Inducing Suprasegmental Structure without Constituency: a Case Study on Southern French*

Eychenne, Julien
(Hankuk University of Foreign Studies)

Eychenne, Julien. 2013. Inducing Suprasegmental Structure without Constituency: a Case Study on Southern French. *The Journal of Studies in Language* 29.1, 97-128. This paper provides a connectionist analysis of the prosodic structure of a non-standard variety of French, using the framework of Dynamic Computational Networks. I develop a corpus-based analysis of schwa in this variety and propose that its behaviour stems from its inability to project from the syllabic to the metrical level. The model shows that a significant portion of the lexical prosodic structure of southern French can be learned from data without needing prosodic constituency. **(Hankuk University of Foreign Studies)**

Key words: schwa, foot, constituency, French, connectionism, dynamic computational networks

1. Introduction

The inference of metrical structure is one of the most fundamental challenges in phonological theory, as it involves inducing hidden structure which is not directly recoverable from the speech signal. Most work in generative phonology assumes that the induction of such covert structure is driven and constrained by a Universal Grammar (Tesar and Smolensky 2000). Optimality Theory (OT, Prince and Smolensky 1993), at least in its standard form, considers that learners are endowed with

* An earlier version of this work was presented at the Congrès Mondial de Linguistique Française (July 4-7, 2012, Lyon). I want to thank the audience for their feedback and constructive criticism. This paper also benefited from the comments and suggestions from three anonymous reviewers. All remaining errors and infelicities are solely mine. This work was supported by the research fund (2013) of Hankuk University of Foreign Studies.

a rich set of innate symbolic constraints (see Soderstrom, Mathis and Smolensky 2006) and that the task of the learner is to find a particular constraint ranking that is consistent with the ambient language. Yet, a growing body of research in computational language learning has questioned the need for such *a priori* knowledge. The pioneering work of the Parallel Distributed Processing group (McClealland and Rumerhart 1986a,b) initiated a fruitful research program, which seeks to model (and ultimately explain) high-level cognition in terms of neurally-inspired architectures[1]. The framework of Harmonic Grammar (Smolensky and Legendre 2006, vol 1: 207-234) envisions grammar as a set of numerically weighted constraints and is an heir to this research tradition. Its properties are currently being actively investigated (see for example Pater 2009; Potts *et al.* 2010). From a different perspective, Alderete *et al.* (2013) showed how a simple connectionist architecture based on a multi-layer perceptron was able to capture constraints on homorganic place restrictions in Arabic roots, suggesting that such constraints need not be available before language acquisition begins. In parallel, Hayes and Wilson (2008), building on previous work by Goldwater and Johnson (2003), have developed a model inspired by principles grounded in information theory, which aims to infer a phonotactic grammar solely on the basis of positive evidence.

This paper is a contribution to this broad research paradigm. It provides a connectionist analysis of the suprasegmental structure of a variety of French spoken in southern France. The framework which is adopted, namely Dynamic Computational Networks (DCN's), is a connectionist model developed by John Goldsmith and Gary Larson in a series of works (Goldsmith 1992, 1993, 1994; Goldsmith and Larson 1990, 1993; Larson 1990, 1993). I show that this type of architecture is able to model the suprasegmental structure of this variety of French and that it is superior to an approach to syllabicity solely based on a traditional sonority scale. Furthermore, I develop an analysis of the neutral vowel schwa ([ə]) in this variety and propose that its behaviour stems from its inability to project from the syllabic to the metrical level. This analysis is based on a large corpus drawn from an authentic sample of

1) See Bechtel and Abrahamsen (2002) for a recent overview of connectionism.

spoken French. Such a corpus-based approach to phonology has two important characteristics: first, it allows us to precisely quantify the predictions of the model, beyond what would be possible by simply examining a few hand-picked examples. Secondly, it enables us to evaluate how much information can be induced from data, which in turn provides insights about how much (or how little) knowledge must be available before learning takes place.

The rest of the paper is organized as follows: the next section provides some background about southern French; Section 3 introduces Dynamic Computational Networks and presents the adjustments that were made to the model; Section 4 discusses the simulations that were run and the results that were obtained, and the last section concludes and discusses a number of issues for future research.

2. Southern French

While it is not possible to treat 'southern French' as a homogenous linguistic system, a number of characteristics have been documented which are shared by most varieties (Armstrong and Unsworth 1999; Durand 1976, 1995; Durand, Slater and Wise 1987; Watbled 1995 *inter alia*). Traditional southern varieties of French have a phonemic schwa which contrasts with zero in word-final position (see 1), an opposition which is no longer found in other varieties, where the pairs in (1) are homophonous.

(1) Minimal pairs involving word-final schwa in southern French

net	[nɛt]	'neat (masc.)'
nette	[nɛtə]	'neat (fem.)'
Paul	[pɔl]	proper name
pôle	[pɔlə]	'pole'
appel	[apɛl]	'call (noun)'
appelle	[apɛlə]	'call (verb)'
golf	[gɔlf]	'golf'
golfe	[gɔlfə]	'gulf'

Early generative analyses of standard French posited an abstract final schwa similar to that found in southern French (Dell 1980; Schane 1968; Selkirk 1978), a view that has found some support until recently (Montreuil 2002). However, Tranel (1981) developed a number of arguments showing that such an approach raised more problems than it solved. Crucially, the arguments in support of such abstract schwas were mostly theory-internal and were not bound empirically. However, even though this type of analysis has been abandoned for northern French, the issue remains relevant in the context of southern French, where the vowel is a genuine segment.

The behaviour of schwa in southern French is tightly bound to that of mid vowels. The contrast which is found in northern French among the mid vowels /e, ɛ, ø, œ, o, ɔ/ (see for instance Walker, 2001) does not exist in southern French. Pairs such as *épée* 'sword' ~ *épais* 'thick', *jeûne* 'fasting' ~ *jeune* 'young', *beauté* 'beauty' ~ *botté* 'kicked', which are minimal in northern hexagonal French, are homophonous in southern French and are realized [e.'pe], ['ʒœ.nə] and [bo.'te], respectively. The distribution of mid (oral) vowels is governed by the *loi de position*, which can be stated as follows:

(2) The *loi de position*: a mid vowel is
  a. close in an open syllable
  b. open in a closed syllable or in an open syllable followed by a schwa-headed syllable
  (after Rizzolo, 2002, p. 11, translation mine)

Examples in (3) illustrate this pattern in word-final position in open syllables (3a), closed syllables (3b), and open syllables followed by schwa (3c), respectively.

(3)	a.	*panacée*	[pa.na.'se]	'panacea'
		paresseux	[pa.re.'sø]	'lazy (masc.)'
		haricot	[a.ri.'ko]	'bean'
	b.	*carrousel*	[ka.ru.'zɛl]	'carousel'
		épagneul	[e.pa.'njœl]	'spaniel (masc.)'
		espagnol	[ɛs.pa.'njɔl]	'Spanish (masc.)'

c.	*varicelle*	[va.ri.'sɛ.lə]	'chicken pox'
	épagneule	[e.pa.'njœ.lə]	'spaniel (fem.)'
	espagnole	[ɛs.pa.'njɔ.lə]	'Spanish (fem.)'

In addition to its lowering effect on a preceding nucleus, schwa has a repelling effect on stress, a pattern which is widely attested cross-linguistically (Crosswhite 2001; van Oostendorp 2000). The stress rule of southern French can be descriptively summarized as follows: stress falls on the last syllable if it is headed by a full (i.e. non-schwa) vowel, otherwise it falls on the second-to-last syllable. Analyses of southern French (Durand 1976, 1995; Watbled 1995) and abstract analyses of standard French (Montreuil 2002; Selkirk 1978; van Oostendorp 2000) consider that all vowels (except schwa) project their own unary foot and that a schwa-headed syllable associates with the previous syllable to form a trochee (Durand 1976, 1995; Selkirk 1976; van Oostendorp 2000; Watbled 1995). Under this view, stress is simply assigned to the head of the rightmost foot (Σ), as in *haricot* [(a)$_{\Sigma}$.(ri)$_{\Sigma}$.'(ko)$_{\Sigma}$], *espagnol* [(ɛs)$_{\Sigma}$.(pa)$_{\Sigma}$.'(njɔl)$_{\Sigma}$], *espagnole* [(ɛs)$_{\Sigma}$.(pa)$_{\Sigma}$.'(njɔ.lə)$_{\Sigma}$].

While the generalization on lexical stress needs to be captured somehow, the postulation of generalized unary feet seems to make southern French a typological oddity[2], since unary feet are generally regarded as degenerate (and thus exceptional) constituents (Hayes 1995: §5.1). I argue that this state of affairs is only the result of considering the metrical structure of this variety through the prism of constituency, but it vanishes as soon as we treat suprasegmental structure in terms of local relative prominence relations. The next section introduces dynamic computational networks which, as we shall see, make such a move possible.

3. Theoretical Framework

As pointed out by Goldsmith (1990), two important schools of thought can be distinguished with respect to syllabification; the first one, the sonority tradition, regards syllabicity as a sonority wave, that is to say

2) This issue is discussed in detail by Andreassen and Eychenne (2013).

a rhythmic alternation of peaks and troughs. The second, more recent tradition is the syntactic approach (see for instance Nespor and Vogel 2007). According to this view, prosodic information is organized in terms of constituents, which can be further decomposed into immediate sub-constituents, down to the terminal nodes (typically timing units or moras). Dynamic Computational Networks belong to the first tradition; they are a type of connectionist architecture which represents sonority as an alternation of crests and troughs[3].

3.1. Dynamic Computational Networks (DCN's)

DCN's were developed to account for syllabification and stress phenomena (Goldsmith 1992, 1993; Goldsmith & Larson 1990, 1993; Laks 1995, 1997; Larson 1990, 1993). Contrary to other connectionist architectures, which are generic and thus able to deal with any kind of input that can be represented by numerical vectors, DCN's were specifically designed to model phonological strings. The fundamental claim that underlies this type of architecture is that the rhythmic profile of a string is the result of the local interaction between an inherent sonority value for each unit of the network and lateral competition between neighbor units. More specifically, a DCN is a single-layer perceptron with bilateral inhibition/excitation connections, in which an input of length n is represented as a network of n formal neurons, each of which is connected to its immediate neighbors. Right-to-left and left-to-right connections are each controlled by a single parameter (or synaptic weight), α and β respectively (see 4 below).

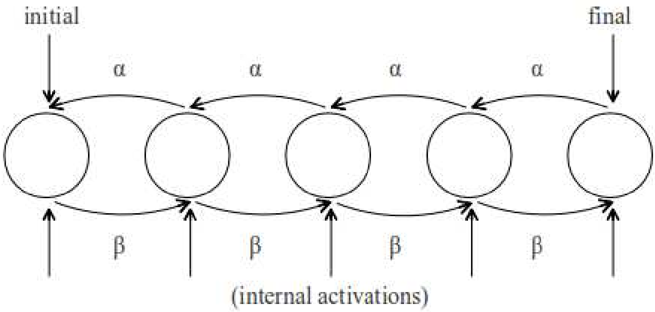
Such a network can model syllabic structure, if units are taken to represent segments, or metrical structure, in which case units represent syllables (or more precisely syllabicity peaks)[4]. In a syllabic network,

3) Lack of space precludes a full discussion of syllabicity in phonological theory, but see Goldsmith (2012) for a recent overview.

4) One of the most interesting results obtained in a DCN model is the analysis of syllabification in Berber (see Goldsmith and Larson 1990), a language known for allowing a wide range of syllabic consonants. As an example, the form /tluat/ (a Berber place name) would be simply syllabificed as [tlwat], with a complex onset, if syllabification were solely based on a traditional sonority scale. Assigning appropriate positive values to the parameters of the

each unit receives as its input a numeric value which corresponds to its own inherent sonority, as well as the output of its immediate neighbors weighted by the synaptic connections α and β.

(4) Dynamic Computational Network



The output of the neuron is the sum of the values received in input: it corresponds to the derived sonority of the unit. (In addition, the model allows for two positional activation values at each edge of the network; their role is discussed in §3.3.) The key insight of this architecture is that the derived sonority profile does not need be identical to the inherent sonority of a string's units. The distinction between inherent and derived sonority roughly corresponds to the distinction between input and output in generative models such as OT. It must be emphasized that the present model does not claim by any means that each phonological unit (a segment in a string) is implemented as one physical neuron in the brain of a speaker. What the model does claim is that phonological units are not symbols on which combinatorial operations can be performed, but that they can be usefully modelled as formal neurons which exhibit computational properties inspired by physical neurons.

When it receives an input, the network dynamically evolves in time

model leads to a derived sonority profile where /l/ is a local sonority peak and /u/ is a local trough (i.e. an onset), which corresponds to the observed syllabification, i.e. the disyllabic form [tl̩.wat] where /l/ has a higher derived sonority that /u/.

until it stabilizes into an equilibrium state in which new updates no longer modify the global state of the system (that is, the derived sonority of the units no longer changes, or remains below an arbitrarily small threshold). The update of the system is described by the following difference equation (after Goldsmith 1992: 223):

(5) $d_i^{t+1} = u_i + \alpha \cdot d_{i+1}^{t} + \beta \cdot d_{i-1}^{t}$

Where u_i and d_i^t are the inherent and derived sonority of unit i at time t, respectively. The variables of the system are summed linearly[5]. The role of the parameters in the model requires further elaboration. They vary from language to language, and each value pair describes a particular linguistic system. However, because these values are continuous, several ranges of values will describe (nearly-)identical linguistic systems. The task of the learner in a DCN is thus to infer a pair of values for α and β which generates the strings of the target language.

In order to better understand the behaviour of a DCN, it is worth considering a simple case. Let M(α,β) be a model whose parameters are α and β, and let **u** be a sonority vector which corresponds to the input values of the network. Figure (6) illustrates the behaviour of a DCN for a model M(0,-0.5) and an input **u** = (0, 0, 0, 0, 0, 0, 0, 0, 1). (This system is an approximation of the lexical stress pattern of French.)

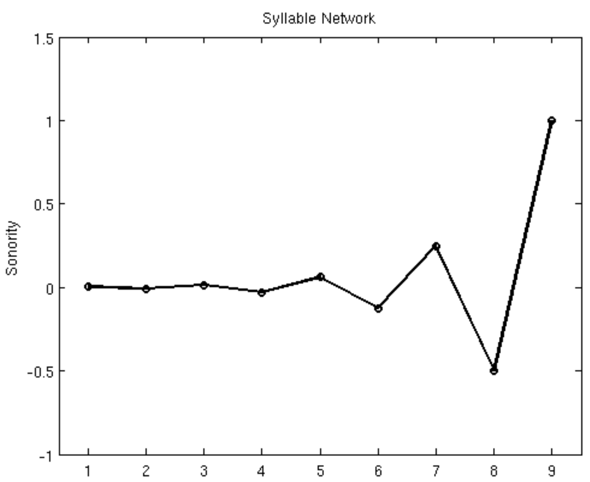
As can be seen, the rightmost unit's synaptic weight spreads leftward in the network, by an exponential factor α^n, where n is the distance between two units. Because the weight is negative, the resulting pattern is an alternation of peaks and troughs of decreasing amplitude. This pattern characterizes rhythmic alternations: in this example, the sonority of even nodes is inhibited, whereas the sonority of odd nodes is enhanced; the magnitude of the change diminishes the further away a unit is from the source node.

This alternation corresponds to the alternation between stressed and unstressed syllables at the metrical level and to the alternation between

5) Prince (1993) demonstrated that a DCN is a discrete approximation of a dynamic linear model.

consonants and vowels at the syllabic level. One of the most compelling aspects of this approach is that these properties are not axioms of the theory. In frameworks such as OT, the tendency to favor alternating rhythm is assumed to be the result of the presence of constraints such as *CLASH ("avoid consecutive peaks") and *LAPSE ("avoid consecutive troughs"), which do not appear to be independently motivated. By contrast, in a DCN, this tendency is the direct consequence of the lateral competition between the units of the string, under certain parameter settings.

(6) Spreading of the sonority wave for **u** = (0, 0, 0, 0, 0, 0, 0, 0, 1), with α = -0.5 et β = 0



This basic architecture was further developed in Larson (1993). Larson postulates three interconnected representational levels, each of which is a network. The autosegmental network (A) specifies the featural structure of segments and computes their inherent sonority; the syllabic network (S) determines the sonority profile of the segmental string; the metrical level (M) determines the stress pattern of the string. The simplest model assumes a feed-forward relation between the 3 levels (A → S → M): the dynamics of the autosegmental level determine the sonority of each unit segment; the syllabic level determines the location of syllabicity peaks (i.e. nuclei), which in turn feed the metrical level. As we shall see in 3.3, the facts of southern French suggest that feedback connections must be allowed between the syllabic and metrical networks. In addition, it is worth mentioning that the autosegmental level has received very little attention so far, and it is often represented as *ad hoc* numerical values (but see Larson 1993: 22-28 for a number of suggestions). Since this paper is primarily concerned with segmental syllabic structure, this approach is also adopted here.

3.2. DCN and French Syllabification

Laks (1995) successfully applied this connectionist architecture to French syllabification. The author analyzed in detail a corpus of 794 words (for a total of 832 syllabified forms) which were judged representative of the types of syllabic structures found in French. For a number of words, the corpus contained several possible syllabifications, such forms with or without schwa, with syneresis or dieresis of high vowels, etc. After training on a subset of the data set, the DCN reached a performance of 99.87% in predicting the syllabic structure.

These results are significant since they show that this kind of non-symbolic architecture is able to learn the syllabic structure of French without the rich *a priori* vocabulary that is usually assumed in symbolic models. Syllabic structure is conceived of as an emergent property of the temporal dynamics of the network. However, Laks' model is a significant departure from Goldsmith and Larson's original proposal. In his model, the parameters α and β are no longer treated as scalar values that control the connections of the network (left and right respectively), but both parameters are indexed on a number of predetermined natural classes of segments. Laks' implementation distinguishes 6 classes, namely vowels (V), glides (G), plosives (O), fricatives (F), nasals (N) and liquids (L). The parameters of the models should thus be represented as two vectors ($\mathbf{a}$ and $\mathbf{b}$) whose dimensions correspond to natural classes of sounds, namely $\mathbf{a} = (\alpha_V, \alpha_G, \alpha_O, \alpha_F, \alpha_N, \alpha_L)$ and $\mathbf{b} = (\beta_V, \beta_G, \beta_O, \beta_F, \beta_N, \beta_L)$.

Several remarks regarding this extension of the 'standard' DCN are in order. First, it appears that the intrinsic sonority of segments (or classes of segments) is encoded twice in this approach: once as the inherent values of the segments, and once in the α and β parameters, since these target specific classes. Furthermore, before learning takes place, vowels are assigned positive α and β parameters and a positive inherent sonority, whereas other segments are assigned negative parameters and inherent sonority (Laks 1995: 63, table 5)[6]. This initial parameter setting constitutes in itself a strong *a priori* bias concerning the asymmetry between syllabic and non-syllabic segments, since its net effect is to favour vowels as peaks. It is thus legitimate to wonder what the performance of the network would be in the absence of such *a priori* knowledge about natural classes. Last but not least, this extension of DCN's introduces more degrees of freedom in the network, which makes it more expressive but also less restrictive. Because the parameters of a standard DCN are only two scalar values, they explore a much more restricted search space, which makes them computationally more tractable and limits, to a certain extent, the possibility of over-generating unattested patterns.

The rest of the paper develops an analysis of the metrical structure of southern French using the original architecture put forth by Goldsmith and Larson. The next sub-section presents the model in detail, paying attention to the interaction between the syllabic and metrical networks.

3.3. Architecture of the Model

Since DCN's do not build constituents, metrical structure is constructed out of syllabicity peaks and troughs. Specifically, the syllabic network must include a recognition layer that is able to identify sonority maxima and minima (Larson 1993: 40-43). In Goldmith and Larson's model, the syllabic recognition layer scans the derived sonority d of each unit i and checks whether it satisfies either of the following conditions:

6) Glides are actually assigned a null inherent sonority and negative α and β.

(7) Identification of maxima and minima in the syllabic recognition layer (adapted from Larson 1993: 41):
Maxima: BOOLEAN($d_i > d_{i-1}$ AND $d_i > d_{i+1}$)
Minima: BOOLEAN($d_i < d_{i-1}$ AND $d_i < d_{i+1}$)

Where BOOLEAN is a function that returns 1 if the condition is true and 0 otherwise. However, an important issue with DCN's, which was clearly identified by Larson (1993: 99-102, 108), is the status of sonority plateaus. This problem is particularly acute is the case of hiatuses, which abound in many languages, including French. Indeed, it is extremely unlikely that two adjacent units in a DCN have exactly the same derived sonority. Therefore, under a strict understanding of the definition of a peak as a local sonority maximum, a DCN is unable to account for hiatuses (cf. southern French *paysage* 'landscape' [pe.i.za.ʒə]). One 'naïve' approach might be to determine a range within which two adjacent nodes are considered as level, but in practice this would be extremely error-prone and the threshold should be determined empirically in an *ad hoc* fashion, which is clearly not satisfactory. Laks (1995) offers a more principled solution to this problem. In his model, all and only the units whose derived sonority is positive are treated as peaks. In other words, 0 is interpreted as a baseline that distinguishes syllabic segments from non-syllabic ones in the sonority plane[7]. Troughs remain identified as local sonority minima.

The artificial learner implemented in this paper builds upon this insight. The interface between the syllabic and metrical networks thus includes the following function to identify the nodes of the metrical network:

(8) Peaks: BOOLEAN($d_i > 0$)

The nodes of the network which are identified as peaks by the syllabic recognition device (according to (8)) become input units to the metrical network. As a result, there is no longer a one-to-one mapping

7) A very similar idea is developed in Klein (1993), although not from a connectionist perspective.

between local sonority maxima and syllabicity peaks. For instance, a hiatus occurs whenever two consecutive units have a positive derived sonority.

As we saw earlier, French stress generally falls on the last syllable of a word unless it is headed by a schwa. Quantity-insensitive languages with demarcative stress (such as French) generally align the main prominence on either edge of the string, with the possibility of marking a number of syllables at that end as extrametrical (Larson 1993: 45). In a DCN, right-edge prominence is achieved by assigning a final positional activation value to the rightmost node in the metrical network (see figure 4). Consider the form *haricot* 'bean' [a.ri.ko]. A proper syllabification of this form will identify 3 peaks, namely [a], [i] and [o]. The input to the metrical network will be a vector $\mathbf{u} = (1, 1, 1)$. Edge enhancement will raise the sonority of the rightmost unit, yielding $\mathbf{u}' = (1, 1, 2)$. Given adequate values for α and β (as in 6), the DCN will assign the highest derived sonority to the rightmost node and will create an alternating pattern of peaks and troughs.

As we saw in Section 2, a key characteristic of southern French is the existence of a lexical schwa. In order to adequately model its prosodic weakness, it is necessary to take into account the interaction between the syllabic and metrical networks. The Boolean recognition of peaks introduces a non-linear relation between the syllabic and metrical networks. Indeed, the input values to the metrical network are not proportional to the output values of the syllabic one; instead, all the units whose value is above a given threshold (namely, 0) are passed as input units to the metrical level, with the same inherent sonority (1). As a result, the metrical network makes no qualitative difference between vowels, and it is not able to distinguish between schwa and full vowels on its own.

It is tempting at first sight to treat French schwa as an extrametrical vowel, but such an option is clearly not workable. Doing so would force one to analyze obstruent+liquid clusters as codas, as in *sobre* 'sober' [(sɔbr)$_{\Sigma}$. ⟨ə⟩] and *socle* 'base' [(sɔkl)$_{\Sigma}$. ⟨ə⟩]. While it may be argued that in varieties where final schwa has been lost, such as North American varieties, these clusters have indeed been reanalyzed as codas and are often simplified (e.g. /sɔbr/ → [sɔb]), we find no independent evidence that these clusters can be analyzed as codas in southern French. They never appear in syllable-final or word-final position, except after schwa deletion in the case of innovative speakers. More critical is the fact that this analysis wrongly predicts that 'extrametrical' schwa cannot occur morpheme-internally due to the peripherality condition, which requires extrametrical constituents to align with either edge of their domain (Hayes, 1995: 57-58). This prediction is falsified by the existence of monomorphemic forms like *céleri* 'celery' [sɛləri] and *écrevisse* 'crayfish' [ɛkrəvisə], where the internal schwa is prosodically weak and triggers mid-vowel lowering, according to the *loi de position* summarized in (2).

In order to account for the prosodic weakness of schwa, I propose that this vowel is simply unable to project from the syllabic to the metrical network. This approach captures the spirit of Selkirk's (1978) analysis and, in an OT context, van Oostendorp's (2000). To model the behavior of schwa in OT, van Oostendorp develops a schema of *projection* constraints which require that the head of certain prosodic constituents (in particular, the foot) dominate certain features. Specifically, the prosodic weakness of schwa is due to the fact that projection constraints prevent it from accessing the head position of a foot because it is featureless. In the connectionist architecture adopted in this paper, this insight is reinterpreted as the fact that schwa is invisible at the metrical level. As we have seen in (8), only the units whose derived sonority is greater than 0 access the metrical level. The metrical invisibility of schwa is therefore the result of it having a non-positive derived sonority. This property is implemented in the learning module as a feedback connection from the metrical network back to the syllabic network. The next section develops an analysis of a corpus of southern French in order to demonstrate the usefulness of the model.

4. Analysis

4.1. The corpus

In order to use realistic data, the corpus was constructed from orthographically transcribed recordings drawn from the project « Phonologie du français contemporain : usages, variétés, structure » (PFC[8]) (Durand, Laks et Lyche 2002, 2009). I used the transcriptions of 10 speakers from a small village in the South West of France (Douzens, in the Languedoc area). The transcribed part of the recordings represents 185 minutes of conversation in total (about three hours). I also added the content of the word list and the read text from the PFC survey protocol, which were specifically designed to study a number of phonological phenomena. The raw corpus initially contained 46,682 tokens, representing 3,314 types. This type list was cleaned up to eliminate noise such as misspelled variants and family names. Since this study is concerned with syllabification, it was decided upon reflection to remove all duplicate homophones whose lemmas were identical, as in inflected forms such as *venu* ~ *venues* [vøny] and *arrivé* ~ *arrivait* [aʁive]. There were many such redundant forms in the data and they would have blurred the results since it would have been difficult to isolate the ability of the network to truly generalize to unseen forms on the hand from its ‘memory’ of forms already encountered in the training data on the other.

The cleaned-up word list was transcribed using a broad IPA transcription according to the general pronunciation of southern French and then syllabified following traditional assumptions about the syllabic structure of southern French (Durand 1995). In particular, the *loi de position* was used as a guide to syllabification for word-internal clusters. By way of example, the word *exemple* ‘example’ was transcribed [ɛg.zaɴ.plə] (with a ‘coda’ /g/) since the first vowel of the word is mid-open. /r+j/ clusters were treated as tautosyllabic (e.g. *intérieur* ‘inside (noun)’ [ɛɴ.te.rjœr]), even though a heterosyllabic pronunciation seems to be spreading (e.g. [ɛɴ.tɛr.jœr]), especially among younger speakers[9].

There is no firm agreement about the status of the vowel written <e> in the initial syllable of polysyllabic forms (e.g. *p<u>e</u>tit* ‘small’ [pøti]).

8) See http://www.projet-pfc.net

9) A proper treatment of /r+j/ sequences would require taking into account the interaction between phonology and morphology, which is beyond the scope of this paper.

Durand, Slater & Wise (1987) argue that this vowel is phonetically and phonologically indistinguishable from the stable vowel written <eu> in a word like *meunier* 'miller' [mønje]. I followed their analysis and transcribed all such <e>'s as stable vowels.

Nasal vowels were transcribed as an oral vowel followed by a nasal appendix, which is their canonical realization in southern varieties of French (Durand 1988). The appendix was transcribed as [ɴ].

The final corpus contains 2,530 forms. It was randomized and divided into three sets, as has become standard practice in machine learning: a training set containing 60% of the data (1,518 words), a validation set and a test set, each containing 20% of the data (506 words). A short excerpt of the corpus is given below[10]:

(9)	*génial*	ʒe.njal	CV.CGVC
	généalogiques	ʒe.ne.a.lo.ʒi.kə	CV.CV.V.CV.CV.CV
	général	ʒe.ne.ral	CV.CV.CVC
	générale	ʒe.ne.ra.lə	CV.CV.CV.CV
	génération	ʒe.ne.ra.sjɔɴ	CV.CV.CV.CGVN
	génétique	ʒe.ne.ti.kə	CV.CV.CV.CV

The corpus contains 6,440 syllables in total, corresponding to 27 different syllabic types. The most frequent ones are summarized in table (10). Unsurprisingly, the CV pattern is by far the most frequent, accounting for over half of the corpus. Even though southern French tolerates complex onsets, it is clear that complex codas tend to be avoided. As a matter of fact, many complex codas found in standard French (such as /rn/ in *alterne* 'alternate' [al.tɛʁn]) are not found because of the presence of a schwa ([al.tɛʁ.nə]).

(10) Main syllabic types in the corpus

type	CV	CVN	CVC	CCV	V	CGV	VN	CGVN
tokens	3619	572	552	425	383	242	133	131
%	56.2	8.9	8.6	6.6	5.9	3.8	2.1	2.0

10) C = consonant (including nasal stops), V = vowel, G = glide and N = nasal appendix.4

4.2. Learning Protocol

The learning algorithm that was used, originally developed by Goldsmith and Larson (see Larson 1993: chap. 6), is inspired by a learning procedure known as "simulated annealing" (see the appendix for a more detailed overview)[11]. The model is a supervised learning algorithm, which relies on a parameter called "temperature", initially set to a very high value. The learner is presented with items from the lexicon, one at a time. Each time a new form is presented, the temperature decreases according to a cooling schedule, whether or not the network successfully predicts the correct form.

Each presentation of the whole training constitutes one *epoch*. At the beginning of each new epoch, the lexicon is randomized so as to avoid any sequence effect. When the temperature of the system falls below a predefined threshold, the system is considered frozen in a stable state and no further learning takes place. The intuition behind this learning algorithm is that the magnitude of the change for α and β depends on the temperature of the system: the hotter the system, the bigger the change may be. The magnitude of the change asymptotically decreases towards 0 as the system cools down.

It must be born in mind that because the algorithm is stochastic, learning takes place in a non-deterministic way and the outcome thus differs on each new trial. In each simulation, the inherent sonority of the units is re-initialized randomly before learning starts. This encodes the linguistic hypothesis that the learner has no direct access to the underlying sonority of the segments of the target language; it only has access to the rhythmic profile of the strings to which it is exposed. As in all supervised learning algorithms, the learner tries to iteratively approximate the target forms of the language by comparing its predictions to the forms observed in the training data. Since the output of a DCN is a vector of continuous sonority values, rather than constituents, inputs and outputs need to be processed so as to make

11) The learning algorithm was implemented in the Python programming language, using the Numpy library for numerical computing; see http://www.python.org and http://www.numpy.org. The source code and the data set may be obtained from the author upon request.

them directly comparable. The sonority wave of the syllabic network is transformed into a sequence of peaks and troughs, according to three degrees of prominence: H (high) corresponds to a local sonority maximum; L (low) corresponds to a local sonority minimum and all other segments are labeled O (other). For example, the word *blessé* 'hurt' [ble.se] will be assigned the following sonority profile: LOHLH. Furthermore, we will use the notation 〈H〉 to represent a local sonority maximum whose derived sonority is non-positive, as in the case of schwa. For instance, the word *pâte* 'pasta' [patə] will be assigned the following profile: LHL 〈H〉 . The syllabification of each word in the corpus was converted to a sonority curve according to these premises.

Previous research in DCN's only used a training set and (sometimes) a test set to train their models. Although this was a common practice in the early nineties, this is now considered problematic in machine learning; since the learning algorithm contains a number of hyper-parameters which must be adjusted to the type of data on which the network is trained, there is a risk of overfitting the model to the data. The procedure that was followed was to train and adjust the hyper-parameters of the model on the training set, and then use the validation set to check the accuracy of the model on unseen data. Once the hyper-parameters were set so as to not overfit the training data, the test set was used to measure the predictive power of the model. We now turn to the results of the simulations.

4.3. Results

This section reports on the results obtained with the architecture laid out above (and detailed in the appendix), focusing on the syllabic network. In order to avoid any bias in the selection of the model, 1,000 simulations were run so as to get a fair representation of the average performance of the architecture. The mean performance of the network on the training set was 99.38%. The mean prediction score on the validation set was slightly lower at 98.48%, ranging from a minimum of 96.84% to a maximum of 98.81%. Unsurprisingly, the network was rather slow to converge since the inherent sonority values and the parameters were initially assigned randomly and the network had to find an optimal setting for all of them. It takes on average 18.6 epochs for the network to stabilize. There is however a very high variance across simulations (standard deviation $\sigma = 19.5$). The fastest simulation converged in only 3 epochs, whereas the slowest one took 329 epochs. Importantly, it must be stressed that none of the simulations failed and that the model always converged.

To get an overall picture of the behaviour of the architecture, I computed an average model over the 1,000 simulations. This model simply averages over all the values obtained for the parameters α and β as well as for the learned inherent sonority of the segments. The mean values of α and β are given in (11):

(11) mean values for α and β over 1,000 simulations for the syllabic network:
$\alpha = 0.185$
$\beta = 0.005$

Let's now turn to the inherent sonority values that were learned. Average results are reported in (12), sorted in descending order from left to right, top down:

(12) Inherent sonority learned by the model (averaged over 1,000 simulations)

a : 4.73	ɛ : 3.78	y : 3.57
ɔ : 3.55	i : 3.37	ø : 2.31
e : 3.07	o : 2.86	u : 2.21
œ : 2.05	ə : −1.29	ɥ : −1.41
w : −1.99	j : −2.33	ɲ : −3.02
ʀ : −3.64	r : −4.01	l : −4.26
ʃ : −5.17	ʒ : −5.16	z : −5.26
g : −6.55	n : −6.96	k : −7.23
v : −7.24	f : −7.51	m : −7.81
b : −7.86	s : −8.19	d : −8.56
p : −9.37	t : −11.05	

A few remarks regarding the inherent sonority that are assigned to the segments are in order. First, all vowels (except schwa) are assigned a positive value, which does not come as a surprise given the architecture of the model (values in the positive half-plane are syllabicity peaks). Schwa is assigned a negative sonority value, but its sonority is higher than all the consonants. The network assigns a fairly high sonority in the negative half-plane to the three glides. This is consistent with the fact that glides are considered intermediate segments between vowels and consonants and are often the result of the weakening of high vowels. Obstruents, on the other hand, are all assigned a low negative sonority, with the voiceless stops [p] and [t] at the very bottom of the scale. The derived sonority values for the liquids /l/ and /r/ are very close to each other, which is certainly the result of their shared distributional properties (e.g. both can appear as the second member of a complex onset, as in *table* [ta.blə]).

The inherent sonority of nasals is much less consistent, but their distribution is not uniform. The segments [m] and [n] can both appear in onset and coda position. The consonant [ɲ] is restricted to onset positions, and it is nowadays realized as [nj] by many speakers. As such, it seems to lie outside of the consonantal system. Finally, the relatively high sonority (between glides and liquids) of [ɴ], the nasal appendix, is consistent with the fact that it only appears in post-vocalic position.

In order to test the predictive power of the average model, it was run once against the test data set, which was never seen during training. The model made 6 prediction errors out of 506 data points, which corresponds to a performance of 98.81%. To fully appreciate the performance of DCN, I run a baseline model with alpha and beta set to 0 (which means that there is no lateral influence of units on one another) and setting the inherent value of the units according to a traditional sonority scale (Goldsmith 1990: 111), with stops at the bottom of the scale (-5) and open vowels at the top (+5). Such a model obtained a performance of 92.49% on the test set. This shows that, even though southern French syllabicity mostly conforms to the sonority sequencing principle, inherent sonority alone is not sufficient and DCN's offer a compelling architecture to model syllabicity in cases where syllabification does not follow sonority.

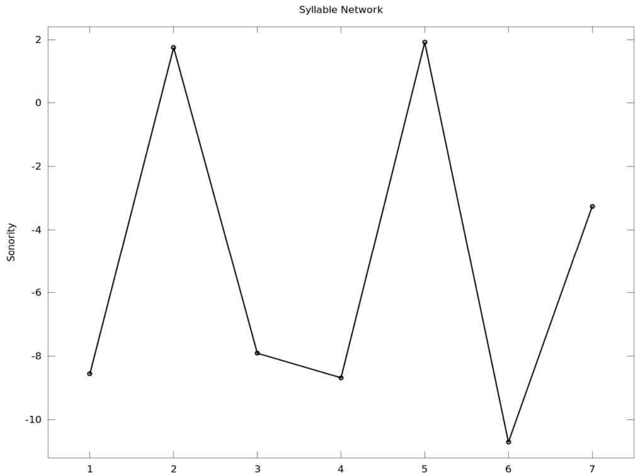
To show the importance of the dynamics of the network in such cases, let's consider the form *discute* 'chat' /diskytə/, whose underlying structure is represented in the average model by the sonority vector **u** = (−8.56, 3.37, −8.19, −7.23, 3.57, −11.05, −1.29). On the sole basis of the inherent sonority values, the network would predict the sonority profile *LHLOHL ⟨H⟩, corresponding to the syllabification *[di.sky.tə]. Remember however that according to the *loi de position*, the trough should be /k/, since /s/ has a lowering effect on a preceding vowel when it is mid. Figure (13) shows the predicted derived sonority curve of this word in the average model, which corresponds to the vector **d** = (−8.55, 1.75, −7.91, −8.68, 1.91, −10.71, −3.27). This syllabic curve is consistent with the effect of the *loi de position*. Even though the derived sonority of /s/ is quite low and only slightly higher than /k/, it is high enough that it is not a local minimum in the output form. This post-peak/pre-trough position corresponds precisely to the traditional notion of coda, but no constituent structure needs be posited.

Let's now turn to the behaviour of schwa in more detail. As we have seen in §3.3, the properties of this vowel in this approach stems from the fact that it is prosodically invisible in the metrical network. Let's consider a word with only an alternation of vowels and consonants but containing a schwa, such as *viticole* 'wine-making' [vitikɔlə]. The average model assigns the derived sonority **d** = (−7.23, 1.98, −10.68, 1.36, −6.97, 2.24, −3.86, −2.00), which corresponds to the sonority curve LHLHLHL ⟨H⟩. Figure (14) plots the derived sonority predicted by the network for this form, along with the 0 baseline for the sake of clarity.

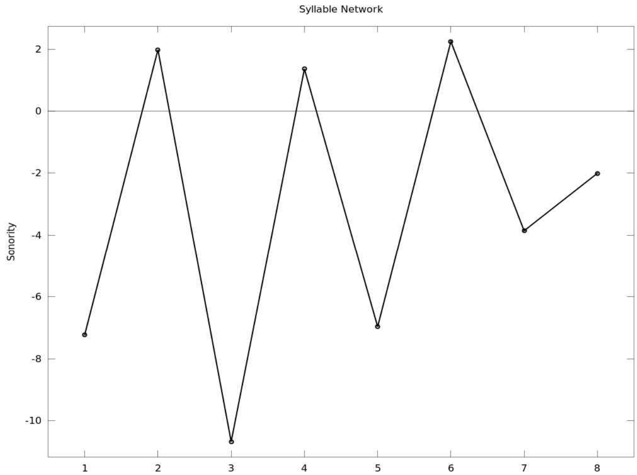
The three peaks above the baseline represent the three full vowels. Schwa, which corresponds to the eighth node, is a local sonority maximum; however, because of the non-linear thresholding function in the recognition layer (see (8)), it fails to be projected to the metrical network.

As a result, the node corresponding to the vowel [ɔ] becomes the rightmost node and receives the main lexical prominence by virtue of positional activation. Under this interpretation of the prosodic properties of schwa, the trochaic foot becomes unnecessary to account for the suprasegmental structure of southern French. The sonority pattern exhibited by the network in (14) would more closely match the concepts of supersyllable or recursive syllable which have been put forth in the literature (see Hall 2006: 403 for an overview).

(13) derived sonority of *discute*



(14) derived sonority of *viticole*



I believe, however, that the metrical grid (Prince 1983) is a far better and more direct symbolic approximation of a DCN architecture, as has been argued by Laks (1997). The metrical structure of *viticole* may thus be represented as in (15) as the layering of several levels of prominence. Crucially, though, these layers are not understood to represent pre-wired, innate symbolic categories but emerge from the dynamics of the architecture that has been developed.

(15) Metrical grid representation of *viticole*

Metrical prominence:						×		
Peaks:				×		×		×
Sonority maxima:		×		×		×		×
	v	i	t	i	k	ɔ	l	ə

To conclude this overview of the behaviour of our artificial learner, it is worth considering the prediction errors of the average model on the test set (the set which was never seen in training) from a qualitative point of view. These are listed in (16):

(16)

form	**expected curve**	**predicted curve**
islamikə	HOLHLHL ⟨H⟩	HLOHLHL ⟨H⟩
spɔr	LOHL	HLHL
spɔrtivə	LOHOLHL ⟨H⟩	HLHOLHL ⟨H⟩
spɛktaklə	LOHOLHLO ⟨H⟩	HLHOLHLO ⟨H⟩
staʒə	LOHL ⟨H⟩	HLHL ⟨H⟩
staɴd	LOHOL	HLHOL

It is striking to observe that all the prediction errors involve /sC/ clusters, and 5 out of 6 are in word-initial position. The cluster /sl/ is not a native cluster in French, and there was only one form in the training corpus that contained it (*slip* ‘underpants’ [slip]), moreover in word-initial position. The lack of integration of this cluster within the phonological grammar of southern French is supported by the fact that many speakers voice the /s/ (*slip* [zlip]; *islamique* [izlamikə]). Anyway, *islamique* is an assimilated borrowing in French and the prediction error of the network seems to simply reflect the performance of the model in the presence of an illicit cluster. Furthermore, the status of word-initial /sC/ clusters is a well-known problem in phonological theory[12]. The Romance prosthesis (Alkire and Rosen 2010: §2.1) in word-initial /sC/ clusters lends support to the idea that these clusters are indeed special (see Latin *schola* > **iscola* > French *école,* Spanish *escuela* 'school'). It might well be the case that the network's predictions in these cases are not errors, but reflect the fact that /s/ is actually a local maximum of sonority in this environment, at least for some languages such as French.

5. Conclusion

In this paper, I have developed a connectionist analysis of the suprasegmental structure of southern French, in a model based on dynamic computational networks. Building upon previous work by Larson (1993) and Laks (1995) in particular, I have presented a two-layer architecture to model syllabic and metrical structure. Because of the computational properties of DCN's, the interface between the syllabic and metrical networks was designed so that all and only the nodes whose derived sonority is positive are visible in the metrical network. While this property was implemented so as to be able to deal with sonority plateaus, it was shown to be useful to model the behaviour of schwa: in this architecture, schwa is simply a local sonority maximum in the negative half-plane of the syllabic network. This connectionist architecture was shown to successfully model the suprasegmental structure of southern French, and it was shown that the errors observed in the test set had to do with a very specific phonotactic pattern, namely sequences of /s/ + consonant, mostly word-initially.

This study has shown that a non-trivial part of the suprasegmental structure of southern French can be modelled without prosodic constituency. As mentioned in §2, the metrical structure of southern French, while very simple, seems to be a typological exception: in all

12) See Kaye (1992) for analysis of /sC/ clusters as heterosyllabic based on evidence drawn from Italian, Portuguese, Ancient Greek and British English.

the analyses I am aware of, the trochaic foot seems necessary to account for the interaction of schwa and stress on the one hand, and schwa and mid-vowels on the other; yet all other feet appear to be degenerate unary feet, a pattern which seems to be highly uncommon (see Andreassen and Eychenne 2013 for a fuller discussion of this issue). The present approach solves this problem by abandoning constituency entirely. The foot and the syllable receive no explicit theoretical status as such: they are an epiphenomenon of the interaction between two rhythmic levels which share the same architecture and computational properties.

If the approach to schwa laid out in this paper is correct, it opens interesting perspectives for the analysis of other languages. Van Oostendorp (2003) distinguishes three types of schwas from a typological perspective: schwas that alternate with zero (often epenthetic schwas), schwas that are the result of vowel reduction and stable schwas. Schwa in southern French is an instance of the latter category, but the DCN might prove an insightful framework to model other types as well. For instance, reduction to schwa in languages such as English might be captured as a lowering of the derived sonority of a full vowel under the influence of stress, which might be implemented as a form of backpropagation from the metrical network to the syllabic network. Similarly, this architecture could be useful in modelling glide-formation phenomena, such as those found in northern varieties of French (e.g. tuer 'to kill' /tye/ → [tɥe]).

I believe that the full potential of DCN's for modelling suprasegmental structure has yet to be explored. Even though a number of quantity-insensitive systems have been successfully modelled in DCN's (Goldsmith 1994), quantity-sensitive systems have received almost no attention. An analysis of a broad range of languages in this framework using large corpora will prove invaluable in testing this architecture on a larger scale and refining it. Further progress may also be achieved by extending the model, perhaps by adding one or several additional layers, to move beyond the lexical level and model connected speech.

References

Alderete, J., P. Tupper and S.A. Frisch. 2013. Phonological Constraint Induction in a Connectionist Network: learning OCP-Place constraints from data. *Language Sciences* 37, 52-69.

Alkire, T. and C Rosen. 2010. *Romance Languages: A Historical Introduction*. Cambridge: Cambridge University Press.

Andreassen, H. N. and J. Eychenne. 2013. The French Foot Revisited. *Language Sciences*. http://dx.doi.org/10.1016/j.langsci.2013.02.016

Armstrong, N. S. Unsworth. 1999. Sociolinguistic Variation in Southern French Schwa. *Linguistics* 37.1, 127-56.

Bechtel, W. and A. Abrahamsen. 2002. *Connectionism and the Mind. Parallel Processing, Dynamics, and Evolution in Networks*, Oxford: Blackwell Publishing.

Crosswhite, K. 2001. *Vowel Reduction in Optimality Theory*. New York and London: Routledge.

Dell, F. 1980. *Generative Phonology and French Phonology*. Cambridge: Cambridge University Press.

Durand, J. 1976. Generative Phonology, Dependency Phonology and southern French. *Lingua e Stile* 11.1, 3-23.

Durand, J. 1995. Alternances vocaliques en français du Midi et phonologie du gouvernement. *Lingua* 95, 27-50.

Durand, J. 1988. Les phénomènes de nasalité en français du Midi: phonologie de dépendance et sous-spécification. *Recherches Linguistiques* 17, 29-54.

Durand, J., B. Laks and C. Lyche. 2002. La phonologie du français contemporain: usages, variétés et structure. In *Romanistische Korpuslinguistik-Korpora und gesprochene Sprache/Romance Corpus Linguistics-Corpora and Spoken Language*, eds. C. Pusch and W. Raible, 93-106. Tübingen: Gunter Narr Verlag.

Durand, J., B. Laks and C. Lyche. 2009. Le projet PFC : une source de données primaires structurées. In *Phonologie, variation et accents du français*, eds. J. Durand, B. Laks and C. Lyche, 19-61. Paris: Hermès.

Durand, J., C. Slater and H. Wise. 1987. Observations on Schwa in Southern French. *Linguistics* 25.2, 983-1004.

Goldsmith, J. 1990. *Autosegmental & Metrical Phonology*. Oxford: Blackwell.

Goldsmith, J. 1992. Local Modeling in Phonology. In *Connectionism: Theory and Practice*, ed. S. Davis, 229-246. Oxford: Oxford University Press.

Goldsmith, J. 1993. Harmonic Phonology. In *The Last Phonological Rule*, ed. J. Goldsmith, 21-60. Chicago and London: The University of Chicago Press.

Goldsmith, J. 1994. A Dynamic Computational Theory of Accent Systems. In

Perspectives in Phonology, eds. J. Cole and C. Kisseberth, 1-28. Stanford: Center for the Study of Language and Information.

Goldsmith, J. 2011. The Syllable. In *The handbook of phonological theory,* eds. J. Goldsmith, J. Riggle and C. L. Alan, 164-196. Oxford: Wiley-Blackwell.

Goldsmith, J. and G. Larson. 1990. Local Modelling and Syllabification. *Papers from the 26th Annual Meeting of the Chicago Linguistic Society: Parasession on the syllable in phonetics and phonology* 26.2, 129-142. Chicago: Chicago Linguistic Society.

Goldsmith, J. and G. Larson. 1993. Using Networks in a Harmonic Phonology. In *Papers from the 28th Annual Meeting of the Chicago Linguistic Society*, vol. 2, 94-125. Chicago: Chicago Linguistics Society.

Goldwater, S. and M. Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Workshop on Variation within Optimality Theory.*

Hall, Nancy. 2006. Cross-linguistic Patterns of Vowel Intrusion. *Phonology* 23.3, 387-429.

Hannahs, S. 1995. *Prosodic Structure and French Morphophonology*, Tübingen: Niemeyer.

Hayes, B. 1995. *Metrical Stress Theory. Principles and case studies*. Chicago: The University of Chicago Press.

Hayes, B. and C. Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning, *Linguistic Inquiry* 39.3, 379-440.

Kaye, J. 1992. Do You Believe in Magic? The Story of s+C Sequences. *SOAS Working Papers in Linguistics and Phonetics* 2, 293-313.

Klein, Marc. 1993. La syllabe comme interface de la production et de la réception phoniques. In *De natura sonorum. Essais de phonologie*, eds. B. Laks and M. Plénat, 101-142. Saint-Denis: Presses Universitaires de Vincennes.

Laks, B. 1995. A Connectionist Account of French Syllabification, *Lingua* 95, 51-76.

Laks, B. 1997. *Phonologie accentuelle. Métrique, autosegmentalité, constituance*. Paris: CNRS Éditions.

Larson, G. 1990. Local Computational Networks and the Distribution of Segments in the Spanish Syllable. *Papers from the 26th Annual Meeting of the Chicago Linguistic Society: Parasession on the syllable in phonetics and phonology.*

Larson, G. 1993. Dynamic Computational Networks, and the Representation of Phonological Information, Unpublished PhD thesis, University of Chicago.

McClealland J., D. Rumerhart eds. 1986a. *Parallel Distributed Processing. Explorations in the microstructure of Cognition. Volume 1: Foundations*. Cambridge and London: MIT Press.

McClealland J., D. Rumerhart eds. 1986b. *Parallel Distributed Processing. Explorations in the microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge and London: MIT Press.

Montreuil, J.-P. 2002. Vestigial Feet in French. *Proceedings of the 2002 Texas Linguistic Society Conference on Stress in Optimality Theory*. University of Texas at Austin, Austin.

Nespor, M., Vogel, I. 2007 [1986]. *Prosodic phonology: With a new foreword*. Berlin and New York: Mouton de Gruyter.

Pater, J. 2009. Weighted Constraints in Generative Linguistics. *Cognitive Science* 33, 999-1035.

Potts, C., J. Pater, K. Jesney, R. Bhatt and M. Becker. 2010. Harmonic Grammar with Linear Programming: from linear systems to linguistic typology. *Phonology* 27, 77-117.

Prince, A. 1983. Relating to the Grid. *Linguistic Inquiry* 14.1, 19-100.

Prince, A. 1993. In Defense of the Number i. Anatomy of a Linear Dynamical Model of Linguistic Generalizations, Technical Report #1, Rutgers University Center for Cognitive Science. Available from http://ruccs.rutgers.edu/~prince/.

Prince, A. and P. Smolensky. 1993. Optimality Theory: constraint interaction in Generative Grammar, New Brunswick: Rutgers University Center for Cognitive Science.

Rizzolo, O. 2002. Du leurre phonétique des voyelles moyennes en français et du divorce entre licenciement et licenciement pour gouverner. Unpublished PhD thesis, Université de Nice-Sophia Antipolis.

Schane, S. 1968. *French Phonology and Morphology*. Cambridge, MA: MIT Press.

Selkirk, E. 1978. The French Foot: on the status of ‘mute’ e. *Studies in French Linguistics* 1.2, 141-150.

Soderstrom, M., D.W. Mathis and P. Smolensky. 2006. Abstract Genomic Encoding of Universal Grammar in Optimality Theory. In Smolensky and Legendre (eds), 403-471.

Smolensky, P. and G. Legendre eds. 2006. *The Harmonic Mind. From neural computation to optimality-theoretic grammar*. 2 volumes. Cambridge, MA and London: MIT Press.

Tesar, B. and P. Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge: MIT Press.

Tranel, B. 1981. *Concreteness in Generative Phonology: Evidence from French*. Berkeley: University of California Press.

van Oostendorp, M. 2000. *Phonological Projection: a theory of feature content and prosodic structure*. Berlin: Mouton De Gruyter.

van Oostendorp, M. 2003. Schwa in Phonological Theory. In *The Second Glot International State-of-the-Article Book*, eds. L. Cheng and R. Sybesma, 431-461. Berlin: Mouton de Gruyter.

Walker, D.C., 2001. *French Sound Structure*. Calgary: University of Calgary Press.

Watbled, J.-P. 1995. Segmental and Suprasegmental Structure in Southern French. In *Linguistic Theory and the Romance Languages*, eds. J.C. Smith and M. Maiden, 181-200. Amsterdam: John Benjamins.

Appendix: Implementation of the model

The goal of learning in a DCN is to assign a sonority value to the segments of the language and to find a pair of parameters (α,β) such that the network correctly predicts all the strings of the target language. A DCN accomplishes this task by iteratively computing the derived sonority *d* of each unit at time *t*, until the difference in derived sonority for each unit between *t* and *t*+1 is null or, in practice, lower than an arbitrary threshold δ (i.e. $|d^t - d^{t+1}| < \delta$). Prince (1993: 20) proved that this learning problem could be tackled more directly using linear algebra. Equation (5) can be formulated as (17) in vectorized form:

(17) $\mathbf{d} \leftarrow \mathbf{W}_n \mathbf{d} + \mathbf{u}$

Where **d** is the derived sonority vector of dimension *n* (initially set to 0), **u** is a vector representing the inherent sonority of the string and **W** a tridiagonal matrix of dimension $n \times n$, whose superdiagonal is set to α and whose subdiagonal is set to β. This equation can be usefully rewritten as (18):

(18) $\mathbf{d} = (\mathbf{I} - \mathbf{W}_n)^{-1} \mathbf{u}$

Where **I** is the identity matrix. The implementation of the model used in this paper is based on equation (18). This closed-form solution was chosen over the iterative approach for several reasons: it allows us to do without the threshold parameter δ, it is computationally more efficient than the iterative method and it yields an exact solution rather than an approximation (whose precision depends on the value of δ).

If we regard a DCN as a model of phonological acquisition, it is of course necessary to circumscribe the search space, which is determined by the parameters α and β. Since they are continuous numerical values, they could take on a theoretically infinite number of values. Prince (1993: 49-53) formally proved that any DCN which satisfies the condition $|\alpha\beta| \leq \frac{1}{4}$ converged, regardless of its size. In other words, for any such DCN, an exact solution to the learning problem is guaranteed to exist. While this still means that there exists an infinite number of possible converging values (Larson 1993: 36), in practice, the search space for α and β is restricted to the interval [-0.5 0.5] and it is assumed that the learning problem amounts to finding an optimal value for each of the parameters within this interval.

For each simulation, the initial values of α and β were assigned randomly from a uniform distribution within the interval [-0.3 0.3].

In order to train the model, I used an algorithm (originally developed by Goldsmith and Larson) inspired by a learning procedure known as "simulated annealing". This supervised learning algorithm relies of a parameter τ called "temperature", which is initially very high (τ = 1.0). Each time a new form is presented to the learner, the temperature decreases according to a cooling schedule ($\Delta\tau$), whether or not the network successfully predicts the correct form. For example, if $\Delta\tau$ = 0.99, the temperature decreases by 1% after each iteration. When the system fails to predict the correct form, the parameters α and β are randomly modified by a factor drawn from a normal distribution with mean 0 and variance τ (the temperature of the system), normalized by a constant κ. The temperature of the system is also increased by the Euclidean distance of the change in α and β; intuitively, this means that big changes in α and β will increase the temperature more than small ones. Finally, the inherent sonority of the segments whose derived sonority was wrongly predicted is modified, by a factor which depends on the temperature of the system and a normalization constant λ. When the temperature becomes lower that a predefined threshold (Θ), the system is said to be frozen in a stable state and no further learning takes place.

It must be emphasized that all these parameters (except α and β) are those of the learning algorithm, not of the DCN. Such parameters are often called "hyper-parameters" in the machine learning literature. For the sake of clarity, the 5 hyper-parameters of the learning algorithm are summarized below:

τ: the temperature of the model (initially set to 1.0 in all the simulations)

- $\Delta\tau$: the cooling schedule
- Θ: the freezing threshold
- κ: the parameters which controls the magnitude of the change for α and β in case of error
- λ: the parameters which controls the magnitude of the change for the inherent sonority of a unit in case of error

Eychenne, Julien
Dept. of Linguistics and Cognitive Science
Hankuk University of Foreign Studies
Mohyeon, Yongin, Gyeonggi 449-791, Korea
E-mail: jeychenne@hufs.ac.kr
Phone: 031-330-4998

Received: 2013. 03. 17
Accepted: 2013. 04. 30