

Gradient Symbolic Representations in Grammar: The case of French Liaison

Paul Smolensky & Matthew Goldrick

SHORT ABSTRACT

Longstanding theoretical debates about whether structure A or structure B is the correct analysis of phenomenon X are commonplace. For example, at the juncture of two words W_1 and W_2 , French liaison consonants alternate with zero. Theories of French phonology have long debated whether the consonant is associated with W_1 or W_2 . In this work, we argue for an alternative approach. Phenomena X is not accounted for by either A or B, but rather a *conjunctive blend* of structures A and B. This notion of ‘blend of structures’ is formalized using *Gradient Symbolic Representations*, symbol structures in which a particular position is generally occupied by a sum of gradient symbols, each symbol having a partial degree of presence: its *activity*. The grammatical consequences of a Gradient Symbolic Representation are the sum of the consequences of all the symbols blended to form it; the consequences of a symbol — e.g., the costs of constraint violations — are proportional to its activity. The proposed grammatical computation consists of optimization with respect to a numerical weighting of familiar phonological constraints from Optimality Theory and Harmonic Grammar, straightforwardly extended to evaluate Gradient Symbolic Representations. We apply this general framework to French liaison consonants, blending together elements of previous proposals to give a single analysis that covers a wide range of data not previously explicable within a single theory.

LONG ABSTRACT

At the foundation of the work reported here is the following hypothesis: theoretical debates about whether structure A or structure B is the correct analysis of phenomenon X can persist indefinitely because, in fact, the mental representation supporting X is a *conjunctive blend* (not a disjunctive probabilistic mixture) of structures A and B. The notion ‘blend of structures’ is formalized using *Gradient Symbolic Representations*, symbol structures in which each individual position is generally occupied by a sum of *gradient symbols*, each symbol having a partial degree of presence: its *activity*. The grammatical consequences of a Gradient Symbolic Representation are the sum of the consequences of all the symbols blended to form it; the consequences of a symbol — e.g., the costs of constraint violations — are proportional to its activity. Gradient Symbolic Representations enable formal expression of important gradient theoretical intuitions.

Our test case is a well-studied phenomenon of French phonology, liaison consonants, which alternate with zero. Theoretical intuitions expressed in the extensive existing literature concerning the structures underlying liaison — intuitions previously assumed to be in conflict — are blended to give a single analysis that covers a wide range of data not previously explicable within a single theory. The analysis formalizes the following intuitions about a liaison consonant \mathcal{L} which may appear at the juncture of two consecutive words $W_1 W_2$: \mathcal{L} is simply a consonant that is weak; different \mathcal{L} s are not all equally weak; \mathcal{L} is underlyingly final in W_1 ; \mathcal{L} is underlyingly initial in W_2 ; the greater the cohesion between W_1 and W_2 — i.e., the smaller the minimal prosodic unit containing both W_1 and W_2 — the greater the likelihood that \mathcal{L} will appear. The proposed grammatical computation consists of optimization with respect to a numerical weighting of familiar constraints from Optimality-Theoretic phonology, straightforwardly extended to evaluate Gradient Symbolic Representations.

Gradient Symbolic Representations constitute the data of *Gradient Symbolic Computation* (GSC), a general computational architecture for cognition developed over the past several decades. The microstructure of a GSC system is a stochastic neural network deploying continuous, distributed representations; however this paper addresses only the macrostructure of the proposed GSC analysis, which is a *Probabilistic Harmonic Grammar* defined over Gradient Symbolic Representations. Although the network-level description is not utilized in this paper, the proposed analysis operates over gradiently active symbols which are ultimately emergent from patterns of gradiently active model neurons.

0. INTRODUCTION AND SYNOPSIS

There are many fairly well-recognized aspects of continuous, numerical gradience in grammar: relative degrees of acceptability, probabilistic mixtures of grammars during language change, probabilistic expectations of uncertain upcoming linguistic material in on-line processing, not to mention the obvious continuous aspects of spoken and signed linguistic signals. A different type of gradience is the topic of the research presented here: continuous, numerical gradience internal to linguistic representations themselves. It is not the kind of continuous pictorial representations proposed in certain theories of cognitive linguistics that we study, but rather a kind of gradience within structured representations built of discrete symbols. In the *Gradient Symbolic Representations* we examine, symbols are discrete but their *degree of presence* in a given linguistic representation is continuously gradient. Thus the last position in a phonological string might contain a segment *t* that is *partially present*: its degree of presence, or *activity*, might for example be 0.5. This does not mean that the last segment is /t/ with probability a half: it means that the last segment is a /t/ with consequences half as strong as a standard, fully-present /t/. This means that constraints targeting that position will be satisfied or violated to a degree one-half as large as would be the case for a standard /t/.

Furthermore, a particular position in a given Gradient Symbol Structure may host multiple symbols, each present (or *active*) to a continuously variable degree: this is a *gradient blend*. Thus the first position in a phonological string might contain a blend of 3 symbols, each partially present to a certain degree — say, *t*, *z*, and *n*, each with an activity level of 0.3. If the remaining segments in the string are standard symbols *a*, *m*, *i*, then we write such a string: $(0.3 \cdot t + 0.3 \cdot z + 0.3 \cdot n)ami$. This is a string with 4 positions, the first of which hosts a blend of 3 gradient symbols.

The example application we address here is French liaison, in which certain underlying consonants surface (are pronounced) only before vowels. These consonants are analyzed through Gradient Symbolic Representations as being literally weak: they are only partially present in underlying forms — active to a degree less than 1. The bold underlined **t** in the pronunciation $.pø.ti.t̥a.mi$ of *petit ami* (lit., ‘little friend’, MASC) disappears in the pronunciation of *petit copain* (lit., ‘little friend’, MASC), $.pø.ti.ko.pɛ̃$. The *t* associated with *petit* is intuitively ‘weak’ in the sense that it only appears under optimal syllabic conditions: when it can serve as a (universally favored) syllable onset. This contrasts with the standard (not weak) **t** of $.pø.tit.ko.pin$ *petite copine* (lit., ‘little friend’, FEM); this underlying **t** forces its way to the surface even in this non-optimal context where it must serve as a (universally disfavored) syllable coda. In the analysis proposed below, while the underlying form of *petite* ends in a standard **t** (activity level 1), the underlying form of *petit* has in its final position a gradient **t** with an activity level of only 0.5: it is $/pøti(0.5 \cdot t)/$. Furthermore, it is proposed that the underlying form of *ami* is exactly the string mentioned above — $/(0.3 \cdot t + 0.3 \cdot z + 0.3 \cdot n)ami/$; this has in its first position a blend of the 3 productive liaison consonants of French.

There are 4 larger contexts in which the proposed work must be situated (elaborated in Sec. 1 below). Most obviously, the work needs to be located in the landscape of the existing massive literature on liaison, which is one of the most-analyzed phenomena of phonology. The analysis proposed here posits a lexicon for French that can be viewed as a gradient blend of previous proposals; this introduces a second larger context: that of linguistic phenomena — of which there are many — that have received multiple analyses, none fully satisfactory. Gradient Symbolic Representations hold the promise, illustrated with liaison, of resolving long-standing disputes by blending together analyses previously viewed as competitors. Such a general perspective is the third larger context, a perspective that was proposed by David Dowty (2003) for the dispute between analyses of certain PPs as complements vs. as adjuncts. Dowty motivates his blending-of-multiple-structures picture through a sketch of acquisition which we here instantiate, in preliminary form, for the case of liaison (Sec. 5). And finally, the study of Gradient Symbolic Representations is motivated not only by linguistic considerations, but also by research addressing the extremely general context of the computational architecture of

cognition. These representations are the data structure over which **Gradient Symbolic Computation (GSC)** is carried out: GSC is a cognitive architecture, in development for nearly 3 decades, that integrates structured symbolic computation such as that customary in linguistics with neural-network computation originating in psychology and neuroscience. GSC provides a framework not only for competence theory, but also for performance theory, which is touched upon briefly below. The concept of the activity level of a gradient symbol arises directly out of the concept of the activation level of units in a neural network (although a single symbol corresponds to an extended pattern of activity distributed over many network units, the same units hosting the patterns for all the symbols; these patterns are superimposed upon each other in the representation of a given symbol structure). Aside from a few high-level remarks, we will not discuss the neural- (or ‘connectionist’-) network foundations of GSC here (on which see [3], [60]).

An outline of the paper is provided in (1), which gives the corresponding Section numbers.

- (1) Outline of the paper (with Section numbers)
 1. Context of the work
 2. Gradient Symbolic Computation in grammar: An informal nano-introduction
 3. The phonological phenomenon: Liaison in French
 4. A GSC analysis
 - 4.1. The intuition
 - 4.2. The formal analysis of liaison
 - 4.3. A meta-analysis of the liaison analysis
 - 4.4. A visualization of the GSC Analysis
 - 4.5. Restrictiveness of the account
 5. Acquisition: Speculations on formalizing Dowty’s sketch in GSC
 6. Contextual factors in liaison — The role of prosody: Tentative suggestions
 7. Extensions
 8. Summary

The absence or presence of liaison consonants — both those generally accepted as ‘correct’ forms and those regarded as ‘errors’ — is known to be sensitive to a wide range of diverse factors. Many of these are treated at least partially in the proposed GSC analysis; they are listed in (2).

- (2) Factors contributing to the complete liaison pattern
 - a. Treated here
 - i. syllable structure well-formedness
 - ii. morphosyntactic or prosodic context
 - iii. lexical exceptionality of multiple sorts
 - iv. linguistic register
 - v. frequency of lexical item combinations
 - vi. prosodic breaks
 - vii. stage of acquisition
 - b. Not treated here
 - i. word length
 - ii. speaker’s linguistic generation — language change in progress

1. CONTEXT OF THE WORK

1.1. General overview of GSC

The analysis discussed in this paper explores the value for grammatical theory of Gradient Symbolic Representations, the foundation of Gradient Symbolic Computation (GSC). In GSC, one and the same representation can be described in two ways: it can be described at a more abstract level as a symbol structure or at a more fine-grained level as a list of numbers. Putting all these numbers into one long list gives an *activation vector*; this aggregates the activation values, at a given time, of all the neurons in the underlying network. These representations are called *Tensor Product Representations* because the tensor (or generalized outer) product is used to bind the activation vector encoding each symbol to the activation vector encoding the role that that symbol plays in the structure as a whole.

The GSC research program has been in development for three decades and stands in contrast to the two research frameworks which have dominated cognitive science since the mid-1980s. One framework takes as an axiom that the whole enterprise of describing mental representations as symbol structures is ill-conceived — the brain doesn't have symbols, *of course*, and the mind is *clearly* too fluid to be captured by rigid symbolic rule systems. The competing approach is entirely sold on the symbolic level — so obviously there's no value for *cognitive* (as opposed to neuro-) science in trying to descend to a level closer to the neural level. But recently there has been growing interest in integrating neural networks and symbolic computation (including symbolic grammars).¹ In the last 10–15 years neural networks have led to dramatic advances in Natural Language Processing; they have improved speech recognition and text processing systems enormously [10]. In many new systems, these neural network computations are coupled with other components performing symbolic computation. Getting these two very different types of computation to work together coherently is clearly a challenge — which GSC is aimed at addressing. In this approach, we don't have neurons here, and symbols there: we have just one system which can be described at one level as a neural network and at another level as (a quite novel kind of) symbolic computation.

In GSC, *knowledge* takes the form of *gradient constraints*, elements of a *Probabilistic Harmonic Grammar*. A constraint in this type of grammar can be described at the symbolic level, as standardly in linguistics, but it can also be implemented as a group of connections within a neural network; these connections drive the system to create maximally well-formed, or *optimal*, representations.

There are two classes of network that have developed in this general approach; they are used for somewhat different purposes.

- (multi-)linear feed-forward neural networks
- stochastic feed-back (higher-order) neural networks

The former type of network has been used to demonstrate that TPRs can be used in neural computation to compute complex symbolic functions of central interest to cognitive science and NLP [50] [57]. It is the latter type of network that underlies the work discussed in this paper.

The adequacy of the GSC architecture has been evaluated over the years in many ways. On the symbolic side — testing whether GSC does adequate justice to symbolic computation — it has been shown that GSC systems can compute: recursive functions; beta reduction (function application) in the lambda calculus; tree adjoining; and some kinds of logical inference. In GSC we can precisely specify, and compute asymptotically, formal languages at all levels of complexity; and as for natural languages, Harmonic Grammar and especially Optimality

¹ After initial work in the 1990's [22] [26] [38] [43] [44], there was little research aimed at such integration until quite recently.

Theory have been used by many linguists with numerous successful results at levels ranging from phonology to semantics and pragmatics [33] [35] [50].

On the neural network side, the computational adequacy of the GSC framework is attested (i) primarily by formal results: theorems about convergence of the processing to the global maximum of Harmony, which is the optimal state which we desire to compute; and (ii) in recent work with colleagues at Microsoft Research, applications to inference and question answering [29] [57].

On the biological neural network side, the adequacy of GSC is far from established; there is some limited positive evidence but ongoing work aims to test directly the hypothesis that tensor product representations are used to encode combinatorial linguistic structures in the brain.

As for the integrated use of both symbolic and neural levels of GSC, current work aims to develop psycholinguistically adequate models of sentence production and comprehension. Of central importance in that work is the complex interaction between crucial gradient and strong structure-sensitivity. Some initial applications of this work have accounted for gradient in sound structure errors in speech production [18] and the distribution of gradient structures in multilingual code switching [19] [21].

All of this is some evidence that the GSC architecture gives us a lot of the power that symbolic computation and neural computation provide, and derives new power from their unification.

1.2. The key general prediction

But most important here is a certain kind of very general prediction that comes out of GSC, something that has not been attended to in the research program until relatively recently. It's not just discrete structures that populate the space of representations in GSC; there are also **non-discrete** Gradient Symbolic Representations; in fact almost all representations are of this type. So the very general prediction is that ***non-discrete Gradient Symbolic Representations should play important roles in cognition***. It's the current focus of the GSC research program to understand what those roles might be.

For now, we can take a Gradient Symbolic Representation to be a discrete structure — perhaps a syntactic tree or an autosegmental multi-tiered phonological structure — in which the symbols in these representations are gradient: they have numerical **activity** levels which indicate the degree to which that symbol is present in the structure. A given position in the structure is occupied by a **blend** of such gradient symbols. Thus in a phonological string, for example, at a single string position there might be a blend of multiple gradient phones, each present to some degree. A standard, **discrete** Gradient Symbolic Representation is one in which, at each non-empty position in the structure, exactly one symbol has non-zero activity, and its activity value is 1.0.

The analysis presented here will be couched exclusively at the symbolic level; the neural level will not be addressed. We'll be exploring how Gradient Symbolic Representations in *lexical* representations can give us some leverage on the classic problem in French phonology, the behavior of liaison consonants (introduced below: (3)).

Why go beyond classical discrete symbol structures in grammatical theory? Because of the following fundamental — quite frustrating — issue which we believe is just a fact of life. Symbolic analyses in linguistics provide a lot of insight, but typically *they don't quite work*. The very general hypothesis we're exploring is that with Gradient Symbolic Representations we can resolve long-standing theoretical disputes that arise because no single analysis does all the work that linguistics would like to see done. Pervasively in the linguistics literature we find analysis A that explains an important part of some empirical pattern X, and a competing analysis B that explains a different but also important part of the pattern X. These competing analyses can survive for decades, with no long-lasting resolution.

The hypothesis we offer is that *such a theoretical impasse persists because, in fact, X arises from a blend of the structure proposed by Analysis A and the structure proposed in Analysis B*: neither alone captures all the critical components of the structure responsible for X. Here, X will be the behavior in French of liaison consonants. Liaison consonants alternate with zero: they can appear when preceding a vowel (3a) but disappear when preceding a consonant (3b).

(3) Most basic consonant liaison alternation (‘.’ = syllable boundary)

- | | | | | |
|----|---------------------|-----------------------------|----------------------------|-----------|
| a. | <i>petit ami</i> | (literally) ‘little friend’ | pronounced: .pø.ti.ʔa.mi. | ʔ present |
| b. | <i>petit copain</i> | (literally) ‘little friend’ | pronounced: .pø.ti.ko.pɛ̃. | ʔ absent |

“Analysis A” follows the orthography: it asserts that the lexical entry for *petit* has a final /t/ — but one that is somehow deficient, and therefore does not always surface (i.e., get pronounced). In Analysis A the lexical entry for *ami* is just /ami/.

The alternative “Analysis B” follows the syllabification rather than the orthography; it posits that the lexical entry for *petit* is² /pøti/, with no final consonant; rather, the liaison consonant [t] is initial in the relevant lexical entry for *ami*, which is /tami/. This is just one of the multiple allomorphs of *ami*; the lexicon also contains /zami/, /nami/, and even /ami/: these are the forms that appear in sequences such as *les ami* ‘the friend’ [lɛ.za.mi.], *un ami* ‘a friend’ [ɛ̃.na.mi.], *joli ami* ‘pretty friend’ [ʒo.li.a.mi.]. It is the word preceding *ami* that selects the appropriate allomorph (*petit* ‘small’ selects /tami/).

As already mentioned, the proposal explored in this paper is (literally, formally) a blend of Analyses A and B; the hypothesis is that such a single blended analysis can account for *both* the body of data accounted for by Analysis A and that accounted for by Analysis B.³ In support of this hypothesis, we show below how the GSC analysis can account for 14 input → output grammatical mappings capturing not only the core behavior — illustrated in part in (3) — but also a diverse range of peripheral mappings.

In the proposed GSC analysis, a *partially-active* /t/ is present at the end of the lexical representation of *petit*, AND a *partially-active* /t/ is present at the beginning of the lexical representation of *ami*. The entire analysis revolves around the *simultaneous* presence of these two partially-active segments. Blends in GSC are *conjunctive*: partially-active elements are co-present. This is in contrast to a corresponding *probabilistic mixture* in which with some probability p (a fully-active, discrete) /t/ would be present in one of these positions OR with probability $1-p$ it would be present at the other position; the probability is 0 that it would be partially present at both positions simultaneously.⁴ In the proposed GSC analysis, the consonant’s being pronounced will be the *joint consequence* of the two blended partially-active consonants in the lexical representations of the two words.⁵

We must make it clear that the goal of this paper is to illustrate the *potential* of Gradient Symbolic Representations (GSRs) to provide novel enlightening accounts of many of the phenomena that have been *claimed* to occur in the rich scope of liaison — putting aside the many divergent views on the actual empirical status of these alleged phenomena. New

² IPA ø is approximately a fronted, rounded schwa-like vowel; it is sometimes written simply ə.

³ Special thanks to Jennifer Culbertson for pointing out that liaison provides an excellent testing ground for such a hypothesis.

⁴ Of course, some non-zero probability could be assigned to the structure in which one /t/ is present in one of these positions and another /t/ is present at the other. But these would be two fully discrete tokens of the same symbol type, not a mixture of partially-present symbols.

⁵ Although developed independently, the current proposal instantiates a conception of Jorge Hankamer [23]: “we must give up the assumption that two or more conflicting analyses cannot be simultaneously correct for a given phenomenon” (pp. 583–4); “such constructions have both analyses at once (in the conjunctive sense)” (p. 592). Thanks to Eric Baković for pointing out this remarkable 1977 paper.

empirical work currently underway is designed to test the robustness of the alleged phenomena, setting the stage for a truly empirical argument for the adequacy of a GSC analysis. For the moment, however, we pursue the less ambitious (and indeed logically prior) objective of testing how GSRs can *in principle* shed light of the wide range of evidence that undergirds the conflict between the two existing (discrete) analyses of liaison, Analyses A and B.

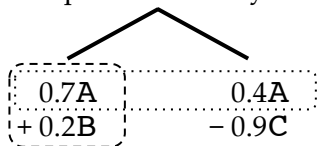
As already mentioned, the approach to liaison we pursue here takes some inspiration from a provocative 2003 paper by David Dowty [11] in which he sketches a radical approach to a classic case of conflicting evidence for two structural analyses of a phenomenon, the complement vs. adjunct analyses of certain PPs. Dowty's proposal is that children initially form a simple, maximally general analysis — the purely compositional adjunct analysis — but that over time adults develop a more complex, specialized analysis — the complement analysis, in which the semantic contribution of the PP is to a significant extent idiosyncratically dictated by the verb. Crucially for our purposes, Dowty posits that the child's adjunct analysis *persists* into adulthood, where it functions jointly with the complement analysis, "in some subtle psychological way, in on-line processing—though in a way that only connectionism or some other future theories of the psychology of language can explain." [antepenultimate paragraph] It is indeed a connectionist-based gradient blend of two discrete analyses that we are proposing in this paper, although we consider the proposal to lie primarily in the realm of competence rather than performance. In Sec. 5 we return to Dowty 2003 for some speculation on the acquisition of the blend that is our proposed adult grammar.

2. GRADIENT SYMBOLIC COMPUTATION IN GRAMMAR: AN INFORMAL NANO-INTRODUCTION

2.1. A syntactic example

Consider (4), a minimal example of a Gradient Symbolic Representation:

(4) Simple Gradient Symbolic tree



This simple example can be looked at in two ways. First, as suggested by the dashed box outline, this can be viewed as a very simple local tree in which the left-child position contains a blend of two symbols, mostly (0.7) A, but also slightly (0.2) B. This is a perspective that is particularly useful for phonology, where elements tend to stay in place but change their content (e.g., featural composition), say, from B to A.

In a second way view of this local tree, suggested by the dotted box outline, the symbol A occupies a blend of *roles*: mostly left-child (0.7) but also right-child (0.4). This perspective is particularly useful for syntax, where elements tend to change position (or occupy multiple positions) while their content remains largely intact.⁶

[Although it is not directly relevant to the rest of the paper, for those interested in the connection of these Gradient Symbolic Representations to their lower-level neural network encodings, the Tensor Product Representation encoding (4) is shown in (5)]

⁶ For expository purposes we use the local tree (4) even though in any bona fide syntactic representation, a blend of roles occupied by a constituent would not be a blend of sister nodes. An unaccusative clause might, for example, be analyzed with the argument of the intransitive verb occupying a blend of subject and direct object roles.

(5) Neural-network activation vector encoding the Gradient Symbolic Representation (4)

$$\begin{bmatrix} 0.7 \mathbf{A} \otimes \mathbf{r}_{\text{left}} + 0.4 \mathbf{A} \otimes \mathbf{r}_{\text{right}} \\ + 0.2 \mathbf{B} \otimes \mathbf{r}_{\text{left}} - 0.9 \mathbf{C} \otimes \mathbf{r}_{\text{right}} \end{bmatrix} = \begin{bmatrix} (0.7 \mathbf{A} + 0.2 \mathbf{B}) \otimes \mathbf{r}_{\text{left}} \\ \mathbf{A} \otimes (0.7 \mathbf{r}_{\text{left}} + 0.4 \mathbf{r}_{\text{right}}) \end{bmatrix} + (0.4 \mathbf{A} - 0.9 \mathbf{C}) \otimes \mathbf{r}_{\text{right}} + \mathbf{B} \otimes (0.2 \mathbf{r}_{\text{left}}) + \mathbf{C} \otimes (-0.9 \mathbf{r}_{\text{right}})$$

In (5), \mathbf{A} , \mathbf{B} , \mathbf{C} denote the neural activation vectors encoding symbols \mathbf{A} , \mathbf{B} , \mathbf{C} , while \mathbf{r}_{left} , $\mathbf{r}_{\text{right}}$ denote the neural activation vectors encoding left-, right-child position. \otimes denotes the tensor product.]

2.2. A phonological example

(6) is another example of a Gradient Symbolic Representation, used in our GSC liaison analysis.

$$(6) \quad [{}^M \text{pøti}(\lambda \cdot t)] \quad [{}^M (\tau \cdot t + \zeta \cdot z + \nu \cdot n) \text{ami}]$$

This is a string (with morpheme boundaries $[{}^M \]$) consisting of the concatenation of the underlying forms of the two morphemes *petit* and *ami*. In the first four positions of the string $\text{pøti}(\lambda \cdot t)$, which is the lexical form of *petit*, there are standard segments p , \emptyset , t , i , each with the unmarked activity level, 1.0. The last position of the string $\text{pøti}(\lambda \cdot t)$ is a gradient symbol: a gradient t with activity level λ (which will have value 0.5 in the particular analysis presented below).

In the proposed lexical form of *ami*, preceding the string of three standard (activity-1) segments a , m , i , is a blend of gradient forms of the three productive liaison consonants of French; this blend is a single entity which constitutes the first element of the underlying string for *ami*. In the blend, each consonant appears with an activity level denoted by the corresponding Greek letter (in the particular analysis presented below, τ , ζ , and ν will all have the value 0.3).

It is also useful to view (6) as a representation containing a t that occupies two roles: partly, the final segment of *petit*, but also, the initial segment of *ami*.

2.3. GSC — Competence and performance theories: (Probabilistic) Harmonic Grammar

In fact, the momentary state of a Gradient Symbolic Computation is a *probability distribution* over such Gradient Symbolic Representations. In the proposed **competence** theory, we attend only to the most probable representation (which might be a blend structure such as (4)). In the proposed **performance** theory, the probabilities of other representations are relevant as they are the probabilities of different types of errors. These probabilities reflect randomness that is central to the dynamics of computation in the underlying neural network, and so play an important role in GSC real-time processing models. Here we will not treat the dynamics of GSC computation, but will rather exploit the key property of this dynamics: the most probable output of computation is a representation that maximizes a well-formedness measure called **Harmony**. In fact, under ideal processing conditions, the probability that the output will be a given representation r is proportional to $e^{H(r)/T}$, where the ‘computational temperature’ parameter T governs the momentary level of randomness in the neural network (7). Ideally, during the computation of a single output, T drops slowly to zero, with the result that the output probability for all representations goes to zero except for the representation that has the highest Harmony — the optimal representation, which has an output probability of 1⁷. This is the GSC competence theory.

(7) \mathfrak{P}_1 : Probabilistic Harmonic Grammar; GSC performance theory

The probability of a representation r is determined by its Harmony $H(r)$:

$$p(r) \propto e^{H(r)/T}$$

⁷ When there are multiple grammatical outputs (in the context of stable grammatical variation) we assume that speakers maintain a probability distribution over Harmonic Grammars.

But aside from some discussion of errors and variation below, here we will stick to the competence-theoretic idealization.

In GSC, the Harmony of a representation r is the weighted sum of the violations by r of the constraints that constitute the grammar: $H(r) = \sum_k w_k C_k(r)$. Each constraint C_k , then, assesses a numerical degree of satisfaction/violation $C_k(r)$ to r , and has a numerical strength or weight w_k in the grammar: this defines a **Harmonic Grammar**. A positive weight w_k indicates a **positive constraint** C_k , which rewards the Harmony of r in proportion to $C_k(r)$, the degree to which r satisfies C_k . A negative weight w_j indicates a **negative constraint** C_j , which penalizes the Harmony of r in proportion to $C_j(r)$, the degree to which r violates C_k . (Optimality Theory [46] [47] has a corresponding *non-numerical* notion of Harmony derived from a non-numerical notion of constraint strength encoded in a strict dominance hierarchy ranking all constraints from strongest to weakest, each constraint being stronger than the combined strength of all lower-ranked constraints. As in Harmonic Grammar, the output of the grammar is the representation with maximal Harmony – the optimal candidate.)

The lineage of GSC extends back to Harmony Theory, a neural network architecture for general cognition which derived the probability–Harmony relation of \mathfrak{P}_1 and the form of the Harmony function from the Maximum Entropy induction principle [51] [52]. Applying Harmony Theory to grammar gave rise to the Harmonic Grammar competence theory [32] [58: Chs. 6, 11] while the Probabilistic Harmonic Grammar theory was developed largely by Bruce Hayes and Colin Wilson [24], under the name used in computational linguistics, Maxent. Harmonic Grammar enjoyed a revival in phonology largely as a result of theoretical and computational work by Joe Pater and colleagues [42] [45].⁸

What is novel in GSC is that it is not just constraint strengths that are numerically gradient: the very representations evaluated by constraints are built of symbols with numerically gradient degrees of presence, i.e., activity. This is a change in the most fundamental formal category in which linguistic representations lie, a shift from the category of discrete mathematics to that of continuous mathematics. This aspect of GSC also extends back almost 3 decades, with the development of Tensor Product Representations [53] [54] [58: Chs. 5, 9].

In many ways, the numerical interaction of constraints in Harmonic Grammars is inherently more complex and difficult to grasp than the strict-domination interaction of Optimality Theory. Early Harmonic Grammars could be designed, using neural-network learning algorithms, to compute constraint weights so as to successfully account for challenging patterns of data, but those grammars proved extremely difficult to analyze ([58: Ch. 11] [59]). A priority in the current GSC research is developing methods for actually *understanding* the Harmonic Grammars deployed. Towards this objective a visualization technique is presented below (see Sec. 4.3, (24) ff.) and used for meta-analysis – analysis of the proposed analysis of liaison.

3. THE PHONOLOGICAL PHENOMENON: LIAISON IN FRENCH

Expanding on (3), the core phenomena of French liaison are illustrated in (8).

(8) Core phenomena of French liaison

| | | | | |
|-----------------------|------------------|---------------------|----------------------|-------------------|
| a. orthography | <i>petit ami</i> | <i>petit copain</i> | <i>petite copine</i> | <i>petit héro</i> |
| b. pronunciation | .pø.ti.ta.mi. | .pø.ti.ko.pẽ. | .pø.tit.ko.pin. | .pø.ti.e.ʁo. |
| c. alternation t ~ Ø | [t] | no [t] | [t] | no [t] |
| d. syllable structure | no coda, onset | no coda, onset | coda, onset | no coda, no onset |

⁸ An important grammatical framework that uses numerical constraint strengths in a quite different way is Paul Boersma's Stochastic Optimality Theory [3]; in this architecture, outputs always result from a strict Optimality-Theoretic constraint ranking, but there is variability in the rankings; the probability that C_k out-ranks C_j in the production of a particular output is determined by the difference in their numerical strengths, $s_k - s_j$.

Row (8a) gives 4 two-word sequences, written in French orthography. The pronunciations of these sequences are given in row (8b), in IPA; ‘.’ marks the edges of syllables. The first 3 two-word sequences all literally mean ‘little friend’; the third sequence *petite copine* is feminine while the first two are masculine. The fourth sequence *petit héro* means ‘little hero’, masculine. Of interest is row (8c), identifying the alternation between a [t] present (in bold) between the two words in *petit ami* and *petite copine*, as opposed to no [t] in *petit copain* and *petit héro*. This [t] is a liaison consonant.

The generalization is that the liaison consonant [t] associated with *petit* appears only before a vowel, when it is pronounced as the onset (initial consonant) of the first syllable of the second word (standardly; but see (11) ⑥). This contrasts with *petite*, which is pronounced with a final [t] even before a consonant, in which case the [t] appears in the coda (final) position of the final syllable of *petite*; in (8) this is shown for *petite copine*. Finally, while it is true that a [t] appears after *petit* only before a vowel-initial word, it is not true that it appears before every vowel-initial word: in (8), *petit héro* illustrates the case of an *h-aspiré* word (*héro*) before which no [t] appears. (Most *h-aspiré* words do in fact have an initial orthographic *h*, which is not pronounced; French has no [h] sound. Not all orthographically *h*-initial words are *h-aspiré* words, however; this is a lexically-idiosyncratic property.)

Important insight into the alternation shown in (8c) is provided through syllable structure; (8d) identifies the syllable-structure configuration at the juncture of the two words⁹: the underlined portion of the pronunciations in (8b). The basic alternation *petit ami* ([t] present) vs. *petit copain* ([t] absent) can be understood as driven by the universal syllable-structure constraints which in Optimality Theory are called (i) **ONSET** and (ii) **NOCODA** [46]; these state that (i) a syllable with an onset consonant is better-formed than one without an onset consonant, all else equal, and (ii) a syllable with no coda consonant is better-formed than one with a coda, all else equal. This [t] appears when it is needed to provide an onset (for the second word’s initial syllable), and it does not appear when it would form a coda (for the first word’s final syllable): ONSET favors .pø.ti.t̥a.mi. over *.pø.ti.a.mi., and NOCODA favors .pø.ti.ko.pẽ. over *.pø.ti.t̥.ko.pẽ.

While syllable-structure constraints seem to determine the presence or absence of a liaison consonant in *petit ami* vs. *petit copain*, these constraints do not govern the presence of all consonants. The feminine form *petite* is always pronounced with a final [t], even when doing so violates NOCODA, as in *petite copine* (8). And *h-aspiré* words force a violation of ONSET in sequences like *petit héro* (8), where the liaison consonant is not pronounced.

The core empirical patterns in (8) can be summarized abstractly and compactly in terms of the four core input → output mappings ① – ④ using the notation introduced in (9), to be used in the remainder of the paper.

(9) Core mappings

- | | | | | |
|---|---------------------|---------------------------------|-------------------------|------------------|
| ① | v + V | → v.v | <i>joli + ami</i> | → .jo.li.a.mi. |
| ② | v \mathcal{L} + V | → v. $\underline{\mathcal{L}}v$ | <i>peti(t) + ami</i> | → .pø.ti.t̥a.mi. |
| ③ | v \mathcal{L} + c | → v.c | <i>peti(t) + copain</i> | → .pø.ti.ko.pẽ. |
| ④ | v \mathcal{L} + V | → v.V | <i>peti(t) + héro</i> | → .pø.ti.e.ɾo. |
| ⑤ | v \mathcal{F} + c | → v \mathcal{F} .c | <i>petite + copine</i> | → .pø.tit.ko.pin |

Here and henceforth, liaison consonants are denoted by \mathcal{L} ; fixed consonants, by \mathcal{F} . Parentheses are used here (but usually omitted below) in *peti(t)* to indicate that the final *t* of *petit* is a liaison consonant \mathcal{L} . The apparently trivial mapping ① has been stated explicitly because under the proposed analysis it will in fact prove non-trivial. The input configurations given before the

⁹ The remaining possible combination, coda and no onset, completely at odds with the universal syllable structure constraints introduced below, actually appears in the peripheral phenomena treated by the GSC analysis: see (11) ⑥.

arrow correspond to the output material that is underlined. Here and throughout liaison consonants that surface in outputs are bold and heavily underlined, e.g., the **t̥** in (9 ①).

Words that are vowel-initial in isolated, citation form, divide into *h-aspiré* words — which will be notated in configurations with an initial \mathbb{V} , and in pseudo-orthographic form with \mathbb{h} (9 ③) — and standard words, notated with an initial \mathbb{V} (9 ②, ①). (Note however that, as anticipated above in (6), the proposed lexical form for \mathbb{V} -initial words is crucially *not* \mathbb{V} -initial; such forms have in their first position a gradient blend of liaison consonants.)

The pattern in (8)–(9) shows that in the French lexicon, the liaison /t/ of *petit* (which alternates with \emptyset , i.e., nothing) must be distinguished, somehow, from the standard or **fixed** /t/ of *petite* (which does not alternate)¹⁰: this is the analytic task we take up first. Later we will turn to how the French lexicon distinguishes ordinary vowel-initial words like *ami* (which induce the appearance of liaison consonants) from *h-aspiré* words like *héro* (which do not).

3.1. The proposed lexicon: A blend of two competing prior analyses

The underlying contrast between fixed \mathcal{F} and liaison \mathcal{L} consonants is addressed by the second central principle hypothesized by the proposed GSC account of liaison (10).

(10) \mathfrak{B}_2 : underlying \mathcal{L} vs. \mathcal{F}

In lexical entries (underlying form), at the end of a word, a liaison \mathcal{L} = /t/ and a fixed \mathcal{F} = /t/ differ in only one respect: the fixed consonant \mathcal{F} is a standard, discrete segment — i.e., it has activity 1.0 — while the liaison consonant \mathcal{L} is a gradient segment with activity λ , less than 1.0. \mathcal{L} is literally just a weak version of \mathcal{F} .

The intuition is that, unlike the full /t/ ending the lexical form of *petite*, the gradient /t/ at the end of the underlying form of *petit* — /pøti(λ ·t)/ — is too weak to surface on its own: it is *activity-deficient* and can only be pronounced if it gets additional activity from another source.

Henceforth the 2-word sequences under investigation will be notated $W_1 W_2$. So far, following the orthography, we've assumed that a liaison consonant is *final* in the word it follows: (9). This will be called the $\hat{W}_1\mathcal{L}$ (or **Final- \mathcal{L}**) **Analysis**, which will also be assumed to posit *syllabification-driven alternation*: \mathcal{L} appears when, and only when, it is needed to provide a syllable onset — except that it does not appear before *h-aspiré* words. Thus when W_1 is *petit*, on this analysis its underlying form is $\hat{W}_1\mathcal{L}$ where \hat{W}_1 = pøti and \mathcal{L} is a liaison t. (This was named “Analysis A” in Sec. 1.2; “Analysis B” is defined next.)

The Final- \mathcal{L} Analysis is predominant and even taught in schools. However there has long been a group of phonologists who favor a competing analysis: the $\mathcal{L}\hat{W}_2$ (or **\mathcal{L} -Initial**) **Analysis** according to which the liaison consonant is morphologically affiliated with W_2 rather than W_1 (e.g., [40]). This analysis follows the syllabification rather than the orthography. For *petit ami*, pronounced .pø.ti.t̥a.mi., \mathcal{L} appears in the underlying form as /tami/: $\mathcal{L}\hat{W}_2$, where \mathcal{L} = t and \hat{W}_2 = ami. As noted in the synopsis, this entails that the lexical entry for *ami* contains multiple allomorphs: /tami/, /zami/, /nami/, /ami/. The correct allomorph of W_2 for the underlying form of a particular sequence $W_1 W_2$ is selected by W_1 : W_1 = *petit* selects / W_2 / = /tami/, whereas W_1 = *les, un, joli* respectively select / W_2 / = /zami/, /nami/, /ami/.

Some may find the \mathcal{L} -Initial Analysis unparsimonious or inelegant relative to the Final- \mathcal{L} Analysis, but certain empirical phenomena we will present, which go beyond the core of liaison behavior, favor the \mathcal{L} -Initial analysis. Furthermore, the allomorphy posited by the $\mathcal{L}\hat{W}_2$ Analysis is observed directly in child language; this analysis can therefore play the role of the

¹⁰ Since it forms a minimal pair with *petit*, the form *petite* is useful as an instance of a ‘fixed final consonant’ \mathcal{F} , but under the proposed analysis, the final consonant of *petite* is actually underlyingly a liaison consonant which behaves like a fixed consonant because of the effect of the [FEMININE] morpheme (Sec. 4.5.10). In this sense, better examples of \mathcal{F} would be the final consonant of *brut* ‘crude’ or of *juste* ‘just (Adj)’; unlike *petite*, these are monomorphemic.

child ‘grammar’ in a Dowty 2003-style account of the development of the adult liaison ‘grammar’ (as introduced in Sec. 1.2 and developed preliminarily in Sec. 5; for liaison, it is most directly the child vs. adult *lexicon* that is at issue).

There are several non-core liaison phenomena that are problematic for a strictly-syllabification-driven account such as the Final- \mathcal{L} Analysis. We will then consider several further phenomena that are problematic for the Final- \mathcal{L} Analysis but expected on the \mathcal{L} -initial Analysis.

(11) Trouble for strictly syllabification-driven distribution of \mathcal{L}

- ⑤ *Phrase-final \mathcal{L}* . In a few words, have, e.g., *huit*|| \rightarrow *qit*|| (but *vingt*|| \rightarrow *vɛ̃*||)
- ⑥ *Coda \mathcal{L}* . Can get $v\mathcal{L} + V \rightarrow v\mathcal{L}.v$ instead of $v.\mathcal{L}v$ (but never $v\mathcal{L} + C \rightarrow *v\mathcal{L}.c$)
- ⑦ *h -aspiré onset \mathcal{F} (but not \mathcal{L})*. Can get $v\mathcal{F} + \mathbb{V} \rightarrow v.\mathcal{F}\mathbb{V}$ (but not $v\mathcal{L} + \mathbb{V} \rightarrow *v.\mathcal{L}\mathbb{V}$)
- ⑧ *Pre/post-pausal \mathcal{L}* . \mathcal{L} can surface before/after a prosodic break
- ⑨ *Frequency effect*. Where optional, $p(\mathcal{L} \text{ surfaces}) \sim p(W_1W_2)$

For an excellent survey of most of the empirical and theoretical ingredients relevant here, see Marie-Hélène Côté 2011 [8], a review which has been invaluable for the development of the proposed analysis. Other crucial resources include Bernard Tranel 1981 [63], Jacques Durand and Chantal Lyche 2008 [13], and Bernard Laks 2009 [28].

We now spell out the 5 problematic mappings in (11) one by one. Readers eager to get to the proposed GSC analysis can read (12) and (13) and skip the rest of Sec. 3 with little loss of continuity.

3.2. *Phrase-final \mathcal{L}* : (11) ⑤

A few words, such as *huit* ‘eight’, pronounce their liaison consonant \mathcal{L} when they appear at the end of a phonological phrase.¹¹ Since the phonological phrase is the domain over which syllabification takes place in French, the final [t] of *huit*, when it occurs phrase-finally, cannot be syllabified with what follows; thus it must surface as a syllable coda rather than as an onset, the position it is limited to phrase-internally. Pronouncing \mathcal{L} phrase-finally patently violates NOCODA and does not produce satisfaction of ONSET. Note that this behavior is limited to about four exceptional words; e.g., the corresponding liaison consonant of *vingt* ‘twenty’ does not surface phrase-finally. We therefore now have evidence for at least a *3-way contrast* in the lexicon, between fixed consonants \mathcal{F} , normal liaison consonants \mathcal{L} , and exceptional liaison consonants \mathcal{L}^* which surface phrase-finally. From the perspective of the proposed analysis, \mathcal{F} , \mathcal{L} and \mathcal{L}^* will be distinguished (only) by having three different activity levels in the lexicon. This is the tip of a large iceberg of lexical distinctions: as the proposed analysis develops, more distinct activity levels will appear — different degrees of weakness of weak consonants.¹²

¹¹Noam Faust (2016) [15] argues that the environment for this exceptional behavior is more generally those environments in which liaison is prohibited. The GSC Analysis presented below can account for this if the prosodic unit the boundaries of which blocks liaison (the Maximal Phonological Phrase) coincides with the domain of the ALIGN-R^{||} constraint, a domain left unspecified in Sec. 4.5.1.

¹²Kie Zuraw and Bruce Hayes (2016) [66] argue for five degrees of *h-aspiré*-ness of W_2 s as reflected in their resistance to elision (deletion) of W_1 -final schwas (another consequence, beyond resistance to the surfacing of W_1 -final liaison consonants, of *h-aspiré* W_2 s). They also argue for three degrees of W_1 resistance to elision. These conflict with the syllable structure constraints which favor elision (to avoid an onsetless initial syllable in W_2). Zuraw and Hayes implement ‘degrees of resistance’ as distinct weights for lexical-item-tagged Harmonic Grammar constraints. The GSC analysis proposed for liaison here could be extended to handle these gradient elision facts by positing three underlying activity levels for W_1 -final schwas and five underlying activity levels for initial schwas in *h-aspiré* W_2 s; these weak initial Vs in W_2 s are irrelevant for liaison as they cannot coalesce with W_2 -final weak Cs.

3.3. Coda \mathcal{L} : (11) ⑥

There is a particular register of French sometimes used in the speeches of public figures — studied thoroughly by Pierre Encrevé (1988) [14] and further by Bernard Laks (2009) [28] — in which liaison consonants are often produced in the coda position of the final syllable of W_1 , leaving the initial syllable of W_2 without an onset. For example, V. G. d’Estaing: *j’avais un rêve* ‘I had a dream’ pronounced [ʒa.vɛz.ɛ̃.ʁɛv] [14: 32ff.] in which the underlined sequence violates both NOCODA and ONSET. This unusual syllabification is robust and systematic — not a result of production error; in at least one speech of each of four French Presidents (Pompidou, Mitterrand, G. d’Estaing, and Chirac), such syllabification was used in 15–25% of liaison sequences [28]. These liaison consonants appear in coda position before a vowel-initial W_2 only; before a consonant-initial W_2 , the liaison consonant does not appear, even though it could in principle surface in the coda position just as easily as it does before a vowel-initial W_2 . The liaison consonant appears where it *could* provide a needed onset, but it does not surface in the onset position, which is left empty. Clearly this is quite problematic for syllabification-driven distribution of \mathcal{L} .

3.4. *h*-aspiré onset \mathcal{F} (but not \mathcal{L}): (11) ⑦

In violation of prescriptive rules, *h*-aspiré W_2 words do variably acquire onset consonants from W_1 words. For example, in *cinq Hollandais*, or *une hauteur*, W_2 can variably receive from W_1 an onset [k] or [n], respectively [63: 305]. Crucially, this intrusion of an onset into an *h*-aspiré word can sometimes occur when W_1 ends in a fixed consonant \mathcal{F} , but liaison consonants \mathcal{L} do not appear with *h*-aspiré W_2 s [64: 814]. The absence of \mathcal{L} before an *h*-aspiré W_2 cannot therefore arise from a strict prohibition of onset consonants for such words. It would appear that such W_2 s, like ordinary V-initial W_2 s, *do* ‘need’ an onset consonant, in which case syllabification-driven distribution should (incorrectly) predict the appearance of \mathcal{L} to satisfy ONSET.

3.5. Pre/post-pausal \mathcal{L} : (11) ⑧

Liaison consonants sometimes occur before or after a prosodic break (which we denote ||) separating W_1 and W_2 . Attested examples of \mathcal{L} || include: *sans envisager* [sã:z || ɑ̃vizaʒe] ‘without expecting’; *est un* [ɛt || œ̃] ‘is a’, while examples of || \mathcal{L} include: *petites histoires* [pətit || zistwa:ʁ] ‘little stories’; *quelques années* [kɛlkə || zane] ‘several years’ [1: 25].

\mathcal{L} || here is problematic for syllabification-driven distribution of \mathcal{L} for the same reason it is problematic phrase-finally for words such as *huit* ⑤. || \mathcal{L} is likewise problematic for the syllabification-driven $\hat{W}_1\mathcal{L}$ Analysis: a W_1 -final \mathcal{L} is not available to provide an onset for W_2 when a prosodic break cleaves the syllabification of W_1 and W_2 . Note, however, that || \mathcal{L} is perhaps less anomalous for the $\mathcal{L}\hat{W}_2$ Analysis where \mathcal{L} is present for the syllabification of W_2 even after a break. (W_1 ’s selection of the correct allomorph of W_2 must, however, span the break.)

3.6. Frequency effect: (11) ⑨

In those W_1W_2 -configurations where liaison is variable or “optional” (see Sec. 6), the probability that \mathcal{L} surfaces increases as a function of the frequency of the W_1W_2 sequence in the language [1] [5]. (This can be seen as a case of the general correlation, discussed in Sec. 6, between pronunciation of \mathcal{L} and the “cohesion” between W_1 and W_2 — often construed as morphosyntactic closeness, but here rather a kind of lexical cohesion.) This frequency effect entails that, regardless of the lexical affiliation of \mathcal{L} , neither the lexical entry for W_1 nor that of W_2 alone can provide all the information needed to completely capture the behavior of \mathcal{L} : information about the *pair* W_1, W_2 is also needed. (Such frequency dependence motivates certain “usage-” or “construction-” based accounts of liaison; a formalization of a kind of usage-based account will in fact be blended into the proposed account — along with the $\hat{W}_1\mathcal{L}$ and $\mathcal{L}\hat{W}_2$ accounts of the bulk of \mathcal{L} behaviors, which are not clearly frequency-dependent.)

3.7. Grammar and performance

We note that all the problematic behaviors in (11) can easily be put aside as peripheral performance effects, of no concern to a competence theory of core liaison behavior. This move is however rejected by phonologists favoring an alternative to the syllabification-driven Initial- \mathcal{L} Analysis; they seek a grammatical account of both the core mappings (9) *and* the peripheral mappings (11). The analysis proposed here shares this goal.

The liaison phenomena in (12) which we now take up are even more readily amenable to dismissal as irrelevant to competence theory, but again, with the advocates of the \mathcal{L} -Initial Analysis, we adopt the more ambitious goal of explaining as much of the empirical pattern as possible with a grammatical account.

There are a number of types of liaison error that favor an \mathcal{L} -Initial over a Final- \mathcal{L} Analysis.

(12) Errors that are *expected* under the $\mathcal{L}\hat{W}_2$ Analysis but *not* under the $\hat{W}_1\mathcal{L}$ Analysis

- ⑩ *Incorrect \mathcal{L} insertion.* When an incorrect C is substituted for \mathcal{L} , it is another liaison C: * $\mathbf{v}.\mathcal{L}'\mathbf{v}$ for $\mathbf{v}.\mathcal{L}\mathbf{v}$
- ⑪ *Exceptional \mathcal{L} epenthesis.* When what should be V.V is illicitly repaired by C-insertion, it is a liaison C: * $\mathbf{v}.\mathcal{L}'\mathbf{v}$ for $\mathbf{v}.\mathbf{v}$
- ⑫ *Child \mathcal{L} -as- $\mathcal{F}\hat{W}_2$.* $\mathcal{L}\hat{W}_2$ treated as if word $\mathcal{F}\hat{W}_2$ — e.g., *joli 'nami'*

3.8. Incorrect \mathcal{L} insertion: (12) ⑩

In one class of error, in a location where a given \mathcal{L} consonant should appear, an incorrect consonant appears in its place. In such errors, the erroneous consonant is almost always another liaison consonant. Quite a few errors of this type occurred in a reading study of French high-school students reported by David Hornsby (2011) [25]; e.g., the sequence *long apprentissage* ‘long apprenticeship’ was pronounced correctly (with $\mathcal{L} = [g]$) 4 times and incorrectly 20 times — all errors of this type. Across the entire experiment there were 81 errors of incorrect C insertion, and the erroneous consonants were: [z] (56 times), [t] (23 times), or [n] (2 times) — in every case, one of the 3 productive liaison consonants. In the Final- \mathcal{L} Analysis, there is no explanation of why an erroneously inserted consonant should be a liaison consonant. But in the \mathcal{L} -Initial Analysis, such errors are in fact expected: each corresponds to a mis-selected allomorph, $*\mathcal{L}'\hat{W}_2$ instead of $\mathcal{L}\hat{W}_2$. Such mis-selection can insert only a liaison consonant. (When the vowel-initial allomorph $*\hat{W}_2$ is selected in place of correct $\mathcal{L}\hat{W}_2$, the error is simply omission of \mathcal{L} ; this is another common type of error.)

3.9. Exceptional \mathcal{L} epenthesis: (12) ⑪

In another type of error, where hiatus V.V should occur, a consonant is inserted: V.cV (e.g., inserting [t] between W_1 and W_2 in *joli ami*). In such errors, the inserted consonant c is with high probability one of the three productive liaison consonants [t, z, n] (e.g., this is true in all 389 errors made by Sophie between 2;1 and 3;6, with inclusion of quasi-liaison [l] [7: 765]). Again, the Final- \mathcal{L} Analysis offers no explanation of this generalization, whereas the $\mathcal{L}\hat{W}_2$ Analysis provides exactly the same explanation as for the errors in ⑩: allomorph mis-selection — in this case, selecting $*\mathcal{L}'\hat{W}_2$ instead of \hat{W}_2 .

3.10. Child \mathcal{L} -as- $\mathcal{F}\hat{W}_2$: (12) ⑫

The acquisition of French liaison has been well-studied (e.g., [4]). Famously, French-acquiring children (e.g., at around 20 months [2]) actually use *tami*, *zami*, *nami* as free variants of *ami* (producing, e.g., *joli tami*, *peti nami*, etc.). This behavior is the same as adult errors of types ⑪ and ⑫, so it too provides an argument in favor of the \mathcal{L} -Initial Analysis. Furthermore, such an early state in child language is actually *expected* given the strong word segmentation heuristic “the beginning of a word coincides with the beginning of a syllable: [$\mathbf{w}\mathbf{a} = [\sigma]$ ” (e.g. [41]) (This is

the consequence, in comprehension, of the constraint ALIGN-L(Morpheme, Syllable) proposed in Sec. 4.2.1) As previously mentioned, this early stage of child French forms part of the Dowty-2003-inspired sketch of the development of the liaison grammar proposed in Sec. 5.

Thus the peripheral behavior of liaison identified above offers at least 6 behaviors favoring, to at least some degree, the \mathcal{L} -initial Analysis over the syllabification-driven Final- \mathcal{L} Analysis: ⑥ – ⑫ except for ⑨. Yet the \mathcal{L} -initial Analysis has a major complication that is absent in the Final- \mathcal{L} Analysis: multiple W_2 allomorphs, with selection driven by some idiosyncratic information in the lexical entry of W_1 ; to facilitate future reference, this is enumerated as ⑬ in (13).

(13) Challenges for the \mathcal{L} -initial Analysis

⑬ W_2 allomorph selection for $\mathcal{L}\hat{W}_2$ Analysis

⑭ Gender-bending \mathcal{L} . *belle copine* and *belle amie*;
beau copain but **beau ami*: instead *bel ami*.

⑭ identifies another challenge for the \mathcal{L} -initial Analysis, which is equally a challenge for the Final- \mathcal{L} Analysis; it pertains to a small class of adjectives with idiosyncratic masculine/feminine alternation. The feminine form of ‘beautiful’, *belle* [bɛl] has a fixed final consonant $\mathcal{F} = [l]$. For consonant-initial W_2 such as *copain*, the corresponding masculine form is *beau* [bo]. But for vowel-initial W_2 such as *ami*, the masculine expression uses the feminine form (orthographically but not phonologically modified): [.bɛ.la.mi.] written *bel ami*. The consonant alternation is similar to *petit/petite*, where the masculine form *petit* has a liaison consonant $\mathcal{L} = [t]$ while the feminine form *petite* has a fixed consonant $\mathcal{F} = [t]$. But here the vowel also differs between genders.

The GSC analysis we now propose provides a unified analysis of all behaviors ① – ⑭ given in (11)–(13) above. These mappings cover a large majority of all behaviors that have been accounted for by any previous account; none of these accounts handles the full range ① – ⑭.

4. A GSC ANALYSIS

We first give the intuition behind the account, then move to the formal analysis. The proposed underlying representation of liaison consonants is literally a weighted blend of the Final- \mathcal{L} and \mathcal{L} -initial Analyses. (A limited component from usage-based and ‘morphological’ accounts [8] will also enter the proposed GSC Analysis.)

4.1. The intuition

Recall that at the most general level, the hypothesis under exploration is that when linguists can’t agree on whether structure A or structure B explains phenomenon X, it’s because in fact X arises from a gradient blend of A and B. For liaison, A posits that the liaison consonant \mathcal{L} is lexically stored at the end of W_1 while B has it stored at the beginning of W_2 . In the blended analysis, it is partially in both places. Specifically we propose the types of gradient lexical entries given in (14). When a word appears in the W_1 role, its possible lexical entries are divided into two types with respect to the final position; in the W_2 role, three types with respect to the initial position.

(14) Lexical forms in the GSC analysis

Underlying forms in $W_1 W_2$

$(\lambda, \tau, \zeta, \nu) \doteq (0.5, 0.3, 0.3, 0.3)$ are constant across the entire lexicon [to 1st approximation]

a. / W_1 /

| | |
|--|----------------------------|
| = $\hat{W}_1(\lambda \cdot \mathcal{L})$ | <i>peti(t)</i> /pøti(λ·t)/ |
| = $\hat{W}_1(1 \cdot \mathcal{F})$ | <i>juste</i> /ʒys(1·t)/ |
| = \hat{W}_1 | <i>joli</i> /ʒoli/ |

| | | | |
|---------------------------|---------------|---------|--|
| b. /W ₂ / | | | |
| = C \hat{W}_2 | <i>copain</i> | /kopɛ̃/ | |
| = V \hat{W}_2 | <i>héro</i> | /ɛʁo/ | (<i>h-aspiré</i>) |
| = $\mathfrak{L}\hat{W}_2$ | <i>ami</i> | /ami/ | where $\mathfrak{L} \equiv (\tau \cdot t + \zeta \cdot z + \nu \cdot n)$ |

This lexicon blends the Final- \mathcal{L} Analysis, weighted by $\lambda \doteq 0.5$, with the \mathcal{L} -initial Analysis, weighted by $\tau, \zeta, \nu \doteq 0.3$. This implements the third principle hypothesized by the GSC Analysis, (15).

(15) \mathfrak{F}_3 : Liaison consonants underlying occupy a gradient blend of two positions

A liaison consonant \mathcal{L} that surfaces in the sequence $W_1W_2 = \hat{W}_1\mathcal{L}\hat{W}_2$ derives simultaneously from two underlyingly sources: $/W_1/ = /(\lambda \cdot \mathcal{L})/$ and $/W_2/ = /(\gamma \cdot \mathcal{L})\hat{W}_2/$, where W_1 's underlying weak final \mathcal{L} and W_2 's underlying weak initial \mathcal{L} have activity values λ and γ respectively.

[When $\mathcal{L} = t$, we write γ as ' τ '; when $\mathcal{L} = z$, as ' ζ '; and when $\mathcal{L} = n$, as ' ν '.]

The intuition behind the proposed GSC explanation of mapping ① $v\mathcal{L}+V \rightarrow v.\underline{\mathcal{L}}v$ (*petit ami*) is that the weak final liaison consonant of $W_1 = \hat{W}_1(\lambda \cdot \mathcal{L})$ gets the extra activity it needs to surface by coalescing with the corresponding weak initial liaison consonant of $W_2 = \mathfrak{L}\hat{W}_2 = (\tau \cdot t + \zeta \cdot z + \nu \cdot n)$; e.g., for $\mathcal{L} = t$, the total underlying activation for t at the juncture of W_1 and W_2 is $\lambda + \tau \doteq 0.8$, which will be strong enough to pass the relevant activity threshold for surfacing.

By contrast, in the case of mapping ② $v\mathcal{L}+c \rightarrow v.c$ (*petit copain*) there is no initial liaison consonant in W_2 , so the W_1 -final \mathcal{L} does not get the extra activation it needs to surface. The same is true in the case of mapping ③ $v\mathcal{L}+V \rightarrow v.V$. (*petit héro*): an *h-aspiré* W_2 has no initial liaison consonant blend \mathfrak{L} : it is underlying V-initial. And as for the apparently trivial mapping ④ $v+V \rightarrow v.v$ (*joli ami*), because W_2 has an initial consonant blend ($V \equiv \mathfrak{L}V$), there arises the possibility that a consonant may appear at the $W_1 - W_2$ juncture: but it will not in the optimal output, because the consonants in the initial blend \mathfrak{L} are much too weak.

4.2. The formal analysis of liaison

Applied to psycholinguistics or performance theory, GSC deals with Gradient Symbolic Representations as *outputs* of the grammatical system. Here, however, as is typical of theoretical linguistics, the empirical generalizations we are trying to explain are all stated purely in terms of discrete forms, so we will consider only fully discrete outputs: it is only the *inputs*, the lexically stored forms of morphemes, that exhibit (a highly restricted form of) gradience.

We propose a Harmonic Grammar that is built of standard constraints from Optimality-Theoretic phonology, interpreted — as explained below — so as to be applicable to Gradient Symbolic Representations: we will refer to this as a *Gradient Harmonic Grammar*. The Harmony function that encapsulates the grammar can be written in terms of these constraints as in (16).

(16) The proposed Harmonic Grammar: the Harmony of a representation r is

$$H(r) = -10 \cdot C_{\text{DEP}}(r) + 2 \cdot C_{\text{MAX}}(r) + 1 \cdot C_{\text{ALIGN-L}}(r) - 0.9 \cdot C_{\text{ONSET}}(r) - 0.7 \cdot C_{\text{UNIF}}(r)$$

As mentioned above and elaborated below, for the k^{th} constraint C_k , $C_k(r)$ is the degree of violation (or satisfaction) of the constraint by r . In general these degrees of constraint violation/satisfaction need not be integers, i.e., whole numbers — hence the “gradient” in “Gradient Harmonic Grammar”.

The numerical constraint weights in the proposed grammar (16) were derived by hand from the data. It is a testament to the interpretability of the analysis that such hand-computation is possible. In larger problems, automatic methods for determining suitable weights will be needed (e.g., [45]). As mentioned in Sec. 2.32.3, machine-learned constraint weights typically lead to Harmonic Grammar analyses that are extremely difficult to understand; hand-

computation, in contrast, inevitably entails a significant degree of comprehensibility to the analysis and in this sense, at least, is desired where possible.

The Harmonic Grammar tableau in (17) applies this grammar to the input for *petit ami*, $[^M\text{p}\text{ø}\text{ti}(\lambda\cdot\text{t}_1)] [^M(\tau\cdot\text{t}_2+\zeta\cdot\text{z}_3+v\cdot\text{n}_4)\text{ami}]$. The non-integer degrees of violation of the FAITHFULNESS constraints DEP and MAX are clearly shown; as it happens, the three other constraints have integer violations.

4.2.1. Mapping ① (*petit ami*): $v\mathcal{L} + V \rightarrow v.\mathcal{L}V$; configuration $v\mathbf{C}V$

(17) displays the proposed GSC analysis of *petit ami*, which instantiates the first core mapping (9) ①. The input here is $[^M\text{p}\text{ø}\text{ti}(\lambda\cdot\text{t}_1)] [^M(\tau\cdot\text{t}_2+\zeta\cdot\text{z}_3+v\cdot\text{n}_4)\text{ami}]$, the lexical representation of the morpheme *petit* (enclosed in brackets $[^M \]$) followed by that of *ami*. Note that in the mapping notation used here, the input denoted “ $v\mathcal{L} + V$ ” refers to a W_2 that is a standard (not *h-aspiré*) “vowel-initial” word, which according to the proposed analysis actually has in its first position a blend of gradient liaison consonants $\tau\cdot\text{t}+\zeta\cdot\text{z}+v\cdot\text{n} \equiv \mathcal{L}$ (so in effect $V \equiv \mathcal{L}V$). Taking a more theory-neutral descriptive perspective, we can identify the input here as “an intervocalic consonant”, a target element \mathbf{C} in an environment v_v , i.e., the configuration $v\mathbf{C}v$; this compact notation will be used in the meta-analysis below.

(17) Gradient Harmonic Grammar tableau for configuration $v\mathbf{C}v$ (*petit ami*)

| <i>weight:</i> | | -10 | 2 | 1 | -0.9 | -0.7 | |
|---|------------------------------|---------------------------|-----------------------|---------|-------|------|----------|
| $[^M\text{p}\text{ø}\text{ti}(\lambda\cdot\text{t}_1)] [^M(\tau\cdot\text{t}_2+\zeta\cdot\text{z}_3+v\cdot\text{n}_4)\text{ami}]$ | | DEP | MAX | ALIGN-L | ONSET | UNIF | <i>H</i> |
| <i>a</i> | .pø.ti.a.mi. | | | | 1 | | -0.9 |
| <i>b</i> | .pø.ti.t ₁₂ a.mi. | $1-(\lambda+\tau)$ 0.2 | $\lambda+\tau$ 0.8 | 1 | | 1 | -0.1 |
| <i>c</i> | .pø.ti.t ₁ a.mi. | $1-\lambda$ 0.5 | λ 0.5 | | | | -4 |

Three critical candidate outputs for *petit ami* are displayed in (17): *a*, in which no liaison consonant surfaces; *b*, in which the liaison consonant surfaces through coalescence of the two matching gradient consonants in the input at the juncture of W_1 and W_2 ; and *c*, in which the liaison consonant surfaces as the correspondent of only the gradient input liaison consonant at the end of W_1 . Following the intuition behind the analysis sketched in Sec. 4.1, the optimal output should be *b*. Through (17), we now show that this is the case, given the Gradient Harmonic Grammar (16); we take us the constraints one at a time, defining them as we go.

In candidate *a* there is no surface liaison consonant. As a result, the first syllable of $W_2 = \text{ami}$ surfaces as $[\text{.a}]$, a syllable lacking an onset, in violation of the constraint ONSET from the Basic Syllable Structure [46]. The other candidates satisfy ONSET because there is a surface liaison consonant to serve as onset of the first syllable of W_2 . Since the weight of ONSET in this harmonic grammar is -0.9, the onsetless syllable reduces the Harmony of *a* relative to either *b* or *c* by 0.9.

In candidate *b*, there is a surface liaison segment $[\text{t}_{12}]$ which is the correspondent of both the weak $/\text{t}_1/$ at the end of the underlying form of W_1 and the weak $/\text{t}_2/$ at the beginning of the underlying form of W_2 . This is **coalescence**, as formalized in the *Correspondence Theory of Faithfulness* of John McCarthy and Alan Prince (1995) [37]; this particular type of departure from 1-to-1 input-output correspondence violates the constraint UNIFORMITY (UNIF). This constraint violation is registered in the HG Tableau as the ‘1’ in the UNIF column¹³. The other two

¹³ Unlike MAX and DEP, the violations of UNIF are not scaled by the activation of input elements. This is because UNIF does not directly refer to the input elements themselves, but rather to the correspondence relations that hold between elements in the input and output. We assume that these correspondence relationships (along with the output

candidates lack coalescence, so they satisfy UNIF. Because the weight of UNIF in the grammar is -0.7 , this violation lowers the Harmony of b relative to either a or c by 0.7 .

The strongest constraint in this analysis is **DEP**, which in standard discrete Correspondence Theory is violated by a segment in the output with no correspondent in the input; intuitively, rather than being contributed by the input, such a segment must be generated from nothing by the candidate generator *Gen*. The proposed gradient version of DEP is violated in candidate b because, while the surface segment $[t_{12}]$ does have an input correspondent (two, in fact), the total activation of these input segments, $\lambda + \tau = 0.8$, is less than the activity of the output segment, 1.0 . *Gen* does not need to add a new segment beyond what is provided by the input, but it does have to add additional activity to reach the surface activity level of 1.0 . *The greater the activity that Gen must add to the underlying form, the greater the degree of violation of DEP.* For candidate b , *Gen* must add $1.0 - 0.8 = 0.2$; that is the degree of DEP violation indicated in the tableau.

Candidate c has a greater violation of DEP because *Gen* must add activity $1.0 - 0.5 = 0.5$ to the input $/t_1/$ to reach the output $[t_1]$'s activity level of 1.0 . Lacking coalescence, c avoids b 's UNIF violation, which reduces b 's Harmony by 0.7 . (This is a net loss for c , however, for while it avoids the 0.7 penalty from UNIF, DEP, with strength 10 , assesses a penalty of $10 \cdot 0.5 = 5$ to c but only $10 \cdot 0.2 = 2$ to b . So relative to c , b 's coalescence produces a net Harmony benefit of $5 - 2 - 0.7 = 2.3$, i.e., $h(b) - h(c) = [-2 - 0.7] - [5]$ with h denoting the Harmony contributions of only the constraints considered so far, ONSET, UNIF and DEP. Actually coalescence yields further Harmony rewards for b from the constraints MAX and ALIGN-L discussed next.)

The second-strongest constraint in the proposed analysis is **MAX**, which in the discrete Correspondence Theory is violated by an input segment with no output correspondent. The proposed gradient version of MAX rewards each input segment that has an output correspondent, rather than penalizing those that don't. MAX is thus a *positive constraint*, indicated by the positive weight 2 shown above it in (17), and the degree of satisfaction of this gradient constraint is defined to be the total activity of all segments that have output correspondents. Intuitively, MAX rewards underlying activity that makes it to the surface. For perspicacity, in (17) we do not record the MAX reward for all the underlying segments that surface in every candidate shown: these rewards all cancel when comparing the Harmony values of the candidates in order to determine the optimal output. Thus we only record MAX rewards accrued by the gradient input consonants at the end of W_1 and the beginning of W_2 : all other input segments surface in all of the candidates $a-c$. So the degree of satisfaction of MAX shown in tableau (17) is $\lambda + \tau = 0.8$ for b , in which both $/t_1/$ and $/t_2/$ surface (as $[t_{12}]$); $\lambda = 0.5$ for c , in which just $/t_1/$ surfaces (as $[t_1]$); and 0 for a , in which neither $/t_1/$ nor $/t_2/$ surface.

The next-strongest constraint is **ALIGN-L**(Morpheme, Syllable), abbreviated here **ALIGN-L** [46]. As indicated by the positive weight of 1 shown for this constraint in (17), this is a positive constraint: it rewards a morpheme if the output correspondent of a segment at its left edge falls at the left edge of a syllable.¹⁴ Again we don't record rewards shared by all candidates — in this

consonant that is the result of coalescence) are not gradiently active, but have activation 1 ; hence, each UNIF violation incurs a penalty of 1 .

A good reference for how the underlying neural-network level informs the Gradient Symbolic level — e.g., constraint satisfaction computation — is the neural-network implementation of a Language Acquisition Device for the Universal CV Syllable Theory of Optimality Theory in [61]. In this network, the connections implementing UNIF are entirely among output units (in the broad sense, which include correspondence units, since the correspondence relation is part of the output candidate) — hence only binary values appear in the computation of Harmony from these connections, i.e., UNIF violations are not scaled by input activity levels. In contrast, MAX and DEP (called PARSE and FILL in [61], following [46]) are implemented by connections impinging on input units, which have gradient activity levels, leading to gradient Harmony contributions: MAX and DEP violations are scaled by input activity levels.

¹⁴ This interprets ALIGN as binary: reward if boundaries coincide, no reward otherwise. We do *not* incorporate the notion of “gradience” according to which ALIGN is violated to varying degrees depending on the distance between the to-be-aligned edges. It would be less ambiguous therefore to name this constraint COINCIDE [65] [34], but with this categorical interpretation understood, we retain the name ALIGN.

case, the reward from the leftmost segment of $[{}^M\text{p}\text{ø}\text{ti}(\lambda\cdot\text{t}_1)]$ *petit* falling at the left edge of the first syllable of every output. In *b* and *c*, but not *a*, the second input morpheme $[{}^M(\tau\cdot\text{t}_2+\zeta\cdot\text{z}_3+\nu\cdot\text{n}_4)\text{ami}]$ *ami* has an initial segment, $/\text{t}_2/$, which has an output correspondent at the left edge of a syllable: in *b* this is the coalesced segment $[\text{t}_{12}]$, in *c* it is simply $[\text{t}_2]$.

The remaining two constraints in (17), ONSET and UNIF, have already been discussed: here, we take them to be identical to their classical discrete counterparts in McCarthy & Prince (1995) [37].

To determine the optimal output, we compute the Harmony (H) of each output by adding together the degree of violation/satisfaction of each constraint multiplied by its weight in the grammar. So, for example, $H(b) = -10\cdot 0.2 + 2\cdot 0.8 + 1\cdot 1 - 0.9\cdot 1 - 0.7\cdot 1 = -0.1$. This Harmony value is in fact the highest (least negative, in this case) so *b* is, as desired, the output of the grammar: the liaison consonant surfaces, through coalescence.

To show that candidate *b* is better than *any* competitor we can use the Gradient Harmonic Grammar counterpart to the Optimality-Theoretic Method of Mark Eliminability [46: Ch. 7]: we ask, for each constraint that *b* does not optimally satisfy, which among *all possible* candidates do better than *b* on that constraint, and then show that these candidates are worse than *b* overall because of their violations of other constraints.

(18) Demonstration of the optimality of candidate (17*b*)

- a. To do better on DEP, *b* would have to either (i) not parse (i.e., not have output correspondents for) the underlying liaison consonants at all, yielding *a*, which has lower total Harmony, or (ii) parse the underlying weak consonants faithfully, yielding gradient output segments: this is not possible in any candidate produced by our *Gen* which only generates fully discrete candidates.
- b. To do better on MAX, *b* would have to parse more of the underlying liaison consonants, requiring a consonant blend on the surface in the first position of W_2 , which is not possible with our discrete *Gen*. (Given that, e.g., $/\text{t}_2/$ and $/\text{z}_3/$ are both in the initial position of $/W_2/$, parsing them sequentially as in $[\text{t}_2\text{z}_3\text{ami}]$ would fatally violate a gradient version of LINEARITY: $[\text{t}_2]$ precedes $[\text{z}_3]$ in the output but $/\text{t}_2/$ does not precede $/\text{z}_3/$ in the input. This constraint will have a large negative weight as no metathesis is licensed by the French liaison grammar.)
- c. *b* satisfies ALIGN-L maximally, so no candidate can be preferred to *b* on this constraint.
- d. The same is true for ONSET.
- e. To do better on UNIF, *b* would have to have no coalescence, either by (i) parsing no liaison consonants, which yields the less-harmonic *a*, or by (ii) parsing only one liaison consonant, the best of these options being to parse the most active underlying liaison consonant, $/\text{t}_1/$, yielding candidate *c*, which has lower total Harmony than *b*.

4.3. A meta-analysis of the liaison analysis

Moving to the meta-analysis, we now ask, under what numerical values for the parameters (constraint weights, underlying activity levels) is *b* optimal? Optimality of *b* requires that $H(b) > H(a)$, i.e., that $H(b) - H(a) > 0$; and similarly, that $H(b) - H(c) > 0$. When will these relations hold?

Generally, the weight of constraint C in the grammar is denoted w_C , but for readability, we will abbreviate w_{DEP} by D , w_{MAX} by M , and the weights of ALIGN-L, ONSET, and UNIF respectively by A , O and U . Then we can write:

(19) Conditions on parameters under which $H(b) > H(a)$

$$\begin{aligned} H(b) - H(a) &= [(1 - \lambda - \tau)D + (\lambda + \tau)M + A^L + U] - [O] \\ &= (\lambda + \tau)[M - D] + D + A^L + U - O \\ &> 0 \end{aligned}$$

if and only if

$$\begin{aligned} (\lambda + \tau) &> -[D + A^L + U - O]/[M - D] \\ &\doteq -[-10 - 0.7 + 1 - (-0.9)]/[2 - (-10)] \\ &= 0.73 \\ &\equiv \theta(\underline{v}\underline{C}\underline{V}) \end{aligned}$$

In order for b to be optimal, we must also have

(20) Conditions on parameters under which $H(b) > H(c)$

$$\begin{aligned} H(b) - H(c) &= [(1 - \lambda - \tau)D + (\lambda + \tau)M + A^L + U] - [(1 - \lambda)D + \lambda M] \\ &= \tau[M - D] + A^L + U \\ &> 0 \end{aligned}$$

if and only if

$$\begin{aligned} \tau &> -[A^L + U]/[M - D] \\ &\doteq -[-0.7 + 1]/[2 - (-10)] \\ &= -0.025 \end{aligned}$$

This is satisfied in any case, as we assume that $\tau > 0$ (and the same for ζ and v).

This establishes the result in (21).

(21) $\underline{v}\underline{C}\underline{V}$: Condition for consonantal material C to surface intervocalically

The total underlying activity χ of C must exceed the threshold $\theta(\underline{v}\underline{C}\underline{V})$:

$$\chi > \theta(\underline{v}\underline{C}\underline{V}) \equiv -[D + U + A^L - O]/[M - D] \doteq 0.73$$

This result immediately entails (22). The first and last results (22a,d) are the two core mappings instantiating this configuration (9 ① ③), while the middle two results (22b,c) are sanity checks: the posited weak consonants at the beginning of normal V -initial words such as *ami* do not surface – (9 ②) – and fixed (non-liaison) W_1 -final consonants do surface.


(22) Intervocalic consonant behavior

- a. In $v\mathcal{L} + v \equiv v\mathcal{L} + \mathcal{L}v$, \mathcal{L} will surface: the output will be $v.\mathcal{L}v$;
mapping ① $v\mathcal{L} + v \rightarrow v.\mathcal{L}v$; *peti(t) + ami* \rightarrow .pø.ti.t̥a.mi.
 - i. if $\mathcal{L} = /t/$: total underlying activity of intervocalic $/t/ = \lambda + \tau \doteq 0.5 + 0.3 = 0.8 > 0.73$
 - ii. for $\mathcal{L} = /z/$ or $/n/$ replace τ by ζ or v : the same conclusion follows
- b. In $v\mathcal{F} + v \equiv v\mathcal{F} + \mathcal{L}v$, \mathcal{F} will surface: the output will be $v.\mathcal{F}v$ since the total underlying activity χ of intervocalic \mathcal{F} is:
 - i. if $\mathcal{F} = /t/$ or $/z/$ or $/n/$ then $\chi = 1 + \tau$ or $1 + \zeta$ or $1 + v$; all $\doteq 1 + 0.3 = 1.3 > 0.73$
 - ii. if \mathcal{F} is not in $\{t, z, n\}$ then $\chi = 1 > 0.73$
- c. In $v + v \equiv v + \mathcal{L}v$, no consonant C will surface: mapping ②; the output will be $v.v$ since
 - i. if $C = /t/$ or $/z/$ or $/n/$ the total underlying C activity is τ or ζ or $v \doteq 0.3 < 0.73$
 - ii. if C is not in $\{t, z, n\}$ then $\chi = 0 < 0.73$
- d. In $v\mathcal{L} + v \equiv v\mathcal{L} + v$, no consonant will surface; the output will be $v.v$:
mapping ③ $v\mathcal{L} + v \rightarrow v.v$; *peti(t) + h̃éro* \rightarrow .pø.ti.e.ɛ.ɔ.
 ♦ the total underlying activity of the consonant \mathcal{L} is $\lambda \doteq 0.5 < 0.73$

The meta-analysis has established a threshold of underlying activity needed for C to surface in the configuration $v\mathbf{Cv}$: $\theta(v\mathbf{Cv}) \doteq 0.73$; this has established the core mappings ① and ③ of (9). Turning now to the configuration $v\mathbf{Cc}$, we can establish the remaining core mappings ② and ④. A constraint that was inactive in the $v\mathbf{Cv}$ configuration becomes active in this configuration: the positive constraint $\text{ALIGN-R}(\text{Morph}, \text{Syll}) \equiv \text{ALIGN}([\mathbf{M}, [\sigma] \equiv \text{'ALIGN-R'}$, which is not satisfied in any of candidates of (17). Exactly the same type of meta-analysis derives a threshold for this configuration, $\theta(v\mathbf{Cc})$ (23b), from which ② and ④ follow immediately (23c).

(23) Analysis of configuration $v\mathbf{Cc}$

a. Harmonic Grammar tableau

| | <i>weight:</i> | -10 | 2 | -0.2 | 1 | 0.1 | |
|--|--|-----|-----|--------|---------|---------|----------|
| $[\mathbf{M}^{\text{poti}}(\lambda \cdot \mathbf{t}_1)]$ | $[\mathbf{M}^{\text{kop}}\tilde{\epsilon}]$ | DEP | MAX | NOCODA | ALIGN-L | ALIGN-R | <i>H</i> |
| <i>a</i> | <u>.pø.ti.ko.pẽ.</u>  | | | | 1 | | 1 |
| <i>b</i> | .pø.ti <u>ti</u> .ko.pẽ. | 1-λ | λ | 1 | 1 | 1 | -3.2 |
| | | 0.5 | 0.5 | | | | |

b. Computation of threshold $\theta(v\mathbf{Cc})$

$$H(b) - H(a) = [(1 - \lambda)D + \lambda M + N + A^L + A^R] - [A^L]$$

$$= \lambda[M - D] + D + N + A^R > 0$$

$$\text{iff } \lambda > -[D + N + A^R]/[M - D] \equiv \theta(v\mathbf{Cc})$$

$$\doteq -[-(10) - 0.2 + 0.1]/[2 - (-10)] = 0.84$$

c. Conclusion:

\mathcal{L} (activity $\doteq 0.5 + 0.3 = 0.8 < 0.84$) does not surface: mapping ②

\mathcal{F} (activity = 1 > 0.84) does surface: mapping ④

4.4. A visualization of the GSC Analysis

We now introduce a visualization tool (24) for the GSC Analysis. The black horizontal bar labeled θ is an activity scale, increasing to the right. On this black bar, the two activity thresholds derived so far, 0.73 (19) and 0.84 (23b), are marked by white vertical bars, each labeled by the configuration it applies to: $v\mathbf{Cv}$ ($\equiv v\mathbf{C}\mathcal{L}\mathbf{v}$) and $v\mathbf{Cc}$, respectively.

Just below the activity scale are the activity levels of different target elements that may appear in these configurations. E.g., the red $0.8 \doteq \lambda + \tau$ gives the underlying activity of $\underline{\mathbf{C}} = /t/$ in *petit ami*, presented in the red-filled box below '0.8'. In this box the input is written below the output. The pre- W_2 gradient-C-blend \mathcal{L} is marked explicitly in the input, which is otherwise written in French orthography: *petit \mathcal{L} ami*. The output is written in IPA above the input: .pø.tita.mi. Underlined in the input is the configuration in question, here $v\mathbf{Cv}$: *petit \mathcal{L} ami*; the target element is heavily underlined. The corresponding material in the output is also underlined correspondingly. The leftward-pointing arrow in the red box indicates that the target element's activity is to the right of — is greater than — the activity threshold for its configuration: hence the element surfaces. This is mapping ①.

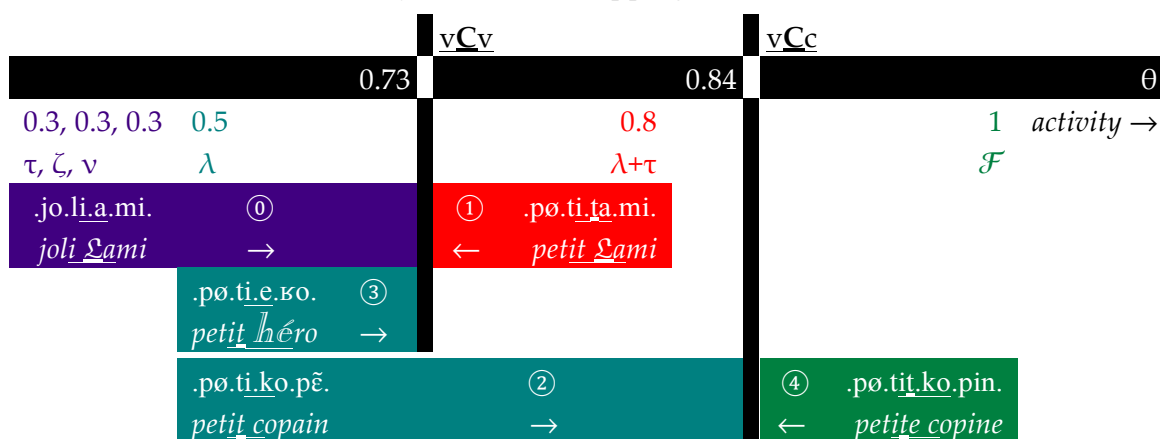
On the other side of this threshold is the case of *joli \mathcal{L} ami*, in purple. Here the target element is just $\underline{\mathbf{C}}$, in which each consonant has activity 0.3 ($\doteq \tau = \zeta = \nu$). As the right-pointing arrow in the purple box indicates, 0.3 is left of (less than) the relevant threshold 0.73 so the element does not surface; the output is .jo.lia.mi. This is mapping ②.

Also on the left side of the vertical black bar marking threshold 0.73 is the case of *petit héros*; here, the consonantal material in the configuration vCv is limited to the final [t] of *petit*, which has total activation $\lambda = 0.5$, the value at which the teal bar for *petit héros* begins. This is mapping ③.

The threshold for configuration vCc , 0.84, is marked by the second (right-most) vertical black bar. The liaison consonant for the case of *petit copain* has activity $\lambda \doteq 0.5$, which value is marked by the left edge of the lower teal rectangle; since this is less than the threshold 0.84, no liaison consonant surfaces. This is mapping ②.

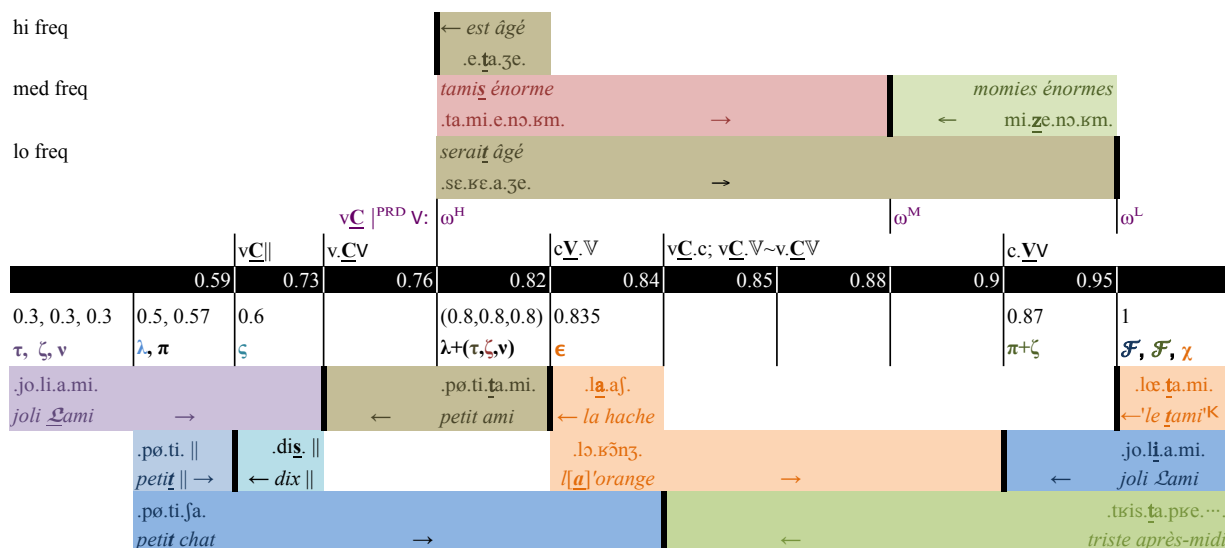
And in the same configuration, *petite copine* has a fixed consonant $\mathcal{F} = [t]$ at the end of W_1 , with activity 1. This is the level marked by the right edge of the green rectangle. Since $1 > 0.84$, \mathcal{F} surfaces. This is mapping ④.

(24) Visualization of the meta-analysis, for core mappings ① – ④



The next Section discusses the non-core liaison mappings. A visualization of the account of many of these mappings is given, without further elaboration, in (25). It shows a considerable range of (nine) distinct thresholds for a variety of configurations and collocation-frequency ranges.

(25) Picture of much of the analysis



4.5. Remaining non-core mappings

The proposed GSC liaison analysis also accounts for the non-core mappings ⑤ – ⑭. For space reasons, here we will simply state for each mapping the key claim in the proposed analysis’s account of that mapping.¹⁵

4.5.1. ⑤ Phrase-final \mathcal{L} .

In the phonological-phrase-final configuration $\underline{v}\underline{C}\|\|$, a constraint that has so far been irrelevant becomes active: $\text{ALIGN-R}(\text{PhonPhrase}, \text{Morph}) \equiv \text{ALIGN-R}\|\|$. We propose that this constraint has weight $w_{\text{ALIGN-R}\|\|} = 3$. This lowers the threshold for the target \underline{C} to surface to $\theta(\underline{v}\underline{C}\|\|) = 0.59$, because now parsing the final \underline{C} satisfies this positive constraint. We propose that the final \underline{t} of $/\text{qi}(\text{c} \cdot \text{t})/ \text{huit}$ has activity = 0.6; this exceeds the threshold, whereas the underlying activity of a standard liaison consonant like the \underline{t} of $\text{petit}\underline{t}$ does not. So $\text{huit}\underline{t}\|\| \rightarrow \cdot\text{qi}\underline{t}\|\|$ but $\text{petit}\underline{t}\|\| \rightarrow \cdot\text{pø.ti}\|\|$. Unlike previous accounts of liaison, the GSC account allows different (gradient) degrees of underlying consonant weakness, enabling an underlying distinction between the final consonants of $\text{petit}\underline{t}$ and $\text{huit}\underline{t}$ which leads to their different behaviors phrase-finally.

4.5.2. ⑥ Coda \mathcal{L} .

The different syllabification of this political speech register of French, $\underline{v}\underline{C}\underline{v} \rightarrow \underline{v}\underline{C}.\underline{v}$ instead of the standard $\underline{v}.\underline{C}\underline{v}$, implicates a (slightly) different grammar. While $w_{\text{ALIGN-R}} = 0.1$ in the standard register’s grammar, we propose that $w_{\text{ALIGN-R}} = 2.5$ in the political speech register; otherwise the two registers’ grammars are identical. Because the syllabification $\underline{v}\underline{C}.\underline{v}$ but not $\underline{v}.\underline{C}\underline{v}$ yields a reward from satisfying ALIGN-R, the increased weight of this constraint in the political register makes the Harmony of the anomalous syllabification greater than that of the standard one.

4.5.3. ⑦ *h*-aspiré onset \mathcal{F} (but not \mathcal{L}).

As we have seen, a liaison consonant \mathcal{L} does not surface before an *h*-aspiré word: $\text{petit } \underline{h}\acute{\text{e}}\text{ro} \rightarrow \cdot\text{pø.ti.e.}\underline{\text{ʁo}}$. But a fixed final consonant such as the $\mathcal{F} = r$ of $/\text{ʃe:}\underline{\text{ʁ}}/ \text{cher}$ ‘dear’ *can* surface; for the $\text{cher } \underline{h}\acute{\text{e}}\text{ro}$ the input $/\text{ʃe:}\underline{\text{ʁ}}_1 \text{e}\underline{\text{ʁo}}/ \rightarrow \cdot\text{ʃe:}\underline{\text{ʁ}}_1.\text{e.}\underline{\text{ʁo}} \sim \cdot\text{ʃe:}\underline{\text{ʁ}}_1.\text{e.}\underline{\text{ʁo}}$: these two syllabifications have the same Harmony and are both optimal.

4.5.4. ⑧ Pre/post-pausal \mathcal{L} .

Production around a prosodic pause is modeled with two optimizations. The first optimization applies before the pause, with the entire input, and produces a pair of outputs: the first is the pre-pausal production, while the second is the content of a buffer (temporary store). Each candidate splits the input string at some “pause point”, treating the portion of the input prior to that point as the input for the first optimization — which yields the pre-pausal production — while the portion after the pause point is stored in the buffer and then used as the input for the second optimization — which yields the post-pausal production. A constraint $*\mathcal{B}$ penalizes all material in the buffer, discouraging the procrastination option of producing material after the pause. When the weak final liaison consonant of W_1 is parsed into the buffer, it serves as the ‘memory’ that carries through the pause W_1 ’s selection of the appropriate ‘allomorph’ of W_2 (26a.ii).

The optimal outcome of the first optimization is a complete production of the entire input: no internal pause. But the highest-Harmony *sub*-optimal outputs of the first optimization predict the most probable pause-containing errors. (Recall from \mathfrak{B}_1 (7) that higher Harmony \Rightarrow higher probability.) Examples are given in (26).

¹⁵ Fuller derivation of the accounts of ⑤ – ⑭ will be provided in a future version of this paper.

(26) Examples of highest-probability errorful productions, in decreasing order of probability

- a. *petit ami* →
- i. .pø.ti. || .a.mi.
 - ii. .pø.ti. || .ta.mi.
 - iii. .pø.tit. || .a.mi.
- b. *cher ami* →
- i. .ʃɛ.ʁ. || .a.mi.
 - ii. .ʃɛ. || .a.mi.
 - iii. .ʃɛ. || .ʁa.mi.

These predictions are the desired ones, except for the problematic case of (26b.ii).

4.5.5. ⑨ *Frequency effect.*

See Sec. 6.1.

4.5.6. ⑩ *Incorrect \mathcal{L} insertion.*

For the input *petit ami*, the optimal output is .pø.ti.ta.mi., so *.pø.ti.za.mi. is an error; such an error is what is referred to as “incorrect \mathcal{L} selection”. To account for such data, we would like the proposed performance theory to assign highest probability to the correct output but second-highest probability to this type of error (substitution of an incorrect liaison consonant). Other types of error, such as substituting a non-liaison consonant for the liaison consonant, should have lower probability. And this will follow from \mathfrak{P}_1 , $p(r) \propto e^{H(r)}$, is the Harmony of substituting an incorrect liaison consonant is higher than that of substituting a non-liaison consonant. The proposed account makes just this prediction, because among the sub-optimal outputs, those with highest Harmony are those in which the erroneously substituted consonant is a liaison consonant. This is because the liaison consonants initial in W_2 that are not selected for by W_1 — those that don’t match W_1 ’s final weak consonant — are nonetheless present in the input and so, unlike non-liaison consonants, their appearance in the output yields some non-zero reward from MAX. (Note that this general account allows our framework to account for a wide variety of error data from language processing across other tasks and populations; see [17], [19] for review and discussion.)

4.5.7. ⑪ *Exceptional \mathcal{L} epenthesis.*

The account of the error $v + v \rightarrow *v.\mathcal{L}'v$ is the same as that given in 4.5.7. The highest-Harmony suboptimal candidates are of this form: e.g., those inserting a non-liaison consonant have lower Harmony.

4.5.8. ⑫ *Child \mathcal{L} -as- $\mathcal{F}\hat{W}_2$.*

See the discussion of acquisition in Sec. 5.

4.5.9. ⑬ *W_2 allomorph selection for $\mathcal{L}\hat{W}_2$ Analysis.*

Literally speaking, there is no “allomorph selection” in the proposed GSC account. There is a unique lexical entry for $[^M W_2]$ as an independent word; it just happens to have a blend of gradient consonants if it is a normal (non-*h*-*aspiré*) “*v*-initial” word. (Furthermore, anticipating the discussion of Sec. 6.1: There is a unique lexical entry for the collocation $[^M W_1 W_2]$, and — given its activity level in the lexicon, a function of its usage frequency — this collocation entry is averaged in a deterministic fashion with the concatenation $/[^M W_1] [^M W_2]/$ to form the input to the grammar for generating $W_1 W_2$.) Functionally speaking, the job performed by allomorph selection is automatically accomplished in the GSC account: of the multiple weak liaison

consonants present initially in $/W_2/$, only the one matching the weak consonant present finally in $/W_1/$ can coalesce with it; the other, non-matching (“non-selected”) consonants contribute to the performance-theoretic explanation of errors (Sec. 4.5.6) but not to the competence-theoretic explanation of nominally correct outputs.

4.5.10. ⑭ Gender-bending \mathcal{L} .

The alternation $o \sim \epsilon l$ is accounted for by borrowing a technique developed in [16]: activation sharing between alternating segments. The underlying form $/b_0(\beta \cdot \epsilon_1)(\omega \cdot \{o_2 l_3\})/$ generates all forms *belle*, *beau*, *bel*: activation ω is shared between $/o_2/$ and $/l_3/$ meaning that in any (input, output) candidate, ω can be split into $\omega = \delta + \psi$, for any positive choice of δ and ψ , and then $\omega \cdot \{o_2 l_3\} = (\delta \cdot o_2)(\psi \cdot l_3)$. Parsing both underlying vowels is suboptimal because the reward is only βw_{MAX} while the penalty is either w_{COMPLEX} — if both vowels are parsed into a single syllable nucleus — $[.b\epsilon_1 o_2 l.]$ or w_{ONSET} — if they are parsed into separate nuclei $[.b\epsilon_1 .o_2 l.]$. Before a V-initial masculine W_2 , or the morpheme [FEMININE] $/\phi \cdot \emptyset/$ (pure activity with no phonological content), it is optimal to parse the $/l_3/$, assigning it all ω units of activity from $/\omega \cdot \{o_2 l_3\}/$, leaving the other vowel $/(\beta \cdot \epsilon_1)/$ to be parsed as the nucleus for onset $/b_0/$: the optimal output is $[.b_0 \epsilon_1 l_3.]$ (orthographically *bel* in the former case, *belle* in the latter). For masculine consonant-initial or *h*-*aspiré* W_2 , it is optimal not to parse $/l_3/$, leaving all ω units of activity for $/o_2/$, so — it being better to parse $/(\omega \cdot o_2)/$ than $/(\beta \cdot \epsilon_1)/$ because the former has greater underlying activation (ω) than the latter (β) — the optimal output is $[.b_0 o_2.]$, orthographic *beau*.

4.6. Restrictiveness of the account

It might seem that — with the array of numerical parameters present in the proposed analysis — just about any pattern of results could be accounted for. While the space of possible predicted behaviors is difficult to determine, the scalar implications in (27) can be identified.

- (27) For a given set of constraint weights, the analysis is restricted at least in the following ways:
- a. No matter the underlying activity of a segment x , if x surfaces in a configuration with a threshold θ , then x must surface in any configuration with a threshold less than θ .
 - b. No matter the threshold of a configuration E , if a segment x with activation a surfaces in E , then a segment x with any activation greater than a must also surface in E .

5. ACQUISITION: SPECULATIONS ON FORMALIZING DOWTY’S SKETCH WITH GSC

The two principal component accounts in the proposed adult blend can be related to Dowty (2003)’s proposal that adult grammars blend the residue of an early child grammar with a more sophisticated, later-acquired grammar. Here the blend is actually of two *lexicons*. The later-acquired component of the adult lexical blend is the standard Final- \mathcal{L} Analysis taught in schools. The early lexicon instantiates the \mathcal{L} -initial Analysis, which as already observed is the lexicon predicted by comprehension-directed optimization [55] (or ‘Robust Interpretive Parsing’ [62]) given the constraint ALIGN-L(Morpheme, Syllable) proposed in Sec. 4.2.1: analyzing $.p\theta.ti.ta.mi.$ as $[^M.p\theta.ti][^M.ta.mi.]$ satisfies this constraint [2: 34], and there is no conflicting constraint plausibly active in early word segmentation. (Later, when paradigmatic learning is operative, there is a potential conflict with UNIQUEEXPONENT, which penalizes the multiple allomorphs that *ami* receives on the \mathcal{L} -initial Analysis.) In addition to ample anecdotal evidence, laboratory studies (e.g., [39]) have documented this prediction that in early Child French, $/ami/$, $/tami/$, $/zami/$, $/nami/$ are treated as allomorphs of *ami* in rather free variation (e.g., at around 20 months [2]).

From this initial state, which realizes the \mathcal{L} -initial Analysis, a possible path to the adult blend state is sketched in (28). The learning procedure is assumed to be error-driven. For concreteness we take the shortcut of assuming that the grammar proposed above is in place; in reality, of

course, learning the grammar and learning the underlying forms must proceed simultaneously – a highly demanding problem computationally (e.g. [27]).

(28) Possible path of acquisition

- a. When the child chooses the free variant /tami/ with *joli*, the resulting output is *[3olitami].
 - i. Comparison with adult productions [3oliami] yields an error signal: the *[t] should not be present.
 - ii. The child's learning procedure weakens the source of *[t]: the initial /t/ of /tami/, say by 0.1; the allomorph /tami/ becomes $/(0.9 \cdot t)ami/$
 - iii. Each time this occurs, the same weakening results, until the allomorph becomes $/(0.7 \cdot t)ami/$, at which point the error signal vanishes because the activity of /t/ has fallen below the grammar's realization threshold (for this configuration) of 0.73: $/3oli(0.7 \cdot t)ami/ \rightarrow [3oliami]$.
- b. When the child happens to choose the correct, /t/-initial allomorph $/(0.7 \cdot t)ami/$ with *petit* /pøti/, the child's output is *[pøtiami] because now the total activity of [t] at the word juncture falls below the grammar's realization threshold of 0.73.
 - i. Comparison with adult productions [pøtiami] yields an error signal: a [t] should be present between [pøti] and [ami]
 - ii. The child's learning procedure adds additional /t/ activation; say, 0.1.
 - iii. This /t/ activation could be added to either the end of *petit* or the beginning of *ami*; no information in the error signal favors one site over the other.
 - iv. Suppose that, as in L^2 -regularized Maxent learning [6] [9], the procedure does the maximum-entropy division of change, splitting the extra 0.1 activation into 0.05 at the end of *petit*, yielding $/pøti(0.05 \cdot t)/$, and 0.05 at the beginning of the allomorph of *ami* in the current input, yielding $/(0.75 \cdot t)ami/$.
- c. Now the situation in (28a) repeats, because $/3oli(0.75 \cdot t)ami/$ again yields output *[3olitami].
 - i. The activation of /t/ drops again by 0.1, yielding $/(0.65 \cdot t)ami/$
 - ii. This correctly produces [3olitami] for *joli ami*, ...
 - iii. ... but now the errorful output *[pøtiami] for *petit ami* returns (total /t/ activation is $0.05 + 0.65 = 0.7 < 0.73$, the relevant realization threshold).
 - iv. Hence (28b) repeats, resulting in lexical entries $/pøti(0.1t)/$ and $/(0.7 \cdot t)ami/$
- d. The result is a gradual shift of /t/ activity from the beginning of what starts out as /tami/ to the end of what starts out as /pøti/
- e. The shift does not go all the way: the final grammar is a blend.
 - i. The final lexicon of $/pøti(0.5 \cdot t)/$, $/(0.3 \cdot t)ami/$, as shown in Sec. 4, produces output forms that match the target adult forms, so the error signal driving learning is zero.
 - ii. To satisfy UNIQUEEXPONENT, some amalgamation process merges the allomorphs to form $/(0.3 \cdot t + 0.3 \cdot z + 0.3 \cdot n)ami/$.
 - iii. The process might instead produce a result $/(τ \cdot t + ζ \cdot z + ν \cdot n)ami/$ in which the activity levels $τ$, $ζ$, and $ν$ are not all identical, but as already stated, the analysis in Sec. 4 takes $τ = ζ = ν$ to be merely a simplifying provisional assumption.

Clearly considerable further development is required before this sketch can become a proper explanation.

6. CONTEXTUAL FACTORS IN LIAISON — THE ROLE OF PROSODY: TENTATIVE SUGGESTIONS

So far we have considered the conditions on W_2 under which a liaison consonant associated with a morpheme W_1 surfaces in the sequence $W_1 W_2$: let us call these conditions **the phonological conditions for liaison**. However these conditions are in fact only *necessary*; they are not in general sufficient. Regardless of the underlying phonological forms of W_1 and W_2 , there are certain sentential environments under which no liaison consonant may appear, the so-called *forbidden liaison contexts*. In the *optional liaison contexts*, when the phonological conditions for liaison are satisfied, there is variation: the liaison consonant will appear in some instances but not others; or both the form with a pronounced liaison consonant, and the form without, are both considered acceptable. The third possibility also exists: *obligatory liaison contexts*, in which a liaison consonant always surfaces whenever the phonological conditions are met.

How to characterize these three types of liaison context has long been controversial. The least but still controversial, virtually theory-neutral, characterization is simply in terms of a list of the *morphosyntactic category* of W_1 : see (29).

(29) Morphosyntactic characterization of liaison contexts (from de Jong [12])

- a. Contexts of obligatory liaison
 - i. articles: *un/les/des/aux*
 - ii. adjectival possessives: *mon/ton/son/mes/tes/ses/nos/vos/leurs*
 - iii. demonstrative adjectives: *ces/cet*
 - iv. indefinite adjectives: *plusieurs/tels/tout/autres/certain*
 - v. interrogative adjectives: *quels/quelles*
 - vi. numerals: *un/deux/trois/vingt/cent*
 - vii. quantifiers: *plusieurs/aucun/tout/quelques/rien*
 - viii. pronominal clitics: *nous/vous/ils/elles/on/les/en*
 - ix. complementizers: *quand/dont*
 - x. *en* introducing a gerund
 - xi. prenominal adjectives
 - xii. modifying adverbs
- b. Contexts of optional liaison
 - i. prepositions
 - ii. the forms of *être* (as a passive/perfective auxiliary or as a copula)
 - iii. the forms of *avoir* (as a perfective auxiliary)
 - iv. modals

Controversy arises particularly in the attempt to provide general characterizations of the three types of liaison context, from which the specific morphosyntactic cases in (29) can be derived. General approaches that have been proposed include syntactic, prosodic, and frequency-based. Our observation — following the general structure of argument — is that none of these approaches alone can account for this phenomenon. Accordingly, we adopt a blend of the prosodically-mediated, syntactically-constrained approach of de Jong (1990) [12] and a frequency-based approach (based on the work of Ågren [1] and Bybee [5]).

All approaches attempt to capture the intuition that liaison is linked to “cohesion”: liaison is more likely the more W_1 and W_2 are “tightly bound”. In the frequency-based approach we partly adopt, “more tightly bound” is defined as ‘a more frequent sequence’. In the prosody-based approach that provides the majority of our account, “more tightly bound” means ‘co-

located in a prosodic constituent lower in the prosodic hierarchy'; or, as we will explicitly adopt, *the lower in the prosodic hierarchy is the highest prosodic boundary separating W_1 and W_2 , the more likely is liaison*.

In de Jong's analysis, the prosodic hierarchy adopted is

(30) Prosodic hierarchy assumed: MPP > SPP > PWd

maximal phonological phrase > small phonological phrase > prosodic word

Since in the GSC analysis, liaison consonants surface only through coalescence, we can formalize the "cohesion" notion through a hierarchy of constraints mirroring the prosodic hierarchy: if \mathfrak{B} is a boundary between constituents higher in the prosodic hierarchy than boundary b , then the constraint 'no coalescence across prosodic boundary \mathfrak{B} ' is stronger than the corresponding constraint for b . Coalescence ensures there is no point between W_1 and W_2 at which there is simultaneously a separation of both the morphemic and the prosodic constituents associated with these words, as shown in (31).

(31) Liaison configuration in *petit ami*;

a. underlying form: $/[m_1 p\theta ti(\lambda \cdot t_1)] [m_2 (\tau \cdot t_2 + \zeta \cdot z_3 + v \cdot n_4) ami]/;$

b. output: (PCat1 $[m_1 \cdot p\theta \cdot ti \cdot PCat1]$) (PCat2 $[m_2 \cdot t_{12} \cdot m_1]$) a.mi. $m_2]$ PCat2) peti.tami

This configuration violates the constraint in (32).

(32) Boundary constraint violated by coalescence

*CROSS(Morph, PCat): $[Morph]$ and $(PCat)$ constituents cannot cross

That is, the following two configurations are banned:

$[Morph (PCat \mu \cdot Morph) PCat]$

$(PCat \mu \cdot [Morph PCat] Morph)$

The Harmony penalty assessed for each violation is $h_{\text{CROSS}} = \mu \cdot w_{*CROSS(Morph, PCat)}$

Here we have anticipated a notable property of the proposed GSC analysis, the gradient strength μ of boundaries — which we can take to be the activity in the Gradient Symbolic Representation of the constituent category symbol (here, 'Morph').

The *CROSS constraint hierarchy supplements UNIF in the grammar; UNIF (equally) penalizes all cases of many-to-one input-output correspondence — coalescence — while the *CROSS family of constraints penalize coalescence differently depending on the level in the prosodic hierarchy of the prosodic boundary involved in the crossed configuration.

In the paper so far, we have implicitly been assuming an obligatory liaison context for which we have implicitly been assuming a w_{*CROSS} equal to 0: meeting the phonological conditions for liaison suffices to entail that the liaison consonant will surface. For an optional liaison context, we posit $w_{*CROSS} = -2.55$.

The generalization about liaison being favored by greater 'cohesion' takes the form (33).

(33) Generalization

- | | | |
|------------------------|-----------------------------------|-------------------------------------|
| a. $(PWd W_1 W_2 PWd)$ | \mathcal{L} always surfaces | within a Prosodic Word |
| b. $W_1 PWd) (PWd W_2$ | \mathcal{L} frequently surfaces | across Prosodic Words |
| c. $W_1 SPP) (SPP W_2$ | \mathcal{L} rarely surfaces | across Small Phonological Phrases |
| d. $W_1 MPP) (MPP W_2$ | \mathcal{L} never surfaces | across Maximal Phonological Phrases |

This generalization is captured by a constraint-strength principle that we posit: (34).

- (34) \mathfrak{F}_4 principle: *CROSS(Morph, PCat) markedness hierarchy parallels the prosodic hierarchy
if PCat' is higher in the prosodic hierarchy than PCat, then

$$w^{*\text{CROSS}(\text{Morph}, \text{PCat}')} > w^{*\text{CROSS}(\text{Morph}, \text{PCat})}$$

The connection to syntactic constituency assumed by de Jong is

- (35) Syntax \leftrightarrow prosody interaction (de Jong 1990 [12])

| | |
|-----------------------------------|--|
| ALIGN-R(X^0 , PWD) | X^0 any head |
| ALIGN-R(\underline{X}^0 , SPP) | \underline{X}^0 a lexical head |
| ALIGN-R(\underline{XP} , MPP) | \underline{XP} the maximal projection of a \underline{X}^0 |

We do not pursue this further here, leaving it to future work to use some account like (35) to derive the morphosyntactic generalizations in (29).

6.1. A frequency effect: (11) ⑨

According to the generalization stated in (11) ⑨, in optional liaison contexts the probability of liaison depends not only on the level in the prosodic hierarchy of the boundary separating W_1 from W_2 : it also depends on the frequency of the collocation W_1W_2 . Building on [1] and [5], GSC provides a natural way to formalize a usage-based approach to explaining this frequency effect, integrating it with the syntactic and prosodic structure effects outlined above.

Suppose that each time the sequence $W_1 W_2$ is processed by a speaker-hearer (say, in comprehension or production), the activity of the *collocation entry* in the speaker-hearer's mental lexicon $[^M W_1 W_2] - a_{[^M W_1 W_2]}$ – increases to a slight degree.¹⁶ (As above, $[^M$ abbreviates $[_{\text{Morph}} \cdot]$.) Following the principle \mathfrak{F}_3 (36), we assume that to produce this sequence, the input to the grammar is a weighted average of $[^M W_1][^M W_2]$, the concatenation of the lexical entries for the two morphemes W_1 and W_2 , and the collocation entry $[^M W_1 W_2]$. The relative weighting of the two components of the input is determined by their activity in the lexicon; the weights are such that the proportion contributed by the collocation entry increases as its activation increases, i.e., as its frequency of usage increases. We can assume for concreteness the weighted average given in the fourth principle hypothesized by our GSC Analysis (36).

- (36) \mathfrak{F}_4 : Input to grammar when multiple lexical entries apply

The input to the grammar is an *activity-weighted average* of the relevant lexical entries.
For the sequence $W_1 W_2$, the input is

$$\mu \cdot [^M W_1][^M W_2] + (1 - \mu) \cdot [^M W_1 W_2] = [^M W_1 \mu \cdot (]^M W_2],$$

where

$$\mu \equiv 1/(1+a) \text{ hence } 1 - \mu = a/(1+a)$$

and $a \equiv a_{[^M W_1 W_2]}$ is the activity in the lexicon of the collocation entry $[^M W_1 W_2]$.

As the number of occurrences of W_1W_2 mounts, so does a , which entails that the activity $\mu \equiv 1/(1+a)$ of the morpheme boundary separating them decreases.

This has two consequences which together derive the desired frequency effect (11) ⑨. First, the optimal prosodic parsing changes as a function of μ because the rewards assessed by the syntax

¹⁶ As in all frequency generalizations, there is the vexing question of “what to count”. Frequency of a ‘lemma’, adding together counts of all morphological variants of a form? Morphosyntactic category, adding together counts of all forms of a particular category, however defined (“pre-nominal adjective”)? For our GSC approach, we must leave such theoretically-vexed questions for latter research. To mention one example, Côté (2013) [8: 162] deploys a combination of these two: the collocation “schema” [*quand* [subject NP]] ‘when NP’.

↔ prosody ALIGNMENT constraints in (35) are proportional to the activity of the heads i.e. of the corresponding morphemes: μ and $(1-\mu)$. When the collocation frequency is low, μ is essentially 1 and the optimal parse assigns prosodic categories in accordance with (35), with W_1 and W_2 each functioning separately as a syntactic head. When the collocation frequency is high, μ is essentially 0, and $W_1 W_2$ functions as a single syntactic head [^M $W_1 W_2$], determining the level of prosodic boundary aligned to its own right edge (the right edge of W_2); there is no prosodic boundary separating W_1 from W_2 – case (33a).

The second consequence of the variation of boundary activity μ with usage frequency concerns the strength with which a boundary between W_1 and W_2 – when such a boundary is optimal – inhibits liaison. With w_{*CROSS} negative, each *CROSS constraint assesses a Harmony penalty to coalescence; the magnitude of this penalty – for a given boundary type – is proportional to the strength of the morpheme boundary crossed by that prosodic category: μ in (36). This means that the penalty shrinks as the usage frequency of the sequence $W_1 W_2$ mounts and μ drops. The effect of the shrinking penalty is an increase in the probability of liaison.

The quantitative relation between the decreasing penalty for coalescence and the increasing probability of liaison can be computed as follows. Let the Harmony associated with the liaison candidate $c_{\mathcal{L}}$ in some given context, omitting the contribution of *CROSS, be $H_{\mathcal{L}}$, and the Harmony of the candidate c_{\emptyset} with no surface liaison consonant (hence no relevant violation of *CROSS) be H_{\emptyset} . According to the principle \mathfrak{P}_1 , under the Probabilistic Harmonic Grammar performance theory, this means that the ratio of probabilities of these two candidates is

$$\begin{aligned} p(c_{\mathcal{L}})/p(c_{\emptyset}) &= e^{H(c_{\mathcal{L}})/T} / e^{H(c_{\emptyset})/T} = e^{[H(c_{\mathcal{L}}) - H(c_{\emptyset})]/T} = e^{(H_{\mathcal{L}} + \mu \cdot w_{*CROSS}) - H_{\emptyset}/T} = e^{[H_{\mathcal{L}} - H_{\emptyset}]/T} e^{\mu \cdot w_{*CROSS}} \\ &= p_0 e^{\mu \cdot w_{*CROSS}} \end{aligned}$$

where $p_0 \equiv e^{[H_{\mathcal{L}} - H_{\emptyset}]/T}$ is the probability ratio of liaison to no-liaison corresponding to no prosodic boundary between W_1 and W_2 ; this corresponds to $\mu = 0$ i.e., $a_{[M W_1 W_2]} = \infty$, the situation after infinitely many occurrences of $W_1 W_2$. Thus, as the number of occurrences of $W_1 W_2$ mounts and μ decreases, this probability ratio increases to p_0 (since $w_{*CROSS} < 0$).

For the prosodic boundary relevant for an optional liaison context – the conflation of contexts (33b–c) – we set $w_{*CROSS} = -2.55$. For low-, medium- and high frequency $W_1 W_2$ collocations we assume morpheme boundary activities of: $\mu^L = 1.0$, $\mu^H = -0.133$, $\mu^M = 0.667$, respectively. Then it turns out to follow that we get the desired predictions in (37).

(37) Frequency-sensitive predictions (see note 16)

- a. *est âgé*, lit. ‘is aged’ (high frequency) → .e.ta.ʒe.
- b. *serait âgé* lit. ‘would be aged’ (low frequency) → .sœ.ʁε.a.ʒe.;
- c. *tamis énorme* ‘enormous sieve’ (medium frequency, SINGULAR) → .ta.mi.e.nœʁm. ‘
- d. *momies énormes* ‘enormous mummies’ (medium frequency, PLURAL) → .mo.mi.ze.nœʁm.

The contrast between (37c) and (37d), driven by the number difference [SINGULAR] vs. [PLURAL], results from two further aspects of the GSC analysis that accounts for the relatively higher rate of liaison associated with the [PLURAL] morpheme /-($\pi \cdot z$)/. First, the underlying activity of the liaison consonant $\mathcal{L} = z$, $\pi \doteq 0.57$, is greater than that of standard W_1 -final consonants, $\lambda \doteq 0.5$. And secondly, when the liaison consonant in question is the $\mathcal{L} = z$ of [PLURAL], failure to parse \mathcal{L} violates the standard OT constraint MAXMORPH, which we assign weight $w_{MAXMORPH} = 0.25$.

7. EXTENSIONS

Before closing we note a few important extensions of the analysis presented above; some have already been carried out (38a) while the others are planned for the near future (38b).

(38) Extensions

- a. Existing extensions omitted here
 - i. [PLURAL] morpheme: $/\pi \cdot z/$
 - ii. eliding vowels: $/\epsilon \cdot V/$
 - iii. [FEMININE] morpheme: $/\phi \cdot \emptyset/$ — pure floating activation (no melodic content)
- b. Planned extensions
 - i. Non-productive liaison consonants: p, g/k, ʁ, l
 - ii. W_1 -final nasal vowels; W_2 -initial glides
 - iii. Variation of gradient activity levels across lexical items W : $\lambda_W, \tau_W, \zeta_W, \nu_W$
 - iv. Variation of mean gradient activity in lexicon across liaison consonants \mathcal{L}
 - v. Theory of the lexicon (new V-initial W_2 s)
 - vi. Derive parameter values through a learning algorithm

While most of these are obvious extensions and self-explanatory, the last two deserve comment.

A theory of the lexicon (38b.v) is needed to account for why all non-*h-aspiré* nominally V-initial words (like *ami*) begin with the same blend of weak consonants $\mathcal{L} \equiv \tau \cdot t + \zeta \cdot z + \nu \cdot n$. Despite the Dowty 2003-esque acquisition sketch of Sec. 5, the claim is not, for example, that a new nominally-V-initial loanword such as *iPhone*, for a given speaker at a given time, has a lexical entry beginning with weak initial consonants \mathcal{C} including only those that have actually been heard preceding *iPhone* by that speaker by that time. The idea is rather that the lexicon is governed by some sort of information-theoretic principle such as Minimal Description Length [48] which entails that the most parsimonious lexicon will posit (i) an actual symbol \mathcal{L} , which refers to $\tau \cdot t + \zeta \cdot z + \nu \cdot n$, and (ii) a lexical constraint that words with a V-initial citation form $[W]$ have an underlying form given by either $/\hat{W}/$ (*h-aspiré*) or $/\mathcal{L}\hat{W}/$.

Finally, we ultimately need a learning procedure implementing principles by which a learner can induce from available data a Harmonic Grammar and lexicon such as that proposed here (38b.vi). Such a learning procedure would ideally be storable at the level of Gradient Symbolic Representations, the level adopted in this paper, and at the underlying neural network level. Candidate algorithms exist but the work reported here has relied instead on lexical and grammatical parameter values determined by hand. The primary reason for this choice is this: automatic parameter-estimation procedures, for models of the complexity of the one proposed here, have a strong tendency to produce results that are extremely difficult to understand. In contrast, the fact that the parameter values posited above were successfully derived manually itself attests to the understandability of the account proposed here.

Eric Rosen (in press) [49] takes some steps in this direction. Building on the analysis reported here, he analyzes the semi-regular application of a process in Japanese compounds (*rendaku* voicing) as reflecting the coalescence of two gradiently activated voicing features from each member of the compound. He presents a gradient descent algorithm for acquiring these activation values. This algorithm accounts for variation across lexical items while restricting gradient activation to a small set of values — minimizing the overall complexity of the lexicon.

8. SUMMARY

The account presented above crucially depends on the use of Gradient Symbolic Representation in a number of respects.

- (39) Crucial use of Gradient Symbolic Representations in the proposed liaison analysis
- a. The adult lexicon is a blend: $0.5 \cdot [\text{Final-}\mathcal{L} \text{ Analysis}] + 0.3 \cdot [\mathcal{L}\text{-initial Analysis}]$
 - b. There are many crucially distinct gradient activity levels for different \mathcal{L} s
 - i. Discussed above:
 - ◆ \mathcal{L} of W_1 (0.5)
 - ◆ \mathcal{L} of W_2 (0.3)
 - ◆ /t/ of *huit* (0.6)
 - ii. In the full analysis but not discussed above:
 - ◆ /z/ of the [PLURAL] morpheme (0.57)
 - ◆ pure floating activity of the [FEMININE] morpheme (0.8)
 - ◆ Vs that elide (0.835)
 - c. The acquisition process gradually shifts activity of \mathcal{L} from $W_2 = \mathcal{L}\hat{W}_2$ to $W_1 = \hat{W}_1\mathcal{L}$
 - d. Lexical entries can have distinct overall activity levels
 - i. The input to the grammar is a weighted average of relevant lexical entries
 - ii. There is a gradual usage-based increase of activity in the lexicon of the collocation $[\text{Morph}W_1W_2]$
 - iii. Therefore there is a gradient boundary separating W_1 and W_2 (which gradually weakens with increasing usage of the collocation W_1W_2)

The GSC analysis crucially depends on Harmonic Grammar's capacity for grammatical computation over Gradient Symbolic Representations. Many optimizations in this account depend crucially on the numerical interaction of constraint violations, as opposed to the strict-domination interaction of Optimality Theory. In our first tableau (17), for example, the candidate that best-satisfies the strongest constraint DEP, candidate *a* with no liaison consonant, is not the optimal candidate.

The proposed analysis can be viewed as a formalization of Dowty 2003's proposal that adult grammars output blends of discrete structures. To the extent that the GSC proposal provides an ultimately empirically adequate analysis of liaison, this can be taken as the first bit of evidence supporting a general hypothesis underlying GSC linguistics research: long-standing disagreements over the (assumed to be unique) correct (discrete) structure arise because the actual structure is a gradient blend of discrete structures.

ACKNOWLEDGEMENTS

For helpful discussion, comments and suggestions we thank the audiences of talks on this material presented at: (2016) Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP) Language Group Seminar; Paris Atelier de Phonologie, University of Edinburgh Linguistics and English Language Department; University of Strasbourg Linguistique, Langues & Parole Seminar; Hungarian Academy of Sciences Research Institute for Linguistics; Aix-Marseille University Brain-Language Research Institute Workshop on Liaison; (2015) Stanford University Linguistics Department; Microsoft Research Redmond; University of Chicago LSA Summer Linguistic Institute Sapir Lecture; Rutgers University RULing-X Linguistics Conference; (2014) Newcastle meeting of the Australian Linguistic Society; Macquarie University Sydney Workshop on the Role of Prosody in Language Learning: Stress, Tone and Intonation; University of California Berkeley Institute of Cognitive and Brain Sciences.

Sincere thanks to the LSCP for hosting a sabbatical visit for PS during which this paper was written, and the Krieger School of Arts and Sciences at Johns Hopkins for support of that sabbatical. We gratefully acknowledge the NSF for INSPIRE grant BCS-1344269 which has partially supported this research. Special appreciation for their intellectual contributions to this work goes to Jennifer Culbertson — whose idea lies at the foundation of the entire analysis — as well as Adam Albright, Alan Prince, John McCarthy, Eric Rosen, Sharon Peperkamp, Emmanuel Dupoux, Sophie Wauquier and Noam Faust and members of the Gradient Symbolic Computation group, Akira Omaki, Kyle Rawlins, Ben Van Durme, Pyeong Whan Cho, Laurel Brehm, Nick Becker, and especially Colin Wilson and Géraldine Legendre.

REFERENCES [ROA = Rutgers Optimality Archive, roa.rutgers.edu]

- [1] Ågren, John. 1973. Etude sur quelques liaisons facultatives dans le français de conversation radiophonique: fréquence et facteurs. Uppsala: Acta Universitatis Upsaliensis.
- [2] Babineau, Mireille, & Shi, Rushen. (2011). Processing of French liaisons in toddlers. *BUCLD 35 Proceedings*. Cascadilla Press, Somerville, MA, 25–37.
- [3] Boersma, P. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Holland Academic Graphics.
- [4] Buerkin-Salgado, Angelica, Culbertson, Jennifer, Legendre, Géraldine & Nazzi, Thierry. 2016. Competing models of liaison acquisition: Evidence from corpus and experimental data. *Language*. In press.
- [5] Bybee, J. (2001). Frequency effects on French liaison. *Typological studies in language*, 45, 337-360.
- [6] Chen, Stanley & Rosenfeld, Ronald. 2000. A survey of smoothing techniques for ME models. *IEEE Transactions on Speech and Audio Processing* 8, 37–50.
- [7] Chevrot, Jean-Pierre & Fayol, Michel. 2001. Acquisition of French liaison and related child errors, in Almgren, M. Barreña, A., Ezeizabarrena, M.J., Idiazabal, I. & MacWhinney, B. (eds), *Research on Child Language Acquisition, vol. 2*, Cascadilla Press, 761-775.
- [8] Côté, Marie-Hélène. 2011. French liaison. In: van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (Eds.), *The Blackwell Companion to Phonology*. Wiley-Blackwell, Malden, pp. 2685–2710.
- [9] Culbertson, Jennifer, Smolensky, Paul & Wilson, Colin. 2013. Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in Cognitive Science* 5, 392–424. DOI: 10.1111/tops.12027. Online May 23. ROA 1166.
- [10] Deng, Li. 2015. Achievements and challenges of deep learning — From Speech Recognition to Language and Multimodal Processing. *APSIPA Transactions on Signal and Information Processing*.
- [11] Dowty, David. 2003. The Dual Analysis of Adjuncts/Complements in Categorical Grammar. In Ewald Lang, Claudia Maienborn, Cathrine Fabricius-Hansen, eds., *Modifying Adjuncts*. pp. 33–66. Mouton de Gruyter.]
- [12] de Jong, Daan. 1990. The syntax-phonology interface and French liaison. *Linguistics* 28, 57–88.
- [13] Durand, Jacques & Lyche, Chantal. 2008. French liaison in the light of corpus data. *French language studies* 18, 33–66.
- [14] Encrevé, Pierre. 1988. *La liaison avec et sans enchaînement, Phonologie tridimensionnelle et usages du français*, Paris : Editions du Seuil.
- [15] Faust, Noam. 2016 Exceptional liaison words in French. In review.
- [16] Faust, Noam & Smolensky, Paul. In preparation.
- [17] Goldrick, Matthew. 2011. Linking speech errors and generative phonological theory. *Language and Linguistics Compass*, 5, 397–412.
- [18] Goldrick, Matthew & Chu, K. 2014. Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein 2010. *Language, Cognition and Neuroscience*, 29, 452–458.
- [19] Goldrick, Matthew & Daland, Robert. 2009. Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology*, 26, 147–185.
- [20] Goldrick, Matthew, Putnam, Michael, & Schwarz, Lara. In press. Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition*.

- [21] Goldrick, Matthew, Putnam, Michael, & Schwarz, Lara. In press. The future of code mixing research: Integrating psycholinguistic and formal grammatical theories. *Bilingualism: Language and Cognition*.
- [22] Halford, Graham S., William H. Wilson, and Steven Phillips. 1998. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences* 21, 803–64.
- [23] Hankamer, Jorge. 1977. Multiple analyses. In Li, Charles (ed.) *Mechanisms of Syntactic Change*, 583–607. University of Texas Press.
- [24] Hayes, Bruce & Wilson, Colin. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
- [25] Hornsby, David. 2011. Getting it wrong: liaison, pataquès, and repair in contemporary French. In Pooley, T. & Laborgette, D. (Eds.), *On linguistic change in French: socio-historical approaches*. Laboratoire Langages, Littératures, Sociétés, Université de Savoie.
- [26] Hummel, J. E., and K. J. Holyoak. 2003. A symbolic-connectionist theory of relational inference and generalization. *Psychological Review* 110, 220–64.
- [27] Jarosz, Gaja. 2013. Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30, 27-71.
- [28] Laks, Bernard. 2009. Dynamiques de la liaison en français. In: Baronian, L., Martineau, F. (Eds.), *Le français d'un continent à l'autre*. Presses de l'Université Laval, Québec, pp. 237–267.
- [29] Lee, Moontae, He, Xiaodong, Yih, Wen-tao, Gao, Jianfeng, Deng, Li. & Smolensky, Paul. 2015. Reasoning in vector space: An exploratory study of question answering. arXiv 1511.06426. (under review for International Conference on Learning Representations, ICLR2016)
- [30] Legendre, Géraldine, Miyata, Yoshiro & Smolensky, Paul. 1991. Unifying syntactic and semantic approaches to unaccusativity: A connectionist approach. In *Proceedings of the Berkeley Linguistics Society* 7.
- [31] Legendre, Géraldine, Miyata, Yoshiro & Smolensky, Paul. 1990. Harmonic Grammar—a formal multi-level connectionist theory of linguistic well-formedness: An application. In *Proceedings of the Cognitive Science Society* 12.
- [32] Legendre, Géraldine, Miyata, Yoshiro & Smolensky, Paul. 1990. Harmonic Grammar—a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Cognitive Science Society* 12.
- [33] Legendre, Géraldine, Putnam, Michael, de Swart, Henriëtte & Zaroukian, Erin. 2016. Optimality Theoretic Syntax, Semantics, and Pragmatics: From Uni-to Bidirectional Optimization. Oxford University Press.
- [34] McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20, 75-138.
- [35] McCarthy, John J., ed. 2004. *Optimality Theory in phonology: A reader*. Blackwell.
- [36] McCarthy, John J. & Prince, Alan. 1993. Generalized alignment. In *Yearbook of morphology*, eds. G. Booij and J. van Marle. Kluwer.
- [37] McCarthy, John J. & Prince, Alan. 1995. Faithfulness and reduplicative identity. In *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*, eds. J. Beckman, L. Walsh Dickey, and S. Urbanczyk. Graduate Linguistic Student Association, University of Massachusetts at Amherst. ROA 60.
- [38] Miikkulainen, Risto. 1993. *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. MIT Press.

- [39] Morel, Éliane. 1994. Le traitement de la liaison chez l'enfant : études expérimentales. *TRANEL (Travaux neuchâtelois de linguistique)* 21, 85–95.
- [40] Morin, Yves Charles. 2005. La Liaison relève-t-elle d'une tendance à éviter les hiatus? Réflexions sur son évolution historique. *Langages* 158. 8-23.
- [41] Norris, Dennis, McQueen, James M., Cutler, Anne, & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology* 34, 191-243.
- [42] Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33, 999–1035.
- [43] Plate, T. A. 1993. Holographic recurrent networks. In C. L. Giles, S. J. Hanson, and J. D. Cowan. (eds.) *Advances in neural information processing systems* 5. Morgan Kaufmann.
- [44] Pollack, Jordan. 1990. Recursive distributed representations. *Artificial Intelligence* 46, 77–105.
- [45] Potts, Christopher, Pater, Joe, Jesney, Karen, Bhatt, Rajesh, & Becker, Michael. 2010. Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27, 77-117.
- [46] Prince, A., and P. Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Technical report, Rutgers University and University of Colorado at Boulder, 1993. ROA 537, 2002. Revised version published by Blackwell, 2004.
- [47] Prince, A., and P. Smolensky. 1997. Optimality: From neural networks to universal grammar. *Science* 275, 1604–10.
- [48] Rissanen, Jorma. 1986. Information theory and neural nets. In Smolensky, Paul, Rumelhart, David E. & Mozer, Michael C. (eds.), *Mathematical Perspectives on Neural Networks*, 567–602. Erlbaum.
- [49] Rosen, Eric. In press. Predicting the unpredictable: Capturing the apparent semi-regularity of rendaku voicing in Japanese through Harmonic Grammar. *Proceedings of the 42nd meeting of the Berkeley Linguistics Society*.
- [50] Rutgers Optimality Archive, roa.rutgers.edu.
- [51] Smolensky, Paul. 1983. Schema selection and stochastic inference in modular environments. In *Proceedings of the National Conference on Artificial Intelligence* 3.
- [52] Smolensky, P. 1986. Information processing in dynamical systems: Foundations of Harmony Theory. In *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1, *Foundations*, D. E. Rumelhart, J. L. McClelland, and the PDP Research Group. MIT Press.
- [53] Smolensky, Paul. 1987. On variable binding and the representation of symbolic structures in connectionist systems. Technical Report CU-CS-355-87, Department of Computer Science, University of Colorado at Boulder. February.
- [54] Smolensky, Paul. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence* 46, 159–216.
- [55] Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27, 720-31. ROA 118.
- [56] Smolensky, Paul. 2012. Symbolic functions from neural computation. *Philosophical Transactions of the Royal Society – A: Mathematical, Physical and Engineering Sciences*, 370, 3543–3569.
- [57] Smolensky, Paul, Lee, Moontae, He, Xiaodong, Yih, Wen-tao, Gao, Jianfeng & Deng, Li. 2016. Basic reasoning with Tensor Product Representations. arXiv 1601.02745. (supplementary material for [3])
- [58] Smolensky, Paul, & Legendre, Géraldine. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Vol. 1: *Cognitive Architecture*; vol. 2: *Linguistic and Philosophical Implications*. Cambridge, MA: MIT Press.

- [59] Smolensky, Paul, Legendre, Géraldine & Miyata, Yoshiro. 1992. Principles for an integrated connectionist/symbolic theory of higher cognition. Technical report CU-CS-600-92, Computer Science Department, and 92-8, Institute of Cognitive Science, University of Colorado at Boulder.
- [60] Smolensky, Paul, Goldrick, Matthew & Mathis, Donald. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, 38, 1102–1138. DOI: 10.1111/cogs.120472013. Online publication: June 26, 2013, DOI: 10.1111/cogs.12047.
- [61] Soderstrom, Melanie, Mathis, Donald W. & Smolensky, Paul. 2006. Abstract genomic encoding of Universal Grammar in Optimality Theory. In [58: Ch. 21]. Vol. 1, 403–471.
- [62] Tesar, Bruce B. & Smolensky, Paul. 2000. *Learnability in Optimality Theory*. MIT Press.
- [63] Tranel, Bernard. 1981. *Concreteness in generative phonology: Evidence from French*. Berkeley, CA: University of California Press.
- [64] Tranel, Bernard. 1995. Current issues in French phonology: Liaison and position theories. In John A. Goldsmith, *The Handbook of Phonological Theory*, 798–816. Oxford : Blackwell.
- [65] Zoll, Cheryl. 1996. Parsing below the segment in a constraint-based framework. PhD dissertation, University of California, Berkeley. ROA-143.
- [66] Zuraw, Kie & Hayes, Bruce. 2016. Intersecting constraint families: An argument for Harmonic Grammar. In review.