria are familiar in inductive grammatical inference.

The third criterion requires that the grammar contain the most "general" factors. Criterion 3 codifies the notion of simplicity that Thagard (1988) deemed the most appropriate for theory choice: preference should be given to theories that make fewer special or ad hoc assumptions. As

Haig (2018) notes, Thagard regarded simplicity as the most important constraint on explanatory breadth. One should not sacrifice simplicity through ad hoc adjustments to a theory in order to enhance its explanatory breadth. Criterion 3 derives this directly through the concept of the factor ideals, and the notion of the partially ordered hypothesis space. In short, a factor is simpler than its superfactors, since it contains less structural information.

As a forbidden structure, the factor has greater explanatory breadth than its superfactors since its extension is larger. This situation can be seen in Figure 3.6. This figure shows the partially ordered hypothesis space, and considers two factors, A and B, whose superfactor filters are in red. With respect to a toy example of attested data, both factors satisfy Criteria 1 and 2. However, Factor B is a superfactor of factor A, meaning it is in A's filter, among many other factors. By criterion 3, factor A will be preferred.

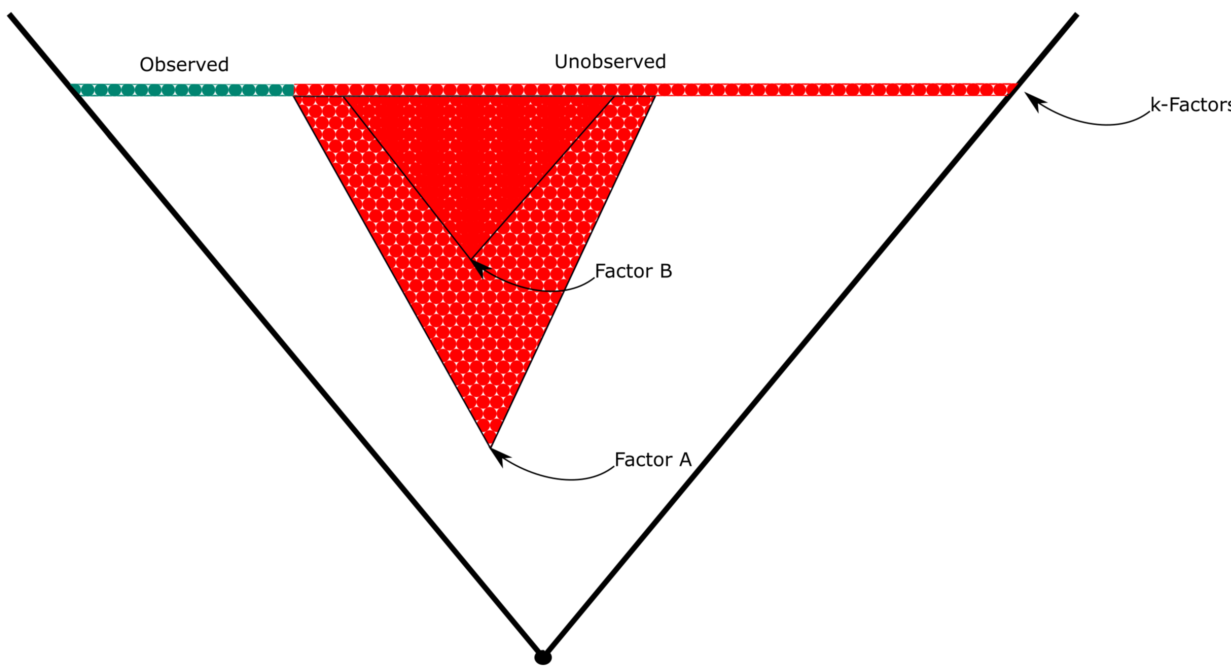


Figure 3.6: Pruning the hypothesis space

An example will help illustrate these points. Consider the grammar $G = \{M^{\triangleleft}(aa), M^{\triangleleft}(bb), M^{\triangleleft}(c)\}$ with $\Sigma = \{a, b, c\}$. $L(G)$ is the same as $L(H)$ where $H = \{M^{\triangleleft}(aa), M^{\triangleleft}(bb), M^{\triangleleft}(ac), M^{\triangleleft}(bc), M^{\triangleleft}(cc), M^{\triangleleft}(ca), M^{\triangleleft}(cb)\}$. In H all the forbidden factors are of size 2, whereas G encapsulates all of the 2-factors in H which include c with a single 1-factor $M^{\lhd}(c)$. Both grammars G and H may satisfy criteria (1) and (2) but H would not satisfy criterion (3) because of G.

3.4 A Bottom-Up Inference Algorithm

With the abductive definition of the learning problem postulated, we may construct a learning algorithm to be evaluated with respect to this paradigm via criteria 1–3. Raedt (2008) identifies two relevant directions of inference: specific-to-general (i.e., 'top-down') and general-to-specific (i.e., 'bottom-up'). Since we are trying to find the most general factors, top-down inference has the potential to consider exponentially many more factors than bottom-up inference. It makes mores sense to traverse bottom-up, that is, from the most general factors possible to the most specific. Additionally, once a factor is identified as an element of the grammar, none of its superfactors (elements of its principal filter) need to be considered further.

A Bottom-Up Factor Inference Algorithm (BUFIA) is shown in Algorithm 1. Its input is a positive data sample of model-theoretic structures D and an integer k that identifies the upper bound on the size of the factors.

The algorithm makes use of a queue Q, which is initialized to contain just the empty structure s_0. It also initializes two empty sets: G, the grammar that will ultimately be returned, and V, the set of 'visited factors'. The factors in Q are considered one at a time, in order, and as each factor s is considered it is added to V. If s is not a factor of the model of any word in the positive sample D (i.e., not contained by any data point in D), then it is added to the grammar G.

If s is a factor of the sample, it is sent to the function `NextSupFact`, which returns a set of *least* superfactors for s. For concreteness, $\texttt{NextSupFact}(s)$ may be defined formally as follows:

$\texttt{NextSupFact}(s) = \{S \in \texttt{Fact}_k(\Sigma^*) \mid s \sqsubseteq S, \neg\exists\, S'[s \sqsubseteq S' \sqsubseteq S]\}$.

Practically `NextSupFact` will be defined constructively so that each factor in $\texttt{Fact}_k(\Sigma^*)$ is constructed only once as needed. Thus, not only will it not be needed to store the whole set $\texttt{Fact}_k(\Sigma^*)$ in memory, but the set V may be excluded from the algorithm as well.

This set of superfactors is then filtered by the following criteria: they must be smaller than $k+1$, they must contain no element of G as a factor, and they must not have been previously considered

Data: positive sample D, empty structure s_0,
max constraint size k
Result: G, a set of constraints

```
Q ← {s_0};
G ← ∅;
V ← ∅;
while Q ≠ ∅ do
    s ← Q.dequeue();
    V ← V ∪ {s};
    if ∃x ∈ D such that s ⊑ x then
        S ← NextSupFact(s);
        S′ ← {s ∈ S | |s| ≤ k ∧ (¬∃g ∈ G)[g ⊑ s] ∧ s ∉ V};
        Q.enqueue(S′);
    end
    else
        G ← G ∪ {s};
    end
end
return G;
```

Algorithm 1: Bottom-up Factor Inference Algorithm (BUFIA)

(i.e., they cannot be in V). Those structures that survive this filter are added to Q. This procedure continues until there are no more structures left to consider in Q.

How does BUFIA behave with respect to the learning problem in Definition 8? In general we may state the following result.

Theorem 1. *For any $L \in \mathcal{L}(M, k)$, and any finite set $P \subseteq L$ provided as input to Algorithm 1, it returns a grammar G satisfying Definition 8.*

Proof. Consider any $x \in D$. Algorithm 1 only adds elements to G that are not factors of x, so $x \notin \texttt{Supfact}(G)$. Thus $x \in L(G)$ and $D \subseteq L(G)$, satisfying Condition (1).

Consider any $L' \in \mathcal{L}$ with $D \subseteq L'$. To show $L = L(G) \subseteq L'$, consider any $w \in L$. Then $\texttt{Fact}(w) \subseteq \texttt{Fact}(D)$ and $\texttt{Fact}(D) \subseteq \texttt{Fact}(L')$ since $D \subseteq L$. Then $\texttt{Fact}(w) \subseteq \texttt{Fact}(L')$. Hence, $w \in L'$, and so $L \subseteq L'$, satisfying Condition (2).

For condition (3), we use the fact that elements in the grammar G were in Q at some point. Suppose s, s' are factors such that $s \in G$, $s' \sqsubseteq s$, and $(\neg\exists x \in D)[s' \sqsubseteq M(x)]$. Since $s \in G$, then at some point $s \in Q$.

If $s' \sqsubseteq s$ then s' will be added to Q before s is generated by `NextSupFact`. Because Q is a queue, s' will also be removed from Q before s is generated by `NextSupFact`. Since s' is not contained by any $M(x)$ with $x \in D$, it will be added to G. When s is generated by `NextSupFact`, it will not pass the filter because it fails the second criterion since $s' \sqsubseteq s$ and $s' \in G$. Then s is never added to Q, and therefore $s \notin G$, contra our original assumption. Thus Condition (3) is satisfied. □

One aspect of BUFIA to highlight is that when a factor g is added to G, it is not added to Q. Consequently, $\texttt{NextSupFact}(g)$ is never added to Q. In this way, finding elements of G prunes the remainder of the space to be searched (see figure 3.6). In general, it is not the case that every element in the principal filter of g will not be generated by `NextSupFact` since some of these elements may belong to $\texttt{NextSupFact}(x)$ for other factors x on the Q. Factors on the 'border' of $\texttt{Supfact}(g)$ to be generated in this way (and then they are filtered out). This pruning, especially when the factors are quite general, can significantly reduce the remaining space to be traversed.

In regard to efficiency, in the worst case, the elements of G are all very specific factors and are greatest elements of $\texttt{Fact}_k(\Sigma^*)$. In this case, every factor $\texttt{Fact}_k(\Sigma^*)$ will be added to Q and the time complexity is thus exponential. However, we are primarily interested in the case when $\texttt{Fact}_k(D)$ are a small proportion of $\texttt{Fact}_k(\Sigma^*)$. This constitutes an example of data sparsity. In this case, we believe the elements of the target grammar will be much 'lower' in the partial order and thus will be found much more quickly. Determining what conditions on $\texttt{Fact}_k(D)$ and $\texttt{Fact}_k(\Sigma^*)$ result in a polynomial time run in the size of D is a focus of current research activity.

One area of flexibility of BUFIA regards the `NextSupFact` function for models with a successor or precedence order relation and arbitary unary relations. The basic idea underlying the bottom-up algorithm is to develop a spanning tree for the poset $\texttt{Fact}(\Sigma^*)$ and to traverse this tree in a breadth-first manner. The function `NextSupFact` helps control this search. Ideally, `NextSupFact` would only generate each factor once, which obviates the need to store visited factors in V. This can be accomplished to some extent in different ways. For incompatible unary relations, like `a` and `b` in the capitalization example, `NextSupFact` can be defined to prevent adding property `a` to a position that already satisfies property `b`.

For compatible unary relations, like a and capital in the capitalization example, an ordering over the unary relations such as $\texttt{a} < \texttt{b} < \texttt{capital}$ can also help eliminate generating the same factor in different ways. For example, if NextSupFact is defined to only add 'lesser' unary relations to positions that already have them then it would only output [capital, a] given the factor [a] as input. On the other hand, when given as input the factor [capital], it could not add any unary relation to this position.

3.5 Further Abductive Principles for Grammar Selection

The previous sections used the structure given to a hypothesis space by the representations in it to derive an abductive inference problem for which there exists a guaranteed solution. In particular, it satisfies criterion 3 purely on a structural basis. While any $L \in \mathcal{L}(M, k)$ may have exponentially many equivalent factor-based grammars, the algorithm prunes out a significant amount based purely on the notions of factor ideals and grammatial entailment.

However, while BUFIA as stated in Chandlee et al. (2019) is guaranteed to find such a grammar, its behavior has a surprising consequence: it finds *any* factor, and thus grammar, which meets these criteria. That is, a factor is added to the grammar if it meets criteria 1-3, without knowing anything more about the grammar as already entailed. This creates a situation of dramatic potential for redundancy. How is the algorithm to know whether a factor which it is proposing covers the data, yet also meets criteria 1-3? In short, what happens to factors whose extensions are all surface-true, but are pairwise incomparable? This situation is shown in Figure 3.7. Here neither Factor A nor B are superfactors of one another, so they meet criteria 1-3. However, they also have overlapping filters, meaning they share some superfactors despite themselves being pairwise incomparable. When considering the most specific factors in their superfactors, those of size k, they may share significant overlap with respect to the data.

This problem is a variant of the problem posited in Figure 3.6, in that there is still a problem of ruling out equivalent factors. Yet, BUFIA succeeds in learning the most general grammar, as shown, so what is BUFIA doing with all of the extra factors? In the BUFIA algorithm as considered, such factors are not pruned out. They are merely added to the current grammar.

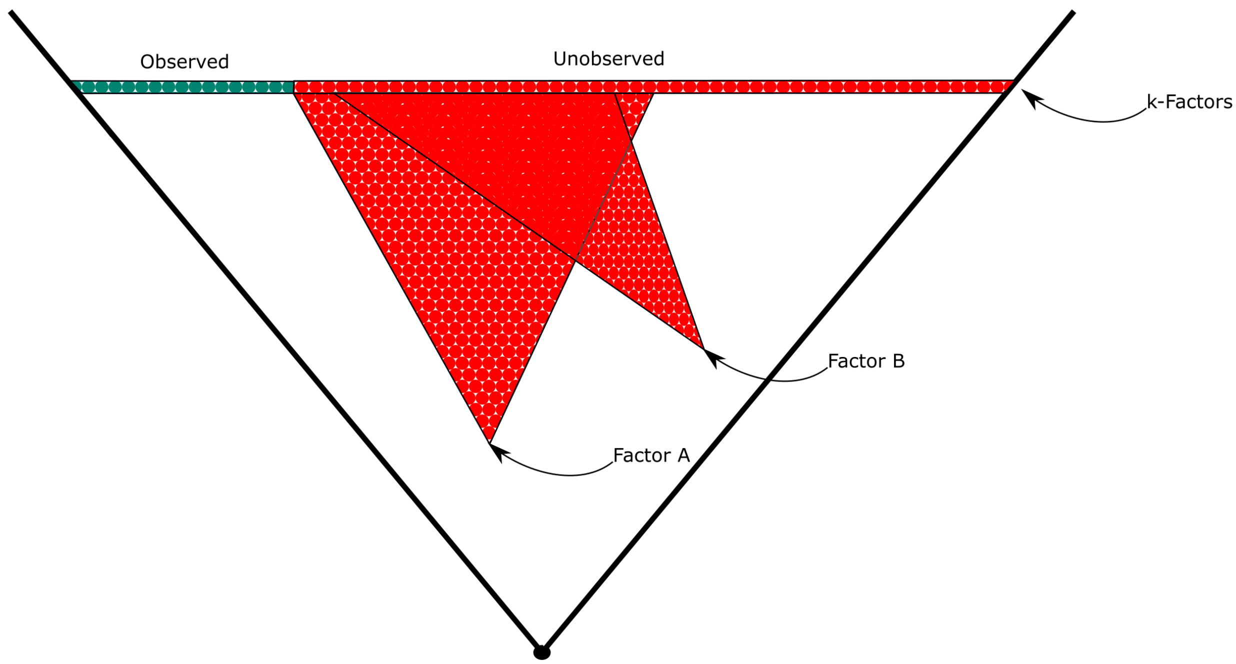


Figure 3.7: Two pairwise incomparable factors with overlapping filters

Thus the potential for dramatic redundancy emerges without much resistance. In this way, though the algorithm is guaranteed to find the most general grammar under the constraints of generality provided in definition 8, it also finds the largest such grammar. In a sense this may not be a problem, and may even be advantageous. Perhaps one would like to know all the possible variant forbidden subfactors a grammar could contain. On the other hand, one may not want to entertain such a large number of forbidden structures if the size of the grammar is large. This situation will be seen on some natural language patterns in the next chapter.

Is there a way to prune out even more factors to avoid this situation? By framing the inference problem as one of abductive constraint satisfaction, the solution to this problem involves changing an existing abductive constraint, or adding a new abductive constraint. In what follows we will consider two additional such constraints.

Let us consider some additional constraints that are extensional rather than intensional, meaning they relate to $L(G)$ rather than properties of G itself. Let us say the learner currently has posited a grammar G and is entertaining a novel k-factor F. As mentioned in Chapter 2, F can also be thought of as a grammar, whose extension is $L(F)$. Thus we may meaningfully compare the symmetric difference between their extensions. Recall the symmetric difference of two sets $A \triangle B$ takes A and B and returns a tuple $\langle \{A - B\}, \{A \cap B\}, \{B - A\} \rangle$. So the symmetric difference between the extensions of the grammars G and F is

$$L(G) \triangle L(F) = \langle \{L(G) - L(F)\}, \{L(G) \cap L(F)\}, \{L(F) - L(G)\} \rangle$$

We may form additional principles out of any of these elements placing requirements on the extensions of the factors. We shall disregard $\{L(G) - L(F)\}$, as we are concerned with the contribution of a factor to a grammar, not the contribution of a grammar to a factor. However, one may conceive of a learning problem for which this is relevant. For example, a learner might entertain all possible permutations of the orders in which factors are presented, and compare the grammars that way. In any case, considering just the last two parts, we can formulate the following additional constraints:

- Constraint 4: $\{L(F) - L(G)\}$ is non-empty
- Constraint 5: $\{L(G) \cap L(F)\}$ is empty

Constraint 4 imposes the requirement that every structure in the grammar must account for some novel bit of the unobserved k-factors. That is, for every structure in the grammar, there is at least one unique k-factor in its filter that is in no other filter of the other structures in the grammar. This scenario is shown visually in Figure 3.8. In this case, Factor B is again pairwise incomparable with Factor A, but now Factor B must necessarily contain some superfactors that are not in the set of Factor A's superfactors. In short, Factor B contributes something novel to the grammar, in that it explains some things that A cannot. Since there are only finitely many subfactors for a particular factor of size k, this requirement will dramatically decrease the size of the output grammar. This decrease will be seen in the case studies in the next chapter.

Constraint 5 ensures that each factor is totally unique, in the sense that it only accounts for what the other factors cannot. This scenario is shown visually in Figure 3.9.

Each of these two abductive constraints affects an aspect of the learning problem that we may lump under the term *factor competition*. If one examines the four scenarios presented visually

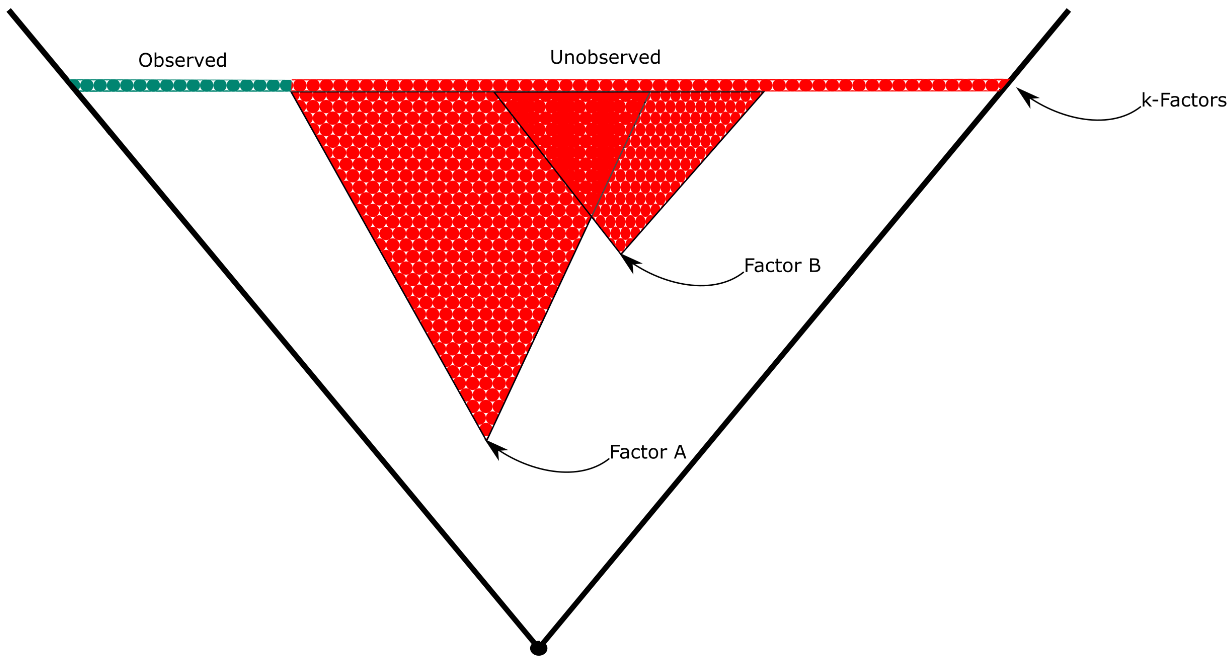


Figure 3.8: Factors with partially overlapping filters

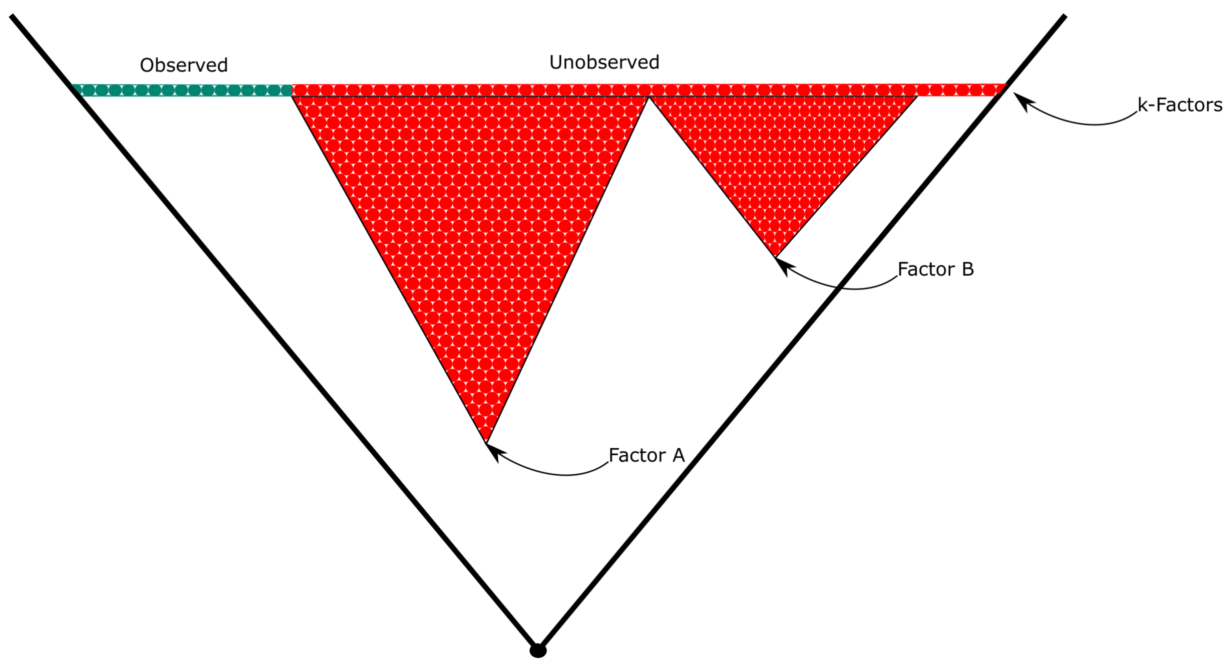


Figure 3.9: Factors with non-overlapping filters

above, the relationship between Factors A and B is one of a continuum of overlap between their filters of superfactors up to length k. They can range from not competing with one another at all in terms of pruning the space of k-factors, which is the scenario that constraint 3 rules out, to totally competing with one another with respect to the space of k-factors, which is the requirement imposed by constraint 5. In between, they may tolerate some non-competition as long as their filters do not completely overlap in terms of k-factors as shown by constraint 4.

In sum, this chapter presented a learning problem which necessitated the use of abductive rather than inductive grammatical inference. The notion of factor ideals introduced the concept of grammatical entailment, which directly led to an abductive statement of the learning problem for traversing and pruning the space of factors to select a grammar. While the BUFIA algorithm presented here may learn the correct grammar, it additionally learns every equivalent grammar, for which additional abductive constraints can be added to the constraint satisfaction problem to further prune the grammar of redundant constraints.

Chapter 4

Abductive Inference of Phonotactic Constraints

The previous chapter presented a learning problem which necessitated the use of abductive rather than inductive grammatical inference. The notion of factor ideals introduced the concept of grammatical entailment, which directly led to an abductive statement of the learning problem for traversing and pruning the space of factors to select a grammar. While the BUFIA algorithm presented here may learn a correct grammar, it learns the largest correct grammar, for which additional abductive constraints can be added to the constraint satisfaction problem to further prune the grammar of redundant factors.

This chapter presents three case studies exploring phonotactic learning with BUFIA using non-canonical string models characterized as phonological feature bundles. Phonotactics, as mentioned in chapter 1, characterize the phenomenon of linguistic speakers to draw well-formedness distinctions of phonological surface forms based on their parts. As such, this is a prime area to test the behavior of the BUFIA algorithm, as BUFIA collects forbidden substructures that, when interpreted as grammars, divide linguistic forms into well-formed and ill-formed.

The first case study is a toy example, where the feature, segment, and training inventories are small and easily interpretable. This example will be used to compare BUFIA's behavior, as well as the contribution of the additional abductive constraints, to the behavior of a well-known statistical inductive algorithm, the UCLA Phonotactic Learner (Hayes and Wilson, 2008). The second case learns the phonotactics of onset clusters in English. The final case study, on non-adjacent dependencies in Quechua, is motivated from Wilson and Gallagher (2018), who used an additional statistical heuristic to inductively prune out equivalent factors, making their situation similar to the goal of this dissertation. In all cases, all three variants of BUFIA successfully learn surface true factors, and the various abductive constraints introduce interesting restrictions on the learned grammars.

4.1 An Inductive Phonotactic Learner Comparison

For the comparisons in this chapter, we will use a popular statistical inductive inference algorithm explicitly designed for the purpose of learning phonotactic factors from data. The UCLA MaxEnt Phonotactic Learner (Hayes and Wilson, 2008), building on work by Goldwater and Johnson (2003), Della Pietra et al. (1997), and others, uses statistical inference techniques to induce phonotactic forbidden factors for sequences that are unattested or underattested in a language. The input to the learner is positive data, a list of attested words encoded as IPA strings, as well as the features that describe the segments of the language.

Their MaxEnt Learning algorithm works as follows. First, it constructs a list of natural classes based on the given features and a list of all possible k-factors constructed from those classes. The learner then randomly constructs a list of hypothetical forms, and uses an iterative scaling algorithm modeled after (Della Pietra et al., 1997) to select forbidden k-factors in the learning data. The learner posits forbidden k-factors for the relevant sequences and uses the principle of Maximum Entropy to weight the factors. This allows it to maximize the probability of the phonotactic distribution given the data. Hayes and Wilson (2008) additionally add several phonologically inspired heuristics to order their factor space. These heuristics include a preference for shorter factors and factors covering larger natural classes over smaller ones. The output grammar is a list of forbidden factors along with their weights.

The UCLA learner is advantageous as a comparison to BUFIA due to its structure, as well as its offshoots (Wilson and Gallagher, 2018; Gouskova and Gallagher, 2020) that we will consider for comparing the abductive additions to BUFIA. The learner takes the phonological feature set defined by the analyst, identifies all the unique natural classes in it (using the shortest featural description of the class), and generates a space of all possible k-factors (up to a certain k) composed of those natural classes. Just as with BUFIA, Hayes & Wilson discuss how the number of possible natural class-based factors grows exponentially with respect to k, making search infeasible.

In order for a factor to be added to the grammar, it must meet or exceed the statistical selection criteria. Hayes & Wilson used the Observed/Expected (O/E) statistic to identify the most promising factors. The O/E statistic calculates the likelihood of a sequence of A and B given the independent probabilities of A and B. This is intended to distinguish phonologically meaningful underattestation from accidental gaps based on the frequency of A or B. Wilson and Obdeyn (2009) note that the O/E metric in the UCLA Phonotactic Learner is position independent which can overestimate negative weights when either A or B is positionally restricted.

4.2 A Simple Artificial Phonotactic Pattern

To demonstrate the behavior of BUFIA and compare to the UCLA Learner, let us consider a very simplified phonological system. This system contains five vowels, parameterized by three distinct features, with binary ± values. This situation is summarized in table 4.1

With this simplified vowel system, we may consider all strings of length 2 which can be created with this system.

	i	u	e	o	a
high	+	+	−	−	−
back	−	+	−	+	−
low	−	−	−	−	+

Table 4.1: A Simple Vowel System

ii	iu	ie	io	ia
ui	uu	ue	uo	ua
ei	eu	ee	eo	ea
oi	ou	oe	oo	oa
ai	ao	ae	ao	aa

Table 4.2: Strings of Length 2

Let us imagine a learning presentation where a learner receives a very small subset of these possible strings as data — say, only two strings aa and ii. Which feature-based forbidden factors will the naive BUFIA algorithm project, and how does it compare to the UCLA Phonotactic Learner trained using the same feature system and training data?

In this case, the factors depend on the model signature the data will be parsed into and the `NextSupFact` function that BUFIA uses to posit new factors. In this case we are considering the successor model signature, and the `NextSupFact` function works as follows. At each iteration, BUFIA may add an empty feature bundled segment before or after the current one, or it may add another feature to an existing feature bundle. if an additional feature is added, it must respect the orders of the existing features. In our case this order is alphabetical $a < b < c$, and BUFIA must add only strictly greater features. The output of BUFIA and the UCLA learner is shown in Table 4.3.

One striking effect that can be observed about BUFIA and the UCLA MaxEnt Learner is that they both select the same forbidden factors. In this case, there are four forbidden factors that are sufficient to describe the surface true patterns in the data as given. However, BUFIA's behavior is qualitatively different. It also selects six other forbidden factors as describing the data. Closer inspection of these additional factors reveals several facts with respect to the data. The first is that all of the factors are pairwise incomparable, meaning that none of them are superfactors of the other. In this sense, each of these factors are maximally general, and BUFIA will collect them all.

BUFIA	UCLA Learner	MaxEnt Score
[+ back]	[+ back]	6.186
[-low, -high]	[-low, -high]	2.162
[-low][-high]	[-low][-high]	5.766
[-high] [-low]	[-high] [-low]	5.766
[+ low][+ high]		
[+high][−high]		
[−high][+high]		
[+low][+high]		
[+low][−low]		
[−low][−low]		

Table 4.3: Comparison of BUFIA and UCLA Phonotactic Learner

The second fact is that BUFIA has no reason to privilege any of these factors over the others. In fact there is no need: they all cover the data equally well.

Here it is easy to see in practice the behavior described in the previous chapter. Because BUFIA has no way to discriminate between forbidden factors which are maximally general yet equally describe the data, it will collect *all* such factors into the grammar it outputs. The grammar inevitably contains redundancy. Compare this to the behavior of the UCLA MaxEnt learner. It selects four forbidden factors which describe the data, yet does not add the others to the list. In a sense, there is no need, since those four will perform the function. As Hayes and Wilson (2008) report, the grammar is stochastic, so multiple runs are likely to introduce some variation into the grammar that is learned. However, the grammar will never collect *all* of the equivalent factors. This introduces a question: what is causing the MaxEnt grammar to decide on these factors and not others?

To unpack this question a bit further, we may ask about the behavior of BUFIA when it is augmented with Constraint 4, which, recall from the previous chapter, prohibited the situation where the superfactors of two factors A & B contained the same k factors. In short under Constraint 4, BUFIA is required to select factors which add something new. The behavior of BUFIA with Constraint 4 (see section 56) is summarized in Table 4.4. Again, several consequences emerge. The first is that, like the UCLA Phonotactic Learner, this variant of BUFIA outputs only 4 forbidden factors. More are unnecessary. However, several of these factors differ from those supposed by the MaxEnt learning strategy, [+high][−high] & [−high][+high]. Each of these two factors are also selected by BUFIA without Constraint 4, but this constraint privileges them over other equivalent factors.

BUFIA + Constraint 4
[+ back]
[-low, -high]
[+high][−high]
[−high][+high]

Table 4.4: Factors output by BUFIA + Constraint 4

It seems that, contrary to the UCLA Phonotactic Learner's inductive nature, that there are some abductive constraints at work pruning out certain factors in order to keep the space small. What are they? Hayes and Wilson (2008) describe several. The UCLA phonotactic Learner "shorter [factors] (fewer matrices) are treated as more general than longer ones", and "…we suggest that the value of a [factor] is proportional to the number of segments contained in its classes, and our metric sorts [factors] of a given length on this basis." BUFIA also prefers more general factors derived solely from the structure of the factor space. It also orders features to sort same-length factors but does so intrinsically as opposed to extensionally.

Since the addition of an additional abductive constraint pruned the factor space significantly, we might wonder about the behavior of BUFIA when Constraint 5 (see section 56) is added, which mandated that each factor contain no overlap in its superfactors with respect to the others. In other words, each factor added to the grammar must only account for previously accounted for data points. In this sense, we should expect that certain cfactors of equal length, or equal feature specificity, are not selected just like in Constraint 4, since there is significant potential for overlap. In fact, this is exactly what we observe when we augment BUFIA with Constraint 5, as summarized below in table 4.5. Here we see that Constraint 5 has resulted in BUFIA again allowing some of the factors present in its naive implementation, but diverging significantly from others and from those allowed by Constraint 4. In particular, while the k-value is still capped at 2, each of the 2-factors is much more featurally specific, and there are more of them. This is consistent with the expectation that the non-intersection requirement of Constraint 5 privileges "higher-up", more specific factors, which necessarily cover less space, requiring more to successfully meet the other requirements of data coverage.

BUFIA + Superfactor Non-intersection (Constraint 5)
[+ back]
[-back][−back, -high, -low]
[+low][-back, +high]
[−back, +high][+low]
[−back, -high, -low][+low]
[−back, -high, -low][−back, +high]

Table 4.5: Factors output by BUFIA + Constraint 5

The results of this toy example demonstrate the behavior of BUFIA and its variants, allowing a qualitative comparison to statistical inductive factor inference algorithms like the UCLA Phonotactic Learner. The next two section will consider two more naturalistic phonotactic patterns.

4.3 English Onsets

The inventory of syllable onsets in English is an ideal empirical domain for the testing of phonotactic learning models. The basic generalizations have been extensively studied (Bloomfield 1933,Whorf 1940, O'Connor and Trim 1953, Fudge 1969, Selkirk 1982, Clements and Keyser 1983,Hammond 1999), and available experimental data showcase similar behavior. Hayes and Wilson (2008) created a corpus of English onsets by mining word-initial onsets from the online CMU Pronouncing Dictionary (http://www.speech.cs.cmu.edu/). They additionally removed onsets they judged "exotic". As the UCLA learner requires a valid feature matrix, they created the following, reproduced below in Table 4.6.

	B	CH	D	DH	F	G	JH	K	P	S	SH	T	TH	V	Z	ZH	HH	M	N	NG	L	R	W	Y	#
consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
approximant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-
sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-
continuant	-	-	-	+	+	-	-	-	-	+	+	-	+	+	+	+	+	0	0	0	0	0	0	0	-
nasal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	0	0	0	0	-
voice	+	-	+	+	-	+	+	-	-	-	-	-	-	+	+	+	-	0	0	0	0	0	0	0	-
spread	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	-
labial	+	0	0	0	+	0	0	0	+	0	0	0	0	+	0	0	0	+	0	0	0	0	+	0	-
coronal	0	+	+	+	0	0	+	0	0	+	+	+	+	0	+	+	0	0	+	0	+	+	0	0	-
anterior	0	-	+	+	0	0	-	0	0	+	-	+	+	0	+	-	0	0	+	0	+	-	0	0	-
strident	0	+	-	-	0	0	+	0	0	+	+	-	-	0	+	+	0	0	-	0	-	-	0	0	-
lateral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	-
dorsal	0	0	0	0	0	+	0	+	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	-
high	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	-
back	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	-	-
boundary	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Table 4.6: Feature chart for English consonants

They then ran their MaxEnt Phonotactic learner over the corpus. A noticeable difference between the corpus as presented below for these simulations and the one used by Hayes & Wilson is that the one here refers only to the types, rather than the tokens of each form. That is, here there is only information about whether a form occurred at all, and not the frequency with which it occurred, simply because this information is irrelevant for the purposes we are considering here.

Readers are referred to Hayes and Wilson (2008) for further information. The complete types of the English onset corpus are as follows.

English onset learning data

K, R, D, S, M, P, B, L, F, HH, T, PR, W, N, V, G, JH, ST, TR, KR, SH, GR, CH, BR, SP, FL, KL, SK, Y, FR, PL, BL, SL, DR, KW, STR, TH, SW, GL, HHW, SN, SKR, Z, SM, THR, SKW, TW, SPR, SHR, SPL, DH, DW, GW, THW, SKL

Hayes & Wilson report the following 23 forbidden factors, which are repeated for reference below. For clarity, their factors notation is preserved for this list, which allow exactly one of the matrices of a factor to be modified by the complementation operator ∧. In this way, $[\wedge\alpha F, \beta G, \ldots]$ refers to any segment not a member of the natural class $[\alpha F, \beta G, \ldots]$. For example, a forbidden factor limiting prenasal segments to [s] would be formulated as *[∧−voice,+ant,+strid][+nas].

UCLA Phonotactic Grammar (from Hayes and Wilson (2008))

[+son, +dors]

[+cont,+voice,-ant]

[∧ -voice,+ant,+strid][-approx]

[][+cont]

[][+cont]

[][+voice]

[+son][]

[-strid][+cons]

[][+strid]

[+lab][∧+approx,+cor]

[-ant][∧+approx,-ant]

[+cont,+voice][ ]

[-cont,-ant][ ]

[][-back]

[+ant,+strid][-ant]

[+spread][∧+back]

[+cont,+voice,+cor]

[+voice][∧+approx,+cor]

[+cont,-strid][∧+approx,-ant]

[][∧-cont,-voice,+lab]

[][+cor][∧+approx,-ant]

[+cont,-ant]

[+strid][-ant]

[-cont,-voice,cor][∧+approx,-ant]

We may ask how the vanilla BUFIA algorithm behaves over this English onset corpus, using the same successor model signature and `NextSupFact` function as before. The results, as in the toy example, behave as expected. The vanilla BUFIA algorithm finds 469 forbidden factors with a k value capped at 2, under the feature-based successor model signature. For ease of exposition, the full list of forbidden factors can be found in the Appendix. Again, what is worth noting is not whether BUFIA finds correct, general factors; it does. Each factor is surface true, and represents a structure which occurs zero times in the corpus. What is notable is that it finds *all* correct, general factors. This is due to the significant possibility of overlap, just as in the previous example.

Can adding the abductive constraints mitigate this possibility somewhat? Again, as in the toy example, the answer is yes. Augmenting BUFIA with Constraint 4 (see section. 3.5) dramatically reduces the output grammar size, from 469 factors to 32 factors. The full set of these factors is listed below. Compare this with the UCLA MaxEnt learner's 23 factors. In the previous toy example the grammar sizes exactly matched, albeit with some minor differences in the substance of the factors. Here the size is slightly larger, purely due to the nature of the abductive principles at play in both the UCLA learner and BUFIA. However, unlike the UCLA learner, BUFIA does not operate stochastically. It will return these same 32 factors every time the learner is run.

BUFIA + Constraint 4 Grammar, Successor Model, $k = 2$

[+anterior][+continuant]

[+anterior][+voice]

[+anterior][+approximant, -back]

[+approximant][+anterior]

[+approximant][+approximant]

[+approximant][+consonantal]

[+boundary][+boundary]

[+consonantal][+continuant]

[+consonantal][+voice]

[+consonantal][+approximant, -back]

[+dorsal][+dorsal]

[+dorsal][+nasal]

[+dorsal][+anterior, -approximant]

[+dorsal][-approximant, +consonantal]

[+labial][+back]

[+labial][+dorsal]

[+labial][+labial]

[+labial][+nasal]

[+labial][+anterior, -approximant]

[+nasal][+anterior]

[+nasal][+approximant]

[+nasal][+consonantal]

[+voice][+dorsal]

[+voice][+nasal]

[+voice][+anterior, -approximant]

[+voice][-approximant, +consonantal]

[+anterior, -continuant][+anterior]

[+anterior, -continuant][+consonantal]

[+anterior, +strident][-anterior, +approximant]	[+continuant, +voice][+anterior]
[+anterior, +voice][+anterior]	[+continuant, +voice][+approximant]

4.4 Quechua

Descriptively, South Bolivian Quechua (henceforth Quechua; (Bills et al., 1969; Laime Ajacopa, 1992; Gallagher, 2016) has three phonemic vowels /i u a/ with allophonic lowering of /i u/ to [eo] in the vicinity of uvulars /q q^h q'/. Mid vowels occur immediately following or preceding a uvular, or preceding a uvular across an intervening coda. High vowels occur in all other consonantal environments.

(1) Uvular contexts:[e o] *[i u]

a. q'epij (*q'ipij) 'to carry'

b. q'oɲi (*q'uɲi) 'hot'

c. noqa (*nuqa) 'I'

d. wesq'aj (*wisq'aj) 'to close'

e. peqaj (*piqaj) 'to grind'

f. toLqa (*tuLqa) 'son-in-law'

(2) Elsewhere:[i u] *[e o]

a. misi (*mese) 'cat'

b. kuʎku (*koʎko) 'type of bird'

As Wilson and Gallagher (2018) state, a traditional analysis of this pattern would assign the high vowels elsewhere status and specify the contexts in which mid vowels occur. The distribution of high vowels is handled by a pair of forbidden factors like *QI or *IQ, where Q is a uvular consonant and I is a high vowel. However, for the mid-vowels, the distribution requires multiple forbidden factors. There must be several surface restrictions against mid vowels in the exhaustive set of "nonuvular" environments. Wilson and Gallagher (2018) give an analysis where restrictions are again stated in terms of a dorsal tier containing velar and uvular consonants and all vowels.

They use this tier projection to deal with segmentally non local interactions between uvular and non-high vowels across an intervening coda (e.g.,[orqo] ‘mountain’). They state the following posited factors, following their notation (Q = uvular, K = velar, I = high, E = mid, V = any vowel, C = any consonant).

(3) Surface-true factors on high vowels in uvular contexts

a. *QI, *IQ

(4) Surface-true factors on mid vowels in “nonuvular”contexts*

a. *#EK, *KEK *VEK, *#E#, *KE#, *VE#, *#EV, *KEV, *VEV

(5) Surface true factors on dorsal consonants

a. *K. . . Q, *Q. . . K

(6) Surface true factors on Laryngeal cooccurrence restrictions

a. *VV, *CCC, *wu, *wo

The main issue for Wilson and Gallagher (2018) is the distinction between systematic and accidental gaps. The unattested legal 3-grams typically contain rare segments. They note, as an example, that the Quechua sequence [eq^{h} o] is legal but unattested (accidental), reflecting the fact that aspirated dorsals are infrequent in medial position generally and that [e] is the least common surface vowel in the language. However, not all unattested sequences containing rare parts are accidental gaps. For example, [k^{h}] is among the rarest segments in Quechua, and the sequence [k^{h} ek] has zero frequency like [eq^{h} o], but in this case the gap is systematic.

For this reason, as mentioned at the beginning of Chapter 3, they advocate for additional statistical methods to distinguish between such factors, and to prune out factors which will reify accidental gaps. Their method is to preserve the inductive character of the UCLA Phonotactic Learner, but to add an additional statistical “gain" criterion (Della Pietra et al., 1997). Della Pietra et al. characterize gain as "the improvement [a factor] brings to the model when it has weight $[\text{w}]''$.

Definition 9. *The* Gain *of a factor* $\text{Gain}_G(w, F) = D(\tilde{p}|G) - D\left(\tilde{p}|G_{wF}\right)$*, where F is the forbidden factor with the weight w, D is the Kullback-Leibler divergence, $\tilde{p}$ is the probability distribution of the data, and G is the current grammar.*

Della Pietra et al. explain the reason for this method of calculating gain intuitively as follows (replacements with respect to BUFIA added in brackets):

> We approximate the improvement due to adding a single candidate [factor], measured by the reduction in Kullback-Leibler divergence, by adjusting only the weight of the [factors] and keeping all of the other parameters of the [grammar] fixed. In general this is only an estimate, since it may well be that adding a [factor] will require significant adjustments to all of the parameters in the new model. From a computational perspective, approximating the improvement in this way can enable the simultaneous evaluation of thousands of candidate [factors], and makes the algorithm practical.
>
> (Della Pietra et al., 1997, pg. 4)

By adding this criterion to the UCLA MaxEnt learner, Wilson and Gallagher (2018) are able to collect forbidden factors which account for the data robustly, and from this they conclude that statistical inductive generalization is a necessary and sufficient ingredient for this problem. We may then ask how BUFIA behaves on this similar dataset. We use the same corpus from Wilson and Gallagher. This is a lexicon of 2400 words from the 1,104 actual roots compiled from the Laime Ajacopa (2007) dictionary, 1,104 of which are confirmed by a native speaker. The feature matrix used for the data is given below in Table 4.7

	p	t	tʃ	k	q	p’	t’	tʃ’	k’	q’	p^h	t^h	$tʃ^h$	k^h	q^h	s	ʃ	x	h	m	n	ŋ	r	l	ʎ	j	w	a	i	u	o	e	#
syllabic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-
sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
continuant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	+	+	+	+	+	+	+	+	+	+	-
cg	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	-
sg	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	-
lateral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	+	+	-	-	0	0	0	0	0	-
nasal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	-	-	-	-	-	0	0	0	0	0	-
labial	+	0	0	0	0	+	0	0	0	0	+	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	-
coronal	0	+	+	0	0	0	+	+	0	0	0	+	+	0	0	+	+	0	0	0	+	+	+	+	+	0	0	0	0	0	0	0	-
anterior	0	+	-	0	0	0	+	-	0	0	0	+	-	0	0	+	-	0	0	0	+	-	+	+	-	0	0	0	0	0	0	0	-
dorsal	0	0	0	+	+	0	0	0	+	+	0	0	0	+	+	0	0	+	0	0	0	0	0	0	0	0	0	+	+	+	+	+	-
RTR	-	-	-	-	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-
low	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	+	-	-	-	-	-
back	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	+	+	-	+	+	-	-
mb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Table 4.7: Feature chart for Quechua segments. ±cg means constricted glottis, ±sg means spread glottis, ±RTR means retracted tongue root, and ±mb means morpheme boundary

As before, we may ask how the vanilla BUFIA algorithm fares with respect to the corpus. In this case, both lcal and non-local factor information may be relevant, so it is meaningful to understand

BUFIA's behavior using multiple model signatures, in this case the feature-based successor model and the feature-based precedence model. In both cases, we set the maximum k-value of a forbidden factor to 2, which has the effect of ensuring that the accidental gap situation does not obtain, as well as eliminating the possibility of reifying an accidental gap by simply memorizing forms. The output feature-based successor grammar contains 1913 forbidden factors, while the feature-based precedence model contains 320 forbidden factors. The full output grammars are reported in the appendix.

In both cases, as with the toy example and the English onset case, the output grammars are prohibitively large. However, just as in the previous cases, this is entirely due to the fact that they find multiple redundant, pairwise incomparable, but equally correct, forbidden factors. To emulate the behavior of the gain criterion used by Wilson and Gallagher (2018), we additionally ran simulations using BUFIA augmented with Constraint 4 (see section. 3.5), which required a measure of novel information when adding an additional factor. The results, just as in the previous cases, drastically reduced the number of factors: $1913 \rightarrow 89$ for the successor-based model, and $320 \rightarrow 32$ for the precedence-based model. The full list of forms for both grammars are reported below.

BUFIA + Constraint 4 Grammar, Precedence Model, $k = 2$

[-coronal]

[+anterior, +lateral]

[-anterior, +continuant, -sonorant]

[+RTR][-RTR, +sg]

[+anterior][-cg, +continuant]

[-cg][-cg, +continuant]

[+RTR, +back][+continuant,+dorsal,-sonorant]

[+RTR,-back][+continuant,+dorsal,-sonorant]

[+anterior, +nasal][-back, -lateral]

[+anterior,+nasal][+continuant,+dorsal,-sonorant]

[-anterior, -lateral][+lateral]

[-anterior, -lateral][-anterior, -cg]

[-anterior, -lateral][-cg, +continuant]

[-back, -lateral][-cg, +continuant]

[+continuant, -sonorant][-cg, +continuant]

[-RTR, -back, +dorsal][+RTR, +back]

[-RTR, -back, +dorsal][+RTR, -back]

[-RTR, -cg, +dorsal][+RTR, -back]

[-RTR, +dorsal, -sonorant][+RTR, -back]

[+anterior, -cg, -sg][-anterior, -lateral]

[+continuant,+dorsal,-sonorant][+RTR,+back]

[+continuant,+dorsal,-sonorant][+continuant,+dorsal,-sonora

BUFIA + Constraint 4 Grammar, Successor Model, $k = 2$

[+anterior, +lateral]
[-anterior, +continuant, -sonorant]
[+RTR][-RTR, +sg]
[+RTR][+continuant, +dorsal, -sonorant]
[+anterior][-cg, +continuant]
[+anterior][+continuant, +dorsal, -sonorant]
[-anterior][-coronal]
[-anterior][+continuant, +dorsal, -sonorant]
[+back][+RTR, +back]
[-back][+RTR, -back]
[-cg][-coronal]
[-cg][+continuant, +dorsal, -sonorant]
[-continuant][+continuant, +dorsal, -sonorant]
[+coronal][-cg, +continuant]
[+labial][-coronal]
[+labial][+anterior, -cg]
[+labial][+anterior, +nasal]
[+labial][+back, -lateral]
[+labial][-cg, +continuant]
[+labial][+labial, -lateral]
[+lateral][+lateral]
[+lateral][+anterior, -lateral]
[+lateral][-anterior, -lateral]
[+lateral][-back, -lateral]
[-lateral][+continuant, +dorsal, -sonorant]
[+low][+low]
[+low][+syllabic]
[+nasal][+anterior, +nasal]
[+nasal][+labial, -lateral]
[+sg][+anterior]
[+sg][-anterior]
[+sg][-cg]
[+sg][-continuant]
[+sg][-lateral]
[-sonorant][+continuant, +dorsal, -sonorant]
[+syllabic][+low]
[+syllabic][+syllabic]
[+RTR, -back][-RTR, +dorsal]
[+RTR, -back][-anterior, +nasal]
[+RTR, -cg][+anterior]
[+RTR, -cg][-anterior]
[+RTR, -cg][-cg]
[+RTR, -cg][-continuant]
[+RTR, -cg][-lateral]
[+RTR, -cg][-RTR, +back, -low]
[-RTR, +dorsal][+RTR, -back]
[+anterior, -cg][+anterior]
[+anterior, -cg][-anterior]
[+anterior, -cg][-cg]
[+anterior, -cg][-nasal]
[+anterior, +continuant][-coronal]
[+anterior, -continuant][-anterior, +nasal]
[+anterior, -continuant][-cg, +labial]
[+anterior,-continuant][-back,+continuant,-lateral]

[+anterior, -nasal][+anterior, -nasal]
[+anterior, -sonorant][+continuant, -sonorant]
[-anterior, +coronal][+lateral]
[-anterior, +coronal][+anterior, -lateral]
[-anterior, +coronal][-anterior, -lateral]
[-anterior, +coronal][-back, -lateral]
[-anterior, +nasal][+anterior]
[-anterior, +nasal][-anterior]
[-anterior, +nasal][-cg]
[-anterior, +nasal][-lateral]
[+back, -lateral][-anterior, +nasal]
[+back, -lateral][+back, -lateral]
[+back, -lateral][-cg, +continuant]
[-back, +continuant][-back, +dorsal]
[-cg, +coronal][+anterior]
[-cg, +coronal][-anterior]
[-cg, +coronal][-cg]
[-cg, +coronal][-nasal]
[-cg, +dorsal][+anterior]
[-cg, +dorsal][-anterior]
[-cg, +dorsal][-cg]
[-cg, +dorsal][-continuant]
[-cg, +dorsal][-lateral]
[-cg, +labial][+anterior]
[-cg, +labial][-anterior]
[-cg, +labial][-cg]
[-cg, +labial][-lateral]
[+continuant, +dorsal][+low]
[+continuant, +dorsal][+syllabic]
[+continuant, -lateral][-cg, +continuant]
[+continuant, -sonorant][-cg, +continuant]
[-RTR, -back, +dorsal][+RTR]
[-RTR, +dorsal, -low][+RTR]
[-anterior, -cg, +coronal][-continuant]
[-back,+continuant,-lateral][-back,+continuant]

Taken together, the grammar resulting from the application of BUFIA with Constraint 4 brings us significantly closer to the generalizations mentioned by Wilson and Gallagher (2018). The reduction in factors given by just this less-restrictive abductive principle appears to filter out an enormous amount of the redundancy in the grammar. In a sense, the results of this case study can be taken to mean that both the gain criterion and our Constraint 4 play a very similar role, albeit under different frameworks. Constraint 4 as stated is clearly an abductive constraint. However, the gain criterion also plays the role of an abductive constraint, albeit within a statistical formulation. Looking at the form of the gain criterion allows one to see that it is performing an extremely similar function as our Constraint 4, namely pruning out factors based on their informativity with respect to the forbidden structures the grammar wishes to entertain.

Another striking fact that emerges when comparing the naive BUFIA runs to the runs with BUFIA+Constraint 4 is that the factors that the algorithm chooses to output are quite different, and rather surprising with respect to the analysis given by Wilson & Gallagher. Yet, with respect to the corpus, they are surface true. How can this be? The obvious answer is that there is a combination of factors that are taking the place of some other combination of factors, just as in the previous case studies. This is almost certainly a direct result of certain features being entertained before other ones. If a particular form is ruled out by both the feature [-Back] and the feature [+High], but [-Back] is entertained first, then there is no reason for the algorithm to entertain [+High] unless that accounts for some novel information in the data.

This presents an intriguing advantage of BUFIA, and of the constraint-satisfaction approach to learning given by the abductive inference paradigm. The naive BUFIA collects all the relevant factors, barring those that are direct superfactors of another factor. In this sense, it serves as a kind of baseline for an abductive phonotactic learner. Each additional abductive constraint controls the types of grammars BUFIA is allowed to entertain, in a structurally transparent way. This is because BUFIA's judgments for adding forbidden factors to the grammar are purely structural. They depend on the structure of factor space, due to the ideals given by the representations themselves. This means that further abductive principles, of which there are likely to be many, will also be structure-dependent, casting a different but important light on the contribution of structural inference, which is exactly the characteristic of abductive inference.

The extensions of these results are immediate. The key problem of course is to understand the contributions of BUFIA on various sets of linguistic data. The key separation here is between the method of collecting the structures, a variant of parsing, and the method of inferring the right grammar over those structures. For the string-based modes we have considered here, this parsing problem is straightforward. For other structures, like autosegmental graphs, the method generally lifts (Jardine and Heinz, 2015). For trees, the primary representation underlying syntactic structure, the parsing is not so trivial, since almost the entirety of the structure is hidden or latent.

Additionally, the behavior of BUFIA should be checked against grammatical behavior whenever possible. For example, every factor output by BUFIA over string-based models has a corresponding interpretation as a finite-state automaton. The performance of the finite-state machine may be checked against the behavior of, say, human subjects rating the acceptability of certain forms, or nonce forms derived from the “correct" grammar for a particular pattern.

The main lesson from this discussion is not that there is a clear winner to be found from statistical vs structural inference, unlike what Wilson and Gallagher (2018) claim. The fact that phonotactic learning can succeed purely on the basis of structural inference is striking, because it means that in a variety of learning scenarios, statistical generalization is not necessary. What then can we take away? The lesson is instead that statistical inference and structural inference have their own roles to play. The case studies above showed that the constraints we considered for BUFIA, and the statistical heuristics considered for Maximum Entropy learners, are each a type of abductive inference. We have shown many aspects of the MaxEnt learner, and its results are driven by structure, not statistics. The problem of induction looms large, as all parties agree. Abduction, instead, subverts the inductive problem, since each of these constraints constitute a form of abductive inference.

As a final, more general point, much of grammatical inference, particularly in linguistic learning, is taken to be a problem of induction, as laid out in Chapter 1. However, the results presented here provide a good case for treating problems of grammatical inference instead as abduction. This has several advantages. First, it shifts the problem to a constraint-satisfaction paradigm, which is useful both for the analyst, and for the cognitive scientist considering the learning mode. A learning paradigm, and its metrics of success, are dependent on how the constraint satisfaction problem is defined. While many grammatical inference paradigms define their success based on finding a single, canonical grammar, it may be the case that there are many, and the constraint-satisfaction perspective allows this to become a feature, not a bug.

Another advantage is that abduction brings to the fore the key issue in linguistic inference. Not only does a learner generalize, but they generalize in a highly structured way. The paradigm for which BUFIA is a result depends entirely on structure. Structure exists in the representations, how the representations are organized into a space, how an algorithm entertains the structures, and how an algorithm decides to attend to or ignore the structures. This sharply diverges from many inductive approaches, where structure is added only when necessary, and the algorithm is refined more and more as an “ideal guesser". Abduction, and inference to the best explanation more generally, rejects this notion, and shows exactly whether and how structure matters. In this case, it seems, structure matters quite a lot.

Chapter 5

Distributed Representations of Subregular Constraints

The previous chapters used the model-theoretic view of linguistic representations to derive a notion of grammatical structure that made it ideally suited for a variant of abductive learning problem. Constraints over model-theoretic structures, as well as translation between one structure and another, can be described using statements in mathematical logic. Such constraints and transformations express in an elegant way the relationship between grammars and representation. In particular, various types of logic over specific model-theoretic representations (say, strings and trees) yield particular classes of grammars (Rogers, 1996).

Following the introduction, it is important to understand how these structures may be realized cognitively. Here we will take a somewhat strong position, and attempt to find a mathematical translation from the symbolic notion of a model theoretic structure, and constraints over those structures, to data structures and a computational architecture more closely matching the level of neural operation. As Poeppel (2012) puts it, “commitment to an algorithm or computation in this domain commits one to representations of one form or another with increasing specificity and also provides clear constraints for what the neural circuitry must accomplish".

This chapter will demonstrate how model-theoretic structures, and constraints on them corresponding to grammars, may be characterized geometrically. Geometric approaches to language and symbolic cognition in general have become increasingly popular during the last two decades. There is work dealing with conceptual spaces for sensory representations (Gardenfors, 2004), multilinear representations for compositional semantics (Blutner, 2009; Aerts, 2009), and dynamical systems for modeling language processes (Beim Graben et al., 2008; Tabor, 2009).

In this vein, Smolensky and Legendre (2006b,a) propose the Integrated Connectionist-Symbolic (ICS) architecture. The crucial innovation is tensor networks, or tensorial computation. Tensorial computation is isomorphic to key aspects of symbolic computation. Tensorial computation is also parallel, distributed connectionist computation, plausibly isomorphic to types of neural computation. The result is a computational theory that reduces symbolic computation to connectionist computation: it provides a formal realization mapping from one to the other.

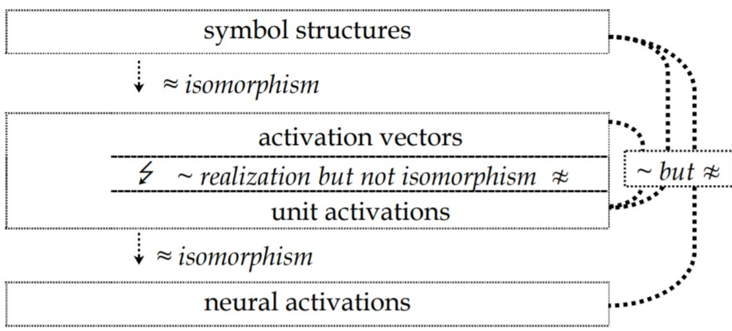


Figure 5.1: Interlevel Relations in ICS (adapted from (Smolensky and Legendre, 2006a, p.)

The central structure in the ICS architecure are Tensor Product Representations (Smolensky, 1990). Here, subsymbolic dynamics of neural activation patterns in a vector space description become interpreted as symbolic cognitive computations at a higher-level description by means of “filler/role" bindings via tensor products. These tensor product representations form the symbolic foundation of Harmonic Grammar and Optimality Theory, and have been successfully employed for phonological and syntactic computations. (Smolensky and Legendre, 2006b).

Tensors can be considered as a multidimensional array. More generally, an n-th-order tensor is an element of the tensor product of n vector spaces, each of which has its own coordinate system[1]. Tensors and their decompositions originally appeared in (Hitchcock, 1927), but did not play a large role in computer science until the late 20th century (Sidiropoulos et al., 2017). Increasing

[1]This notion of tensors is not to be confused with tensors in physics and engineering (such as stress tensors), which are generally referred to as tensor fields in mathematics

computing capacity and a better understanding of multilinear algebra, especially during the last decade, allowed tensors to be applied to other domains, like machine learning, signal processing, computer vision, numerical analysis, data mining, graph analysis, neuroscience, and more (Kolda and Bader, 2009).

In tensor product representations, symbolic structures are decomposed into structural roles and fillers, bound together using the tensor product. For example, strings can be decomposed into a tensor realizing string positions, each of which is bound to a tensor realizing different symbols in some alphabet. Similarly, tree structures can be represented recursively, with a tensor product representation using tree node position as the structural role, and using an entire subtree as an alphabet filler symbol. Hale and Smolensky (2001) use this tree representation to describe a Harmonic Grammar for context-free languages, and Beim Graben and Gerth (2012) use them to formalize Stabler (1997)'s Minimalist Grammars.

While tensor product representations form a powerful method for geometrically interpreting symbolic structures, explicitly modeling the relational structures given by model-theory, as well as the logical constraints characteristic of subregular languages, is an open issue. This paper provides this connection, by translating model-theoretic structures into vector spaces and describing logical grammatical constraints over them using tensors.

There has been some recent work on embedding logical calculi using tensors. Grefenstette (2013) introduces tensor-based predicate calculus that realizes logical operations. Yang et al. (2014) introduce a method of mining Horn clauses from relational facts represented in a vector space. Serafini and Garcez (2016) introduce logic tensor networks that integrate logical deductive reasoning and data-driven relational learning. Sato (2017) formalizes Tarskian semantics of first-order logic in vector spaces. Here we apply Sato's method for translating model-theoretic representations and first-order logic into tensors.

Finally, tensor methods and subregular grammars/automata have been used to evaluate and interpret neural networks (see (Rabanser et al., 2017)). Avcu et al. (2017) tested the generalization capacity of LSTM networks on the Strictly Local and Strictly Piecewise languages. McCoy et al. (2018) showed that recurrent neural networks (RNNs) implicitly encode tensor product representations, and Weiss et al. (2017) used regular languages to test the generalization capacity of RNNs. Explicit translation of subregular languages into tensors over various representations thus allows model-theoretic linguistics to study neural nets in a principled way.

5.1 Tensor Product Representations

Tensors are the mathematical objects dealt with in multilinear algebra just as vectors and matrices are the objects dealt with in linear algebra. In fact, tensors can be seen as generalisations of vectors and matrices by introducing the notion of tensor rank or order. Let the rank of a tensor be the number of indices required to describe a vector/matrix-like object in sum notation. We can view familiar objects like vectors and matrices this way

- Order 1 — vector: $\vec{v} \in A = \sum_i C_i^v \overrightarrow{a_i}$
- Order 2 — matrix: $M \in A \otimes B = \sum_{ij} C_{ij}^M \overrightarrow{a_i} \otimes \overrightarrow{b_j}$
- Order 3 — Cuboid: $R \in A \otimes B \otimes C = \sum_{ijk} C_{ijk}^R \overrightarrow{a_i} \otimes \overrightarrow{b_j} \otimes \overrightarrow{c_k}$

Smolensky (1990) proposes a decomposition of the notion of “structure" as a set of bindings of various structural roles to their alphabetic fillers. Structural roles may be very general, as may fillers. For example, a set of structural roles may be string positions, or tree positions. The class of a structure is determined by its roles; the set of length-2 strings Σ^2 is determined, for example, by the positional roles r_1, r_2. These roles are variables, or slots, or attributes, which must be bound to particular values in order to individuate a particular structure within the general class (e.g., the particular string AB) The set of fillers may be symbols from an alphabet Σ, or even an entire structure, if the structure is to be recursive. Smolensky's crucial definition is that this filler/role binding is isomorphic to the notion of tensor product of vector spaces.

Definition 10 (Tensor Product Binding)**.** *The binding f/r of a filler f to a role r is realized as a vector* $\mathbf{f}/\mathbf{r}$ *that is the tensor product of a vector* $\mathbf{f}$ *realizing f with a vector* $\mathbf{r}$ *realizing r :* $\mathbf{f}/\mathbf{r} = \mathbf{f} \otimes \mathbf{r}$

The tensor product binding allows us to see the internal structure of a binding of a symbol A in the string 'AB' as A_1: it is the tensor product of the vector $\mathbf{A}$ and the vector $r_1 = (1, 0)$, which realizes the first-element role r_1. This tensor product binding is realized realized explicitly in

Figure 5.2. Here the Vector $\mathbf{A} = (1, 0, ..., 0, 0)$ and the vector $\mathbf{r_1} = (1, 0)$ produce an order two tensor (matrix) which realizes a 1 in exactly the cell that corresponds to position 1 and symbol A.

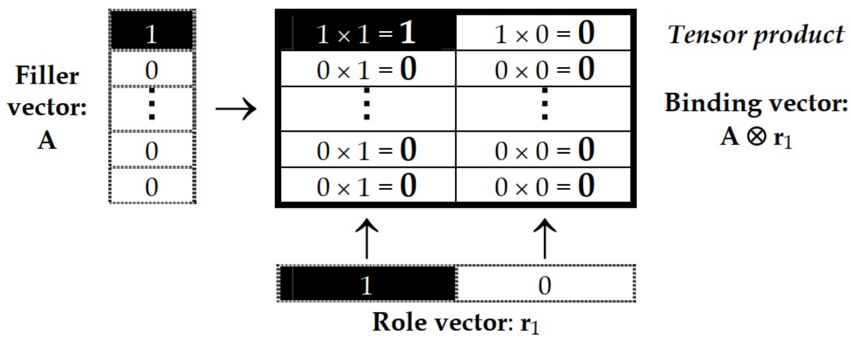


Figure 5.2: Tensor Product Binding (adapted from (Smolensky and Legendre, 2006b)

The explicit tensor product binding means that symbolic structures consisting of tensor products of sets of symbols as vectors and sets of roles as vectors may combine into sets of filler/role bindings as vectors (tensors). Smolensky shows that one may represent specific structures by means of superposition of vectors, realized as sums. So a representation of a string like 'AB' is the superposition of the filler/role bindings for A_1 and B_2. This is explicitly noted in Figure 5.3. Here, a notation similar to that of neural networks is used. Here a filled circle is a value of 1 or active and an empty circle is a value of 0, or inactive.

The contribution of **A** to the vector **AB** that realizes the string 'AB' is distinguishable from its contribution to the vector **BA** that realizes 'BA'. The simplest example of such a difference is shown in the contrast between the first lines of panels A and A of Figure 5.3 (the rectangle enclosing two groups of five units each). There are two horizontally separated groups of units, one hosting the realization of the first symbol in the string, the other, the second. In each group, there is a single unit dedicated to the realization of each type of symbol: the first for realizing A, the second for B, and so on. In terms of activation vectors, the activity pattern over the first group of units is (1, 0, 0, 0, 0) = **A** for A, (0, 1, 0, 0, 0) = **B** for B, and so on. The vector **AB** for 'AB' is now expressable as the superposition of the **A** and **B**, shown below

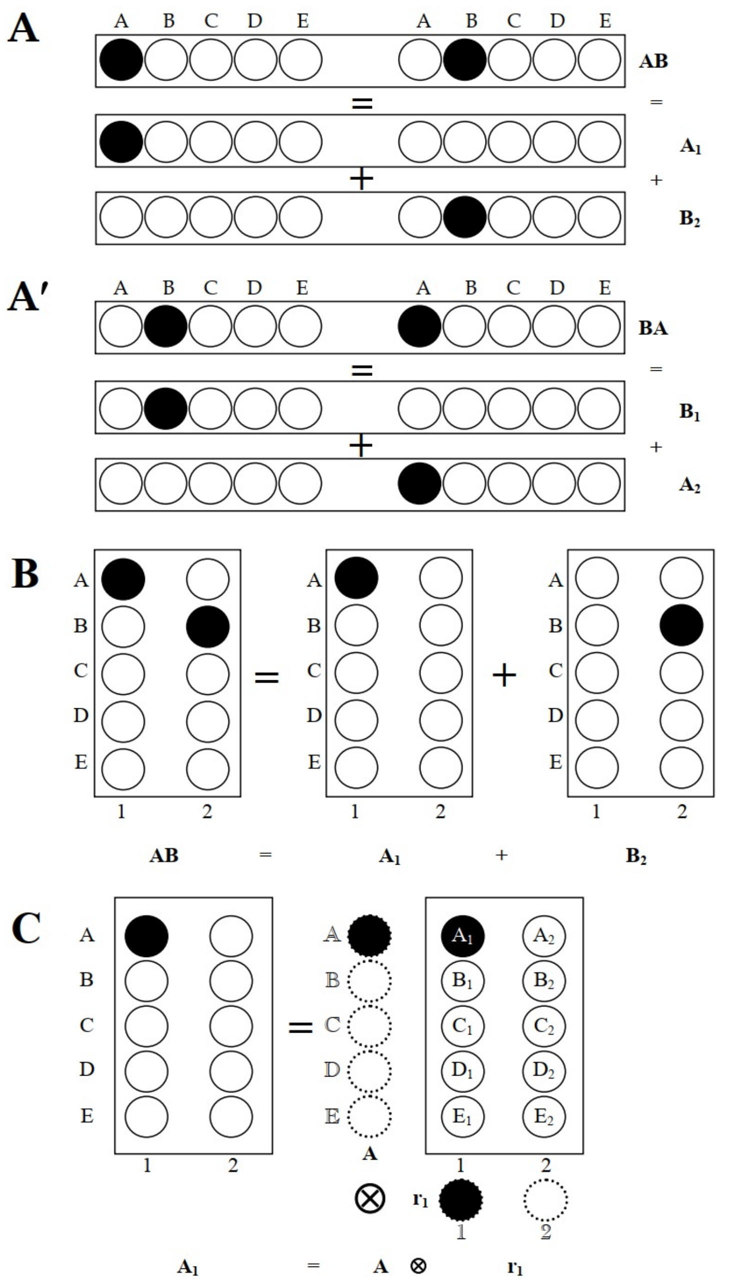


Figure 5.3: Local realization of Symbol Strings (adapted from (Smolensky and Legendre, 2006b, p.167)

$$
\begin{aligned}
\mathbf{AB} = (\mathbf{A};\mathbf{B})& \\
&=(1,0,0,0,0;0,1,0,0,0) \\
&=(1,0,0,0,0;0,0,0,0,0)+(0,0,0,0,0;0,1,0,0,0) \\
&=(\mathbf{A};0)+(0;\mathbf{B}) \\
&=\mathbf{A_1}+\mathbf{B_2}
\end{aligned}
$$

Tensor Product Bindings may also be “semilocal", meaning that a vector is a weighted combination of its basis vectors, or a “distributed representation", as shown in Figure 5.4

While tensor product representations are a flexible way of realizing symbolic structures, the independence assumption that underlies the structural roles (namely, that the vectors must be linearly independent) is limiting. In particular, it does not allow the flexibility in relation among structural elements that characterizes the flexibility given by the model-theoretic approach. In other words, there is a structural bias present in the tensor product binding. Roles positions are absolute, and there is no notion of explicit relations among the elements of a structure, which is a core element of the model-theoretic view. Additionally, different structures have wildly different roles, which seems odd from the perspective of a factor. For example, Smolensky’s way of representing finite trees is by using recursive role vectors, where a filler vector is actually an entire tensor product binding. This chapter enriches the tensor product representation to accommodate model-theoretic structures of any data structure type, in a unified way. I now turn to a description of such model-theoretic structures.

5.2 Tensor Representations of First-Order Logic

This section describes how to embed a model domain and signature into a vector space, using tensors to encode relational information. Scalars are denoted with lower case letters like a. Vectors mean column vectors and are denoted by boldface lower case letters like **a** and **a**’s components

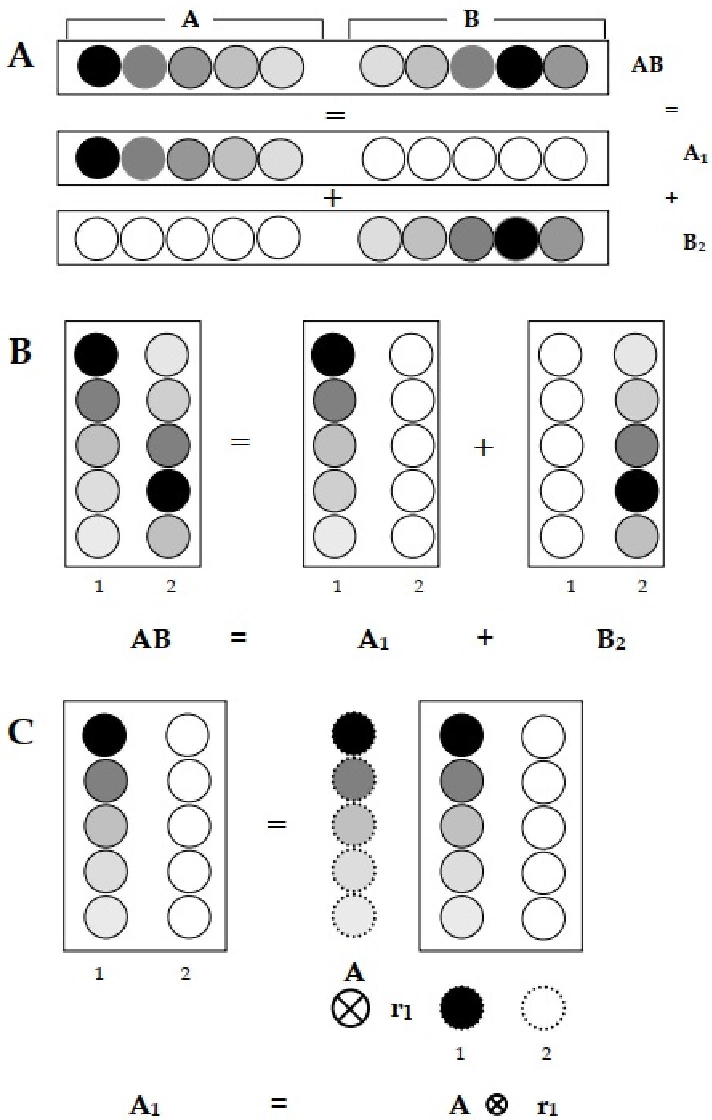


Figure 5.4: Semilocal realization of Symbol Strings (adapted from (Smolensky and Legendre, 2006b, p.169)

by a_i. $\mathcal{D}' = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is the standard basis of N -dimensional Euclidean space $\mathbb{R}^N$ where $\mathbf{e}_i = (0 \cdots, 1, \cdots, 0)^T$ is a vector that has one at the i -th position and zeros elsewhere. Such vectors are called one-hot vectors. $\mathbf{1}$ is a vector of all ones. We assume square matrices, written by boldface upper case letters like $\mathbf{A}$. $\mathbf{I}$ is an identity matrix. Order-p tensors $\mathcal{A} \in \mathbb{R}^{D^p}$, are also denoted by $\{a_{i_1,\ldots,i_p\}}\}$ $(1 \leq i_1, \ldots, i_p \leq N)$. $\mathcal{A}$'s component $a_{i_1,\ldots,i_p}$ is also written as $(\mathcal{A})_{i_1,\ldots,i_p}$. $(\mathbf{a} \bullet \mathbf{b}) = \mathbf{a}^T\mathbf{b}$ is the inner product of $\mathbf{a}$ and $\mathbf{b}$ whereas $\mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{b}^T$ is their outer product. $\mathbf{1} \circ \cdots \circ \mathbf{1}$ is a k-order tensor, and $\mathbf{1} \circ \cdots \circ \mathbf{1}\,(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}) = (\mathbf{1} \bullet \mathbf{e}_{i_1}) \cdots (\mathbf{1} \bullet \mathbf{e}_{i_k}) = 1$.

There exists an isomorphism between tensors and multilinear maps (Bourbaki, 1989), such that any curried multilinear map

$$f : V_1 \to \ldots \to V_j \to V_k$$

can be represented as a tensor $\mathcal{T}_f \in V_k \otimes V_j \otimes \ldots \otimes V_1$. This means that tensor contraction acts as function application. In particular, the isomorphism guarantees that there exists such a tensor $\mathcal{T}^f$ for every f, such for any $v_1 \in V_1, \ldots, v_j \in V_j$:

$$f\mathbf{v}_1 \ldots \mathbf{v}_j = \mathbf{v}_k = \mathcal{T}^f \times \mathbf{v}_1 \times \ldots \times \mathbf{v}_j \tag{5.1}$$

We first isomorphically map a model M to a model M' in $\mathbb{R}^N$. We map entities $e_i \in D$ to one-hot vectors $\mathbf{e_i}$.

So D is mapped to $D' = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$, the basis of $\mathbb{R}^N$. We next map a k-ary relation r in M to a k-ary relation r' over D' which is computed by an order-k tensor $\mathcal{R} = \{r_{i_1,\ldots,i_k}\}$, whose truth value $[\![r(e_{i_1}, \ldots, e_{i_k})]\!]$ in M is given by

$$\begin{aligned}
&[\![r(e_{i_1}, \ldots, e_{i_k})]\!] \\
&= \mathcal{R}(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}) \\
&= \mathcal{R} \times_{11} \mathbf{e}_{i_1} \times_{12} \cdots \times_{1,i_k} \mathbf{e}_{i_k} \\
&= r_{i_1,\ldots,i_k} \in \{1, 0\}\ (\forall i_1, \ldots, i_k \in \{1, \ldots, N\})
\end{aligned} \tag{5.2}$$

We identity r' with $\mathcal{R}$ so that $\mathcal{R}$ encodes the M-relation r. Let M' be a model (D', I') in $\mathbb{R}^N$ such that I' interprets entities by $I'(e_i) = e_i (1 \leq i \leq N)$ and relations r by $I'(r) = \mathcal{R}$.

For example, taking the successor model of the word *abba* above, we may consider the domain elements to be the basis vectors of a 4-dimensional vector space

$$D = \{1,2,3,4\} \Rightarrow \mathbf{1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{3} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \mathbf{4} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Over these domain elements, we may define tensors for each unary labeling relation and the binary successor relation:

$$\mathcal{R}_a = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \mathcal{R}_b = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \mathcal{R}_{\triangleleft} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

If we were to change the successor relation to the general precedence relation, only the binary tensor would change:

$$\mathcal{R}_{<} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

For the purposes of this paper, we restrict ourselves to binary relations and predicates. When r is a binary predicate, the corresponding tensor $\mathcal{R}$ is a bilinear map and represented by an adjacency matrix **R** as follows:

$$[\![(e_i, e_j)]\!] = (\mathbf{e}_i \cdot \mathbf{R}\mathbf{e}_j) = \mathbf{e}_i^T \mathbf{R} \mathbf{e}_j = r_{ij} \in \{1, 0\} \tag{5.3}$$

Note that when $r(x, y)$ is encoded by $\mathcal{R}$ as $(\mathbf{x} \bullet \mathbf{R}\mathbf{y})$, $r(y, x)$ is encoded by $\mathbf{R}^T$, since $(\mathbf{y} \bullet \mathbf{R}\mathbf{x}) = (\mathbf{x} \bullet \mathbf{R}^T\mathbf{y})$ holds

We next inductively define the evaluation $[\![F]\!]_{I',a'}$ of a formula F in M. Let a be an assignment in M and a' the corresponding assignment in M', so $a(x) = e_i$ iff $a'(x) = \mathbf{e}_i$. For a ground atom

$r(e_{i_1}, ..., e_{i_k})$, define

$$[\![r(e_{i_1}, \ldots, e_{i_k}]\!]' = \underline{\mathbf{R}}(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k})(\forall i_1, \ldots, i_k \in \{1, \ldots, N\})$$

where $\mathcal{R} = \{r_{i_1,\ldots,i_1}\}$ is a tensor encoding the M-relation r in M. By definition $[\![F]\!]_{I,a} = [\![F]\!]_{I,a}$ holds for any atom F. Negative literals are evaluated using $\neg\mathcal{R}$ defined as

$$\begin{gathered}[\![\neg r\,(e_{i_1}, \ldots, e_{i_k}]\!]' = \neg\mathcal{R}\,(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}) \\ \text{where } \neg\mathcal{R} \stackrel{def}{=} \overbrace{\mathbf{1} \circ \cdots \circ \mathbf{1}}^{k} - \mathcal{R}\end{gathered} \tag{5.4}$$

$\neg\mathcal{R}$ encodes an M-relation $\neg r_1$. Negation other than negative literals, conjunction, disjunction, and quantifiers are evaluated in M' as follows.

$$[\![\neg F]\!]' = 1 - [\![F]\!]' \tag{5.5}$$

$$[\![F_1 \wedge \cdots \wedge F_h]\!]' = [\![F_1]\!]' \cdots [\![F_h]\!]' \tag{5.6}$$

$$[\![F_1 \vee \cdots \vee F_h]\!]' = \min_1([\![F_1]\!]' + \ldots + [\![F_h]\!]') \tag{5.7}$$

$$[\![\exists y F]\!]' = \min_1(\sum_{i=1}^{N} [\![F_{y \leftarrow e_i}]\!]') \tag{5.8}$$

Here the operation $\min_1(x) = \min(x, 1) = x$ if $x < 1$, otherwise 1, as componentwise application. $F_{y \leftarrow e_i}$ means replacing every free occurrence of y in F with e_i. We treat universal quantification as $\forall x F = \neg\exists x \neg F$.

5.3 Subregular Grammars via Tensor Algebra

As mentioned in Chapter 2, there are several well-known connections between logical statements and languages classes. Most famous is Büchi (1960)'s result that languages characterizable by finite-state machines, the regular languages, are equivalent to statements in Monadic Second-Order Logic over the precedence model for strings (and successor, since precedence is MSO-definable from successor). Within the regular languages, many well-known subregular classes can be characterized by weakening the logic (McNaughton and Papert, 1971; Rogers et al., 2013; Thomas, 1997).

Here we restrict ourselves to first-order logic, as it is the lightest restriction corresponding to properly subregular languages. First-order statements over particular model signatures define distinct language classes. For the successor model, Thomas (1997) characterizes FO($\lhd$) in terms of Local Threshold Testability, equivalence in terms of the multiplicity of k-subfactors up to some fixed finite threshold t.

Theorem 2. *((Thomas, 1997)) A set of strings is First-order definable over* $\langle D; \lhd, [R_\sigma]_{\sigma \in \Sigma} \rangle$ *iff it is Locally Threshold Testable.*

Correspondingly, First-order formulas over the precedence model characterize the Non-Counting or Star-Free class of languages.

Theorem 3. *((McNaughton and Papert, 1971)) Languages that are first-order definable over* $\langle D; \prec, [R_\sigma]_{\sigma \in \Sigma} \rangle$ *are Non-Counting.*

Non-Counting languages are those languages definable in linear temporal logic, and with aperiodic syntactic monoids. Since the Successor relation ($\lhd$) is first-order definable from Precedence ($\prec$), the Non-Counting class properly includes the Locally Threshold Testable class.

Further restrictions of the logic over these signatures to propositional logic or conjunctions of negative literals characterize subclasses of the LTT and Star Free languages, yielding the Local and Piecewise hierarchies. (Rogers et al., 2013).

Sato (2017) presents an algorithm for compiling any first-order formula into a tensor embedding without grounding. The algorithm works by converting a formula into prenex normal form. Each quantified statement is put in conjunctive or disjunctive normal form, depending on the quantifier, and each formula is then converted to the appropriate tensor realization.

Here we restrict ourselves to formulas with binary predicates, which as stated above may be represented as adacency matrices since their corresponding tensor is a bilinear map. Sato shows that in these cases, we can often "optimize" compilation by directly compiling a formula F using matrices.

5.3.1 A Simple First-Order Constraint

As an illustrative example, consider a common constraint over the alphabet a, b like “every word must contain a b". It is rather easy to construct a first-order definable statement describing this constraint over word models:

$$F_{\text{one-}b} = \exists x(R_b(x))$$

Relying on the method in the previous section, we may translate this statement into an equivalent tensor over the vector space embedding of a model:

$$\mathcal{T}_{\text{one-}B} = \min_{1} \Big(\sum_{i=1}^{N} \mathcal{R}_b(\mathbf{d}_i) \Big)$$

Let us consider how the model behaves over a particular model of a word like $abba$. In the case of the precedence model signature with unary labeling symbols, we will have a model as below:

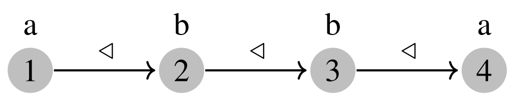


Figure 5.5: Successor Word model for $abba$

The semantic evaluation of the tensor statement $\mathcal{T}_{\text{one-}B}$ of the formula $F_{\text{one-}b}$ is as follows: each of the four domain elements is mapped to a unique basis vector. The tensor $\mathcal{R}_b(\mathbf{d}_i)$ is applied to each of these basis vectors to test whether they possess the property of being a b, each decision of which returns a scalar. These are then summed, and the result is passed through the $\min$ operation to yield a unique Boolean decision. In this case, there are exactly two b’s in the word, which occur at positions 2 and 3. The minimization flattens this to a 1 and return that the statement is true of this particular word model. This process is visualized below.

$$\min_{1} \Big(\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}^{T} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}^{T} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}^{T} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}^{T} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Big)$$

$$= \min_1(0+1+1+0) = \min_1(2) = 1$$

We may contrast this example with an evaluation over a word which contains no b's at all, such as $aaaa$ in which case it is not part of the language defined by $F_{\text{one-}b}$ and thus $\mathcal{T}_{\text{one-}B}$ should map its well-formedness to 0.

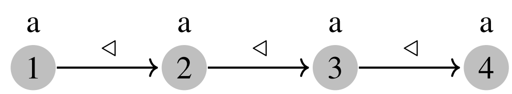


Figure 5.6: Successor model for $aaaa$

In this case the evaluation proceeds exactly as before, except that there are exactly no domain elements possessing the b property, and correspondingly the tensor contraction of the $\mathcal{R}_b(\mathbf{d}_i)$ tensor returns a 0 for each of them, which is preserved under summation and minimization to return a 0 for the entire formula. This signifies that the model for this word is ill-formed. This process is depicted below.

$$\min_1 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

$$= \min_1(0+0+0+0) = \min_1(0) = 0$$

5.3.2 Compiling a Locally Threshold Testable Formula

Here we demonstrate a formula which is properly First Order over the successor model for strings, characterizing a Locally Threshold Testable Language. With the ability to distinguish distinct occurrences of a symbol we can define a formula which is satisfied by strings containing exactly one occurrence of some symbol b. Such a system is seen in phonological stress patterns in the world's languages, which often mandate exactly one primary stress in a word. We do this by asserting that there is some position in which b occurs $\exists x(R_b(x))$, and that there are no other positions in which b occurs $\wedge(\forall y)[R_b(y) \rightarrow (x = y)]]$. The conjunction of these two gives the FO formula

$$F_{\text{one-}B} = (\exists x \forall y)[R_b(x) \wedge [R_b(y) \rightarrow (x = y)]] \tag{5.9}$$

Converting this into prenex normal form we get

$$\exists x \forall y (R_B(x) \wedge [\neg R_B(y) \vee (x = y)] \tag{5.10}$$

Compiling this formula into tensor notation is rather straightforward.

$$\mathcal{T}_{\text{one-}B} = min_1\Big(\sum_{i=1}^{N} 1 - min_1\Big(\sum_{j=1}^{N} \mathcal{R}^b \mathbf{e}_i \bullet [(1 - \mathcal{R}^b \mathbf{e}^j) + (\mathbf{e}_i \bullet \mathbf{e}_j)]\Big)\Big) \tag{5.11}$$

Intuitively, this formula checks whether for any domain elements, two domain elements are the same via the inner product, and if both domain elements have the property of being a b, then the formula evaluates to 1. If either formula is not a b, or if two different domain elements are a b, the formula evaluates to 0. We can apply this to to the successor model for $abba$ in (**??**) by defining each of the relational tensors in the formula over the domain and relations in the model. Doing so, it is intuitive to see that the formula evaluates to 0 (false) due to domain elements 2 and 3, which are distinct and each have the property of being a b.

To concretely illustrate the procedure, consider the following two examples of a derivation using two minimally different words, each of length two: one satisfying the model (*ba*) and one not (*bb*). First I show the model for *ba*.

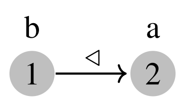


Figure 5.7: successor model of ba.

$$D = \{1, 2\} \Rightarrow \mathbf{1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathcal{R}_a = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathcal{R}_b = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathcal{R}_{\triangleleft} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

The first model satisfies the constraint, since there is exactly one b in the word, and so the tensor contraction should end up with a scalar value of 1. A derivation of the corresponding formula, when the domain and relations are appropriately converted to vectors and matrices, is below. Tensor algebra can get horribly long and cluttered to the point of madness (see Beim Graben and Gerth (2012) for extreme examples). Thus for the first derivation step I apply the multilinear map ($\mathcal{R}\mathbf{e}_i$) beforehand for each tensor and domain vector.

$$\begin{aligned}
\mathcal{T}^{ba}_{one-B} &= min((1 - min((1 - (1 \bullet [0+1])) + (1 - (1 \bullet [1+0])))) \\
&\quad + (1 - min((1 - (0 \bullet [1+0])) + (1 - (0 \bullet [0+1]))))) \\
&= min((1 - min((1-1) + (1-1))) + (1 - min((1-0) + (1-0)))) \\
&= min((1-0) + (1-1)) \\
&= min(1+0) \\
&= 1
\end{aligned}$$

The model of the second word, *bb*, does not satisfy the constraint, since there is more than one b in the word.

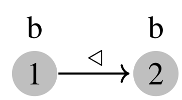


Figure 5.8: Visualization of successor model of bb.

$$D = \{1,2\} \Rightarrow \mathbf{1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathbf{2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathcal{R}_a = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \mathcal{R}_b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mathcal{R}_\triangleleft = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

In this case, the tensor contraction computes a scalar valued 0. A derivation of the corresponding formula, when the domain and relations are appropriately converted to vectors and matrices, is as follows, again first applying the multilinear map ($\mathcal{R}\mathbf{e}_i$) for each tensor and domain vector:

$$\mathcal{T}^{bb}_{one-B} = min((1 - min((1 - (1 \bullet [(1-1)+1])) + (1 - (1 \bullet [(1-1)+0])))$$
$$+ (1 - min((1 - (1 \bullet [(1-1)+0])) + (1 - (1 \bullet [(1-1)+1]))))$$
$$= min((1 - min(0+1)) + (1 - min((1+0))))$$
$$= min((1-1) + (1-1))$$
$$= min(0+0)$$
$$= 0$$

5.3.3 Compiling a properly Non-Counting Formula

Next I demonstrate a formula which is properly First Order over the precedence model for strings, characterizing a Non-Counting language. I motivate this formula using a phonological pattern from Latin, in which in certain cases an l cannot follow another l unless an r intervenes, no matter the distance between them (Jensen, 1974; Heinz, 2010b). This can be seen in the $-alis$ adjectival suffix which appears as $-aris$ if the word it attaches to already contains an l, except in cases where there is an intervening r, in which it appears again as -alis.

Example 1

a. navalis	'naval'
b. episcopalis	'episcopal'
c. infinitalis	'negative'
d. solaris	'solar' (*solalis)
e. lunaris	'lunar' (*lunalis)
f. militaris	'military' (*militalis)
g. floralis	'floral'
h. sepulkralis	'funereal'
i. litoralis	'of the shore'

The blocking effects in this non-local alternating pattern require the use of quantifiers, and is properly Non-Counting. We can represent it with the following first-order formula:

$$F_{diss} \; = \; \forall x \forall y [R_l(x) \; \wedge \; R_l(y) \; \wedge \; R_{\prec}(x,y)] \;\; \rightarrow \;\; \exists z [R_r(z) \; \wedge \; R_{\prec}(x,z) \; \wedge \; R_{\prec}(z,y)] \quad (5.12)$$

The tensor compilation of this formula is again quite straightforward:

$$\mathcal{T}_{diss} = min_1\Big(\sum_{i=1}^{N} min_1\Big(\sum_{j=1}^{N}\Big(\mathbf{1} - \Big[(\mathcal{R}^l\mathbf{e}_i) \bullet (\mathcal{R}^l\mathbf{e}_j) \bullet (\mathbf{e}_i\mathcal{R}^{\prec}\mathbf{e}_j)\Big]\Big) + \\ + min_1\Big(\sum_{k=1}^{N}\Big[(\mathcal{R}^z\mathbf{e}_k) \bullet (\mathbf{e}_i\mathcal{R}^{\prec}\mathbf{e}_k) \bullet (\mathbf{e}_k\mathcal{R}^{\prec}\mathbf{e}_j)\Big]\Big)\Big)\Big) \quad (5.13)$$

Intuitively, this formula tests whether, for any two domain elements labeled l and which precede each other, there is another element labeled r which comes between them. The use of the precedence relation here shows that this can happen anywhere in the word, and can thus handle the Latin dissimilation patterns above. Note that the negation applies componentwise to each multilinear map in the antecedent.

As with the locally threshold testable formula above, I present two minimally different examples to showcase the evaluation procedure. Consider a precedence model of the word *lrl*.

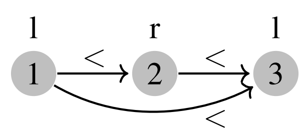


Figure 5.9: Visualization of the precedence model for lrl.

$$D = \{1,2\} \Rightarrow \mathbf{1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mathbf{3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathcal{R}_l = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \mathcal{R}_r = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mathcal{R}_< = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The precedence model for this word model satisfies the constraint, since there are two ls but there is an intervening r. The corresponding derivation is straightforward. I apply the multilinear map $(\mathcal{R}\mathbf{e}_i)$ beforehand for each tensor and domain vector. Additionally, for the first step of the derivation, I apply the componentwise negation in the antecedent $(\mathbf{1} - \Big[(\mathcal{R}^l\mathbf{e}_i) \bullet (\mathcal{R}^l\mathbf{e}_j) \bullet (\mathbf{e}_i\mathcal{R}^{\prec}\mathbf{e}_j)\Big])$.

$$
\begin{aligned}
\mathcal{T}^{lrl}_{diss} &= min(min(\\
&\quad ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [1 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
&\quad + ([0 \bullet 1 \bullet 0] + min([0 \bullet 0 \bullet 1] + [1 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
&\quad + ([0 \bullet 0 \bullet 0] + min([0 \bullet 0 \bullet 1] + [1 \bullet 1 \bullet 1] + [0 \bullet 1 \bullet 0])) \\
&\quad + min(([1 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [1 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0])) \\
&\quad + ([1 \bullet 1 \bullet 1] + min([0 \bullet 0 \bullet 1] + [1 \bullet 0 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
&\quad + ([1 \bullet 0 \bullet 0] + min([0 \bullet 0 \bullet 1] + [1 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0])) \\
&\quad + min(([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [1 \bullet 0 \bullet 0] + [0 \bullet 0 \bullet 0])) \\
&\quad + ([0 \bullet 1 \bullet 1] + min([0 \bullet 0 \bullet 1] + [1 \bullet 0 \bullet 0] + [0 \bullet 0 \bullet 0])) \\
&\quad + ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 1] + [1 \bullet 0 \bullet 1] + [0 \bullet 0 \bullet 0])))) \\
&= min(min(\\
&\quad ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
&\quad + ([0 \bullet 1 \bullet 0] + min(0 + 0 + 0)) \\
&\quad + ([0 \bullet 0 \bullet 0] + min(0 + 1 + 0)) \\
&\quad + min(([1 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
&\quad + ([1 \bullet 1 \bullet 1] + min(0 + 0 + 0)) \\
&\quad + ([1 \bullet 0 \bullet 0] + min(0 + 0 + 0)) \\
&\quad + min(([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
&\quad + ([0 \bullet 1 \bullet 1] + min(0 + 0 + 0)) \\
&\quad + ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)))) \\
&= min(min((0 + 0) + (0 + 0) = (0 + 1)) + min((0 + 0) + (1 + 0) + (0 + 0)) + \\
&\quad + min((0 + 0) + (0 + 0) + (0 + 0))) \\
&= min(min(0 + 0 + 1) + min(0 + 1 + 0) + min(0 + 0 + 0) \\
&= min(1 + 1 + 0) \\
&= min(2) = 1
\end{aligned}
$$

Interestingly, the two parts of the conditional play off each other numerically, allowing the satisfaction of the model to proceed geometrically. To better see this, consider a minimally different word model, *lll*. The precedence model of this word does not satisfy the dissimilation constraint, since there are more than 1 ls, but no r intervenes.

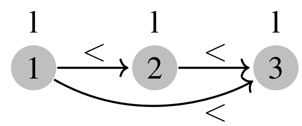


Figure 5.10: Visualizations of precedence model of lll.

$$D = \{1, 2\} \Rightarrow \mathbf{1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mathbf{3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathcal{R}_l = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mathcal{R}_r = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mathcal{R}_< = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The corresponding derivation is below. Again, in the first step of the derivation I apply the multilinear map for each tensor and domain vector, and apply the componentwise negation in the antecedent.

$$
\begin{aligned}
\mathcal{T}_{diss}^{lll} = & \, min(min(\\
& ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
& + ([0 \bullet 0 \bullet 0] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0]))+ \\
& + ([0 \bullet 0 \bullet 0] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 1] + [0 \bullet 1 \bullet 0])) \\
& + min(([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
& + ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
& + ([0 \bullet 0 \bullet 0] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
& + min(([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 0] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0])) \\
& + ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 0] + [0 \bullet 1 \bullet 0]))+ \\
& + ([0 \bullet 0 \bullet 1] + min([0 \bullet 0 \bullet 1] + [0 \bullet 1 \bullet 1] + [0 \bullet 1 \bullet 0])))) \\
= & \, min(min(\\
& ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
& + ([0 \bullet 0 \bullet 0] + min(0 + 0 + 0))+ \\
& + ([0 \bullet 0 \bullet 0] + min(0 + 0 + 0)) \\
& + min(([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
& + ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
& + ([0 \bullet 0 \bullet 0] + min(0 + 0 + 0)) \\
& + min(([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)) \\
& + ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0))+ \\
& + ([0 \bullet 0 \bullet 1] + min(0 + 0 + 0)))) \\
= & \, min(min((0 + 0) + (0 + 0) = (0 + 0)) + min((0 + 0) + (0 + 0) + (0 + 0))+ \\
& + min((0 + 0) + (0 + 0) + (0 + 0))) \\
= & \, min(min(0 + 0 + 0) + min(0 + 0 + 0) + min(0 + 0 + 0) \\
= & \, min(0 + 0 + 0) \\
= & \, min(0) = 0
\end{aligned}
$$

Each of the preceding derivations illustrates how the tensor evaluation works. In each case, the numerical activation mimics the logical evaluation over the structure, tracing the computation required for computing well-formedness over each structure. In this way, we preserve the fundamental characteristic of the Tensor Product representation, that of symbolic structure and computation at a high, descriptive level, with spreading activation and numerical computation on a lower level.

5.4 Conclusion

This chapter provided a method for geometrically characterizing model-theoretic structures and subregular constraints over them via vector spaces. Model-theoretic descriptions of relational structures were embedded in Euclidean vector spaces, and statements in first-order logic over these structures were compiled into tensor formulas. Semantic evaluation was given via tensor contraction over tensors implementing a specific model. This method can easily be extended to consider other relational structures, and to other logics. Another application is to consider logical translations between model signatures, which define mappings between structures (Courcelle, 1994), another area relevant for linguistics. The analytical power given by multilinear algebra, combined with the representational flexibility given by finite model theory and mathematical logic, provides a powerful combination for analyzing the nature of linguistic structures and cognition, and for exploring the cognitive instantiation of linguistic computation generally.

Chapter 6

Conclusion

This dissertation examined the relationship between structure and learning in natural language. Structure was found to take several forms: in the definition of a learning problem, on restrictions to the problem in the form of abductive constraints, and as representational parses which constitute evidence for the learner.

First, we mathematically describe a unified model-theoretic notion of structural representational information, and use it to define several linguistically relevant structural representations, which then defines a notion of grammar whose components are sets factors of these representations. The chapter then showed how the inherent model-theoretic properties of these representations structure the space of possible components of a grammar into a partial order.

Chapter 3 considered the question of learning from a constraint-based perspective. The partial order structure that the components of the grammar possess directly leads to a property called "grammatical entailment" which allows a learner to prune out vast swathes of possible components of grammars given evidence. These entailments directly enabled a constraint-based abductive algorithm, BUFIA, which traverses this partial order of factors to select the most general grammar according to the constraints on the problem. depending on the representations, BUFIA's outputs may be highly redundant, and additional abductive principles are introduced to constrain the candidate solutions further.

Chapter 4 demonstrates the effectiveness of the variants of BUFIA on a variety of well-understood phonotactic data. Each of these case studies allowed the different behavior of the learning algorithm to come out. Its behavior was compared to the UCLA Maximum Entropy Phonotactic Learner (Hayes and Wilson, 2008), a statistical inductive inference learning algorithm. In the cases studied, the main result is that the statistical generalization is unnecessary, as BUFIA correctly finds the grammars from data and does not contain any notion of statistical inference. In the weakest case, we may say that structural methods play the same inferential role as statistical ones. In a stronger case, we may say that the aspects of the MaxEnt learner that enable it to succeed are necessarily structural, dressed in the guise of statistics. The similarity between the additional abductive principles we considered and the and the additional statistical measures considered for MaxEnt suggest this as well.

Finally, Chapter 5 explored the cognitive instantiation of the model-theoretic view on representation. In particular, it discusses how to embed the structured discrete symbolic representations into vector spaces via tensors, and defines several constrained classes of grammars as operations using tensor calculus. In this way, the unified notion of structure gains another interpretation, one closer to the notions of distributed computation that emerged from the connectionist literature over the past decades.

One overarching message from this dissertation is the unavoidable relationship between structure and successful learning. Structure is present at each and every level of learning. It is definitionally necessary, and as was shown, is sufficient for a variety of learning cases in phonotactics. Additional aesthetic or feasibility constraints, such as those related to size of the grammar or redundancy, can be formulated non-statistically, in contrast to algorithms whose core is built on statistical inference.

It was further argued, and is worth repeating, that statistical inference is not a bogeyman. There is merely a complex relationship between inference based on structure and inference based on statistics. However, in the comparisons between BUFIA and the UCLA learner, the structure, defined by the constraints placed on the algorithms, not the statistics, was the determinant of the structure of the output grammar. The gain-based selection criteria imposed on UCLA, a statistically defined distance metric, also has non-statistical variants, and this is desirable. All of these comprise various abductive constraints.

The most direct way to explore the behavior of BUFIA further is to run more simulations on a variety of linguistic corpora. It could be that there is some representational information, say autosegmental graph structures that are used to learn tonal grammars, that affects the learning in an as-yet-unknown way.

Additionally, expansion of the algorithm to consider non-phonological structures, such as tree structures encoding syntactic or semantic information, are also important to pursue. As we saw, the parsing is separate from, but intimately relevant for, BUFIA's behavior. In an important sense, parsing goes a long way to solving the learning problem. The reason syntactic explorations are necessary is that this structural information is not present at all in the signal, further evidence of the hidden structure problem. It is even unclear what the best syntactic parsing strategy is, or what the relevant data structure is. BUFIA is an excellent place to test, since the data structures it considers are entirely up to the analyst. Results similar to the phonotactic learning experiments will present an interesting hypothesis about modularity across linguistic subsystems like syntax and phonology.

Finally, it is desirable to compare the performance of BUFIA to phonotactic experiments on human speakers whenever possible. Of particular relevance are experimental cases where the decision is between more general of more specific structural well-formedness judgments. However, there is interesting recent experimental evidence showing that humans in phonotactic learning experiments have a simplicity bias in the sense we considered here (Durvasula and Liter, 2020). This preference persists even when entertaining simpler factors means storing more of them, when a single superfactor would do.

In general, as noted in the introduction, structure is a part of the cognitive structure of human learning and knowledge. The characterization of such knowledge and learning as abductive inference represents a new road into the nature of learning from a mathematical perspective, and for the cognitive implications of such mathematical results. Structure in humans is ever-present yet largely unobservable. While it is indeed interesting to remove as much structure from the learning problem as one can, and see the results, this may in the end turn out to be shortsighted. In the cognitive study of human linguistic knowledge, the relationship between structure and learning is deep and intimate. Further understanding in this area is a promising avenue to unlocking the many remaining mysteries in linguistic and cognitive inquiry.

Appendices

Appendix A

Output BUFIA Grammars

A.1 Forbidden Substructure Grammars for English Consonant Clusters

BUFIA, Successor Model, $k = 2$

-coronal

[-dorsal]

[-high]

[-labial]

[-lateral]

[-nasal]

[-spread]

[+dorsal, +nasal]

[+dorsal, +sonorant]

[+anterior][-back]

[+anterior][+continuant]

[+anterior][+spread]

[+anterior][+strident]

[+anterior][+voice]

[+anterior][-anterior, -approximant]

[+anterior][-anterior, +consonantal]

[+anterior][-anterior, -continuant]

[+anterior][-anterior, -sonorant]

[+anterior][-anterior, -voice]

[-anterior][+anterior]

[-anterior][-approximant]

[-anterior][+back]

[-anterior][-back]

[-anterior][+consonantal]

[-anterior][+continuant]

[-anterior][-continuant]

[-anterior][+dorsal]

[-anterior][+high]
[-anterior][+labial]
[-anterior][+lateral]
[-anterior][+nasal]
[-anterior][-sonorant]
[-anterior][+spread]
[-anterior][+strident]
[-anterior][+voice]
[-anterior][-voice]
[+approximant][+anterior]
[+approximant][-anterior]
[+approximant][+approximant]
[+approximant][-approximant]
[+approximant][+back]
[+approximant][-back]
[+approximant][-boundary]
[+approximant][+consonantal]
[+approximant][-consonantal]
[+approximant][+continuant]
[+approximant][-continuant]
[+approximant][+coronal]
[+approximant][+dorsal]
[+approximant][+high]
[+approximant][+labial]
[+approximant][+lateral]
[+approximant][+nasal]
[+approximant][+sonorant]
[+approximant][-sonorant]
[+approximant][+spread]
[+approximant][+strident]
[+approximant][-strident]
[+approximant][+voice]
[+approximant][-voice]
[-approximant][-back]
[-approximant][+continuant]
[-approximant][+spread]
[-approximant][+strident]
[-approximant][+voice]
[-approximant][-anterior, -approximant]
[-approximant][-anterior, +consonantal]
[-approximant][-anterior, -continuant]
[-approximant][-anterior, -sonorant]
[-approximant][-anterior, -voice]
[+back][+anterior]
[+back][-anterior]
[+back][+approximant]
[+back][-approximant]
[+back][+back]
[+back][-back]
[+back][-boundary]
[+back][+consonantal]
[+back][-consonantal]
[+back][+continuant]
[+back][-continuant]
[+back][+coronal]
[+back][+dorsal]
[+back][+high]
[+back][+labial]
[+back][+lateral]
[+back][+nasal]
[+back][+sonorant]
[+back][-sonorant]

[+back][+spread]
[+back][+strident]
[+back][-strident]
[+back][+voice]
[+back][-voice]
[-back][+anterior]
[-back][-anterior]
[-back][+approximant]
[-back][-approximant]
[-back][+back]
[-back][-back]
[-back][-boundary]
[-back][+consonantal]
[-back][-consonantal]
[-back][+continuant]
[-back][-continuant]
[-back][+coronal]
[-back][+dorsal]
[-back][+high]
[-back][+labial]
[-back][+lateral]
[-back][+nasal]
[-back][+sonorant]
[-back][-sonorant]
[-back][+spread]
[-back][+strident]
[-back][-strident]
[-back][+voice]
[-back][-voice]
[+boundary][+boundary]
[-boundary][-back]
[-boundary][+continuant]
[-boundary][+spread]
[-boundary][+strident]
[-boundary][+voice]
[-boundary][-anterior, -approximant]
[-boundary][-anterior, +consonantal]
[-boundary][-anterior, -continuant]
[-boundary][-anterior, -sonorant]
[-boundary][-anterior, -voice]
[+consonantal][-back]
[+consonantal][+continuant]
[+consonantal][+spread]
[+consonantal][+strident]
[+consonantal][+voice]
[+consonantal][-anterior, -approximant]
[+consonantal][-anterior, +consonantal]
[+consonantal][-anterior, -continuant]
[+consonantal][-anterior, -sonorant]
[+consonantal][-anterior, -voice]
[-consonantal][+anterior]
[-consonantal][-anterior]
[-consonantal][+approximant]
[-consonantal][-approximant]
[-consonantal][+back]
[-consonantal][-back]
[-consonantal][-boundary]
[-consonantal][+consonantal]
[-consonantal][-consonantal]
[-consonantal][+continuant]
[-consonantal][-continuant]
[-consonantal][+coronal]

[-consonantal][+dorsal]
[-consonantal][+high]
[-consonantal][+labial]
[-consonantal][+lateral]
[-consonantal][+nasal]
[-consonantal][+sonorant]
[-consonantal][-sonorant]
[-consonantal][+spread]
[-consonantal][+strident]
[-consonantal][-strident]
[-consonantal][+voice]
[-consonantal][-voice]
[+continuant][-back]
[+continuant][+continuant]
[+continuant][+spread]
[+continuant][+strident]
[+continuant][+voice]
[+continuant][-anterior, -approximant]
[+continuant][-anterior, +consonantal]
[+continuant][-anterior, -continuant]
[+continuant][-anterior, -sonorant]
[+continuant][-anterior, -voice]
[-continuant][-approximant]
[-continuant][-back]
[-continuant][+continuant]
[-continuant][-continuant]
[-continuant][+dorsal]
[-continuant][+nasal]
[-continuant][-sonorant]
[-continuant][+spread]
[-continuant][+strident]

[-continuant][+voice]
[-continuant][-voice]
[-continuant][-anterior, +consonantal]
[-continuant][+consonantal, +labial]
[+coronal][-back]
[+coronal][+continuant]
[+coronal][+spread]
[+coronal][+strident]
[+coronal][+voice]
[+coronal][-anterior, -approximant]
[+coronal][-anterior, +consonantal]
[+coronal][-anterior, -continuant]
[+coronal][-anterior, -sonorant]
[+coronal][-anterior, -voice]
[+dorsal][-approximant]
[+dorsal][-back]
[+dorsal][+continuant]
[+dorsal][-continuant]
[+dorsal][+dorsal]
[+dorsal][+nasal]
[+dorsal][-sonorant]
[+dorsal][+spread]
[+dorsal][+strident]
[+dorsal][+voice]
[+dorsal][-voice]
[+dorsal][-anterior, +consonantal]
[+dorsal][+consonantal, +labial]
[+high][+anterior]
[+high][-anterior]
[+high][+approximant]
[+high][-approximant]

[+high][+back]
[+high][-back]
[+high][-boundary]
[+high][+consonantal]
[+high][-consonantal]
[+high][+continuant]
[+high][-continuant]
[+high][+coronal]
[+high][+dorsal]
[+high][+high]
[+high][+labial]
[+high][+lateral]
[+high][+nasal]
[+high][+sonorant]
[+high][-sonorant]
[+high][+spread]
[+high][+strident]
[+high][-strident]
[+high][+voice]
[+high][-voice]
[+labial][-approximant]
[+labial][+back]
[+labial][-back]
[+labial][+continuant]
[+labial][-continuant]
[+labial][+dorsal]
[+labial][+high]
[+labial][+labial]
[+labial][+nasal]
[+labial][-sonorant]
[+labial][+spread]
[+labial][+strident]
[+labial][+voice]
[+labial][-voice]
[+labial][-anterior, +consonantal]
[+lateral][+anterior]
[+lateral][-anterior]
[+lateral][+approximant]
[+lateral][-approximant]
[+lateral][+back]
[+lateral][-back]
[+lateral][-boundary]
[+lateral][+consonantal]
[+lateral][-consonantal]
[+lateral][+continuant]
[+lateral][-continuant]
[+lateral][+coronal]
[+lateral][+dorsal]
[+lateral][+high]
[+lateral][+labial]
[+lateral][+lateral]
[+lateral][+nasal]
[+lateral][+sonorant]
[+lateral][-sonorant]
[+lateral][+spread]
[+lateral][+strident]
[+lateral][-strident]
[+lateral][+voice]
[+lateral][-voice]
[+nasal][+anterior]
[+nasal][-anterior]
[+nasal][+approximant]

[+nasal][-approximant]

[+nasal][+back]

[+nasal][-back]

[+nasal][-boundary]

[+nasal][+consonantal]

[+nasal][-consonantal]

[+nasal][+continuant]

[+nasal][-continuant]

[+nasal][+coronal]

[+nasal][+dorsal]

[+nasal][+high]

[+nasal][+labial]

[+nasal][+lateral]

[+nasal][+nasal]

[+nasal][+sonorant]

[+nasal][-sonorant]

[+nasal][+spread]

[+nasal][+strident]

[+nasal][-strident]

[+nasal][+voice]

[+nasal][-voice]

[+sonorant][+anterior]

[+sonorant][-anterior]

[+sonorant][+approximant]

[+sonorant][-approximant]

[+sonorant][+back]

[+sonorant][-back]

[+sonorant][-boundary]

[+sonorant][+consonantal]

[+sonorant][-consonantal]

[+sonorant][+continuant]

[+sonorant][-continuant]

[+sonorant][+coronal]

[+sonorant][+dorsal]

[+sonorant][+high]

[+sonorant][+labial]

[+sonorant][+lateral]

[+sonorant][+nasal]

[+sonorant][+sonorant]

[+sonorant][-sonorant]

[+sonorant][+spread]

[+sonorant][+strident]

[+sonorant][-strident]

[+sonorant][+voice]

[+sonorant][-voice]

[-sonorant][-back]

[-sonorant][+continuant]

[-sonorant][+spread]

[-sonorant][+strident]

[-sonorant][+voice]

[-sonorant][-anterior, -approximant]

[-sonorant][-anterior, +consonantal]

[-sonorant][-anterior, -continuant]

[-sonorant][-anterior, -sonorant]

[-sonorant][-anterior, -voice]

[+spread][+anterior]

[+spread][-anterior]

[+spread][-approximant]

[+spread][-back]

[+spread][+consonantal]

[+spread][+continuant]

[+spread][-continuant]

[+spread][+coronal]
[+spread][+dorsal]
[+spread][+lateral]
[+spread][+nasal]
[+spread][-sonorant]
[+spread][+spread]
[+spread][+strident]
[+spread][-strident]
[+spread][+voice]
[+spread][-voice]
[+strident][-back]
[+strident][+continuant]
[+strident][+spread]
[+strident][+strident]
[+strident][+voice]
[+strident][-anterior, -approximant]
[+strident][-anterior, +consonantal]
[+strident][-anterior, -continuant]
[+strident][-anterior, -sonorant]
[+strident][-anterior, -voice]
[-strident][+anterior]
[-strident][-approximant]
[-strident][-back]
[-strident][+consonantal]
[-strident][+continuant]
[-strident][-continuant]
[-strident][+dorsal]
[-strident][+lateral]
[-strident][+nasal]
[-strident][-sonorant]
[-strident][+spread]
[-strident][+strident]
[-strident][+voice]
[-strident][-voice]
[+voice][-approximant]
[+voice][-back]
[+voice][+continuant]
[+voice][-continuant]
[+voice][+dorsal]
[+voice][+nasal]
[+voice][-sonorant]
[+voice][+spread]
[+voice][+strident]
[+voice][+voice]
[+voice][-voice]
[+voice][-anterior, +consonantal]
[+voice][+consonantal, +labial]
[-voice][-back]
[-voice][+continuant]
[-voice][+spread]
[-voice][+strident]
[-voice][+voice]
[-voice][-anterior, -approximant]
[-voice][-anterior, +consonantal]
[-voice][-anterior, -continuant]
[-voice][-anterior, -sonorant]
[-voice][-anterior, -voice]
[+anterior, -continuant][+anterior]
[+anterior, -continuant][+consonantal]
[+anterior, -continuant][+lateral]
[+anterior, +strident][-anterior]
[+anterior, +strident][-consonantal, +coronal]

[+anterior, +strident][-consonantal, -strident]
[+anterior, +voice][+anterior]
[+anterior, +voice][+consonantal]
[+anterior, +voice][+lateral]
[-anterior, -continuant][-anterior]
[-anterior, -continuant][+approximant]
[-anterior, -continuant][-boundary]
[-anterior, -continuant][-consonantal]
[-anterior, -continuant][+coronal]
[-anterior, -continuant][+sonorant]
[-anterior, -continuant][-strident]
[-anterior, -strident][-anterior]
[-anterior, -strident][+approximant]
[-anterior, -strident][-boundary]
[-anterior, -strident][-consonantal]
[-anterior, -strident][+coronal]
[-anterior, -strident][+sonorant]
[-anterior, -strident][-strident]
[-anterior, +voice][-anterior]
[-anterior, +voice][+approximant]
[-anterior, +voice][-boundary]
[-anterior, +voice][-consonantal]
[-anterior, +voice][+coronal]
[-anterior, +voice][+sonorant]
[-anterior, +voice][-strident]
[+continuant, +voice][+anterior]
[+continuant, +voice][-anterior]
[+continuant, +voice][+approximant]
[+continuant, +voice][+back]
[+continuant, +voice][-boundary]
[+continuant, +voice][+consonantal]
[+continuant, +voice][-consonantal]
[+continuant, +voice][+coronal]
[+continuant, +voice][+high]
[+continuant, +voice][+labial]
[+continuant, +voice][+lateral]
[+continuant, +voice][+sonorant]
[+continuant, +voice][-strident]
[-continuant, +coronal][+anterior]
[-continuant, +coronal][+consonantal]
[-continuant, +coronal][+lateral]
[-continuant, +strident][+anterior]
[-continuant, +strident][-anterior]
[-continuant, +strident][+approximant]
[-continuant, +strident][+back]
[-continuant, +strident][-boundary]
[-continuant, +strident][+consonantal]
[-continuant, +strident][-consonantal]
[-continuant, +strident][+coronal]
[-continuant, +strident][+high]
[-continuant, +strident][+labial]
[-continuant, +strident][+lateral]
[-continuant, +strident][+sonorant]
[-continuant, +strident][-strident]
[+coronal, +voice][+anterior]
[+coronal, +voice][+consonantal]
[+coronal, +voice][+lateral]
[+strident, +voice][+anterior]
[+strident, +voice][-anterior]
[+strident, +voice][+approximant]
[+strident, +voice][+back]
[+strident, +voice][-boundary]

[+strident, +voice][+consonantal]

[+strident, +voice][-consonantal]

[+strident, +voice][+coronal]

[+strident, +voice][+high]

[+strident, +voice][+labial]

[+strident, +voice][+lateral]

[+strident, +voice][+sonorant]

[+strident, +voice][-strident]

A.2 Forbidden Substructure Grammars for Quechua

A.2.1 Forbidden Precedence Structures

BUFIA, Successor Model, $k = 2$
-coronal

[-dorsal]

[-labial]

[+mb]

[+anterior, +lateral]

[-anterior, -back]

[-anterior, -low]

[-back, -cg]

[-back, -continuant]

[-back, -sg]

[-back, -sonorant]

[-cg, -lateral]

[-cg, -low]

[-cg, -nasal]

[-continuant, -low]

[-continuant, -nasal]

[-lateral, -sg]

[-lateral, -sonorant]

[-low, -sg]

[-low, -sonorant]

[-nasal, -sg]

[-nasal, -sonorant]

[+RTR, +back, -syllabic]

[+RTR, -back, -syllabic]

[+RTR, -cg, +continuant]

[+RTR, +continuant, +sg]

[+RTR, +continuant, -sonorant]

[+RTR, +continuant, -syllabic]

[+RTR, -continuant, +sonorant]

[+RTR, -low, -syllabic]

[+RTR, +sonorant, -syllabic]

[+anterior, -cg, +continuant]

[+anterior, +continuant, +sg]

[-anterior, -cg, +continuant]

[-anterior, +continuant, -lateral]

[-anterior, +continuant, +sg]

[-anterior, +continuant, -sonorant]

[-anterior, -lateral, -nasal]

[+back, +dorsal, -syllabic]

[-back, +dorsal, -syllabic]
[-cg, +continuant, +coronal]
[-cg, +continuant, +dorsal]
[+continuant, +coronal, +sg]
[+continuant, +dorsal, +sg]
[-continuant, +dorsal, +sonorant]
[+dorsal, -low, -syllabic]
[+dorsal, +sonorant, -syllabic]
[+RTR][-RTR, +cg]
[+RTR][-RTR, +sg]
[+RTR][+anterior, +cg]
[+RTR][+anterior, +sg]
[+RTR][-anterior, +cg]
[+RTR][-anterior, +sg]
[+RTR][+cg, +coronal]
[+RTR][+cg, +labial]
[+RTR][-cg, +continuant]
[+RTR][+continuant, +sg]
[+RTR][+coronal, +sg]
[+RTR][+labial, +sg]
[+anterior][-cg, +continuant]
[+anterior][+continuant, +sg]
[+cg][+cg]
[+cg][+RTR, +sg]
[+cg][+anterior, +sg]
[+cg][-anterior, +sg]
[+cg][-continuant, +sg]
[+cg][+coronal, +sg]
[+cg][+dorsal, +sg]
[+cg][+labial, +sg]
[-cg][+anterior, +sg]
[-cg][-anterior, +sg]
[-cg][-cg, +continuant]
[-cg][+continuant, +sg]
[-cg][+coronal, +sg]
[+sg][-RTR, +sg]
[+sg][+anterior, +sg]
[+sg][-anterior, +sg]
[+sg][-cg, +continuant]
[+sg][+continuant, +sg]
[+sg][+coronal, +sg]
[+sg][+labial, +sg]
[-sg][+cg]
[-sg][+RTR, +sg]
[-sg][+anterior, +sg]
[-sg][-anterior, +sg]
[-sg][+coronal, +sg]
[-sonorant][+anterior, +sg]
[-sonorant][-anterior, +sg]
[-sonorant][+coronal, +sg]
[+RTR, +back][+continuant, +dorsal, -sonorant]
[+RTR, +back][+continuant, +dorsal, -syllabic]
[+RTR, -back][+continuant, +dorsal, -sonorant]
[+RTR, -back][+continuant, +dorsal, -syllabic]
[+RTR, +cg][+sg]
[+RTR, +cg][-RTR, -cg, +dorsal]
[+RTR, +cg][-RTR, -continuant, +dorsal]
[+RTR, +cg][-RTR, +dorsal, -sg]
[+RTR, -cg][+cg]
[+RTR, -cg][+sg]
[+RTR,+continuant][+continuant,+dorsal,-sonorant]
[+RTR,+continuant][+continuant,+dorsal,-syllabic]

[+RTR, -continuant][+cg]

[+RTR, -continuant][+sg]

[+RTR,-low][+continuant, +dorsal, -sonorant]

[+RTR,-low][+continuant, +dorsal, -syllabic]

[+RTR, +sg][+cg]

[+RTR, +sg][+sg]

[+RTR, -sg][+sg]

[+RTR, +sonorant][+continuant, +dorsal, -sonorant]

[+RTR, +sonorant][+continuant, +dorsal, -syllabic]

[+RTR, -sonorant][+cg]

[+RTR, -sonorant][+sg]

[+RTR, +syllabic][+continuant, +dorsal, -sonorant]

[+RTR, +syllabic][+continuant, +dorsal, -syllabic]

[+RTR, -syllabic][+cg]

[+RTR, -syllabic][+sg]

[+anterior, +cg][+sg]

[+anterior, -cg][+cg]

[+anterior, -cg][+sg]

[+anterior, +continuant][+anterior, +sg]

[+anterior, +continuant][-anterior, +sg]

[+anterior, +continuant][+coronal, +sg]

[+anterior, -continuant][+cg, +labial]

[+anterior, -continuant][+labial, +sg]

[+anterior, +nasal][-back, -lateral]

[+anterior, +nasal][-back, -nasal]

[+anterior, +nasal][-back, -syllabic]

[+anterior, +nasal][+cg, +labial]

[+anterior, +nasal][+labial, +sg]

[+anterior, +nasal][+continuant, +dorsal, -sonorant]

[+anterior, +nasal][+continuant, +dorsal, -syllabic]

[+anterior, -nasal][+anterior, +sg]

[+anterior, -nasal][-anterior, +sg]

[+anterior, -nasal][+coronal, +sg]

[+anterior, +sg][+cg]

[+anterior, +sg][+sg]

[+anterior, +sg][-anterior, -cg]

[+anterior, +sg][-anterior, -sg]

[+anterior, +sg][-anterior, -sonorant]

[+anterior, -sg][+sg]

[-anterior, -cg][+cg]

[-anterior, -cg][+RTR, +sg]

[-anterior, -cg][+dorsal, +sg]

[-anterior, -continuant][-anterior, +sg]

[-anterior, -continuant][-RTR, +dorsal, +sg]

[-anterior, -lateral][+lateral]

[-anterior, -lateral][-anterior, -cg]

[-anterior, -lateral][-anterior, +continuant]

[-anterior, -lateral][-anterior, -nasal]

[-anterior, -lateral][-anterior, +sg]

[-anterior, -lateral][-cg, +continuant]

[-anterior, -lateral][+continuant, +sg]

[-anterior, -lateral][+labial, +sg]

[-anterior, -lateral][-RTR, +dorsal, +sg]

[-anterior, +nasal][+lateral]

[-anterior, +nasal][-anterior, -cg]

[-anterior, +nasal][-anterior, +continuant]

[-anterior, +nasal][-anterior, -nasal]
[-anterior, +nasal][-anterior, +sg]
[-anterior, +nasal][-cg, +continuant]
[-anterior, +nasal][+continuant, +sg]
[-anterior, +nasal][+labial, +sg]
[-anterior, +nasal][-RTR, +dorsal, +sg]
[-anterior, +sg][+cg]
[-anterior, +sg][+sg]
[-anterior, +sg][-anterior, -lateral]
[-anterior, +sg][-anterior, +nasal]
[-anterior, +sg][-anterior, -continuant, +sonorant]
[-anterior, -sg][+dorsal, +sg]
[-anterior, -sonorant][+cg]
[-anterior, -sonorant][+RTR, +sg]
[-anterior, -sonorant][+dorsal, +sg]
[-back, -lateral][+anterior, +sg]
[-back, -lateral][-anterior, +sg]
[-back, -lateral][+cg, +labial]
[-back, -lateral][-cg, +continuant]
[-back, -lateral][+continuant, +sg]
[-back, -lateral][+coronal, +sg]
[-back, -nasal][+anterior, +sg]
[-back, -nasal][-anterior, +sg]
[-back, -nasal][+cg, +labial]
[-back, -nasal][-cg, +continuant]
[-back, -nasal][+continuant, +sg]
[-back, -nasal][+coronal, +sg]
[-back, -syllabic][+anterior, +sg]
[-back, -syllabic][-anterior, +sg]
[-back, -syllabic][+cg, +labial]
[-back, -syllabic][-cg, +continuant]
[-back, -syllabic][+continuant, +sg]
[-back, -syllabic][+coronal, +sg]
[+cg, +labial][+sg]
[+cg, +labial][-cg, +continuant]
[-cg, +continuant][-RTR, +sg]
[-cg, +continuant][+labial, +sg]
[-cg, -continuant][+cg]
[-cg, -continuant][+RTR, +sg]
[-cg, +coronal][+cg]
[-cg, +coronal][+RTR, +sg]
[-cg, +coronal][+dorsal, +sg]
[-cg, +dorsal][+cg]
[-cg, +dorsal][+RTR, +sg]
[-cg, +dorsal][+labial, +sg]
[-cg, +labial][+cg]
[-cg, +labial][+sg]
[+continuant, -sonorant][-cg, +continuant]
[+continuant, -sonorant][+continuant, +sg]
[-continuant, +dorsal][+cg]
[-continuant, +dorsal][+RTR, +sg]
[-continuant, +dorsal][+anterior, +sg]
[-continuant, +dorsal][-anterior, +sg]
[-continuant, +dorsal][+coronal, +sg]
[-continuant, +dorsal][+labial, +sg]
[-continuant, +sg][+cg]
[-continuant, +sg][+sg]
[-continuant, -sonorant][+cg]
[-continuant, -sonorant][+RTR, +sg]
[+coronal, -lateral][-cg, +continuant]
[+coronal, -lateral][+continuant, +sg]
[+coronal, +nasal][-cg, +continuant]

[+coronal, +nasal][+continuant, +sg]

[+coronal, +nasal][+labial, +sg]

[+coronal, +sg][+cg]

[+coronal, +sg][+sg]

[+coronal, -sg][+dorsal, +sg]

[+dorsal, +sg][+cg]

[+dorsal, +sg][+sg]

[+dorsal, -sg][+labial, +sg]

[+dorsal, -sonorant][+RTR, +sg]

[+dorsal, -sonorant][+cg, +labial]

[+dorsal, -sonorant][+labial, +sg]

[+dorsal, -sonorant][-RTR, +cg, +dorsal]

[+dorsal, -syllabic][+RTR, +sg]

[+dorsal, -syllabic][+anterior, +sg]

[+dorsal, -syllabic][-anterior, +sg]

[+dorsal, -syllabic][+cg, +labial]

[+dorsal, -syllabic][+coronal, +sg]

[+dorsal, -syllabic][+labial, +sg]

[+dorsal, -syllabic][-RTR, +cg, +dorsal]

[+labial, +sg][+cg]

[+labial, +sg][+sg]

[+labial, -sg][+sg]

[+labial, -sg][-cg, +continuant]

[+labial, -sonorant][+cg]

[+labial, -sonorant][+sg]

[+labial, -sonorant][-cg, +continuant]

[-RTR, -back, +dorsal][+RTR, +back]

[-RTR, -back, +dorsal][+RTR, -back]

[-RTR, -back, +dorsal][+RTR, +continuant]

[-RTR, -back, +dorsal][+RTR, -low]

[-RTR, -back, +dorsal][+RTR, +sg]

[-RTR, -back, +dorsal][+RTR, +sonorant]

[-RTR, -back, +dorsal][+RTR, +syllabic]

[-RTR, -back, +syllabic][+RTR, +back]

[-RTR, -back, +syllabic][+RTR, -back]

[-RTR, -back, +syllabic][+RTR, +continuant]

[-RTR, -back, +syllabic][+RTR, -low]

[-RTR, -back, +syllabic][+RTR, +sg]

[-RTR, -back, +syllabic][+RTR, +sonorant]

[-RTR, -back, +syllabic][+RTR, +syllabic]

[-RTR, +cg, +dorsal][+RTR]

[-RTR, -cg, +dorsal][+RTR, -back]

[-RTR, -continuant, +dorsal][+RTR, -back]

[-RTR, +dorsal, -low][+RTR, +sg]

[-RTR, +dorsal, +sg][+RTR]

[-RTR, +dorsal, +sg][+continuant, +dorsal, -sonorant]

[-RTR, +dorsal, +sg][+continuant, +dorsal, -syllabic]

[-RTR, +dorsal, -sg][+RTR, -back]

[-RTR, +dorsal, -sonorant][+RTR, -back]

[-RTR, +dorsal, -syllabic][+RTR, -back]

[-RTR, -low, +syllabic][+RTR, +sg]

[+anterior, -cg, -sg][-anterior, -lateral]

[+anterior, -cg, -sg][-anterior, +nasal]

[+anterior, -cg, -sg][-anterior, -continuant, +sonorant]

[+anterior, -continuant, -lateral][-back, -lateral]

[+anterior, -continuant, -lateral][-back, -nasal]

[+anterior, -continuant, -lateral][-back, -syllabic]

[+anterior, -continuant, -lateral][+continuant, +dorsal, -sonorant]

[+anterior, -continuant, -lateral][+continuant, +dorsal, -syllabic]

[+anterior, -continuant, +sonorant][-back, -lateral]

[+anterior, -continuant, +sonorant][-back, -nasal]

[+anterior, -continuant, +sonorant][-back, -syllabic]

[+anterior, -continuant, +sonorant][+continuant, +dorsal, -sonorant]

[+anterior, -continuant, +sonorant][+continuant, +dorsal, -syllabic]

[+anterior, -continuant, -sonorant][+sg]

[-anterior, -continuant, +sonorant][+lateral]

[-anterior, -continuant, +sonorant][-anterior, -cg]

[-anterior, -continuant, +sonorant][-anterior, +continuant]

[-anterior, -continuant, +sonorant][-anterior, -nasal]

[-anterior, -continuant, +sonorant][-cg, +continuant]

[-anterior, -continuant, +sonorant][+continuant, +sg]

[-anterior, -continuant, +sonorant][+labial, +sg]

[+continuant, +coronal, -lateral][+anterior, +sg]

[+continuant, +coronal, -lateral][-anterior, +sg]

[+continuant, +coronal, -lateral][+coronal, +sg]

[+continuant, +dorsal, -sonorant][+sg]

[+continuant, +dorsal, -sonorant][+RTR, +back]

[+continuant, +dorsal, -sonorant][+RTR, -back]

[+continuant, +dorsal, -sonorant][+RTR, +continuant]

[+continuant, +dorsal, -sonorant][+RTR, -low]

[+continuant, +dorsal, -sonorant][+RTR, +sonorant]

[+continuant, +dorsal, -sonorant][+RTR, +syllabic]

[+continuant, +dorsal, -sonorant][+continuant, +dorsal, -sonorant]

[+continuant, +dorsal, -sonorant][+continuant, +dorsal, -syllabic]

[+continuant, +dorsal, -syllabic][+sg]

[+continuant, +dorsal, -syllabic][+RTR, +back]

[+continuant, +dorsal, -syllabic][+RTR, -back]

[+continuant, +dorsal, -syllabic][+RTR, +continuant]

[+continuant, +dorsal, -syllabic][+RTR, -low]

[+continuant, +dorsal, -syllabic][+RTR, +sonorant]

[+continuant, +dorsal, -syllabic][+RTR, +syllabic]

[+continuant, +dorsal, -syllabic][-cg, +continuant]

[+continuant, +dorsal, -syllabic][+continuant, +dorsal, -sonorant]

[+continuant, +dorsal, -syllabic][+continuant, +dorsal, -syllabic]

[-continuant, +coronal, -lateral][+labial, +sg]

[-continuant, +coronal, +sonorant][-cg, +continuant]

[-continuant, +coronal, +sonorant][+continuant, +sg]

[-continuant, +coronal, +sonorant][+labial, +sg]

[-continuant, +coronal, -sonorant][+dorsal, +sg]

[+coronal, -lateral, -nasal][+anterior, +sg]

[+coronal, -lateral, -nasal][-anterior, +sg]

[+coronal, -lateral, -nasal][+coronal, +sg]

A.2.2 Forbidden Successor Structures

[+anterior, +lateral]

[+RTR, +back, -syllabic]

[+RTR, -back, -cg]

[+RTR, -back, -continuant]

[+RTR, -back, -sg]

[+RTR, -back, -sonorant]

[+RTR, -back, -syllabic]

[+RTR, -cg, +continuant]

[+RTR, -cg, -low]

[+RTR, +continuant, +sg]

[+RTR, +continuant, -sonorant]

[+RTR, +continuant, -syllabic]

[+RTR, -continuant, -low]

[+RTR, -continuant, +sonorant]

[+RTR, -low, -sg]

[+RTR, -low, -sonorant]

[+RTR, -low, -syllabic]

[+RTR, +sonorant, -syllabic]

[+anterior, -cg, +continuant]

[+anterior, -cg, -lateral]

[+anterior, -cg, -nasal]

[+anterior, +continuant, +sg]

[+anterior, -continuant, -nasal]

[+anterior, -lateral, -sg]

[+anterior, -lateral, -sonorant]

[+anterior, -nasal, -sg]

[+anterior, -nasal, -sonorant]

[-anterior, -back, +continuant]

[-anterior, -back, -mb]

[-anterior, -back, +sonorant]

[-anterior, -cg, +continuant]

[-anterior, +continuant, -lateral]

[-anterior, +continuant, -low]

[-anterior, +continuant, +sg]

[-anterior, +continuant, -sonorant]

[-anterior, -low, -mb]

[-anterior, -low, +sonorant]

[+back, +dorsal, -syllabic]

[-back, -cg, +continuant]

[-back, -cg, +dorsal]

[-back, -cg, -mb]

[-back, +continuant, -sonorant]

[-back, -continuant, +dorsal]

[-back, -continuant, -mb]

[-back, -continuant, +sonorant]

[-back, +dorsal, -sg]

[-back, +dorsal, -sonorant]

[-back, +dorsal, -syllabic]

[-back, -mb, -sg]

[-back, -mb, -sonorant]

[-cg, +continuant, +coronal]

[-cg, +continuant, +dorsal]

[-cg, +continuant, -lateral]

[-cg, +continuant, -low]

[-cg, +continuant, -nasal]

[-cg, +coronal, -lateral]

[-cg, +coronal, -nasal]

[-cg, +dorsal, -low]

[-cg, +labial, -lateral]

[-cg, -lateral, -mb]

[-cg, -low, -mb]

[-cg, -mb, -nasal]

[+continuant, +coronal, +sg]

[+continuant, +dorsal, +sg]

[+continuant, -lateral, -sonorant]

[+continuant, -low, -sonorant]

[+continuant, -nasal, -sonorant]

[-continuant, +coronal, -nasal]

[-continuant, +dorsal, -low]

[-continuant, +dorsal, +sonorant]

[-continuant, -low, -mb]

[-continuant, -low, +sonorant]

[-continuant, -mb, -nasal]

[-continuant, -nasal, +sonorant]

[+coronal, -lateral, -sg]

[+coronal, -lateral, -sonorant]

[+coronal, -nasal, -sg]

[+coronal, -nasal, -sonorant]

[+dorsal, -low, -sg]

[+dorsal, -low, -sonorant]

[+dorsal, -low, -syllabic]

[+dorsal, +sonorant, -syllabic]

[+labial, -lateral, -sg]

[+labial, -lateral, -sonorant]

[-lateral, -mb, -sg]

[-lateral, -mb, -sonorant]

[-low, -mb, -sg]

[-low, -mb, -sonorant]

[-mb, -nasal, -sg]

[-mb, -nasal, -sonorant]

[+RTR][-RTR, +cg]

[+RTR][-RTR, +sg]

[+RTR][+anterior, +cg]

[+RTR][+anterior, +sg]

[+RTR][-anterior, +cg]

[+RTR][-anterior, +sg]

[+RTR][+cg, +coronal]

[+RTR][+cg, +labial]

[+RTR][-cg, +continuant]

[+RTR][+continuant, +sg]

[+RTR][+coronal, +sg]

[+RTR][+labial, +sg]

[+RTR][+continuant, +dorsal, -sonorant]

[+RTR][+continuant, +dorsal, -syllabic]

[+anterior][-cg, +continuant]

[+anterior][+continuant, +sg]

[+anterior][+continuant, +dorsal, -sonorant]

[+anterior][+continuant, +dorsal, -syllabic]

[-anterior][-coronal]

[-anterior][-dorsal]

[-anterior][-labial]

[-anterior][+mb]

[-anterior][-anterior, -back]

[-anterior][-anterior, -low]

[-anterior][-back, -cg]
[-anterior][-back, -continuant]
[-anterior][-back, -sg]
[-anterior][-back, -sonorant]
[-anterior][-cg, -lateral]
[-anterior][-cg, -low]
[-anterior][-cg, -nasal]
[-anterior][-continuant, -low]
[-anterior][-continuant, -nasal]
[-anterior][-lateral, -sg]
[-anterior][-lateral, -sonorant]
[-anterior][-low, -sg]
[-anterior][-low, -sonorant]
[-anterior][-nasal, -sg]
[-anterior][-nasal, -sonorant]
[-anterior][-anterior, -lateral, -nasal]
[-anterior][+continuant, +dorsal, -sonorant]
[-anterior][+continuant, +dorsal, -syllabic]
[+back][+RTR, +back]
[-back][+RTR, -back]
[+cg][+anterior]
[+cg][-anterior]
[+cg][+cg]
[+cg][-cg]
[+cg][-continuant]
[+cg][+coronal]
[+cg][-coronal]
[+cg][-dorsal]
[+cg][+labial]
[+cg][-labial]
[+cg][+lateral]
[+cg][-lateral]
[+cg][+mb]
[+cg][+nasal]
[+cg][-nasal]
[+cg][+sg]
[+cg][-sg]
[+cg][-sonorant]
[+cg][-syllabic]
[-cg][-coronal]
[-cg][-dorsal]
[-cg][-labial]
[-cg][+mb]
[-cg][-anterior, -back]
[-cg][-anterior, -low]
[-cg][-back, -cg]
[-cg][-back, -continuant]
[-cg][-back, -sg]
[-cg][-back, -sonorant]
[-cg][-cg, -lateral]
[-cg][-cg, -low]
[-cg][-cg, -nasal]
[-cg][-continuant, -low]
[-cg][-continuant, -nasal]
[-cg][-lateral, -sg]
[-cg][-lateral, -sonorant]
[-cg][-low, -sg]
[-cg][-low, -sonorant]
[-cg][-nasal, -sg]
[-cg][-nasal, -sonorant]
[-cg][-anterior, -lateral, -nasal]
[-cg][+continuant, +dorsal, -sonorant]

[-cg][+continuant, +dorsal, -syllabic]
[-continuant][+continuant, +dorsal, -sonorant]
[-continuant][+continuant, +dorsal, -syllabic]
[+coronal][-cg, +continuant]
[+coronal][+continuant, +sg]
[+coronal][+continuant, +dorsal, -sonorant]
[+coronal][+continuant, +dorsal, -syllabic]
[-coronal][-coronal]
[-coronal][-dorsal]
[-coronal][-labial]
[-coronal][+mb]
[-coronal][+RTR, -back]
[-coronal][-anterior, -back]
[-coronal][-anterior, -low]
[-coronal][-back, -cg]
[-coronal][-back, -continuant]
[-coronal][-back, -sg]
[-coronal][-back, -sonorant]
[-coronal][-cg, -lateral]
[-coronal][-cg, -low]
[-coronal][-cg, -nasal]
[-coronal][-continuant, -low]
[-coronal][-continuant, -nasal]
[-coronal][-lateral, -sg]
[-coronal][-lateral, -sonorant]
[-coronal][-low, -sg]
[-coronal][-low, -sonorant]
[-coronal][-nasal, -sg]
[-coronal][-nasal, -sonorant]
[-coronal][-anterior, -lateral, -nasal]
[-coronal][+continuant, +dorsal, -sonorant]
[-coronal][+continuant, +dorsal, -syllabic]
[-dorsal][-coronal]
[-dorsal][-dorsal]
[-dorsal][-labial]
[-dorsal][+mb]
[-dorsal][+RTR, -back]
[-dorsal][-anterior, -back]
[-dorsal][-anterior, -low]
[-dorsal][-back, -cg]
[-dorsal][-back, -continuant]
[-dorsal][-back, -sg]
[-dorsal][-back, -sonorant]
[-dorsal][-cg, -lateral]
[-dorsal][-cg, -low]
[-dorsal][-cg, -nasal]
[-dorsal][-continuant, -low]
[-dorsal][-continuant, -nasal]
[-dorsal][-lateral, -sg]
[-dorsal][-lateral, -sonorant]
[-dorsal][-low, -sg]
[-dorsal][-low, -sonorant]
[-dorsal][-nasal, -sg]
[-dorsal][-nasal, -sonorant]
[-dorsal][-anterior, -lateral, -nasal]
[-dorsal][+continuant, +dorsal, -sonorant]
[-dorsal][+continuant, +dorsal, -syllabic]
[+labial][-coronal]
[+labial][-dorsal]
[+labial][-labial]
[+labial][+mb]
[+labial][+RTR, +cg]

[+labial][+RTR, +sg]

[+labial][+anterior, -cg]

[+labial][+anterior, +nasal]

[+labial][+anterior, +sg]

[+labial][-anterior, -back]

[+labial][-anterior, -low]

[+labial][-anterior, +sg]

[+labial][+back, -lateral]

[+labial][+back, -nasal]

[+labial][+back, -syllabic]

[+labial][-back, -cg]

[+labial][-back, -continuant]

[+labial][-back, -sg]

[+labial][-back, -sonorant]

[+labial][-cg, +continuant]

[+labial][-cg, -lateral]

[+labial][-cg, -low]

[+labial][-cg, -nasal]

[+labial][+continuant, +sg]

[+labial][-continuant, -low]

[+labial][-continuant, -nasal]

[+labial][+coronal, +sg]

[+labial][+labial, -lateral]

[+labial][+labial, +nasal]

[+labial][+labial, +sonorant]

[+labial][-lateral, -sg]

[+labial][-lateral, -sonorant]

[+labial][-low, -sg]

[+labial][-low, -sonorant]

[+labial][-nasal, -sg]

[+labial][-nasal, -sonorant]

[+labial][+anterior, -continuant, -lateral]

[+labial][+anterior, -continuant, +sonorant]

[+labial][-anterior, -lateral, -nasal]

[+labial][+continuant, +dorsal, -sonorant]

[+labial][+continuant, +dorsal, -syllabic]

[-labial][-coronal]

[-labial][-dorsal]

[-labial][-labial]

[-labial][+mb]

[-labial][+RTR, -back]

[-labial][-anterior, -back]

[-labial][-anterior, -low]

[-labial][-back, -cg]

[-labial][-back, -continuant]

[-labial][-back, -sg]

[-labial][-back, -sonorant]

[-labial][-cg, -lateral]

[-labial][-cg, -low]

[-labial][-cg, -nasal]

[-labial][-continuant, -low]

[-labial][-continuant, -nasal]

[-labial][-lateral, -sg]

[-labial][-lateral, -sonorant]

[-labial][-low, -sg]

[-labial][-low, -sonorant]

[-labial][-nasal, -sg]

[-labial][-nasal, -sonorant]

[-labial][-anterior, -lateral, -nasal]

[-labial][+continuant, +dorsal, -sonorant]

[-labial][+continuant, +dorsal, -syllabic]

[+lateral][-coronal]

[+lateral][-dorsal]
[+lateral][-labial]
[+lateral][+lateral]
[+lateral][+mb]
[+lateral][+RTR, +sg]
[+lateral][+anterior, +cg]
[+lateral][+anterior, -lateral]
[+lateral][+anterior, +nasal]
[+lateral][+anterior, -nasal]
[+lateral][+anterior, +sonorant]
[+lateral][-anterior, -back]
[+lateral][-anterior, +continuant]
[+lateral][-anterior, -lateral]
[+lateral][-anterior, -low]
[+lateral][-anterior, +nasal]
[+lateral][-anterior, -nasal]
[+lateral][-anterior, +sonorant]
[+lateral][-back, -cg]
[+lateral][-back, -continuant]
[+lateral][-back, -lateral]
[+lateral][-back, -nasal]
[+lateral][-back, -sg]
[+lateral][-back, -sonorant]
[+lateral][-back, -syllabic]
[+lateral][-cg, +continuant]
[+lateral][-cg, -lateral]
[+lateral][-cg, -low]
[+lateral][-cg, -nasal]
[+lateral][+continuant, +sg]
[+lateral][-continuant, -low]
[+lateral][-continuant, -nasal]
[+lateral][+coronal, -lateral]
[+lateral][+coronal, +nasal]
[+lateral][+coronal, -nasal]
[+lateral][+coronal, +sonorant]
[+lateral][-lateral, -sg]
[+lateral][-lateral, -sonorant]
[+lateral][-low, -sg]
[+lateral][-low, -sonorant]
[+lateral][-nasal, -sg]
[+lateral][-nasal, -sonorant]
[+lateral][+continuant, +dorsal, -sonorant]
[+lateral][+continuant, +dorsal, -syllabic]
[-lateral][+continuant, +dorsal, -sonorant]
[-lateral][+continuant, +dorsal, -syllabic]
[+low][+low]
[+low][+syllabic]
[+low][+RTR, +back]
[+low][+RTR, -back]
[+low][+RTR, +continuant]
[+low][+RTR, -low]
[+low][+RTR, +sonorant]
[+low][-anterior, +sg]
[+low][+back, +dorsal]
[+low][-back, +dorsal]
[+low][+dorsal, -low]
[+low][+dorsal, +sonorant]
[+mb][-coronal]
[+mb][-dorsal]
[+mb][-labial]
[+mb][+mb]
[+mb][+RTR, -back]

[+mb][-anterior, -back]

[+mb][-anterior, -low]

[+mb][-back, -cg]

[+mb][-back, -continuant]

[+mb][-back, -sg]

[+mb][-back, -sonorant]

[+mb][-cg, -lateral]

[+mb][-cg, -low]

[+mb][-cg, -nasal]

[+mb][-continuant, -low]

[+mb][-continuant, -nasal]

[+mb][-lateral, -sg]

[+mb][-lateral, -sonorant]

[+mb][-low, -sg]

[+mb][-low, -sonorant]

[+mb][-nasal, -sg]

[+mb][-nasal, -sonorant]

[+mb][-anterior, -lateral, -nasal]

[+mb][+continuant, +dorsal, -sonorant]

[+mb][+continuant, +dorsal, -syllabic]

[+nasal][+anterior, +nasal]

[+nasal][-cg, +continuant]

[+nasal][+continuant, +sg]

[+nasal][+labial, -lateral]

[+nasal][+labial, +nasal]

[+nasal][+labial, +sonorant]

[+nasal][+anterior, -continuant, -lateral]

[+nasal][+anterior, -continuant, +sonorant]

[+nasal][+continuant, +dorsal, -sonorant]

[+nasal][+continuant, +dorsal, -syllabic]

[-nasal][+continuant, +dorsal, -sonorant]

[-nasal][+continuant, +dorsal, -syllabic]

[+sg][+anterior]

[+sg][-anterior]

[+sg][+cg]

[+sg][-cg]

[+sg][-continuant]

[+sg][+coronal]

[+sg][-coronal]

[+sg][-dorsal]

[+sg][+labial]

[+sg][-labial]

[+sg][+lateral]

[+sg][-lateral]

[+sg][+mb]

[+sg][+nasal]

[+sg][-nasal]

[+sg][+sg]

[+sg][-sg]

[+sg][-sonorant]

[+sg][-syllabic]

[-sg][-coronal]

[-sg][-dorsal]

[-sg][-labial]

[-sg][+mb]

[-sg][-anterior, -back]

[-sg][-anterior, -low]

[-sg][-back, -cg]

[-sg][-back, -continuant]

[-sg][-back, -sg]

[-sg][-back, -sonorant]

[-sg][-cg, -lateral]

[-sg][-cg, -low]
[-sg][-cg, -nasal]
[-sg][-continuant, -low]
[-sg][-continuant, -nasal]
[-sg][-lateral, -sg]
[-sg][-lateral, -sonorant]
[-sg][-low, -sg]
[-sg][-low, -sonorant]
[-sg][-nasal, -sg]
[-sg][-nasal, -sonorant]
[-sg][-anterior, -lateral, -nasal]
[-sg][+continuant, +dorsal, -sonorant]
[-sg][+continuant, +dorsal, -syllabic]
[-sonorant][+continuant, +dorsal, -sonorant]
[-sonorant][+continuant, +dorsal, -syllabic]
[+syllabic][+low]
[+syllabic][+syllabic]
[+syllabic][+RTR, +back]
[+syllabic][+RTR, -back]
[+syllabic][+RTR, +continuant]
[+syllabic][+RTR, -low]
[+syllabic][+RTR, +sonorant]
[+syllabic][+back, +dorsal]
[+syllabic][-back, +dorsal]
[+syllabic][+dorsal, -low]
[+syllabic][+dorsal, +sonorant]
[-syllabic][+continuant, +dorsal, -sonorant]
[-syllabic][+continuant, +dorsal, -syllabic]
[+RTR, +back][+low]
[+RTR, +back][+syllabic]
[+RTR, +back][+RTR, -back]
[+RTR, +back][+RTR, +continuant]
[+RTR, +back][+RTR, -low]
[+RTR, +back][+RTR, +sonorant]
[+RTR, +back][+back, +dorsal]
[+RTR, +back][-back, +dorsal]
[+RTR, +back][+continuant, +dorsal]
[+RTR, +back][+dorsal, -low]
[+RTR, +back][+dorsal, +sonorant]
[+RTR, -back][+low]
[+RTR, -back][+syllabic]
[+RTR, -back][+RTR, +back]
[+RTR, -back][+RTR, +continuant]
[+RTR, -back][+RTR, -low]
[+RTR, -back][+RTR, +sonorant]
[+RTR, -back][-RTR, +dorsal]
[+RTR, -back][-anterior, +nasal]
[+RTR, -back][+back, +dorsal]
[+RTR, -back][-back, +dorsal]
[+RTR, -back][+continuant, +dorsal]
[+RTR, -back][+dorsal, -low]
[+RTR, -back][+dorsal, +sonorant]
[+RTR, -back][-anterior, -continuant, +sonorant]
[+RTR, -back][-anterior, +coronal, -lateral]
[+RTR, -back][-anterior, -lateral, -mb]
[+RTR, -back][-anterior, -lateral, +sonorant]
[+RTR, +cg][-RTR, +back, -low]
[+RTR, -cg][+anterior]
[+RTR, -cg][-anterior]
[+RTR, -cg][+cg]
[+RTR, -cg][-cg]
[+RTR, -cg][-continuant]

[+RTR, -cg][+coronal]

[+RTR, -cg][+labial]

[+RTR, -cg][+lateral]

[+RTR, -cg][-lateral]

[+RTR, -cg][+nasal]

[+RTR, -cg][-nasal]

[+RTR, -cg][+sg]

[+RTR, -cg][-sg]

[+RTR, -cg][-sonorant]

[+RTR, -cg][-syllabic]

[+RTR, -cg][-RTR, +back, -low]

[+RTR, +continuant][+low]

[+RTR, +continuant][+syllabic]

[+RTR, +continuant][+RTR, +back]

[+RTR, +continuant][+RTR, -back]

[+RTR, +continuant][+RTR, +continuant]

[+RTR, +continuant][+RTR, -low]

[+RTR, +continuant][+RTR, +sonorant]

[+RTR, +continuant][+back, +dorsal]

[+RTR, +continuant][-back, +dorsal]

[+RTR, +continuant][+continuant, +dorsal]

[+RTR, +continuant][+dorsal, -low]

[+RTR, +continuant][+dorsal, +sonorant]

[+RTR, -continuant][+anterior]

[+RTR, -continuant][-anterior]

[+RTR, -continuant][+cg]

[+RTR, -continuant][-cg]

[+RTR, -continuant][-continuant]

[+RTR, -continuant][+coronal]

[+RTR, -continuant][-coronal]

[+RTR, -continuant][-dorsal]

[+RTR, -continuant][+labial]

[+RTR, -continuant][-labial]

[+RTR, -continuant][+lateral]

[+RTR, -continuant][-lateral]

[+RTR, -continuant][+mb]

[+RTR, -continuant][+nasal]

[+RTR, -continuant][-nasal]

[+RTR, -continuant][+sg]

[+RTR, -continuant][-sg]

[+RTR, -continuant][-sonorant]

[+RTR, -continuant][-syllabic]

[+RTR, -continuant][-RTR, +back, -low]

[+RTR, -low][+low]

[+RTR, -low][+syllabic]

[+RTR, -low][+RTR, +back]

[+RTR, -low][+RTR, -back]

[+RTR, -low][+RTR, +continuant]

[+RTR, -low][+RTR, -low]

[+RTR, -low][+RTR, +sonorant]

[+RTR, -low][+back, +dorsal]

[+RTR, -low][-back, +dorsal]

[+RTR, -low][+continuant, +dorsal]

[+RTR, -low][+dorsal, -low]

[+RTR, -low][+dorsal, +sonorant]

[+RTR, +sg][-RTR, -back]

[+RTR, +sg][-RTR, -low]

[+RTR, -sg][+anterior]

[+RTR, -sg][-anterior]

[+RTR, -sg][+cg]

[+RTR, -sg][-cg]

[+RTR, -sg][-continuant]

[+RTR, -sg][+coronal]
[+RTR, -sg][+labial]
[+RTR, -sg][+lateral]
[+RTR, -sg][-lateral]
[+RTR, -sg][+nasal]
[+RTR, -sg][-nasal]
[+RTR, -sg][+sg]
[+RTR, -sg][-sg]
[+RTR, -sg][-sonorant]
[+RTR, -sg][-syllabic]
[+RTR, -sg][-RTR, +back, -low]
[+RTR, +sonorant][+low]
[+RTR, +sonorant][+syllabic]
[+RTR, +sonorant][+RTR, +back]
[+RTR, +sonorant][+RTR, -back]
[+RTR, +sonorant][+RTR, +continuant]
[+RTR, +sonorant][+RTR, -low]
[+RTR, +sonorant][+RTR, +sonorant]
[+RTR, +sonorant][+back, +dorsal]
[+RTR, +sonorant][-back, +dorsal]
[+RTR, +sonorant][+continuant, +dorsal]
[+RTR, +sonorant][+dorsal, -low]
[+RTR, +sonorant][+dorsal, +sonorant]
[+RTR, -sonorant][+anterior]
[+RTR, -sonorant][-anterior]
[+RTR, -sonorant][+cg]
[+RTR, -sonorant][-cg]
[+RTR, -sonorant][-continuant]
[+RTR, -sonorant][+coronal]
[+RTR, -sonorant][-coronal]
[+RTR, -sonorant][-dorsal]
[+RTR, -sonorant][+labial]
[+RTR, -sonorant][-labial]
[+RTR, -sonorant][+lateral]
[+RTR, -sonorant][-lateral]
[+RTR, -sonorant][+mb]
[+RTR, -sonorant][+nasal]
[+RTR, -sonorant][-nasal]
[+RTR, -sonorant][+sg]
[+RTR, -sonorant][-sg]
[+RTR, -sonorant][-sonorant]
[+RTR, -sonorant][-syllabic]
[+RTR, -sonorant][-RTR, +back, -low]
[+RTR, +syllabic][+continuant, +dorsal]
[+RTR, -syllabic][+anterior]
[+RTR, -syllabic][-anterior]
[+RTR, -syllabic][+cg]
[+RTR, -syllabic][-cg]
[+RTR, -syllabic][-continuant]
[+RTR, -syllabic][+coronal]
[+RTR, -syllabic][-coronal]
[+RTR, -syllabic][-dorsal]
[+RTR, -syllabic][+labial]
[+RTR, -syllabic][-labial]
[+RTR, -syllabic][+lateral]
[+RTR, -syllabic][-lateral]
[+RTR, -syllabic][+mb]
[+RTR, -syllabic][+nasal]
[+RTR, -syllabic][-nasal]
[+RTR, -syllabic][+sg]
[+RTR, -syllabic][-sg]
[+RTR, -syllabic][-sonorant]

[+RTR, -syllabic][-syllabic]

[+RTR, -syllabic][-RTR, +back, -low]

[-RTR, +dorsal][+RTR, -back]

[+anterior, -cg][+anterior]

[+anterior, -cg][-anterior]

[+anterior, -cg][+cg]

[+anterior, -cg][-cg]

[+anterior, -cg][+coronal]

[+anterior, -cg][+lateral]

[+anterior, -cg][-nasal]

[+anterior, -cg][+sg]

[+anterior, -cg][-sg]

[+anterior, -cg][-sonorant]

[+anterior, -cg][+RTR, -continuant]

[+anterior, -cg][+RTR, -syllabic]

[+anterior, -cg][+back, -lateral]

[+anterior, -cg][+back, -syllabic]

[+anterior, -cg][-back, -lateral]

[+anterior, -cg][-back, -syllabic]

[+anterior, -cg][+continuant, -lateral]

[+anterior, -cg][+continuant, -syllabic]

[+anterior, -cg][-continuant, +dorsal]

[+anterior, -cg][+dorsal, -syllabic]

[+anterior, -cg][-lateral, -low]

[+anterior, -cg][-low, -syllabic]

[+anterior, +continuant][-coronal]

[+anterior, +continuant][-dorsal]

[+anterior, +continuant][-labial]

[+anterior, +continuant][+mb]

[+anterior, +continuant][+anterior, +sg]

[+anterior, +continuant][-anterior, -back]

[+anterior, +continuant][-anterior, +cg]

[+anterior, +continuant][-anterior, -low]

[+anterior, +continuant][-anterior, +sg]

[+anterior, +continuant][-back, -cg]

[+anterior, +continuant][-back, -continuant]

[+anterior, +continuant][-back, -sg]

[+anterior, +continuant][-back, -sonorant]

[+anterior, +continuant][-cg, -lateral]

[+anterior, +continuant][-cg, -low]

[+anterior, +continuant][-cg, -nasal]

[+anterior, +continuant][-continuant, -low]

[+anterior, +continuant][-continuant, -nasal]

[+anterior, +continuant][+coronal, +sg]

[+anterior, +continuant][-lateral, -sg]

[+anterior, +continuant][-lateral, -sonorant]

[+anterior, +continuant][-low, -sg]

[+anterior, +continuant][-low, -sonorant]

[+anterior, +continuant][-nasal, -sg]

[+anterior, +continuant][-nasal, -sonorant]

[+anterior, +continuant][-anterior, -lateral, -nasal]

[+anterior, -continuant][+anterior, +nasal]

[+anterior, -continuant][-anterior, +nasal]

[+anterior, -continuant][+cg, +labial]

[+anterior, -continuant][-cg, +labial]

[+anterior, -continuant][+coronal, +nasal]

[+anterior, -continuant][+labial, +sg]

[+anterior, -continuant][+labial, -sg]

[+anterior, -continuant][+labial, -sonorant]

[+anterior, -continuant][+anterior, -continuant, -lateral]

[+anterior, -continuant][+anterior, -continuant, +sonorant]

[+anterior, -continuant][-anterior, -continuant, +sonorant]

[+anterior, -continuant][-anterior, +coronal, -lateral]

[+anterior, -continuant][-anterior, -lateral, -mb]

[+anterior, -continuant][-anterior, -lateral, +sonorant]

[+anterior, -continuant][-back, +continuant, -lateral]

[+anterior, -continuant][-back, +continuant, -nasal]

[+anterior, -continuant][-back, +continuant, -syllabic]

[+anterior, -continuant][-back, -lateral, -mb]

[+anterior, -continuant][-back, -lateral, +sonorant]

[+anterior, -continuant][-back, -mb, -nasal]

[+anterior, -continuant][-back, -mb, -syllabic]

[+anterior, -continuant][-back, -nasal, +sonorant]

[+anterior, -continuant][-back, +sonorant, -syllabic]

[+anterior, -continuant][-continuant, +coronal, -lateral]

[+anterior, -continuant][-continuant, +coronal, +sonorant]

[+anterior, -lateral][+labial, +sg]

[+anterior, +nasal][+labial]

[+anterior, +nasal][+nasal]

[+anterior, +nasal][-continuant, +sonorant]

[+anterior, +nasal][-anterior, +coronal, -lateral]

[+anterior, +nasal][-anterior, -lateral, -mb]

[+anterior, +nasal][-anterior, -lateral, +sonorant]

[+anterior, +nasal][-back, +continuant, -lateral]

[+anterior, +nasal][-back, +continuant, -nasal]

[+anterior, +nasal][-back, +continuant, -syllabic]

[+anterior, +nasal][-back, -lateral, -mb]

[+anterior, +nasal][-back, -lateral, +sonorant]

[+anterior, +nasal][-back, -mb, -nasal]

[+anterior, +nasal][-back, -mb, -syllabic]

[+anterior, +nasal][-back, -nasal, +sonorant]

[+anterior, +nasal][-back, +sonorant, -syllabic]

[+anterior, +nasal][-continuant, +coronal, -lateral]

[+anterior, +nasal][-continuant, -lateral, -mb]

[+anterior, -nasal][-coronal]

[+anterior, -nasal][-dorsal]

[+anterior, -nasal][-labial]

[+anterior, -nasal][+mb]

[+anterior, -nasal][+anterior, +cg]

[+anterior, -nasal][+anterior, -nasal]

[+anterior, -nasal][+anterior, +sg]

[+anterior, -nasal][-anterior, -back]

[+anterior, -nasal][-anterior, +cg]

[+anterior, -nasal][-anterior, -low]

[+anterior, -nasal][-anterior, +sg]

[+anterior, -nasal][-back, -cg]

[+anterior, -nasal][-back, -continuant]

[+anterior, -nasal][-back, -sg]

[+anterior, -nasal][-back, -sonorant]

[+anterior, -nasal][+cg, +coronal]

[+anterior, -nasal][-cg, -lateral]

[+anterior, -nasal][-cg, -low]

[+anterior, -nasal][-cg, -nasal]

[+anterior, -nasal][-continuant, -low]

[+anterior, -nasal][-continuant, -nasal]

[+anterior, -nasal][+coronal, +sg]

[+anterior, -nasal][+labial, +sg]

[+anterior, -nasal][-lateral, -sg]

[+anterior, -nasal][-lateral, -sonorant]

[+anterior, -nasal][-low, -sg]

[+anterior, -nasal][-low, -sonorant]

[+anterior, -nasal][-nasal, -sg]

[+anterior, -nasal][-nasal, -sonorant]

[+anterior, -nasal][+anterior, +continuant, -lateral]

[+anterior, -nasal][+anterior, +continuant, +sonorant]

[+anterior, -nasal][-anterior, -lateral, -nasal]

[+anterior, -nasal][+continuant, +coronal, -lateral]

[+anterior, -nasal][+coronal, -lateral, -nasal]

[+anterior, +sg][+RTR, -back]

[+anterior, -sg][+anterior]

[+anterior, -sg][-anterior]

[+anterior, -sg][+cg]

[+anterior, -sg][-cg]

[+anterior, -sg][+coronal]

[+anterior, -sg][+lateral]

[+anterior, -sg][-nasal]

[+anterior, -sg][+sg]

[+anterior, -sg][-sg]

[+anterior, -sg][-sonorant]

[+anterior, -sg][+RTR, -continuant]

[+anterior, -sg][+RTR, -syllabic]

[+anterior, -sg][+back, -lateral]

[+anterior, -sg][+back, -syllabic]

[+anterior, -sg][-back, -lateral]

[+anterior, -sg][-back, -syllabic]

[+anterior, -sg][+continuant, -lateral]

[+anterior, -sg][+continuant, -syllabic]

[+anterior, -sg][-continuant, +dorsal]

[+anterior, -sg][+dorsal, -syllabic]

[+anterior, -sg][-lateral, -low]

[+anterior, -sg][-low, -syllabic]

[+anterior, +sonorant][+labial, +sg]

[+anterior, -sonorant][-coronal]

[+anterior, -sonorant][-dorsal]

[+anterior, -sonorant][-labial]

[+anterior, -sonorant][+mb]

[+anterior, -sonorant][+RTR, +sg]

[+anterior, -sonorant][+anterior, +sg]

[+anterior, -sonorant][-anterior, -back]

[+anterior, -sonorant][-anterior, +cg]

[+anterior, -sonorant][-anterior, -low]

[+anterior, -sonorant][-anterior, +sg]

[+anterior, -sonorant][-back, -cg]

[+anterior, -sonorant][-back, -continuant]

[+anterior, -sonorant][-back, -sg]

[+anterior, -sonorant][-back, -sonorant]

[+anterior, -sonorant][-cg, -lateral]

[+anterior, -sonorant][-cg, -low]

[+anterior, -sonorant][-cg, -nasal]

[+anterior, -sonorant][+continuant, -sonorant]

[+anterior, -sonorant][-continuant, -low]

[+anterior, -sonorant][-continuant, -nasal]
[+anterior, -sonorant][+coronal, +sg]
[+anterior, -sonorant][-lateral, -sg]
[+anterior, -sonorant][-lateral, -sonorant]
[+anterior, -sonorant][-low, -sg]
[+anterior, -sonorant][-low, -sonorant]
[+anterior, -sonorant][-nasal, -sg]
[+anterior, -sonorant][-nasal, -sonorant]
[+anterior, -sonorant][-anterior, -lateral, -nasal]
[-anterior, +continuant][+lateral]
[-anterior, +continuant][+RTR, +sg]
[-anterior, +continuant][+anterior, +cg]
[-anterior, +continuant][+anterior, -lateral]
[-anterior, +continuant][+anterior, +nasal]
[-anterior, +continuant][+anterior, -nasal]
[-anterior, +continuant][+anterior, +sonorant]
[-anterior, +continuant][-anterior, +continuant]
[-anterior, +continuant][-anterior, -lateral]
[-anterior, +continuant][-anterior, +nasal]
[-anterior, +continuant][-anterior, -nasal]
[-anterior, +continuant][-anterior, +sonorant]
[-anterior, +continuant][-back, -lateral]
[-anterior, +continuant][-back, -nasal]
[-anterior, +continuant][-back, -syllabic]
[-anterior, +continuant][-cg, +continuant]
[-anterior, +continuant][+continuant, +sg]
[-anterior, +continuant][+coronal, -lateral]
[-anterior, +continuant][+coronal, +nasal]
[-anterior, +continuant][+coronal, -nasal]
[-anterior, +continuant][+coronal, +sonorant]
[-anterior, +coronal][+lateral]
[-anterior, +coronal][+RTR, +sg]
[-anterior, +coronal][+anterior, +cg]
[-anterior, +coronal][+anterior, -lateral]
[-anterior, +coronal][+anterior, +nasal]
[-anterior, +coronal][+anterior, -nasal]
[-anterior, +coronal][+anterior, +sonorant]
[-anterior, +coronal][-anterior, +continuant]
[-anterior, +coronal][-anterior, -lateral]
[-anterior, +coronal][-anterior, +nasal]
[-anterior, +coronal][-anterior, -nasal]
[-anterior, +coronal][-anterior, +sonorant]
[-anterior, +coronal][-back, -lateral]
[-anterior, +coronal][-back, -nasal]
[-anterior, +coronal][-back, -syllabic]
[-anterior, +coronal][+coronal, -lateral]
[-anterior, +coronal][+coronal, +nasal]
[-anterior, +coronal][+coronal, -nasal]
[-anterior, +coronal][+coronal, +sonorant]
[-anterior, -low][+RTR, -back]
[-anterior, -mb][+lateral]
[-anterior, -mb][+RTR, +sg]
[-anterior, -mb][+anterior, +cg]
[-anterior, -mb][+anterior, -lateral]
[-anterior, -mb][+anterior, +nasal]
[-anterior, -mb][+anterior, -nasal]
[-anterior, -mb][+anterior, +sonorant]
[-anterior, -mb][-anterior, +continuant]
[-anterior, -mb][-anterior, -lateral]
[-anterior, -mb][-anterior, +nasal]
[-anterior, -mb][-anterior, -nasal]
[-anterior, -mb][-anterior, +sonorant]

[-anterior, -mb][-back, -lateral]
[-anterior, -mb][-back, -nasal]
[-anterior, -mb][-back, -syllabic]
[-anterior, -mb][-cg, +continuant]
[-anterior, -mb][+continuant, +sg]
[-anterior, -mb][+coronal, -lateral]
[-anterior, -mb][+coronal, +nasal]
[-anterior, -mb][+coronal, -nasal]
[-anterior, -mb][+coronal, +sonorant]
[-anterior, +nasal][+anterior]
[-anterior, +nasal][-anterior]
[-anterior, +nasal][+cg]
[-anterior, +nasal][-cg]
[-anterior, +nasal][-continuant]
[-anterior, +nasal][+coronal]
[-anterior, +nasal][+labial]
[-anterior, +nasal][+lateral]
[-anterior, +nasal][-lateral]
[-anterior, +nasal][+nasal]
[-anterior, +nasal][-nasal]
[-anterior, +nasal][+sg]
[-anterior, +nasal][-sg]
[-anterior, +nasal][-sonorant]
[-anterior, +nasal][-syllabic]
[-anterior, +sonorant][+lateral]
[-anterior, +sonorant][+RTR, +sg]
[-anterior, +sonorant][+anterior, +cg]
[-anterior, +sonorant][+anterior, -lateral]
[-anterior, +sonorant][+anterior, +nasal]
[-anterior, +sonorant][+anterior, -nasal]
[-anterior, +sonorant][+anterior, +sonorant]
[-anterior, +sonorant][-anterior, +continuant]
[-anterior, +sonorant][-anterior, -lateral]
[-anterior, +sonorant][-anterior, +nasal]
[-anterior, +sonorant][-anterior, -nasal]
[-anterior, +sonorant][-anterior, +sonorant]
[-anterior, +sonorant][-back, -lateral]
[-anterior, +sonorant][-back, -nasal]
[-anterior, +sonorant][-back, -syllabic]
[-anterior, +sonorant][-cg, +continuant]
[-anterior, +sonorant][+continuant, +sg]
[-anterior, +sonorant][+coronal, -lateral]
[-anterior, +sonorant][+coronal, +nasal]
[-anterior, +sonorant][+coronal, -nasal]
[-anterior, +sonorant][+coronal, +sonorant]
[+back, +dorsal][+low]
[+back, +dorsal][+syllabic]
[+back, +dorsal][+RTR, -back]
[+back, +dorsal][+RTR, +continuant]
[+back, +dorsal][+RTR, -low]
[+back, +dorsal][+RTR, +sonorant]
[+back, +dorsal][+back, +dorsal]
[+back, +dorsal][-back, +dorsal]
[+back, +dorsal][+dorsal, -low]
[+back, +dorsal][+dorsal, +sonorant]
[+back, -lateral][+anterior, +sg]
[+back, -lateral][-anterior, +nasal]
[+back, -lateral][+back, -lateral]
[+back, -lateral][+back, -nasal]
[+back, -lateral][+back, -syllabic]
[+back, -lateral][+cg, +labial]
[+back, -lateral][-cg, +continuant]

[+back, -lateral][+continuant, +sg]

[+back, -lateral][+labial, +sg]

[+back, -lateral][-anterior, -continuant, +sonorant]

[+back, -lateral][-anterior, +coronal, -lateral]

[+back, -lateral][-anterior, -lateral, -mb]

[+back, -lateral][-anterior, -lateral, +sonorant]

[+back, -nasal][+anterior, +sg]

[+back, -nasal][-anterior, +nasal]

[+back, -nasal][+back, -lateral]

[+back, -nasal][+back, -nasal]

[+back, -nasal][+back, -syllabic]

[+back, -nasal][+cg, +labial]

[+back, -nasal][-cg, +continuant]

[+back, -nasal][+continuant, +sg]

[+back, -nasal][+labial, +sg]

[+back, -nasal][-anterior, -continuant, +sonorant]

[+back, -nasal][-anterior, +coronal, -lateral]

[+back, -nasal][-anterior, -lateral, -mb]

[+back, -nasal][-anterior, -lateral, +sonorant]

[+back, -syllabic][+anterior, +sg]

[+back, -syllabic][-anterior, +nasal]

[+back, -syllabic][+back, -lateral]

[+back, -syllabic][+back, -nasal]

[+back, -syllabic][+back, -syllabic]

[+back, -syllabic][+cg, +labial]

[+back, -syllabic][-cg, +continuant]

[+back, -syllabic][+continuant, +sg]

[+back, -syllabic][+labial, +sg]

[+back, -syllabic][-anterior, -continuant, +sonorant]

[+back, -syllabic][-anterior, +coronal, -lateral]

[+back, -syllabic][-anterior, -lateral, -mb]

[+back, -syllabic][-anterior, -lateral, +sonorant]

[-back, +continuant][-anterior, +sg]

[-back, +continuant][-back, +dorsal]

[-back, +continuant][-back, +syllabic]

[-back, -continuant][-coronal]

[-back, -continuant][-dorsal]

[-back, -continuant][-labial]

[-back, -continuant][+mb]

[-back, -continuant][-anterior, -back]

[-back, -continuant][-anterior, -low]

[-back, -continuant][-back, -cg]

[-back, -continuant][-back, -continuant]

[-back, -continuant][-back, -sg]

[-back, -continuant][-back, -sonorant]

[-back, -continuant][-cg, -lateral]

[-back, -continuant][-cg, -low]

[-back, -continuant][-cg, -nasal]

[-back, -continuant][-continuant, -low]

[-back, -continuant][-continuant, -nasal]

[-back, -continuant][-lateral, -sg]

[-back, -continuant][-lateral, -sonorant]

[-back, -continuant][-low, -sg]

[-back, -continuant][-low, -sonorant]

[-back, -continuant][-nasal, -sg]

[-back, -continuant][-nasal, -sonorant]

[-back, -continuant][-anterior, -lateral, -nasal]

[-back, +dorsal][+low]

[-back, +dorsal][+syllabic]

[-back, +dorsal][+RTR, +back]

[-back, +dorsal][+RTR, +continuant]
[-back, +dorsal][+RTR, -low]
[-back, +dorsal][+RTR, +sonorant]
[-back, +dorsal][-anterior, +sg]
[-back, +dorsal][+back, +dorsal]
[-back, +dorsal][-back, +dorsal]
[-back, +dorsal][+dorsal, -low]
[-back, +dorsal][+dorsal, +sonorant]
[-back, -mb][-anterior, +sg]
[-back, -mb][-back, +dorsal]
[-back, -mb][-back, +syllabic]
[-back, +sonorant][-anterior, +sg]
[-back, +sonorant][-back, +dorsal]
[-back, +sonorant][-back, +syllabic]
[-back, -sonorant][-coronal]
[-back, -sonorant][-dorsal]
[-back, -sonorant][-labial]
[-back, -sonorant][+mb]
[-back, -sonorant][-anterior, -back]
[-back, -sonorant][-anterior, -low]
[-back, -sonorant][-back, -cg]
[-back, -sonorant][-back, -continuant]
[-back, -sonorant][-back, -sg]
[-back, -sonorant][-back, -sonorant]
[-back, -sonorant][-cg, -lateral]
[-back, -sonorant][-cg, -low]
[-back, -sonorant][-cg, -nasal]
[-back, -sonorant][-continuant, -low]
[-back, -sonorant][-continuant, -nasal]
[-back, -sonorant][-lateral, -sg]
[-back, -sonorant][-lateral, -sonorant]
[-back, -sonorant][-low, -sg]
[-back, -sonorant][-low, -sonorant]
[-back, -sonorant][-nasal, -sg]
[-back, -sonorant][-nasal, -sonorant]
[-back, -sonorant][-anterior, -lateral, -nasal]
[-back, +syllabic][-anterior, +sg]
[-cg, +continuant][+anterior]
[-cg, +continuant][-anterior]
[-cg, +continuant][+cg]
[-cg, +continuant][-cg]
[-cg, +continuant][-continuant]
[-cg, +continuant][+coronal]
[-cg, +continuant][+labial]
[-cg, +continuant][+lateral]
[-cg, +continuant][-lateral]
[-cg, +continuant][+nasal]
[-cg, +continuant][-nasal]
[-cg, +continuant][+sg]
[-cg, +continuant][-sg]
[-cg, +continuant][-sonorant]
[-cg, +continuant][-syllabic]
[-cg, +coronal][+anterior]
[-cg, +coronal][-anterior]
[-cg, +coronal][+cg]
[-cg, +coronal][-cg]
[-cg, +coronal][+coronal]
[-cg, +coronal][+lateral]
[-cg, +coronal][-nasal]
[-cg, +coronal][+sg]
[-cg, +coronal][-sg]
[-cg, +coronal][-sonorant]

[-cg, +coronal][+RTR, -continuant]
[-cg, +coronal][+RTR, -syllabic]
[-cg, +coronal][+back, -lateral]
[-cg, +coronal][+back, -syllabic]
[-cg, +coronal][-back, -lateral]
[-cg, +coronal][-back, -syllabic]
[-cg, +coronal][+continuant, -lateral]
[-cg, +coronal][+continuant, -syllabic]
[-cg, +coronal][-continuant, +dorsal]
[-cg, +coronal][+dorsal, -syllabic]
[-cg, +coronal][-lateral, -low]
[-cg, +coronal][-low, -syllabic]
[-cg, +dorsal][+anterior]
[-cg, +dorsal][-anterior]
[-cg, +dorsal][+cg]
[-cg, +dorsal][-cg]
[-cg, +dorsal][-continuant]
[-cg, +dorsal][+coronal]
[-cg, +dorsal][+labial]
[-cg, +dorsal][+lateral]
[-cg, +dorsal][-lateral]
[-cg, +dorsal][+nasal]
[-cg, +dorsal][-nasal]
[-cg, +dorsal][+sg]
[-cg, +dorsal][-sg]
[-cg, +dorsal][-sonorant]
[-cg, +dorsal][-syllabic]
[-cg, +labial][+anterior]
[-cg, +labial][-anterior]
[-cg, +labial][+cg]
[-cg, +labial][-cg]
[-cg, +labial][-continuant]
[-cg, +labial][+coronal]
[-cg, +labial][+labial]
[-cg, +labial][+lateral]
[-cg, +labial][-lateral]
[-cg, +labial][+nasal]
[-cg, +labial][-nasal]
[-cg, +labial][+sg]
[-cg, +labial][-sg]
[-cg, +labial][-sonorant]
[-cg, +labial][-syllabic]
[-cg, -lateral][+RTR, -back]
[-cg, -low][+RTR, -back]
[-cg, -mb][+anterior]
[-cg, -mb][-anterior]
[-cg, -mb][+cg]
[-cg, -mb][-cg]
[-cg, -mb][+coronal]
[-cg, -mb][+lateral]
[-cg, -mb][-nasal]
[-cg, -mb][+sg]
[-cg, -mb][-sg]
[-cg, -mb][-sonorant]
[-cg, -mb][+RTR, -continuant]
[-cg, -mb][+RTR, -syllabic]
[-cg, -mb][+back, -lateral]
[-cg, -mb][+back, -syllabic]
[-cg, -mb][-back, -lateral]
[-cg, -mb][-back, -syllabic]
[-cg, -mb][+continuant, -lateral]
[-cg, -mb][+continuant, -syllabic]

[-cg, -mb][-continuant, +dorsal]
[-cg, -mb][+dorsal, -syllabic]
[-cg, -mb][-lateral, -low]
[-cg, -mb][-low, -syllabic]
[-cg, -nasal][+RTR, -back]
[+continuant, +coronal][-coronal]
[+continuant, +coronal][-dorsal]
[+continuant, +coronal][-labial]
[+continuant, +coronal][+mb]
[+continuant, +coronal][-anterior, -back]
[+continuant, +coronal][-anterior, -low]
[+continuant, +coronal][-back, -cg]
[+continuant, +coronal][-back, -continuant]
[+continuant, +coronal][-back, -sg]
[+continuant, +coronal][-back, -sonorant]
[+continuant, +coronal][-cg, -lateral]
[+continuant, +coronal][-cg, -low]
[+continuant, +coronal][-cg, -nasal]
[+continuant, +coronal][-continuant, -low]
[+continuant, +coronal][-continuant, -nasal]
[+continuant, +coronal][-lateral, -sg]
[+continuant, +coronal][-lateral, -sonorant]
[+continuant, +coronal][-low, -sg]
[+continuant, +coronal][-low, -sonorant]
[+continuant, +coronal][-nasal, -sg]
[+continuant, +coronal][-nasal, -sonorant]
[+continuant, +coronal][-anterior, -lateral, -nasal]
[+continuant, +dorsal][+low]
[+continuant, +dorsal][+syllabic]
[+continuant, +dorsal][+RTR, +back]
[+continuant, +dorsal][+RTR, -back]
[+continuant, +dorsal][+RTR, +continuant]
[+continuant, +dorsal][+RTR, -low]
[+continuant, +dorsal][+RTR, +sonorant]
[+continuant, +dorsal][+back, +dorsal]
[+continuant, +dorsal][-back, +dorsal]
[+continuant, +dorsal][+dorsal, -low]
[+continuant, +dorsal][+dorsal, +sonorant]
[+continuant, -lateral][+anterior, +sg]
[+continuant, -lateral][-cg, +continuant]
[+continuant, -lateral][+continuant, +sg]
[+continuant, -lateral][+labial, +sg]
[+continuant, -nasal][-cg, +continuant]
[+continuant, -nasal][+continuant, +sg]
[+continuant, -sonorant][+RTR, +sg]
[+continuant, -sonorant][+anterior, +sg]
[+continuant, -sonorant][-anterior, +sg]
[+continuant, -sonorant][-cg, +continuant]
[+continuant, -sonorant][+continuant, +sg]
[+continuant, -sonorant][+coronal, +sg]
[+continuant, -syllabic][-cg, +continuant]
[+continuant, -syllabic][+continuant, +sg]
[-continuant, +coronal][+anterior, +nasal]
[-continuant, +coronal][-anterior, +nasal]
[-continuant, +coronal][+cg, +labial]
[-continuant, +coronal][-cg, +labial]
[-continuant, +coronal][+coronal, +nasal]
[-continuant, +coronal][+labial, +sg]
[-continuant, +coronal][+labial, -sg]
[-continuant, +coronal][+labial, -sonorant]
[-continuant, +coronal][+anterior, -continuant, -lateral]

[-continuant, +coronal][+anterior, -continuant, +sonorant]

[-continuant, +coronal][-anterior, -continuant, +sonorant]

[-continuant, +coronal][-anterior, +coronal, -lateral]

[-continuant, +coronal][-anterior, -lateral, -mb]

[-continuant, +coronal][-anterior, -lateral, +sonorant]

[-continuant, +coronal][-back, +continuant, -lateral]

[-continuant, +coronal][-back, +continuant, -nasal]

[-continuant, +coronal][-back, +continuant, -syllabic]

[-continuant, +coronal][-back, -lateral, -mb]

[-continuant, +coronal][-back, -lateral, +sonorant]

[-continuant, +coronal][-back, -mb, -nasal]

[-continuant, +coronal][-back, -mb, -syllabic]

[-continuant, +coronal][-back, -nasal, +sonorant]

[-continuant, +coronal][-back, +sonorant, -syllabic]

[-continuant, +coronal][-continuant, +coronal, -lateral]

[-continuant, +coronal][-continuant, +coronal, +sonorant]

[-continuant, +dorsal][+anterior]

[-continuant, +dorsal][-anterior]

[-continuant, +dorsal][+cg]

[-continuant, +dorsal][-cg]

[-continuant, +dorsal][-continuant]

[-continuant, +dorsal][+coronal]

[-continuant, +dorsal][-coronal]

[-continuant, +dorsal][-dorsal]

[-continuant, +dorsal][+labial]

[-continuant, +dorsal][-labial]

[-continuant, +dorsal][+lateral]

[-continuant, +dorsal][-lateral]

[-continuant, +dorsal][+mb]

[-continuant, +dorsal][+nasal]

[-continuant, +dorsal][-nasal]

[-continuant, +dorsal][+sg]

[-continuant, +dorsal][-sg]

[-continuant, +dorsal][-sonorant]

[-continuant, +dorsal][-syllabic]

[-continuant, -low][-coronal]

[-continuant, -low][-dorsal]

[-continuant, -low][-labial]

[-continuant, -low][+mb]

[-continuant, -low][+RTR, -back]

[-continuant, -low][-anterior, -back]

[-continuant, -low][-anterior, -low]

[-continuant, -low][-back, -cg]

[-continuant, -low][-back, -continuant]

[-continuant, -low][-back, -sg]

[-continuant, -low][-back, -sonorant]

[-continuant, -low][-cg, -lateral]

[-continuant, -low][-cg, -low]

[-continuant, -low][-cg, -nasal]

[-continuant, -low][-continuant, -low]

[-continuant, -low][-continuant, -nasal]

[-continuant, -low][-lateral, -sg]

[-continuant, -low][-lateral, -sonorant]

[-continuant, -low][-low, -sg]

[-continuant, -low][-low, -sonorant]

[-continuant, -low][-nasal, -sg]

[-continuant, -low][-nasal, -sonorant]

[-continuant, -low][-anterior, -lateral, -nasal]

[-continuant, -mb][+anterior, +nasal]

[-continuant, -mb][-cg, +continuant]

[-continuant, -mb][+continuant, +sg]

[-continuant, -mb][+anterior, -continuant, -lateral]

[-continuant, -mb][+anterior, -continuant, +sonorant]

[-continuant, -nasal][-coronal]

[-continuant, -nasal][-dorsal]

[-continuant, -nasal][-labial]

[-continuant, -nasal][+mb]

[-continuant, -nasal][+RTR, -back]

[-continuant, -nasal][-anterior, -back]

[-continuant, -nasal][-anterior, -low]

[-continuant, -nasal][-back, -cg]

[-continuant, -nasal][-back, -continuant]

[-continuant, -nasal][-back, -sg]

[-continuant, -nasal][-back, -sonorant]

[-continuant, -nasal][-cg, -lateral]

[-continuant, -nasal][-cg, -low]

[-continuant, -nasal][-cg, -nasal]

[-continuant, -nasal][-continuant, -low]

[-continuant, -nasal][-continuant, -nasal]

[-continuant, -nasal][-lateral, -sg]

[-continuant, -nasal][-lateral, -sonorant]

[-continuant, -nasal][-low, -sg]

[-continuant, -nasal][-low, -sonorant]

[-continuant, -nasal][-nasal, -sg]

[-continuant, -nasal][-nasal, -sonorant]

[-continuant, -nasal][-anterior, -lateral, -nasal]

[-continuant, +sonorant][+anterior, +nasal]

[-continuant, +sonorant][-cg, +continuant]

[-continuant, +sonorant][+continuant, +sg]

[-continuant, +sonorant][+labial, -lateral]

[-continuant, +sonorant][+labial, +nasal]

[-continuant, +sonorant][+labial, +sonorant]

[-continuant, +sonorant][+anterior, -continuant, -lateral]

[-continuant, +sonorant][+anterior, -continuant, +sonorant]

[-continuant, -sonorant][-coronal]

[-continuant, -sonorant][-dorsal]

[-continuant, -sonorant][-labial]

[-continuant, -sonorant][+mb]

[-continuant, -sonorant][-anterior, -back]

[-continuant, -sonorant][-anterior, -low]

[-continuant, -sonorant][-back, -cg]

[-continuant, -sonorant][-back, -continuant]

[-continuant, -sonorant][-back, -sg]

[-continuant, -sonorant][-back, -sonorant]

[-continuant, -sonorant][-cg, -lateral]

[-continuant, -sonorant][-cg, -low]

[-continuant, -sonorant][-cg, -nasal]

[-continuant, -sonorant][-continuant, -low]

[-continuant, -sonorant][-continuant, -nasal]

[-continuant, -sonorant][-lateral, -sg]

[-continuant, -sonorant][-lateral, -sonorant]

[-continuant, -sonorant][-low, -sg]
[-continuant, -sonorant][-low, -sonorant]
[-continuant, -sonorant][-nasal, -sg]
[-continuant, -sonorant][-nasal, -sonorant]
[-continuant, -sonorant][-anterior, -lateral, -nasal]
[+coronal, -lateral][+labial, +sg]
[+coronal, +nasal][+labial]
[+coronal, +nasal][+nasal]
[+coronal, +nasal][-continuant, +sonorant]
[+coronal, +nasal][-anterior, +coronal, -lateral]
[+coronal, +nasal][-anterior, -lateral, -mb]
[+coronal, +nasal][-anterior, -lateral, +sonorant]
[+coronal, +nasal][-back, +continuant, -lateral]
[+coronal, +nasal][-back, +continuant, -nasal]
[+coronal, +nasal][-back, +continuant, -syllabic]
[+coronal, +nasal][-back, -lateral, -mb]
[+coronal, +nasal][-back, -lateral, +sonorant]
[+coronal, +nasal][-back, -mb, -nasal]
[+coronal, +nasal][-back, -mb, -syllabic]
[+coronal, +nasal][-back, -nasal, +sonorant]
[+coronal, +nasal][-back, +sonorant, -syllabic]
[+coronal, +nasal][-continuant, +coronal, -lateral]
[+coronal, +nasal][-continuant, -lateral, -mb]
[+coronal, -nasal][-coronal]
[+coronal, -nasal][-dorsal]
[+coronal, -nasal][-labial]
[+coronal, -nasal][+mb]
[+coronal, -nasal][+anterior, +cg]
[+coronal, -nasal][+anterior, -nasal]
[+coronal, -nasal][-anterior, -back]
[+coronal, -nasal][-anterior, -low]
[+coronal, -nasal][-back, -cg]
[+coronal, -nasal][-back, -continuant]
[+coronal, -nasal][-back, -sg]
[+coronal, -nasal][-back, -sonorant]
[+coronal, -nasal][-cg, -lateral]
[+coronal, -nasal][-cg, -low]
[+coronal, -nasal][-cg, -nasal]
[+coronal, -nasal][-continuant, -low]
[+coronal, -nasal][-continuant, -nasal]
[+coronal, -nasal][-lateral, -sg]
[+coronal, -nasal][-lateral, -sonorant]
[+coronal, -nasal][-low, -sg]
[+coronal, -nasal][-low, -sonorant]
[+coronal, -nasal][-nasal, -sg]
[+coronal, -nasal][-nasal, -sonorant]
[+coronal, -nasal][+anterior, +continuant, -lateral]
[+coronal, -nasal][+anterior, +continuant, +sonorant]
[+coronal, -nasal][-anterior, -lateral, -nasal]
[+coronal, -nasal][+continuant, +coronal, -lateral]
[+coronal, -nasal][+coronal, -lateral, -nasal]
[+coronal, -sg][+anterior]
[+coronal, -sg][-anterior]
[+coronal, -sg][+cg]
[+coronal, -sg][-cg]
[+coronal, -sg][+coronal]
[+coronal, -sg][+lateral]
[+coronal, -sg][-nasal]
[+coronal, -sg][+sg]
[+coronal, -sg][-sg]
[+coronal, -sg][-sonorant]

[+coronal, -sg][+RTR, -continuant]
[+coronal, -sg][+RTR, -syllabic]
[+coronal, -sg][+back, -lateral]
[+coronal, -sg][+back, -syllabic]
[+coronal, -sg][-back, -lateral]
[+coronal, -sg][-back, -syllabic]
[+coronal, -sg][+continuant, -lateral]
[+coronal, -sg][+continuant, -syllabic]
[+coronal, -sg][-continuant, +dorsal]
[+coronal, -sg][+dorsal, -syllabic]
[+coronal, -sg][-lateral, -low]
[+coronal, -sg][-low, -syllabic]
[+coronal, -sonorant][-coronal]
[+coronal, -sonorant][-dorsal]
[+coronal, -sonorant][-labial]
[+coronal, -sonorant][+mb]
[+coronal, -sonorant][+RTR, +sg]
[+coronal, -sonorant][+anterior, +sg]
[+coronal, -sonorant][-anterior, -back]
[+coronal, -sonorant][-anterior, +cg]
[+coronal, -sonorant][-anterior, -low]
[+coronal, -sonorant][-anterior, +sg]
[+coronal, -sonorant][-back, -cg]
[+coronal, -sonorant][-back, -continuant]
[+coronal, -sonorant][-back, -sg]
[+coronal, -sonorant][-back, -sonorant]
[+coronal, -sonorant][-cg, -lateral]
[+coronal, -sonorant][-cg, -low]
[+coronal, -sonorant][-cg, -nasal]
[+coronal, -sonorant][+continuant, -sonorant]
[+coronal, -sonorant][-continuant, -low]
[+coronal, -sonorant][-continuant, -nasal]
[+coronal, -sonorant][+coronal, +sg]
[+coronal, -sonorant][-lateral, -sg]
[+coronal, -sonorant][-lateral, -sonorant]
[+coronal, -sonorant][-low, -sg]
[+coronal, -sonorant][-low, -sonorant]
[+coronal, -sonorant][-nasal, -sg]
[+coronal, -sonorant][-nasal, -sonorant]
[+coronal, -sonorant][-anterior, -lateral, -nasal]
[+dorsal, -low][+low]
[+dorsal, -low][+syllabic]
[+dorsal, -low][+RTR, +back]
[+dorsal, -low][+RTR, -back]
[+dorsal, -low][+RTR, +continuant]
[+dorsal, -low][+RTR, -low]
[+dorsal, -low][+RTR, +sonorant]
[+dorsal, -low][+back, +dorsal]
[+dorsal, -low][-back, +dorsal]
[+dorsal, -low][+dorsal, -low]
[+dorsal, -low][+dorsal, +sonorant]
[+dorsal, -sg][+anterior]
[+dorsal, -sg][-anterior]
[+dorsal, -sg][+cg]
[+dorsal, -sg][-cg]
[+dorsal, -sg][-continuant]
[+dorsal, -sg][+coronal]
[+dorsal, -sg][+labial]
[+dorsal, -sg][+lateral]
[+dorsal, -sg][-lateral]
[+dorsal, -sg][+nasal]
[+dorsal, -sg][-nasal]

[+dorsal, -sg][+sg]
[+dorsal, -sg][-sg]
[+dorsal, -sg][-sonorant]
[+dorsal, -sg][-syllabic]
[+dorsal, +sonorant][+low]
[+dorsal, +sonorant][+syllabic]
[+dorsal, +sonorant][+RTR, +back]
[+dorsal, +sonorant][+RTR, -back]
[+dorsal, +sonorant][+RTR, +continuant]
[+dorsal, +sonorant][+RTR, -low]
[+dorsal, +sonorant][+RTR, +sonorant]
[+dorsal, +sonorant][+back, +dorsal]
[+dorsal, +sonorant][-back, +dorsal]
[+dorsal, +sonorant][+dorsal, -low]
[+dorsal, +sonorant][+dorsal, +sonorant]
[+dorsal, -sonorant][+sg]
[+dorsal, -sonorant][+cg, +labial]
[+dorsal, -sonorant][-cg, +continuant]
[+dorsal, -sonorant][-RTR, +cg, +dorsal]
[+dorsal, -syllabic][+sg]
[+dorsal, -syllabic][+cg, +labial]
[+dorsal, -syllabic][-cg, +continuant]
[+dorsal, -syllabic][-RTR, +cg, +dorsal]
[+labial, -sg][+anterior]
[+labial, -sg][-anterior]
[+labial, -sg][+cg]
[+labial, -sg][-cg]
[+labial, -sg][-continuant]
[+labial, -sg][+coronal]
[+labial, -sg][+labial]
[+labial, -sg][+lateral]
[+labial, -sg][-lateral]
[+labial, -sg][+nasal]
[+labial, -sg][-nasal]
[+labial, -sg][+sg]
[+labial, -sg][-sg]
[+labial, -sg][-sonorant]
[+labial, -sg][-syllabic]
[+labial, -sonorant][+anterior]
[+labial, -sonorant][-anterior]
[+labial, -sonorant][+cg]
[+labial, -sonorant][-cg]
[+labial, -sonorant][-continuant]
[+labial, -sonorant][+coronal]
[+labial, -sonorant][+labial]
[+labial, -sonorant][+lateral]
[+labial, -sonorant][-lateral]
[+labial, -sonorant][+nasal]
[+labial, -sonorant][-nasal]
[+labial, -sonorant][+sg]
[+labial, -sonorant][-sg]
[+labial, -sonorant][-sonorant]
[+labial, -sonorant][-syllabic]
[-lateral, -mb][-cg, +continuant]
[-lateral, -mb][+continuant, +sg]
[-lateral, -sg][+RTR, -back]
[-lateral, +sonorant][-cg, +continuant]
[-lateral, +sonorant][+continuant, +sg]
[-lateral, -sonorant][-coronal]
[-lateral, -sonorant][-dorsal]
[-lateral, -sonorant][-labial]
[-lateral, -sonorant][+mb]

[-lateral, -sonorant][+RTR, -back]
[-lateral, -sonorant][-anterior, -back]
[-lateral, -sonorant][-anterior, -low]
[-lateral, -sonorant][-back, -cg]
[-lateral, -sonorant][-back, -continuant]
[-lateral, -sonorant][-back, -sg]
[-lateral, -sonorant][-back, -sonorant]
[-lateral, -sonorant][-cg, -lateral]
[-lateral, -sonorant][-cg, -low]
[-lateral, -sonorant][-cg, -nasal]
[-lateral, -sonorant][-continuant, -low]
[-lateral, -sonorant][-continuant, -nasal]
[-lateral, -sonorant][-lateral, -sg]
[-lateral, -sonorant][-lateral, -sonorant]
[-lateral, -sonorant][-low, -sg]
[-lateral, -sonorant][-low, -sonorant]
[-lateral, -sonorant][-nasal, -sg]
[-lateral, -sonorant][-nasal, -sonorant]
[-lateral, -sonorant][-anterior, -lateral, -nasal]
[-low, -sg][+RTR, -back]
[-low, -sonorant][-coronal]
[-low, -sonorant][-dorsal]
[-low, -sonorant][-labial]
[-low, -sonorant][+mb]
[-low, -sonorant][+RTR, -back]
[-low, -sonorant][-anterior, -back]
[-low, -sonorant][-anterior, -low]
[-low, -sonorant][-back, -cg]
[-low, -sonorant][-back, -continuant]
[-low, -sonorant][-back, -sg]
[-low, -sonorant][-back, -sonorant]
[-low, -sonorant][-cg, -lateral]
[-low, -sonorant][-cg, -low]
[-low, -sonorant][-cg, -nasal]
[-low, -sonorant][-continuant, -low]
[-low, -sonorant][-continuant, -nasal]
[-low, -sonorant][-lateral, -sg]
[-low, -sonorant][-lateral, -sonorant]
[-low, -sonorant][-low, -sg]
[-low, -sonorant][-low, -sonorant]
[-low, -sonorant][-nasal, -sg]
[-low, -sonorant][-nasal, -sonorant]
[-low, -sonorant][-anterior, -lateral, -nasal]
[-mb, -nasal][-cg, +continuant]
[-mb, -nasal][+continuant, +sg]
[-mb, -sg][+anterior]
[-mb, -sg][-anterior]
[-mb, -sg][+cg]
[-mb, -sg][-cg]
[-mb, -sg][+coronal]
[-mb, -sg][+lateral]
[-mb, -sg][-nasal]
[-mb, -sg][+sg]
[-mb, -sg][-sg]
[-mb, -sg][-sonorant]
[-mb, -sg][+RTR, -continuant]
[-mb, -sg][+RTR, -syllabic]
[-mb, -sg][+back, -lateral]
[-mb, -sg][+back, -syllabic]
[-mb, -sg][-back, -lateral]
[-mb, -sg][-back, -syllabic]
[-mb, -sg][+continuant, -lateral]

[-mb, -sg][+continuant, -syllabic]

[-mb, -sg][-continuant, +dorsal]

[-mb, -sg][+dorsal, -syllabic]

[-mb, -sg][-lateral, -low]

[-mb, -sg][-low, -syllabic]

[-mb, -sonorant][+RTR, +sg]

[-mb, -sonorant][+anterior, +sg]

[-mb, -sonorant][-anterior, +sg]

[-mb, -sonorant][-cg, +continuant]

[-mb, -sonorant][+continuant, +sg]

[-mb, -sonorant][+coronal, +sg]

[-mb, -syllabic][-cg, +continuant]

[-mb, -syllabic][+continuant, +sg]

[-nasal, -sg][+RTR, -back]

[-nasal, +sonorant][-cg, +continuant]

[-nasal, +sonorant][+continuant, +sg]

[-nasal, -sonorant][-coronal]

[-nasal, -sonorant][-dorsal]

[-nasal, -sonorant][-labial]

[-nasal, -sonorant][+mb]

[-nasal, -sonorant][+RTR, -back]

[-nasal, -sonorant][-anterior, -back]

[-nasal, -sonorant][-anterior, -low]

[-nasal, -sonorant][-back, -cg]

[-nasal, -sonorant][-back, -continuant]

[-nasal, -sonorant][-back, -sg]

[-nasal, -sonorant][-back, -sonorant]

[-nasal, -sonorant][-cg, -lateral]

[-nasal, -sonorant][-cg, -low]

[-nasal, -sonorant][-cg, -nasal]

[-nasal, -sonorant][-continuant, -low]

[-nasal, -sonorant][-continuant, -nasal]

[-nasal, -sonorant][-lateral, -sg]

[-nasal, -sonorant][-lateral, -sonorant]

[-nasal, -sonorant][-low, -sg]

[-nasal, -sonorant][-low, -sonorant]

[-nasal, -sonorant][-nasal, -sg]

[-nasal, -sonorant][-nasal, -sonorant]

[-nasal, -sonorant][-anterior, -lateral, -nasal]

[+sonorant, -syllabic][-cg, +continuant]

[+sonorant, -syllabic][+continuant, +sg]

[-RTR, -back, +continuant][+RTR, +sg]

[-RTR, -back, +dorsal][+RTR]

[-RTR, -back, -mb][+RTR, +sg]

[-RTR, -back, +sonorant][+RTR, +sg]

[-RTR, -back, +syllabic][+RTR]

[-RTR, +cg, +dorsal][+RTR]

[-RTR, +dorsal, -low][+RTR]

[-RTR, +dorsal, +sg][+RTR]

[-RTR, -low, +syllabic][+RTR]

[+anterior, +continuant, -lateral][+anterior, +cg]

[+anterior, +continuant, -lateral][+anterior, -nasal]

[+anterior, +continuant, -lateral][+cg, +coronal]

[+anterior, +continuant, -lateral][+anterior, +continuant, -lateral]

[+anterior, +continuant, -lateral][+anterior, +continuant, +sonorant]

[+anterior, +continuant, -lateral][+continuant, +coronal, -lateral]

[+anterior, +continuant, -lateral][+coronal, -lateral, -nasal]

[+anterior, +continuant, +sonorant][+anterior, +cg]

[+anterior, +continuant, +sonorant][+anterior, -nasal]

[+anterior, +continuant, +sonorant][+cg, +coronal]

[+anterior, +continuant, +sonorant][+anterior, +continuant, -lateral]

[+anterior, +continuant, +sonorant][+anterior, +continuant, +sonorant]

[+anterior, +continuant, +sonorant][+continuant, +coronal, -lateral]

[+anterior, +continuant, +sonorant][+coronal, -lateral, -nasal]

[+anterior, -continuant, -lateral][+labial]

[+anterior, -continuant, -lateral][+nasal]

[+anterior, -continuant, -lateral][-continuant, +sonorant]

[+anterior, -continuant, -lateral][-continuant, -lateral, -mb]

[+anterior, -continuant, +sonorant][+labial]

[+anterior, -continuant, +sonorant][+nasal]

[+anterior, -continuant, +sonorant][-continuant, +sonorant]

[+anterior, -continuant, +sonorant][-continuant, -lateral, -mb]

[+anterior, -continuant, -sonorant][+anterior]

[+anterior, -continuant, -sonorant][-anterior]

[+anterior, -continuant, -sonorant][+cg]

[+anterior, -continuant, -sonorant][-cg]

[+anterior, -continuant, -sonorant][+coronal]

[+anterior, -continuant, -sonorant][+lateral]

[+anterior, -continuant, -sonorant][-nasal]

[+anterior, -continuant, -sonorant][+sg]

[+anterior, -continuant, -sonorant][-sg]

[+anterior, -continuant, -sonorant][-sonorant]

[+anterior, -continuant, -sonorant][+RTR, -continuant]

[+anterior, -continuant, -sonorant][+RTR, -syllabic]

[+anterior, -continuant, -sonorant][+back, -lateral]

[+anterior, -continuant, -sonorant][+back, -syllabic]

[+anterior, -continuant, -sonorant][-back, -lateral]

[+anterior, -continuant, -sonorant][-back, -syllabic]

[+anterior, -continuant, -sonorant][+continuant, -lateral]

[+anterior, -continuant, -sonorant][+continuant, -syllabic]

[+anterior, -continuant, -sonorant][-continuant, +dorsal]

[+anterior, -continuant, -sonorant][+dorsal, -syllabic]

[+anterior, -continuant, -sonorant][-lateral, -low]

[+anterior, -continuant, -sonorant][-low, -syllabic]

[-anterior, -cg, +coronal][-continuant]

[-anterior, -cg, +coronal][+labial]

[-anterior, -cg, +coronal][-lateral]

[-anterior, -cg, +coronal][+nasal]

[-anterior, -cg, +coronal][-syllabic]

[-anterior, -cg, -mb][-continuant]
[-anterior, -cg, -mb][+labial]
[-anterior, -cg, -mb][-lateral]
[-anterior, -cg, -mb][+nasal]
[-anterior, -cg, -mb][-syllabic]
[-anterior, -continuant, +coronal][+anterior]
[-anterior, -continuant, +coronal][-anterior]
[-anterior, -continuant, +coronal][+cg]
[-anterior, -continuant, +coronal][-cg]
[-anterior, -continuant, +coronal][-continuant]
[-anterior, -continuant, +coronal][+coronal]
[-anterior, -continuant, +coronal][+labial]
[-anterior, -continuant, +coronal][-lateral]
[-anterior, -continuant, +coronal][+nasal]
[-anterior, -continuant, +coronal][-nasal]
[-anterior, -continuant, +coronal][+sg]
[-anterior, -continuant, +coronal][-sg]
[-anterior, -continuant, +coronal][-sonorant]
[-anterior, -continuant, +coronal][-syllabic]
[-anterior, -continuant, -mb][+anterior]
[-anterior, -continuant, -mb][-anterior]
[-anterior, -continuant, -mb][+cg]
[-anterior, -continuant, -mb][-cg]
[-anterior, -continuant, -mb][-continuant]
[-anterior, -continuant, -mb][+coronal]
[-anterior, -continuant, -mb][+labial]
[-anterior, -continuant, -mb][-lateral]
[-anterior, -continuant, -mb][+nasal]
[-anterior, -continuant, -mb][-nasal]
[-anterior, -continuant, -mb][+sg]
[-anterior, -continuant, -mb][-sg]
[-anterior, -continuant, -mb][-sonorant]
[-anterior, -continuant, -mb][-syllabic]
[-anterior, -continuant, +sonorant][+anterior]
[-anterior, -continuant, +sonorant][-anterior]
[-anterior, -continuant, +sonorant][+cg]
[-anterior, -continuant, +sonorant][-cg]
[-anterior, -continuant, +sonorant][-continuant]
[-anterior, -continuant, +sonorant][+coronal]
[-anterior, -continuant, +sonorant][+labial]
[-anterior, -continuant, +sonorant][-lateral]
[-anterior, -continuant, +sonorant][+nasal]
[-anterior, -continuant, +sonorant][-nasal]
[-anterior, -continuant, +sonorant][+sg]
[-anterior, -continuant, +sonorant][-sg]
[-anterior, -continuant, +sonorant][-sonorant]
[-anterior, -continuant, +sonorant][-syllabic]
[-anterior, +coronal, -lateral][+anterior]
[-anterior, +coronal, -lateral][-anterior]
[-anterior, +coronal, -lateral][+cg]
[-anterior, +coronal, -lateral][-cg]
[-anterior, +coronal, -lateral][-continuant]
[-anterior, +coronal, -lateral][+coronal]
[-anterior, +coronal, -lateral][+labial]
[-anterior, +coronal, -lateral][-lateral]
[-anterior, +coronal, -lateral][+nasal]
[-anterior, +coronal, -lateral][-nasal]
[-anterior, +coronal, -lateral][+sg]
[-anterior, +coronal, -lateral][-sg]
[-anterior, +coronal, -lateral][-sonorant]
[-anterior, +coronal, -lateral][-syllabic]
[-anterior, +coronal, -sg][-continuant]

[-anterior, +coronal, -sg][+labial]
[-anterior, +coronal, -sg][-lateral]
[-anterior, +coronal, -sg][+nasal]
[-anterior, +coronal, -sg][-syllabic]
[-anterior, +coronal, -sonorant][+anterior]
[-anterior, +coronal, -sonorant][-anterior]
[-anterior, +coronal, -sonorant][+cg]
[-anterior, +coronal, -sonorant][-cg]
[-anterior, +coronal, -sonorant][-continuant]
[-anterior, +coronal, -sonorant][+coronal]
[-anterior, +coronal, -sonorant][+labial]
[-anterior, +coronal, -sonorant][-lateral]
[-anterior, +coronal, -sonorant][+nasal]
[-anterior, +coronal, -sonorant][-nasal]
[-anterior, +coronal, -sonorant][+sg]
[-anterior, +coronal, -sonorant][-sg]
[-anterior, +coronal, -sonorant][-sonorant]
[-anterior, +coronal, -sonorant][-syllabic]
[-anterior, -lateral, -mb][+anterior]
[-anterior, -lateral, -mb][-anterior]
[-anterior, -lateral, -mb][+cg]
[-anterior, -lateral, -mb][-cg]
[-anterior, -lateral, -mb][-continuant]
[-anterior, -lateral, -mb][+coronal]
[-anterior, -lateral, -mb][+labial]
[-anterior, -lateral, -mb][-lateral]
[-anterior, -lateral, -mb][+nasal]
[-anterior, -lateral, -mb][-nasal]
[-anterior, -lateral, -mb][+sg]
[-anterior, -lateral, -mb][-sg]
[-anterior, -lateral, -mb][-sonorant]
[-anterior, -lateral, -mb][-syllabic]
[-anterior, -lateral, -nasal][+RTR, -back]
[-anterior, -lateral, +sonorant][+anterior]
[-anterior, -lateral, +sonorant][-anterior]
[-anterior, -lateral, +sonorant][+cg]
[-anterior, -lateral, +sonorant][-cg]
[-anterior, -lateral, +sonorant][-continuant]
[-anterior, -lateral, +sonorant][+coronal]
[-anterior, -lateral, +sonorant][+labial]
[-anterior, -lateral, +sonorant][-lateral]
[-anterior, -lateral, +sonorant][+nasal]
[-anterior, -lateral, +sonorant][-nasal]
[-anterior, -lateral, +sonorant][+sg]
[-anterior, -lateral, +sonorant][-sg]
[-anterior, -lateral, +sonorant][-sonorant]
[-anterior, -lateral, +sonorant][-syllabic]
[-anterior, -mb, -sg][-continuant]
[-anterior, -mb, -sg][+labial]
[-anterior, -mb, -sg][-lateral]
[-anterior, -mb, -sg][+nasal]
[-anterior, -mb, -sg][-syllabic]
[-anterior, -mb, -sonorant][+anterior]
[-anterior, -mb, -sonorant][-anterior]
[-anterior, -mb, -sonorant][+cg]
[-anterior, -mb, -sonorant][-cg]
[-anterior, -mb, -sonorant][-continuant]
[-anterior, -mb, -sonorant][+coronal]
[-anterior, -mb, -sonorant][+labial]
[-anterior, -mb, -sonorant][-lateral]
[-anterior, -mb, -sonorant][+nasal]
[-anterior, -mb, -sonorant][-nasal]

[-anterior, -mb, -sonorant][+sg]
[-anterior, -mb, -sonorant][-sg]
[-anterior, -mb, -sonorant][-sonorant]
[-anterior, -mb, -sonorant][-syllabic]
[-back, +continuant, -lateral][+RTR, +sg]
[-back, +continuant, -lateral][-back, +continuant]
[-back, +continuant, -lateral][-back, -mb]
[-back, +continuant, -lateral][-back, +sonorant]
[-back, +continuant, -lateral][+cg, +labial]
[-back, +continuant, -lateral][+coronal, +sg]
[-back, +continuant, -nasal][+RTR, +sg]
[-back, +continuant, -nasal][+anterior, +sg]
[-back, +continuant, -nasal][-back, +continuant]
[-back, +continuant, -nasal][-back, -mb]
[-back, +continuant, -nasal][-back, +sonorant]
[-back, +continuant, -nasal][+cg, +labial]
[-back, +continuant, -nasal][+coronal, +sg]
[-back, +continuant, -nasal][+labial, +sg]
[-back, +continuant, -syllabic][+RTR, +sg]
[-back, +continuant, -syllabic][+anterior, +sg]
[-back, +continuant, -syllabic][-back, +continuant]
[-back, +continuant, -syllabic][-back, -mb]
[-back, +continuant, -syllabic][-back, +sonorant]
[-back, +continuant, -syllabic][+cg, +labial]
[-back, +continuant, -syllabic][+coronal, +sg]
[-back, +continuant, -syllabic][+labial, +sg]
[-back, -lateral, -mb][+RTR, +sg]
[-back, -lateral, -mb][+anterior, +sg]
[-back, -lateral, -mb][-back, +continuant]
[-back, -lateral, -mb][-back, -mb]
[-back, -lateral, -mb][-back, +sonorant]
[-back, -lateral, -mb][+cg, +labial]
[-back, -lateral, -mb][+coronal, +sg]
[-back, -lateral, -mb][+labial, +sg]
[-back, -lateral, +sonorant][+RTR, +sg]
[-back, -lateral, +sonorant][+anterior, +sg]
[-back, -lateral, +sonorant][-back, +continuant]
[-back, -lateral, +sonorant][-back, -mb]
[-back, -lateral, +sonorant][-back, +sonorant]
[-back, -lateral, +sonorant][+cg, +labial]
[-back, -lateral, +sonorant][+coronal, +sg]
[-back, -lateral, +sonorant][+labial, +sg]
[-back, -mb, -nasal][+RTR, +sg]
[-back, -mb, -nasal][+anterior, +sg]
[-back, -mb, -nasal][-back, +continuant]
[-back, -mb, -nasal][-back, -mb]
[-back, -mb, -nasal][-back, +sonorant]
[-back, -mb, -nasal][+cg, +labial]
[-back, -mb, -nasal][+coronal, +sg]
[-back, -mb, -nasal][+labial, +sg]
[-back, -mb, -syllabic][+RTR, +sg]
[-back, -mb, -syllabic][+anterior, +sg]
[-back, -mb, -syllabic][-back, +continuant]
[-back, -mb, -syllabic][-back, -mb]
[-back, -mb, -syllabic][-back, +sonorant]
[-back, -mb, -syllabic][+cg, +labial]
[-back, -mb, -syllabic][+coronal, +sg]
[-back, -mb, -syllabic][+labial, +sg]
[-back, -nasal, +sonorant][+RTR, +sg]
[-back, -nasal, +sonorant][+anterior, +sg]
[-back, -nasal, +sonorant][-back, +continuant]
[-back, -nasal, +sonorant][-back, -mb]

[-back, -nasal, +sonorant][-back, +sonorant]

[-back, -nasal, +sonorant][+cg, +labial]

[-back, -nasal, +sonorant][+coronal, +sg]

[-back, -nasal, +sonorant][+labial, +sg]

[-back, +sonorant, -syllabic][+RTR, +sg]

[-back, +sonorant, -syllabic][+anterior, +sg]

[-back, +sonorant, -syllabic][-back, +continuant]

[-back, +sonorant, -syllabic][-back, -mb]

[-back, +sonorant, -syllabic][-back, +sonorant]

[-back, +sonorant, -syllabic][+cg, +labial]

[-back, +sonorant, -syllabic][+coronal, +sg]

[-back, +sonorant, -syllabic][+labial, +sg]

[+continuant, +coronal, -lateral][+anterior, +cg]

[+continuant, +coronal, -lateral][+anterior, -nasal]

[+continuant, +coronal, -lateral][-anterior, +cg]

[+continuant, +coronal, -lateral][-anterior, +sg]

[+continuant, +coronal, -lateral][+cg, +coronal]

[+continuant, +coronal, -lateral][+coronal, +sg]

[+continuant, +coronal, -lateral][+anterior, +continuant, -lateral]

[+continuant, +coronal, -lateral][+anterior, +continuant, +sonorant]

[+continuant, +coronal, -lateral][+continuant, +coronal, -lateral]

[+continuant, +coronal, -lateral][+coronal, -lateral, -nasal]

[+continuant, +coronal, +sonorant][+anterior, +cg]

[+continuant, +coronal, +sonorant][+anterior, -nasal]

[+continuant, +coronal, +sonorant][+anterior, +continuant, -lateral]

[+continuant, +coronal, +sonorant][+anterior, +continuant, +sonorant]

[+continuant, +coronal, +sonorant][+continuant, +coronal, -lateral]

[+continuant, +coronal, +sonorant][+coronal, -lateral, -nasal]

[+continuant, +dorsal, -sonorant][+continuant, +dorsal]

[+continuant, +dorsal, -syllabic][+continuant, +dorsal]

[+continuant, -lateral, -low][+cg, +labial]

[+continuant, -low, -nasal][+anterior, +sg]

[+continuant, -low, -nasal][+cg, +labial]

[+continuant, -low, -nasal][+labial, +sg]

[+continuant, -low, -syllabic][+anterior, +sg]

[+continuant, -low, -syllabic][+cg, +labial]

[+continuant, -low, -syllabic][+labial, +sg]

[-continuant, +coronal, -lateral][+labial]

[-continuant, +coronal, -lateral][+nasal]

[-continuant, +coronal, -lateral][-continuant, +sonorant]

[-continuant, +coronal, -lateral][-continuant, -lateral, -mb]

[-continuant, +coronal, +sonorant][+labial]

[-continuant, +coronal, +sonorant][+nasal]

[-continuant, +coronal, +sonorant][-continuant, +sonorant]

[-continuant, +coronal, +sonorant][-continuant, -lateral, -mb]

[-continuant, +coronal, -sonorant][+anterior]
[-continuant, +coronal, -sonorant][-anterior]
[-continuant, +coronal, -sonorant][+cg]
[-continuant, +coronal, -sonorant][-cg]
[-continuant, +coronal, -sonorant][+coronal]
[-continuant, +coronal, -sonorant][+lateral]
[-continuant, +coronal, -sonorant][-nasal]
[-continuant, +coronal, -sonorant][+sg]
[-continuant, +coronal, -sonorant][-sg]
[-continuant, +coronal, -sonorant][-sonorant]
[-continuant, +coronal, -sonorant][+RTR, -continuant]
[-continuant, +coronal, -sonorant][+RTR, -syllabic]
[-continuant, +coronal, -sonorant][+back, -lateral]
[-continuant, +coronal, -sonorant][+back, -syllabic]
[-continuant, +coronal, -sonorant][-back, -lateral]
[-continuant, +coronal, -sonorant][-back, -syllabic]
[-continuant, +coronal, -sonorant][+continuant, -lateral]
[-continuant, +coronal, -sonorant][+continuant, -syllabic]
[-continuant, +coronal, -sonorant][-continuant, +dorsal]
[-continuant, +coronal, -sonorant][+dorsal, -syllabic]
[-continuant, +coronal, -sonorant][-lateral, -low]
[-continuant, +coronal, -sonorant][-low, -syllabic]
[-continuant, -lateral, -mb][+labial, -lateral]
[-continuant, -lateral, -mb][+labial, +nasal]
[-continuant, -lateral, -mb][+labial, +sonorant]
[-continuant, -mb, -sonorant][+anterior]
[-continuant, -mb, -sonorant][-anterior]
[-continuant, -mb, -sonorant][+cg]
[-continuant, -mb, -sonorant][-cg]
[-continuant, -mb, -sonorant][+coronal]
[-continuant, -mb, -sonorant][+lateral]
[-continuant, -mb, -sonorant][-nasal]
[-continuant, -mb, -sonorant][+sg]
[-continuant, -mb, -sonorant][-sg]
[-continuant, -mb, -sonorant][-sonorant]
[-continuant, -mb, -sonorant][+RTR, -continuant]
[-continuant, -mb, -sonorant][+RTR, -syllabic]
[-continuant, -mb, -sonorant][+back, -lateral]
[-continuant, -mb, -sonorant][+back, -syllabic]
[-continuant, -mb, -sonorant][-back, -lateral]
[-continuant, -mb, -sonorant][-back, -syllabic]
[-continuant, -mb, -sonorant][+continuant, -lateral]
[-continuant, -mb, -sonorant][+continuant, -syllabic]
[-continuant, -mb, -sonorant][-continuant, +dorsal]
[-continuant, -mb, -sonorant][+dorsal, -syllabic]
[-continuant, -mb, -sonorant][-lateral, -low]
[-continuant, -mb, -sonorant][-low, -syllabic]
[+coronal, -lateral, -nasal][+anterior, +sg]
[+coronal, -lateral, -nasal][-anterior, +cg]
[+coronal, -lateral, -nasal][-anterior, +sg]
[+coronal, -lateral, -nasal][+cg, +coronal]

[+coronal, -lateral, -nasal][+coronal, +sg]
[-lateral, -low, -mb][+anterior, +sg]
[-lateral, -low, -mb][+cg, +labial]
[-lateral, -low, -mb][+labial, +sg]
[-lateral, -low, +sonorant][+anterior, +sg]
[-lateral, -low, +sonorant][+cg, +labial]
[-lateral, -low, +sonorant][+labial, +sg]
[-lateral, -mb, -nasal][+anterior, +sg]
[-lateral, -mb, -nasal][+labial, +sg]
[-lateral, -nasal, +sonorant][+anterior, +sg]
[-lateral, -nasal, +sonorant][+labial, +sg]
[-low, -mb, -nasal][+anterior, +sg]
[-low, -mb, -nasal][+cg, +labial]
[-low, -mb, -nasal][+labial, +sg]
[-low, -mb, -syllabic][+anterior, +sg]
[-low, -mb, -syllabic][+cg, +labial]
[-low, -mb, -syllabic][+labial, +sg]
[-low, -nasal, +sonorant][+anterior, +sg]
[-low, -nasal, +sonorant][+cg, +labial]
[-low, -nasal, +sonorant][+labial, +sg]
[-low, +sonorant, -syllabic][+anterior, +sg]
[-low, +sonorant, -syllabic][+cg, +labial]
[-low, +sonorant, -syllabic][+labial, +sg]

Bibliography

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology 53*(5), 314–348.

Angluin, D. (1980a). Finding patterns common to a set of strings. *Journal of Computer and System Sciences 21*, 46–62.

Angluin, D. (1980b). Inductive inference of formal languages from positive data. *Information Control 45*, 117–135.

Angluin, D. and P. Laird (1988). Learning from noisy examples. *Machine Learning 2*, 343–370.

Arnauld, A. and C. Lancelot (1660). *Grammaire générale et raisonnée de Port-Royal/Précédée d'un Essai sur l'origine et les progrès de la langue françoise/par M. Petitot; et suivie du commentaire de M. Duclos, auquel on a ajoutédes notes*. Bossange et Masson.

Avcu, E., C. Shibata, and J. Heinz (2017, November). Subregular complexity and deep learning. Volume 1, Gothenburg, Sweden, pp. 20–33. Department of Philosophy, Linguistics and Theory of Science (FLOV), University of Gothenburg: CLASP, Centre for Language and Studies in Probability.

Baker, C. L. and J. J. McCarthy (1981). *The logical problem of language acquisition*. MIT Press (MA).

Beattie, J. (1788). *The Theory of Language: In Two Parts. Part I. Of the Origin and General Nature of Speech. Part II. Of Universal Grammar*. Number 10. A. Strahan.

Beim Graben, P. and S. Gerth (2012). Geometric representations for minimalist grammars. *Journal of Logic, Language and Information 21*(4), 393–432.

Beim Graben, P., D. Pinotsis, D. Saddy, and R. Potthast (2008). Language processing with dynamic fields. *Cognitive Neurodynamics 2*(2), 79–88.

Bills, G. D. et al. (1969). An introduction to spoken bolivian quechua.

Bird, S. (1995). *Computational phonology: A constraint-based approach*. Studies in Natural Language Processing. Cambridge: Cambridge University Press.

Blum, M. and L. Blum (1975). Towards a mathematical theory of inductive inference. *Information and Control 28*, 125–155.

Blutner, R. (2009). Concepts and bounded rationality: An application of niestegge's approach to conditional quantum probabilities. In *AIP Conference Proceedings*, Volume 1101, pp. 302–310. AIP.

Bourbaki, N. (1989). *Commutative Algebra: Chapters 1-7*. Springer-Verlag.

Büchi, J. R. (1960). Weak second-order arithmetic and finite automata. *Mathematical Logic Quarterly 6*(1-6), 66–92.

Cantor, G. (1892). *Über eine elementare Frage der Mannigfaltigkeitslehre*. Druck und Verlag von Georg Reimer.

Case, J. and S. Moelius (2007). Parallelism increases iterative learning power. In *18th Annual Conference on Algorithmic Learning Theory (ALT07)*, Volume 4754 of *Lecture Notes in Artificial Intelligence*, pp. 49–63. Springer-Verlag, Berlin.

Chandlee, J., R. Eyraud, J. Heinz, A. Jardine, and J. Rawski (2019). Learning with partially ordered representations. Ms.

Chandlee, J. and A. Jardine (2019). Quantifier-free least fixed point functions for phonology. In *Proceedings of the 16th Meeting on the Mathematics of Language (MoL 16)*, Toronto, Canada. Association for Computational Linguistics.

Chater, N., A. Clark, J. A. Goldsmith, and A. Perfors (2015). *Empiricism and language learnability*. OUP Oxford.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and control 2*(2), 137–167.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. The MIT Press.

Chomsky, N. and M. P. Schützenberger (1959). The algebraic theory of context-free languages. In *Studies in Logic and the Foundations of Mathematics*, Volume 26, pp. 118–161. Elsevier.

Clark, A. and S. Lappin (2011a). *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell.

Clark, A. and S. Lappin (2011b). *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell.

Clark, R. (1989). On the relationship between the input data and parameter setting. In J. C. . R.-M. Dèchaine (Ed.), *North Eastern Linguistic Society (NELS) 19*, Amherst, MA, pp. 48–62. University of Massachusetts, Graduate Linguistic Student Association.

Coleman, J. (1998). *Phonological representations: Their names, forms and powers*. Cambridge University Press.

Courcelle, B. (1994). Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science 126*(1), 53–75.

Danis, N. and A. Jardine (2019). Q-theory representations are logically equivalent to autosegmental representations. In *Proceedings of the Society for Computation in Linguistics*, Volume 2, pp. 29–38.

Day, T. and H. Kincaid (1994). Putting inference to the best explanation in its place. *Synthese 98*(2), 271–295.

de la Higuera, C. (2010a). *Grammatical Inference: Learning Automata and Grammars*. Cambridge University Press.

de la Higuera, C. (2010b). *Grammatical Inference: Learning Automata and Grammars*. Cambridge University Press.

Della Pietra, S., V. Della Pietra, and J. Lafferty (1997). Inducing features of random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 19*(4), 380–393.

Dolatian, H. (2020). *Computational locality of cyclic phonology in Armenian*. Ph. D. thesis, State University of New York at Stony Brook.

Dresher, E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry 30*, 27–67.

Dresher, E. and J. Kaye (1990). A computational learning model for metrical phonology. *Cognition 34*, 137–195.

Durvasula, K. and A. Liter (2020). There is a simplicity bias when generalising from ambiguous data. *Phonology 37*(2), 177–213.

Feldman, J. (1972). Some decidability results on grammatical inference and complexity. *Information and Control 20*(3), 244 – 262.

Gallagher, G. (2016). Vowel height allophony and dorsal place contrasts in cochabamba quechua. *Phonetica 73*(2), 101–119.

Gallistel, C. and A. P. King (2009). *Memory and the Computational Brain*. Wiley-Blackwell.

Gallistel, C. R. (1999). The replacement of general-purpose learning models with adaptively specialized learning modules. *The new cognitive neurosciences*, 1179–1191.

Garcia, P., E. Vidal, and J. Oncina (1990). Learning locally testable languages in the strict sense. In *Proceedings of the Workshop on Algorithmic Learning Theory*, pp. 325–338.

Gardenfors, P. (2004). Conceptual spaces as a framework for knowledge representation. *Mind and Matter 2*(2), 9–27.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition 1*(1), 3–55.

Gold, E. M. (1967). Language identification in the limit. *Information and Control 10*, 447–474.

Goldsmith, J. A. (1976). *Autosegmental phonology*, Volume 159. Indiana University Linguistics Club Bloomington.

Goldwater, S. and M. Johnson (2003). Learning ot constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, and O. Dahl (Eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory*, Volume 111–120, Stockholm. Stockholm University.

Gorn, S. (1967, September). Explicit definitions and linguistic dominoes. In J. Hart and S. Takasu (Eds.), *Systems and Computer Science*, Toronto, pp. 77–115. University of Toronto Press.

Gouskova, M. and G. Gallagher (2020). Inducing nonlocal constraints from baseline phonotactics. *Natural Language & Linguistic Theory 38*(1), 77–116.

Graf, T. (2010). Logics of phonological reasoning. Master's thesis, University of California, Los Angeles.

Graf, T. (2013). *Local and transderivational constraints in syntax and semantics*. Ph. D. thesis, University of California, Los Angeles.

Graf, T. (2014). Beyond the apparent: Cognitive parallels between syntax and phonology. UCLA Working Papers in Linguistics.

Grefenstette, E. (2013). Towards a formal distributional semantics: Simulating logical calculi with tensors. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*.

Haig, B. D. (1987). Scientific problems and the conduct of research. *Educational Philosophy and Theory 19*(2), 22–32.

Haig, B. D. (2018). An abductive theory of scientific method. In *Method matters in psychology*, pp. 35–64. Springer.

Hale, J. and P. Smolensky (2001). A parser for harmonic context-free grammars. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pp. 427–432.

Hammarberg, R. (1981). The cooked and the raw. *Journal of Information Science 3*(6), 261–267.

Hansson, G. (2010). *Consonant Harmony: Long-Distance Interaction in Phonology*. Number 145 in University of California Publications in Linguistics. Berkeley, CA: University of California Press. Available on-line (free) at eScholarship.org.

Hayes, B. (2009). *Introductory Phonology*. Wiley-Blackwell.

Hayes, B. and C. Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry 39*, 379–440.

Heinz, J. (2010a). Learning long-distance phonotactics. *Linguistic Inquiry 41*(4), 623–661.

Heinz, J. (2010b). Learning long-distance phonotactics. *Linguistic Inquiry 41*(4), 623–661.

Heinz, J. (2010c, July). String extension learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, pp. 897–906. Association for Computational Linguistics.

Heinz, J. (2016). Computational theories of learning and developmental psycholinguistics. In J. Lidz, W. Synder, and J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics*, Chapter 27, pp. 633–663. Oxford, UK: Oxford University Press.

Heinz, J., C. de la Higuera, and M. van Zaanen (2015a). *Grammatical Inference for Computational Linguistics*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool.

Heinz, J., C. de la Higuera, and M. van Zaanen (2015b). *Grammatical Inference for Computational Linguistics*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool.

Heinz, J., A. Kasprzik, and T. Kötzing (2012, October). Learning with lattice-structured hypothesis spaces. *Theoretical Computer Science 457*, 111–127.

Heinz, J. and J. Rawski (forthcoming). History of phonology: Learnability. In E. Dresher and H. van der Hulst (Eds.), *Oxford Handbook of the History of Phonology*, Chapter 32. Oxford University Press.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics 6*(1-4), 164–189.

Hopcroft, J., R. Motwani, and J. Ullman (1979). *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley.

Hornstein, N. and D. Lightfoot (1981). *Explanation in linguistics. The logical problem of language acquisition*. London: Longman.

Jain, S., D. Osherson, J. S. Royer, and A. Sharma (1999). *Systems That Learn: An Introduction to Learning Theory (Learning, Development and Conceptual Change)* (2nd ed.). The MIT Press.

Jakobson, R., C. Gunnar, M. Fant, and M. Halle (1952). *Preliminaries to Speech Analysis*. MIT Press.

Jardine, A. (2017). The local nature of tone-association patterns. *Phonology 34*, 385–405.

Jardine, A. and J. Heinz (2015, July). A concatenation operation to derive autosegmental graphs. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, Chicago, USA, pp. 139–151.

Jarosz, G. (2006). *Rich Lexicons and Restrictive Grammars – Maximum Likelihood Learning in Optimality Theory.* Ph. D. thesis, Johns Hopkins University.

Jarosz, G. (2013). Learning with hidden structure in optimality theory and harmonic grammar: Beyond robust interpretive parsing. *Phonology 30*(1), 27–71.

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review in Linguistics*.

Jensen, J. (1974). Variables in phonology. *Language 50*(4), 675–686.

Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM review 51*(3), 455–500.

Ladefoged, P. (2000). *A Course in Phonetics* (2nd ed.). Heinle.

Laime Ajacopa, T. (1992). *Diccionario Bilingue, Iskay Simipi Yu-yayk'ancha: Quechua–Castellano, Castellano–Quechua*. LaPaz, Bolivia.

Lambert, D., J. Rawski, and J. Heinz (2021). Typology emerges from simplicity in representations and learning. *Journal of Language Modelling*.

Lenneberg, E. (1967). *Biological foundations of language*. Oxford, England: Wiley.

Lipton, P. (2004). Evidence to the best explanation. 2nd edn london and new york.

Matilal, B. K. (1990). *The Word and the World India's Contribution to the Study of Language*.

McCoy, R. T., T. Linzen, E. Dunbar, and P. Smolensky (2018). Rnns implicitly implement tensor product representations. *arXiv preprint arXiv:1812.08718*.

McCulloch, W. S. and W. Pitts (1990). A logical calculus of the ideas immanent in nervous activity. *Bulletin of mathematical biology 52*(1-2), 99–115.

McNaughton, R. and S. Papert (1971). *Counter-Free Automata*. MIT Press.

Mitchell, T. (2017). Key ideas in machine learning. In *Machine Learning: Second Edition*. (forthcoming) `http://www.cs.cmu.edu/~tom/mlbook/keyIdeas.pdf`.

Morawietz, F. (2003). *Two-Step Approaches to Natural Language Formalism*. Berlin: Mouton de Gruyter.

Nazarov, A. and G. Jarosz (2017). Learning parametric stress without domain-specific mechanisms. In *Proceedings of the Annual Meetings on Phonology*, Volume 4.

Nickles, T. (1981). What is a problem that we may solve it? *Synthese*, 85–118.

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. Cambridge, MA: MIT Press.

Oakden, C. (2020). Notational equivalence in tonal geometry. *Phonology 37*(2), 257–296.

Osherson, D. and S. Weinstein (1983). Formal learning theory. In M. Gazzaniga and G. Miller (Eds.), *Handbook of Cognitive Neurology*. Plenum, New York.

Osherson, D., S. Weinstein, and M. Stob (1986). *Systems that Learn*. Cambridge, MA: MIT Press.

Piccinini, G. and S. Bahar (2013). Neural computation and the computational theory of cognition. *Cognitive science 37*(3), 453–488.

Pitt, L. (1989). Inductive inference, DFAs and computational complexity. In *Proceedings of the International Workshop on Analogical and Inductive Inference*, pp. 18–44. Springer-Verlag. Lecture Notes in Artificial Intelligence (v. 397).

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive neuropsychology 29*(1-2), 34–55.

Post, E. L. (1944). Recursively enumerable sets of positive integers and their decision problems. *bulletin of the American Mathematical Society 50*(5), 284–316.

Potts, C. and G. K. Pullum (2002). Model theory and the content of OT constraints. *Phonology 19*(3), 361–393.

Pullum, G. K. (2007). The evolution of model-theoretic frameworks in linguistics. *Model-theoretic syntax at 10*, 1–10.

Pullum, G. K. and B. C. Scholz (2001). On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In *International Conference on Logical Aspects of Computational Linguistics*, pp. 17–43. Springer.

Putnam, H. (1967). Psychological predicates. *Art, Mind, and reliogion*, 37–48.

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT press.

Rabanser, S., O. Shchur, and S. Günnemann (2017). Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*.

Raedt, L. D. (2008). *Logical and Relational Learning*. Springer-Verlag Berlin Heidelberg.

Rawski, J. and J. Heinz (2019). No free lunch in linguistics or machine learning: Response to pater. *Language 95*(1), e125–e135.

Rogers, J. (1996). Strict lt 2: Regular:: local: recognizable. In *International Conference on Logical Aspects of Computational Linguistics*, pp. 366–385. Springer.

Rogers, J. (1997). "grammarless" phrase structure grammar. *Linguistics and Philosophy 20*(6), 721–746.

Rogers, J. (1998). *A Descriptive Approach to Language-Theoretic Complexity*. Stanford, CA: CSLI Publications.

Rogers, J. (2003a). Syntactic structures as multi-dimensional trees. *Research on Language and Computation 1*(3-4), 265–305.

Rogers, J. (2003b). wMSO theories as grammar formalisms. *Theoretical Computer Science 293*, 291–320.

Rogers, J., J. Heinz, G. Bailey, M. Edlefsen, M. Visscher, D. Wellcome, and S. Wibel (2010). On languages piecewise testable in the strict sense. In C. Ebert, G. Jäger, and J. Michaelis (Eds.), *The Mathematics of Language*, Volume 6149 of *Lecture Notes in Artifical Intelligence*, pp. 255–265. Springer.

Rogers, J., J. Heinz, M. Fero, J. Hurst, D. Lambert, and S. Wibel (2013). Cognitive and sub-regular complexity. In G. Morrill and M.-J. Nederhof (Eds.), *Formal Grammar*, Volume 8036 of *Lecture Notes in Computer Science*, pp. 90–108. Springer.

Rogers, J. and D. Lambert (2019a, September). Extracting Subregular constraints from Regular stringsets. *Journal of Language Modelling 7*(2), 143–176.

Rogers, J. and D. Lambert (2019b, July). Some classes of sets of structures definable without quantifiers. In *Proceedings of the 16th Meeting on the Mathematics of Language*, Toronto, Canada, pp. 63–77. Association for Computational Linguistics.

Rogers, J. and G. K. Pullum (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information 20*, 329–342.

Sato, T. (2017). Embedding tarskian semantics in vector spaces. In *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.

Schmidhuber, J., F. Gers, and D. Eck (2002). Learning nonregular languages: A comparison of simple recurrent networks and lstm. *Neural Computation 14*, 2039–2041.

Serafini, L. and A. S. d. Garcez (2016). Learning and reasoning with logic tensor networks. In *Conference of the Italian Association for Artificial Intelligence*, pp. 334–348. Springer.

Sidiropoulos, N. D., L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing 65*(13), 3551–3582.

Simon, H. (1977). Models of discovery, dordrecht: D. *D. Reidel Publishinag Company*.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence 46*(1-2), 159–216.

Smolensky, P. and G. Legendre (2006a). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, Volume Volume II: Linguistic and Philosophical Implications. MIT Press.

Smolensky, P. and G. Legendre (2006b). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, Volume Volume I: Cognitive Architecture. MIT Press.

Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science Part A 41*(3), 260–270.

Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics*, Volume 1328 of *Lecture Notes in Computer Science*, Berlin, pp. 68–195. Springer.

Strother-Garcia, K. (2019). *Using model theory in phonology: a novel characterization of syllable structure and syllabification*. Ph. D. thesis, University of Delaware.

Strother-Garcia, K., J. Heinz, and H. J. Hwangbo (2016, October). Using model theory for grammatical inference: a case study from phonology. In S. Verwer, M. van Zaanen, and R. Smetsers (Eds.), *Proceedings of The 13th International Conference on Grammatical Inference*, Volume 57 of *JMLR: Workshop and Conference Proceedings*, pp. 66–78.

Tabor, W. (2009). A dynamical systems perspective on the relationship between symbolic and non-symbolic computation. *Cognitive neurodynamics 3*(4), 415–427.

ter Meulen, A. G. (2012). Mathematical linguistics. *Oxford Bibliographies*.

Tesar, B. (2014). *Output-driven phonology: Theory and learning*. Number 139. Cambridge University Press.

Tesar, B. and P. Smolensky (2000). *Learnability in Optimality Theory*. MIT Press.

Thagard, P. (1988). *Computational philosophy of science*. MIT press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton University Press.

Thomas, W. (1982). Classifying regular events in symbolic logic. *Journal of Computer and Systems Sciences 25*, 370–376.

Thomas, W. (1997). Languages, automata, and logic. In *Handbook of Formal Languages*, Volume 3, Chapter 7. Springer.

Turing, A. M. (1937). On computable numbers, with an application to the entscheidungs problem. *Proceedings of the London mathematical society 2*(1), 230–265.

Valiant, L. G. (1984, November). A theory of the learnable. *Communications of the ACM 27*(11), 1134–1142.

van Rooij, I. and G. Baggio (2021, January). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*.

Vu, M. H., A. Zehfroosh, K. Strother-Garcia, M. Sebok, J. Heinz, and H. G. Tanner (2018). Statistical relational learning with unconventional string models. *Frontiers in Robotics and AI 5*, 76.

Weiss, G., Y. Goldberg, and E. Yahav (2017). Extracting automata from recurrent neural networks using queries and counterexamples. *arXiv preprint arXiv:1711.09576*.

Williams, E. S. (1976). Underlying tone in margi and igbo. *Linguistic Inquiry*, 463–484.

Wilson, C. and G. Gallagher (2018). Accidental gaps and surface-based phonotactic learning: A case study of south bolivian quechua. *Linguistic Inquiry 49*(3), 610–623.

Wilson, C. and M. Obdeyn (2009). Simplifying subsidiary theory: statistical evidence from arabic, muna, shona, and wargamay. Johns Hopkins University.

Yang, B., W.-t. Yih, X. He, J. Gao, and L. Deng (2014). Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature communications 10*(1), 1–7.