

# Apprenabilité de la structure syllabique du français méridional dans un réseau calculatoire dynamique standard

Eychenne, Julien

Center for Language and Cognition Groningen, Rijksuniversiteit Groningen  
j.eychenne@rug.nl

## 1 Introduction

Si le rôle et le statut exact de la syllabe continue encore aujourd'hui d'alimenter le débat parmi les phonologues, c'est qu'il s'agit de l'un des objets linguistiques les plus élusifs qui soient. N'ayant aucune manifestation directe dans le signal de parole, sa nature ne peut être appréhendée qu'indirectement et, comme dans tout domaine d'investigation scientifique, son acceptation en tant que construction théorique passe nécessairement par la validation empirique : autrement dit, une théorie de la syllabe n'est aussi bonne que les prédictions qu'elle permet de faire.

Sans vouloir récrire ici l'histoire de la syllabe dans la théorie phonologique (voir Angoujard 1997 ; Blevins 1995 ; Goldsmith 1990 : ch. 3 ; Klein 1993 *inter alia*), le présent travail se propose de revenir sur la question de la syllabe en français en adoptant le principe du rasoir d'Ockham, à savoir que le meilleur modèle est le modèle le plus parcimonieux compatible avec les données. Le modèle théorique que nous adoptons est un modèle non symbolique développé par John Goldsmith et Gary Larson à la fin des années 80 et au début des années 90, à savoir les réseaux calculatoires dynamiques (*dynamic computational networks*). Sur la base d'un corpus de 2530 formes phonétisées et syllabifiées, nous montrons que ce type de modèle est non seulement capable d'apprendre la structure syllabique du corpus, mais il est également capable de généraliser ces prédictions à de nouvelles données, et ce même s'il ne dispose d'aucune information linguistique *a priori*. Le cadre épistémologique général dans lequel nous nous situons est celui du connexionnisme<sup>i</sup> réductionniste, tel qu'il a été défini par Laks (1996 : 50-65) dans la lignée de Smolensky (1988). À la suite de Laks, nous admettons l'existence de trois niveaux d'analyse pertinents : (i) le niveau conceptuel, qui est le niveau de la « description systématique » : c'est le niveau où sont formulées les généralisations (linguistiques) ; (ii) le niveau subsymbolique<sup>ii</sup>, qui est le niveau de la modélisation : ce niveau modélise et propose des explications des phénomènes émergeant au niveau conceptuel dans des architectures neuromimétiques ; (iii) le niveau physique : le niveau proprement causal. Ces trois niveaux correspondent respectivement aux niveaux computationnel, algorithmique et d'implémentation matérielle dans la théorie des systèmes de traitement de l'information de Marr (1982)<sup>iii</sup>. On prendra garde de distinguer ce connexionnisme réductionniste du connexionnisme implémentatif, dont l'objectif est simplement d'implémenter des analyses linguistiques dans des architectures de neurones artificiels (typiquement, des réseaux auto-associatifs), et du connexionnisme éliminativiste, qui nie l'existence d'un niveau conceptuel autonome. Le programme de recherche défendu par le connexionnisme réductionniste, et dans lequel s'inscrit cette contribution, vise à *expliquer* les structures qui émergent au niveau conceptuel en termes de dynamiques s'inspirant de celles que l'on voit à l'œuvre dans les architectures neuronales.

Cet article est organisé comme suit : nous discutons d'abord en (2) la question de la syllabe et le statut de la sonorité ; la section (3) offre une présentation des réseaux calculatoires dynamiques dans leur forme « standard » ainsi que la forme généralisée développée par Laks (1995) ; la section (4) présente le corpus de français méridional que nous avons constitué et nous discutons en (5) enfin les résultats obtenus avec un dispositif d'acquisition biaisé et un dispositif non biaisé.

## 2 Syllabité, constituance et sonorité

Quelle que soit son orientation théorique, tout phonologue s'accordera sur le fait que la constituance est considérée comme un des piliers de la phonologie contemporaine ; en particulier, les avancées formelles dans la modélisation de la structure prosodique représentent l'un des développements les plus significatifs de la période post-SPE<sup>iv</sup>, et il n'est pas exagéré de dire que la phonologie prosodique (Nespor et Vogel 1986, Hannahs 1995) représente aujourd'hui le modèle dominant de la structure suprasegmentale.

Il est toutefois utile de situer cet état de faits dans un contexte plus large, et en particulier du point de vue de la syllabe qui nous occupe ici. Comme le rappelle fort justement Goldsmith (1990), on peut distinguer deux grandes traditions : la tradition ondulatoire, la plus ancienne, considère la sonorité comme une onde, dont les sommets sont identifiés par des pics locaux de *sonorité*. La citation suivante de Bloomfield (1933 : 120-121) l'illustre parfaitement : « An utterance is said to have as many syllables (or natural syllables) as it has syllabics. The ups and downs of syllabification play an important part in the phonetic structure of all languages. » (cité par Goldsmith 1990 : 105). La deuxième tradition, plus récente, et dont la phonologie prosodique est l'un des aboutissements, est la tradition syntagmatique. Dans cette perspective, l'information prosodique est structurée en constituants qui s'analysent en sous-constituants immédiats, les unités chronématiques en étant généralement considérées comme les éléments terminaux.

Si l'approche syntagmatique représente aujourd'hui le courant dominant, force est de constater qu'elle a été attaquée de diverses manières, et ce même dans la lignée des approches logico-symboliques. À titre d'exemple, le modèle déclaratif développé par Angoujard (1997, 2006) réinterprète la syllabe comme le résultat de l'interaction entre la sonorité des segments (calculée sur la base de primitives unaires) et une grille rythmique de type « creux pic (creux) ». La syllabe n'y est donc pas vue comme une primitive théorique mais comme une construction émergeant de l'interaction d'objets phonologiques plus fondamentaux. Dans une perspective différente, Scheer (2008) offre une déconstruction méthodique de la constituance prosodique et défend de manière convaincante l'idée selon laquelle la hiérarchie prosodique a un rôle fondamentalement diacritique. Il montre en particulier que la constituance prosodique est une « prophétie auto-réalisatrice » (2008 : 165) : ainsi, l'argument du non-isomorphisme de la structure syntaxique et de la structure prosodique, souvent cité dans la littérature, ne tient que si l'on envisage l'information suprasegmentale sous le prisme des domaines, et non des frontières comme c'était par exemple le cas dans SPE. En d'autres termes, il s'agit d'un artifice créé par l'analyse. Dans une tout autre perspective, Laks (1997) a lui aussi proposé une analyse critique de la constituance prosodique et a montré que de nombreux systèmes accentuels pouvaient s'analyser dans un modèle par grille (Prince 1983) et dans des réseaux dynamiques calculatoires de même nature que ceux défendus ici. Enfin, Côté (2000) a quant à elle montré que de nombreux phénomènes phonotactiques généralement attribués au constituant « syllabe » pouvaient être expliqués en termes de proéminence perceptuelle. Elle souligne que les transitions formantiques entre voyelle et consonnes, notamment les transitions de type CV, jouent un rôle déterminant dans l'identification perceptuelle des consonnes : par exemple, le fait que les consonnes de coda sont souvent neutralisées ou assimilées s'explique du fait qu'elles se trouvent dans un contexte perceptuel non saillant où elles sont souvent réalisées comme des consonnes implosives et sont par ailleurs dépourvues de transitions formantiques CV.

Ces différents travaux montrent que la constituance, syllabique et prosodique, est loin d'aller de soi. Il est néanmoins un point qui semble partagé par les approches ondulatoires et les approches syntagmatiques : le rôle de la sonorité dans l'identification des proéminences (que celles-ci soient appelées sommets, têtes ou gouverneurs). Bien qu'il n'y ait pas de consensus sur cette question, les chercheurs semblent s'accorder sur l'idée qu'il existe une forme de hiérarchie de sonorité, que celle-ci soit primitive ou dérivée. Nous adoptons ici l'échelle donnée par Larson (1993 : 58), inspirée de Jespersen, et qui est celle qui est la plus couramment admise :

(1) voyelles basses > voyelles hautes > glissantes > liquides > nasales > fricatives voisées > fricatives non voisées > occlusives voisées > occlusives non voisées

On doit à Clements (1990) d'avoir formulé un « principe de séquençage de la sonorité » (*Sonority Sequencing Principle*, SSP) qui permet, étant donnée une échelle de sonorité numérique, d'identifier de

manière algorithmique les pics et creux locaux de sonorité (et donc les syllabes). Ce principe joue un rôle fondamental dans la plupart des théories syllabiques, à tel point qu'il est devenu monnaie courante en théorie de l'optimalité (*Optimality Theory*, OT ; Prince et Smolensky 1993) de postuler une contrainte « SSP » responsable de la syllabation. Ce principe se retrouve également au cœur des réseaux calculatoires dynamiques, que nous présentons dans la section suivante.

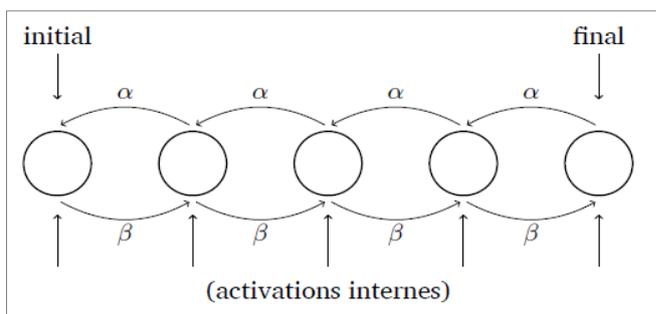
### 3 Les réseaux calculatoires dynamiques

#### 3.1 Présentation générale

Les réseaux calculatoires dynamiques (RCD) ont été développés par Goldsmith et Larson dans une série de travaux (voir en particulier Goldsmith 1992, 1993 ; Goldsmith & Larson 1990, 1993 ; Larson 1990, 1993). Ce modèle propose une architecture connexionniste originale pour l'analyse des phénomènes de syllabation et d'accentuation. L'hypothèse qui sous-tend cette approche est que le profil rythmique de la chaîne parlée est le résultat de l'interaction entre une valeur de sonorité (ou d'activation) inhérente et des dynamiques latérales de coopération/compétition (ou d'excitation/inhibition dans la littérature connexionniste) entre unités voisines. Le modèle général, tel qu'il a été articulé par Larson (1993), comporte trois niveaux : un niveau autosegmental, qui spécifie l'organisation des segments en termes de traits et détermine la sonorité des segments (niveau A) ; un niveau syllabique, qui détermine la courbe de sonorité de la chaîne (niveau S) ; un niveau métrique, qui détermine le patron accentuel de la chaîne (niveau M). Ces trois niveaux s'alimentent mutuellement ( $A \rightarrow S \rightarrow M$ )<sup>v</sup> : la dynamique du niveau A produit en sortie une valeur correspondant à la sonorité de chaque segment ; le niveau S identifie des pics de sonorité locaux et globaux, lesquels servent d'entrée au niveau M pour l'identification des prééminences accentuelles. Les niveaux M et S sont modélisés à l'aide de RCD. Le niveau A peut être modélisé de diverses manières, en fonction des hypothèses linguistiques que l'on souhaite encoder. Il s'agit du niveau qui a été le moins étudié et il est courant de représenter la sortie de ce niveau par des valeurs numériques *ad hoc* (mais voir Larson 1993 : 22-28 pour diverses suggestions). C'est d'ailleurs la stratégie qui sera adoptée ici.

Un RCD est un perceptron unicouche à inhibition bilatérale, dans lequel un input de longueur  $n$  est représenté comme un réseau de  $n$  nœuds, chaque nœud possédant deux connexions synaptiques avec ses voisins adjacents. Les connexions de droite à gauche sont contrôlées par un paramètre (poids synaptique)  $\alpha$  alors que les connexions de gauche à droite sont contrôlées par un paramètre  $\beta$  (cf. Fig. 1).

Figure 1. Réseau calculatoire dynamique



Dans un réseau syllabique, chaque unité reçoit en entrée une valeur correspondant à sa *sonorité inhérente* (qui correspond à la sortie du niveau A pour cette unité), ainsi que la sortie de ses voisins adjacents pondérés par les poids synaptiques  $\alpha$  et  $\beta$ . Elle transmet en sortie la somme des valeurs reçues en entrée, ce qui correspond à la *sonorité dérivée* de l'unité. Le réseau évolue alors dynamiquement dans le temps

jusqu'à se stabiliser dans un état d'équilibre où de nouvelles mises à jour de l'activation des unités ne modifient plus l'état global du système (en d'autres termes, la sonorité dérivée des unités n'évolue plus).

La mise à jour du système est décrite par l'équation suivante (d'après Goldsmith 1992 : 223):

$$(2) d_i^{t+1} = u_i + \alpha \cdot d_{i+1}^t + \beta \cdot d_{i-1}^t$$

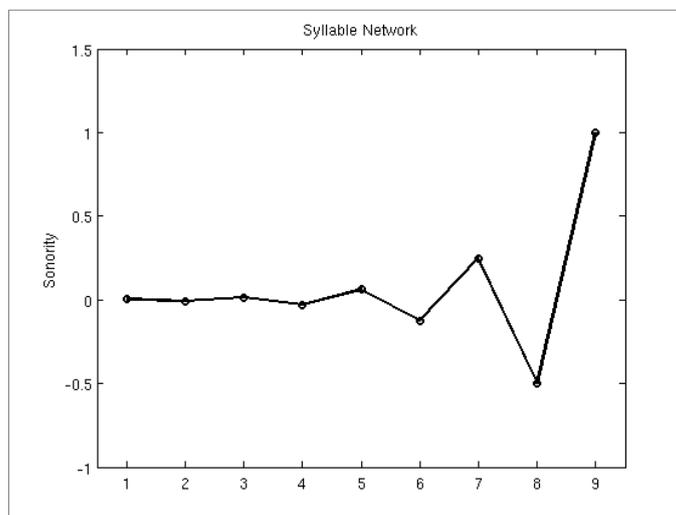
où  $u$  est la sonorité inhérente et  $d_i^t$  représente la sonorité dérivée de l'unité  $i$  à l'instant  $t$ . On remarquera que les variables et constantes du système sont additionnées linéairement. Prince (1993) a ainsi démontré que les RCD sont des approximations discrètes de modèles dynamiques linéaires (*dynamic linear models*, DLM), et les RCD sont également connus sous ce nom. Il nous semble toutefois important de garder les deux appellations distinctes : la linéarité n'est pas une propriété intrinsèque des RCD et il est tout à fait possible d'envisager des RCD non linéaires : il est par exemple possible de transformer l'activation d'une unité selon une fonction non linéaire telle que la fonction sigmoïde (très utilisée en modélisation connexionniste) ou encore d'autoriser des formes de rétropropagation entre les différents niveaux (cf. note v).

Les paramètres  $\alpha$  et  $\beta$  sont sans doute l'aspect le plus « déconcertant » de ce cadre théorique au premier abord. Ces paramètres varient de langue à langue, chaque paire de valeurs décrivant un système linguistique particulier. En pratique, différentes plages de valeurs produisent des systèmes identiques ou similaires. Cet état de faits est tout à fait comparable à la situation en théorie de l'optimalité (OT), où différentes hiérarchies de contraintes produisent les mêmes descriptions structurales. Pour poursuivre l'analogie, on peut considérer que la tâche d'un apprenant est de découvrir une paire de valeurs  $\alpha$  et  $\beta$  qui génère les chaînes de sa langue, de la même manière qu'en OT sa tâche est de découvrir une hiérarchie de contrainte. La différence fondamentale entre les deux approches est que les paramètres des RCD sont des valeurs numériques continues alors que les contraintes OT sont des symboles discrets. Néanmoins, la question de l'apprentissage demeure dans les deux cas un problème d'*optimisation* : étant donnée une fonction (complexe), le rôle du dispositif d'acquisition, qu'il soit spécifique au langage ou non, est de trouver une paramétrisation qui prédise au mieux les données.

Pour mieux comprendre le comportement d'un RCD, intéressons-nous à un cas simple. Acceptons tout d'abord de noter  $M(\alpha, \beta)$  un modèle dont les paramètres sont  $\alpha$  et  $\beta$ , et  $\mathbf{u}$  un vecteur de sonorité correspondant aux valeurs fournies en entrée au réseau. La figure 2 illustre le comportement du réseau pour le Modèle  $M(0, -0,5)$  pour un input  $\mathbf{u} = (0, 0, 0, 0, 0, 0, 0, 0, 1)$ , ce qui peut correspondre à une approximation du système accentuel du français de référence.

Comme nous pouvons le voir, le poids synaptique de l'unité finale se propage de droite à gauche dans le réseau, influençant chaque unité précédente d'un facteur polynomial  $\beta^n$ , où  $n$  est la distance entre les deux unités. La propagation de l'onde a donc pour effet de créer une alternance rythmique : la sonorité des unités paires est inhibée alors que celle des unités impaires est au contraire augmentée, la magnitude de l'influence diminuant progressivement au fur et à mesure que l'on s'éloigne de sa source. Cette alternance pic/creux correspond exactement à l'alternance syllabe accentuée/inaccentuée dans le modèle métrique et à l'alternance consonne/voyelle dans le modèle syllabique. L'un des résultats les plus significatifs des RCD est que ces propriétés (alternance rythmique et alternance CV) ne sont pas des propositions axiomatiques, comme dans les cadres logico-symboliques<sup>vi</sup>, mais découlent toutes deux de la même architecture neuronale dynamique.

Figure 2. Propagation de l'onde de sonorité pour l'input (0, 0, 0, 0, 0, 0, 0, 1), avec  $\alpha = 0$  et  $\beta = -0,5$



### 3.2 RCD et syllabation en français

Laks (1995) a appliqué avec succès ce modèle connexionniste à la syllabation en français. L'auteur y analyse de manière détaillée un corpus de 794 mots (pour 832 formes syllabifiées) représentatifs des structures syllabiques du français. Pour un certain nombre de mots, le corpus contient plusieurs syllabifications possibles, telles que des formes avec ou sans schwa final, des formes avec diérèse ou synérèse, etc. Le réseau, après apprentissage sur un sous-corpus d'entraînement, obtient une performance de 99,87%.

Ces résultats nous semblent importants dans la mesure où ils montrent que ce type d'architecture non symbolique est capable d'apprendre la structure syllabique du français sans être doté *a priori* du riche vocabulaire généralement postulé dans les théories logico-symboliques : la structure syllabique émerge littéralement de la dynamique des interactions dans le réseau. Il faut toutefois mentionner que le modèle proposé par Laks est une généralisation du modèle originel de Goldsmith et Larson, modèle que l'on appellera « standard » faute de mieux. En effet, les paramètres d'un RCD généralisé ne sont plus des valeurs scalaires  $\alpha$  et  $\beta$  contrôlant toutes les connexions du réseau (gauches et droites respectivement) mais des vecteurs  $\mathbf{a}$  et  $\mathbf{b}$  dont chaque dimension correspond à une classe naturelle de sons. Tel qu'il est proposé, le modèle reconnaît 6 classes (voyelles, glissantes, occlusives, fricatives, nasales et liquides), et les paramètres ont donc la forme  $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$  et  $\mathbf{b} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^{vii}$ , ce que l'on peut abrégé en  $M(\mathbf{a}, \mathbf{b})$ .

Plusieurs observations peuvent être faites par rapport à ce modèle. Tout d'abord, il semble que la sonorité intrinsèque des unités soit doublement encodée dans le réseau, d'une part dans la sonorité inhérente, et d'autre part dans les poids synaptiques, puisque ceux-ci sont dorénavant spécifiques à une classe de sons particulière. Ainsi, avant et après apprentissage, les voyelles ont un potentiel latéral positif, ce qui a pour effet d'inhiber la sonorité de ses voisins immédiats lorsque leur sonorité est négative, alors que les segments non syllabiques ont tous un potentiel négatif, ce qui a l'effet inverse. On notera également que le paramétrage initial (voyelles avec sonorité et potentiel latéral positif vs autres segments avec une sonorité et un potentiel latéral négatif) constitue déjà une hypothèse linguistique forte, selon laquelle les voyelles sont considérées comme des sommets privilégiés, au détriment des autres segments. Il est donc légitime de se demander quelle serait la performance d'un système d'apprentissage ne faisant aucune hypothèse sur la distribution initiale des unités. En dernier lieu, on fera remarquer que l'ajout de

dimensions introduit nécessairement davantage de degrés de liberté dans le modèle, ce qui a pour effet de le rendre à la fois plus expressif mais également moins restrictif : parce qu'il est plus parcimonieux, le modèle standard explore nécessairement un espace de recherche plus restreint, ce qui le rend computationnellement plus tractable et limite, dans une certaine mesure, les possibilités de surgénération.

Ces différentes remarques nous conduisent donc à évaluer la performance d'un RCD standard pour la syllabation du français. Sans remettre en cause la portée globale des résultats obtenus par Laks (1995), nous montrerons ici que le modèle standard est à même d'acquiescer la structure syllabique du français, et ce à partir d'hypothèses *a priori* minimales (l'existence de segments).

### 3.3 Implémentation du modèle

#### 3.3.1 Implémentation

Comme nous l'avons souligné plus haut, le problème de l'apprentissage dans un RCD consiste à découvrir une paire de paramètres  $(\alpha, \beta)$  permettant de décrire un ensemble de données observées. Pour ce faire, le réseau calcule récursivement la sonorité dérivée  $d$  de chaque unité à un instant  $t$ , jusqu'à ce que la différence de sonorité pour chaque unité entre l'instant  $t$  et  $t + 1$  soit nulle ou, en pratique, inférieure à un seuil (arbitraire)  $\delta$ , soit  $|d^t - d^{t+1}| < \delta$ . Lorsqu'il est abordé d'un point de vue computationnel, le problème peut être appréhendé de manière plus directe à l'aide de l'algèbre linéaire. Prince (1993 : 20) a en effet montré que l'équation (3) pouvait être écrite sous forme vectorisée :

$$(3) \mathbf{d} \leftarrow \mathbf{W}_n \mathbf{d} + \mathbf{u}$$

où  $\mathbf{d}$  est le vecteur de sonorité dérivée de dimension  $n$  (initialement 0),  $\mathbf{u}$  est le vecteur de sonorité inhérente et  $\mathbf{W}$  une matrice tridiagonale de dimension  $n \times n$  contenant une diagonale  $\alpha$  au dessus de la diagonale principale et une diagonale de paramètres  $\beta$  au dessous. Cette équation peut être utilement réécrite de la manière suivante :

$$(4) \mathbf{d} = (\mathbf{I} - \mathbf{W}_n)^{-1} \mathbf{u}$$

où  $\mathbf{I}$  est la matrice d'identité<sup>viii</sup>. Cette méthode directe est la forme que nous avons retenue dans notre implémentation<sup>ix</sup>, et ce pour trois raisons : elle permet de se passer du paramètre  $\delta$ , elle est computationnellement plus efficace que la méthode récursive, et elle offre une solution numériquement exacte au lieu d'une valeur approchée (dont la précision dépend de la valeur de  $\delta$ ).

Si l'on considère un RCD comme un modèle de l'acquisition phonologique, il est nécessaire de circonscrire l'espace de recherche. Il s'agit en l'occurrence des paramètres  $\alpha$  et  $\beta$  qui, étant des nombres réels, peuvent prendre théoriquement n'importe quelle valeur. Il a été formellement démontré par Prince (1993 : 49-53) que tout RCD satisfaisant la condition  $|\alpha\beta| \leq \frac{1}{4}$  était convergent, indépendamment de sa taille. Cela signifie que pour un modèle satisfaisant cette condition, il est garanti qu'il existe une solution (exacte) au problème d'apprentissage. Cela implique également qu'il existe en théorie une infinité de valeurs convergentes (Larson 1993 : 36). Néanmoins, pour des raisons pratiques, nous avons restreint la valeur de  $\alpha$  et  $\beta$  à l'intervalle  $[-0,5, 0,5]$  et nous supposons que le problème de l'apprentissage consiste à trouver un paramétrage optimal dans cet intervalle. Lors des simulations, les valeurs initiales de  $\alpha$  et  $\beta$  sont choisis aléatoirement à chaque tentative dans l'intervalle  $[-0,3, 0,3]$  selon une distribution uniforme.

Enfin, il nous faut insister sur le fait que la sortie d'un RCD produit des valeurs de sonorité continues et non pas des constituants. Pour interpréter cette onde de sonorité on utilise en général trois niveaux de syllabité : H (*high*), qui correspond aux sommets syllabiques (pics locaux de sonorité) ; L (*low*), qui correspond à l'initiale d'une syllabe ou à une consonne finale (creux local de sonorité) ; O (*other*), qui correspond aux autres segments qui ne sont ni des pics ni des creux (segments non initiaux d'une attaque ou codiques). Par exemple, le mot [sti.lo] correspond au profil LOHLH. Dans notre implémentation, la syllabation originelle du corpus a été transformée en séquences LOH de manière à la rendre directement comparable à la sortie du RCD.

### 3.3.2 Algorithme d'apprentissage

L'algorithme d'apprentissage utilisé, développé par Goldsmith et Larson (voir notamment Larson 1993 : ch. 6), est inspiré de la méthode dite de « recuit simulé »<sup>x</sup> (*simulated annealing*). Le modèle est un modèle d'apprentissage supervisé qui intègre un paramètre  $\tau$  appelé « température », initialement très élevée ( $\tau = 1.0$ ). À chaque présentation d'une nouvelle forme, la température décroît en fonction d'un taux de refroidissement  $\Delta\tau$ , indépendamment du succès ou de l'échec : ainsi, si  $\Delta\tau = 0.99$ , la température décroît d'un pourcent à chaque itération. Lorsque le système prédit une forme erronée, les paramètres  $\alpha$  et  $\beta$  sont modifiés aléatoirement d'un facteur tiré d'une loi normale de moyenne zéro et de variance  $\tau$ , normalisé par une constante  $\kappa$ . La température du système est également augmentée de la distance euclidienne du degré de modification de  $\alpha$  et  $\beta$ . Enfin, la sonorité inhérente des segments dont la sonorité dérivée est incorrectement prédite est elle aussi modifiée, d'un facteur qui dépend de la température et d'une constante de normalisation que nous noterons  $\lambda$ . Lorsque la température du système descend sous un seuil  $\theta$ , on considère que le système est « gelé » dans un état stable et la procédure d'apprentissage s'interrompt.

Ces paramètres sont ceux du dispositif d'acquisition et non du RCD. Afin de bien les distinguer de  $\alpha$  et  $\beta$ , nous les appellerons « hyperparamètres » du modèle, comme c'est parfois le cas dans la littérature. Pour plus de clarté, nous résumons ci-dessous les 5 hyperparamètres de l'algorithme d'apprentissage :

- $\tau$  : la température du système (réglée à 1.0 dans tous nos tests)
- $\Delta\tau$  : le facteur de refroidissement du système
- $\theta$  : le seuil de gel du système
- $\kappa$  : le paramètre contrôlant l'amplitude de la modification de  $\alpha$  et  $\beta$  en cas d'erreur
- $\lambda$  : le paramètre contrôlant l'amplitude de la modification de la sonorité inhérente en cas d'erreur

L'intuition qui sous-tend cet algorithme est que la magnitude du changement de  $\alpha$  et  $\beta$  dépend de la température du système : plus le système est chaud, plus le changement est potentiellement grand, et l'amplitude du changement tend asymptotiquement vers zéro au fur et à mesure que le système refroidit.

### 3.3.3 Dispositif d'acquisition biaisé et non biaisé

Au-delà de la question de l'apprentissage des paramètres, se pose la question de savoir si la sonorité inhérente est connue *a priori* ou apprise. Le choix entre les deux stratégies correspond à deux hypothèses linguistiques distinctes qu'il est utile d'explicitier.

La première approche consiste à attribuer au système une échelle de sonorité *a priori* aux unités, telle que celle donnée en (1). Dans la version forte de cette hypothèse, on peut considérer que l'échelle de sonorité est inaliénable et qu'elle ne peut par conséquent pas évoluer au cours de l'apprentissage. Au plan linguistique, cela peut traduire un positionnement rationaliste, selon lequel la sonorité des segments est déterminée par la Grammaire Universelle (GU), ou un positionnement empiriste « modéré » qui, bien que ne reconnaissant pas de GU, soutiendrait que l'échelle de sonorité traduit une réalité phonétique, qu'elle soit de nature perceptive et/ou articulatoire. Dans cette perspective, une échelle de sonorité donnée constitue une hypothèse précise et unique sur la nature de la hiérarchie de sonorité, et donc sur la GU ou sur la réalité phonétique dont la sonorité est la manifestation linguistique. Il est bien sûr possible d'assouplir cette position en considérant que le dispositif d'acquisition est bien équipé *a priori* d'une échelle de sonorité mais que celle-ci peut évoluer au cours de l'apprentissage. C'est la position adoptée par Laks (1995), ainsi que par Goldsmith et Larson dans plusieurs de leurs travaux. Un dispositif d'acquisition équipé d'une échelle de sonorité *a priori* sera dit *biaisé* en ce sens qu'il est pourvu d'hypothèses préconçues sur la nature du problème à résoudre (voir par exemple Hayes et Wilson 2008).

La deuxième approche consiste à ne faire aucune hypothèse particulière sur la nature de l'échelle de sonorité. De ce point de vue, le dispositif d'acquisition démarre avec une distribution uniforme pour

chaque segment et les valeurs de sonorité sont modifiées au cours de l'apprentissage en réponse à une erreur de prédiction : on peut alors soit légèrement modifier la sonorité (d'une valeur aléatoire), soit introduire une « dose d'intelligence » dans le système en fonction de la direction de l'erreur : par exemple, si un son est prédit comme un sommet alors qu'il ne l'est pas, sa sonorité inhérente sera abaissée d'une certaine valeur (fixe ou aléatoire), et inversement, s'il est erronément prédit comme étant un creux, sa sonorité inhérente sera légèrement augmentée (Larson 1993 : 228). Cette approche, qui fait le moins de présupposés quant à la sonorité des unités sous-jacentes<sup>xi</sup>, traduit au plan linguistique un positionnement théorique empiriste plus radical : le dispositif d'acquisition démarre *tabula rasa* et infère la sonorité des unités à partir des régularités distributionnelles observées dans les données et, dans le cas d'un apprentissage supervisé comme ici, des erreurs de prédiction entre les observables et l'hypothèse qui a généré les prédictions.

Les RCD offrent un cadre formel idéal pour tester les prédictions de ces deux types d'approche, aussi avons-nous jugé utile de les comparer. Nous présentons dans la section suivante le corpus que nous avons construit à cette fin.

## 4 Le corpus

Ce travail s'inscrit dans un projet plus large d'analyse de la phonologie du français méridional dans une perspective subsymbolique. Au lieu de nous baser sur un lexique existant de type dictionnaire, nous avons sélectionné un échantillon d'oral spontané transcrit orthographiquement. Le matériau brut provient de la banque de données du projet « Phonologie du français contemporain : usages, variétés, structure » (PFC<sup>xii</sup>) (Durand, Laks et Lyche 2002, 2009). Nous nous sommes basé sur une enquête de 10 locuteurs effectuée dans un village viticole du Languedoc (Douzens). La partie retranscrite représente environ 185 minutes de conversation (environ trois heures). À cela nous avons ajouté les mots de la liste de mots standard de PFC ainsi que du texte PFC, lesquels ont été conçus pour l'étude de divers phénomènes phonologiques (et phonotactiques en particulier). Le corpus de départ contenait 46 682 tokens, correspondant à 3 314 types. La liste de types brute a ensuite été nettoyée pour éliminer le bruit (notamment les variantes mal épelées d'un type existant et les noms de famille). Notre étude portant sur la syllabation, nous avons jugé utile de ne garder qu'une seule forme pour les homophones dont le lemme était identique, comme dans les paires *venu* ~ *venues* et *arrivé* ~ *arrivait*. Nous avons en revanche décidé de conserver les homonymes non homographes (*jeune* ~ *jeûne*, *et* ~ *est*).

La liste finale contient 2 530 mots. Nous l'avons ensuite transcrite en API (encodage Unicode) dans une transcription large, puis syllabifiée. La transcription phonétique suit la prononciation générale du français méridional. Nous avons considéré que le *e* graphique final correspondait systématiquement à un schwa. Les *e* en syllabe initiale ainsi que ceux pour lesquels il est difficile de dire s'il s'agit d'un schwa ou d'une voyelle stable (ex : *allemand*, *patelin*) ont été distingués dans la transcription mais sont considérés dans cette étude comme des /E/ stables (cf. Durand, Slater et Wise 1987). Les « voyelles nasales » ont été analysées comme des séquences voyelle orale + appendice nasal (voir Durand 1988 pour des arguments à l'appui de cette analyse), l'appendice étant transcrit [N]. Du point de vue de la syllabification, nous avons suivi un découpage syllabique tel qu'il est guidé par l'application de la loi de position (Durand 1995, Rizzolo 2002). Ainsi, le mot *exemple* a été transcrit [eg.zaŋ.plə], puisque le premier /E/ est normalement réalisé mi-ouvert et que le /g/ est par conséquent analysé comme une coda. Enfin, notons que les groupes /r+j/ ont été traités comme des groupes tautosyllabiques (ex : *intérieur* [ɛN.te.rjœr]), bien que la réalisation où les consonnes sont hétérosyllabiques soit très répandue (ex : [ɛN.te.r.jœr]), notamment chez les jeunes locuteurs.

Enfin, pour les besoins de ce travail, nous avons exclu les mots contenant des séquences VV : leur traitement dans les RCD, surtout dans le cas d'un apprenant non biaisé, constitue une question de recherche ouverte, bien que plusieurs pistes aient été suggérées (voir Larson 1993 : 99–102 ainsi que Laks 1995 pour le français).

Le corpus final pour ce travail contient 2 486 formes phonétisées et syllabifiées. Nous l'avons divisé, après randomisation, en un corpus d'entraînement contenant 60% des données (1492 mots), un corpus de validation et un corpus de test contenant chacun 20% (497 mots). Nous donnons en (5) un court extrait du corpus global.

(5)	<i>génial</i>	ʒe.njal	CV.CGVC
	<i>généalogiques</i>	ʒe.ne.a.lo.ʒi.kə	CV.CV.V.CV.CV.CV
	<i>général</i>	ʒe.ne.ral	CV.CV.CVC
	<i>générale</i>	ʒe.ne.ra.lə	CV.CV.CV.CV
	<i>génération</i>	ʒe.ne.ra.sjɔ̃	CV.CV.CV.CGVN
	<i>génétiq</i>	ʒe.ne.ti.kə	CV.CV.CV.CV

Tournons-nous maintenant vers les résultats obtenus à partir de ce corpus pour un dispositif d'acquisition biaisé et un dispositif non biaisé.

## 5 Résultats et discussion

### 5.1 Réglage du dispositif d'acquisition

Comme pour tout problème d'apprentissage supervisé, les résultats obtenus dépendent grandement de la paramétrisation du module d'apprentissage, lequel dépend à son tour de la nature du problème et des données : les réglages optimaux doivent donc être déterminés de manière empirique. Afin de limiter le problème de surapprentissage, nous avons réglé les hyperparamètres du modèle en nous basant sur la performance obtenue (minimale, maximale et moyenne) sur le corpus de validation sur des séries de 100 tentatives.

Nous avons conservé pour les paramètres  $\kappa$  et  $\lambda$  les valeurs déterminées par Larson (1993), à savoir 0,2 et 0,125 respectivement. Les meilleures performances ont été obtenues, moyennant un temps de traitement plus long, avec un taux de déclin très lent ( $\Delta\tau = 0,995$ ) et un seuil de refroidissement très bas ( $\theta = 0,001$ ). La température, comme nous l'avons déjà mentionné, a toujours été réglée à 1,0.

Du point de vue de la sonorité inhérente dans le cas d'un apprenant biaisé, nous avons adopté l'échelle suivante, inspirée de (1) :

(6)	occlusives non voisées	1
	occlusives voisées	2
	fricatives non voisées	3
	fricatives voisées	4
	occlusives nasales	5
	liquides	6
	glissantes	7
	schwa	8
	voyelles hautes	9
	voyelles moyennes	11

Nous avons ignoré le débat qui entoure la nature phonologique du /r/ et celui-ci a été considéré comme une véritable liquide, ce qui est au demeurant valable pour plusieurs des locuteurs du corpus dont les données sont tirées. L'appendice nasal a été considéré comme une glissante (nasale), ce qui nous semble en accord avec la représentation phonologique que lui attribue Durand (1988) en phonologie de dépendance, à savoir le seul élément [N] (ce qui correspond au trait unaire [nasal] dans un cadre géométrique). Dans la mesure où il n'y a pas de contraste phonologique entre les voyelles mi-ouvertes et mi-fermées (cf. Durand 1995), nous avons décidé de leur attribuer une sonorité sous-jacente identique (ceci n'a pas d'incidence sur la question de la syllabation).

Le dispositif d'apprentissage étant stochastique, les résultats varient d'une tentative à l'autre. Pour cette raison, nous avons mesuré deux séries de 1000 tentatives sur le corpus d'entraînement et de validation avec les paramètres optimaux rapportés ci-dessus, l'une pour l'apprentissage biaisé et l'autre pour l'apprentissage non biaisé. Dans les deux cas, nous avons construit un « système moyen » à partir des valeurs obtenues pour  $\alpha$ ,  $\beta$  et le vecteur de sonorité inhérente. Ce système moyen est celui qui nous a semblé refléter le mieux la performance globale du système, et c'est celui-là que nous présenterons dans les deux cas.

## 5.2 Performance du dispositif d'acquisition biaisé

La performance du système est globalement très bonne : la moyenne sur le corpus de validation pour 1000 tentatives se situe à 98,98%, avec un minimum de 96,78% et un maximum de 100%. Il converge en moyenne au bout de 4,8 présentations du lexique (ou époques).

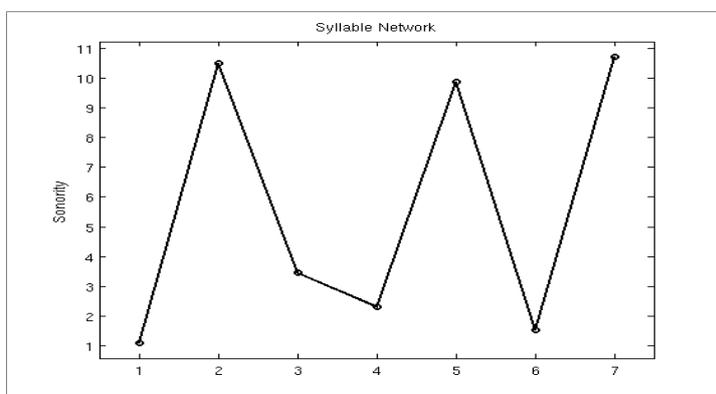
Le modèle se stabilise en moyenne avec  $\alpha = 0,210659$  et  $\beta = -0,015455^{xiii}$ . Les valeurs de sonorité inhérente moyennes après apprentissage (arrondies au centième) sont les suivantes (comparer à (6)):

(7)	a : 14,33	ɔ : 12,48	ɛ : 12,26	e : 12,15	o : 11,33
	œ : 11,20	ø : 11,11	ə : 10,41	i : 10,33	u : 9,59
	y : 9,41	w : 7,75	ɥ : 7,06	j : 7,05	ɳ : 6,36
	r : 5,48	l : 5,24	ɲ : 4,90	n : 3,99	ʒ : 3,69
	z : 3,47	m : 3,46	v : 3,06	ʃ : 2,79	f : 2,42
	g : 1,92	k : 1,73	b : 1,50	s : 1,27	d : 1,24
	p : 0,54	t : -0,40			

La sonorité du /s/ est ici particulièrement remarquable : avec une valeur de 1,27, il fait partie des segments les moins sonores avec /p/, /t/ et /d/, et se situe nettement en dessous des autres fricatives, qui sont comparativement plus proche de leur valeur de départ. Ce résultat mérite d'être souligné dans la mesure où la nature phonologique du /s/ fait débat et que son statut est loin d'être tranché, quelle que soit la perspective théorique que l'on adopte<sup>xiv</sup>. À l'autre bout du spectre, la sonorité du /a/, initialisée à 13,0 est élevée à 14,33 après apprentissage. La sonorité du schwa, de manière plutôt surprenante, a été nettement augmentée et est même supérieure à celle de /i/, /y/ et /u/ : si pour les voyelles la sonorité est corrélée avec l'aperture, la sonorité obtenue rapproche le schwa des voyelles moyennes, ce qui correspond effectivement à sa description articulaire. Il est cependant clair que cette voyelle doit être étudiée à part en tenant compte de l'interaction entre le niveau syllabique et le niveau métrique, chose que nous n'avons pas faite dans cette étude. La sonorité de l'appendice nasal est elle aussi intéressante puisqu'elle demeure très proche de sa valeur initiale (7,0), et reste voisine des glissantes (nous verrons que c'est également le cas pour l'apprentissage non biaisé).

Pour illustrer la dynamique à l'œuvre dans le réseau, prenons la forme discute /diskytə/, dont la représentation sous-jacente est encodée avec le vecteur  $\mathbf{u} = (1.24, 10.33, 1.27, 1.73, 9.41, -0.40, 10.41)$ . Pour rappel, la sonorité inhérente apprise de /s/ < /k/ : sur la seule base de la sonorité inhérente, on prédirait la forme \*[di.sky.tə]. Comme on peut le voir à la Fig. 3, le réseau a prédit la syllabation correcte et a attribué une sonorité dérivée plus élevée à [s] ( $d_s = 3,45$ , unité 3), qu'à [t] ( $d_k = 2,30$ , unité k). Le vecteur de sonorité du mot entier est  $\mathbf{d} = (1.08, 10.50, 3.45, 2.30, 9.87, 1.52, 10.73)$ .

Figure 3. Sonorité dérivée de *discute*



Sur le corpus de validation, le modèle biaisé moyen obtient une performance de 99,80% (1 seule erreur). Cette (excellente) performance n'est pas surprenante puisque les hyperparamètres ont été sélectionnés sur la base de ce corpus. Pour évaluer la capacité de généralisation du modèle de manière équitable, il est donc nécessaire de le faire sur le corpus de test, puisque celui-ci n'a jamais été vu pendant la période d'entraînement et de mise au point. Sur ce corpus, le modèle moyen obtient une performance de 98,59%, soit 7 erreurs. Pour pleinement apprécier cette performance, il est utile de la comparer à celle d'un modèle où la sonorité est prédite uniquement sur la base du seul principe de séquençage de la sonorité, ce qui correspond au modèle  $M(0,0)$  où  $\alpha$  et  $\beta$  sont nuls. Ce modèle obtient une performance de 93,56% sur le corpus de test avec l'échelle de sonorité que nous avons adoptée. Il est donc clair que le RCD obtient une performance très supérieure (d'environ 5%) à ce point de repère. Ces chiffres confirment, si besoin est, que la conception dynamique de la construction de la sonorité, sur la base d'*interactions locales* et évolutives, conception qui est au cœur des réseaux calculatoires dynamiques, a un pouvoir prédictif supérieur à un modèle statique dans lequel la sonorité des unités a une valeur figée.

### 5.3 Performance du dispositif d'acquisition non biaisé

Le modèle non biaisé a été construit avec une sonorité inhérente de départ uniforme, réglée arbitrairement à 1 pour toutes les unités. Sur une série de 1000 présentations, le modèle obtient une performance moyenne sur le corpus de validation de 99,13%, avec un minimum de 97,38% et un maximum de 100%. La performance moyenne sur le corpus d'entraînement est de 99,42%. Sans surprise, ce modèle est computationnellement plus exigeant que le modèle biaisé : il converge en moyenne au bout de 7,8 époques.

Nous voyons que la performance moyenne sur le corpus de validation est supérieure à celle obtenue avec le modèle biaisé. Un examen visuel de la densité de la distribution des taux de réussite sur les deux modèles fait apparaître que les deux distributions sont asymétriques et ne sont donc pas normalement

distribuées (elles sont plutôt étalées vers la droite, vers les 100%). Ceci est confirmé par un test de Shapiro-Wilk pour la normalité ( $p < 0,001$  dans les deux cas). Nous avons donc comparé les deux distributions à l'aide d'un test de Wilcoxon pour la médiane, lequel met en évidence le fait que la médiane du modèle non biaisé est significativement supérieure à celle du modèle biaisé ( $p < 0,001$ ).

Il ressort de ces résultats que le modèle non biaisé est, toutes choses égales par ailleurs, plus performant que le modèle biaisé sur le corpus de validation. Sa capacité de généralisation sur le corpus de test est très bonne et légèrement supérieure à celle du modèle biaisé, à une forme près (98,79%, soit 6 erreurs). Nous pouvons donc affirmer que, même sous l'interprétation la plus conservatrice, une échelle de sonorité donnée *a priori* ne favorise pas l'apprentissage de la structure syllabique du corpus. C'est donc bien l'architecture du réseau qui explique à elle seule l'apprentissage qui a lieu, et ce sans nécessiter aucun savoir linguistique préétabli quant à la sonorité inhérente des unités. À nouveau, il nous faut rappeler qu'un algorithme basé sur la seule échelle de sonorité ne peut prédire que 93,56% des formes de ce corpus ; or c'est bien ce type d'algorithme qui sous-tend la quasi-totalité des approches symboliques.

Passons maintenant aux valeurs apprises par le modèle non biaisé moyen. Les paramètres moyens qui émergent sont  $\alpha = 0,204843$  et  $\beta = -0,020909$ , ce qui est très proche des valeurs obtenues avec le modèle biaisé (voir aussi note xiii). Les valeurs de sonorité moyennes après apprentissage sont données en (8) :

(8)	a : 4,58	ɔ : 3,99	ɛ : 3,84	ə : 3,76	e : 3,62
	i : 3,33	o : 2,71	u : 2,67	y : 2,45	œ : 2,44
	ø : 2,05	w : 1,98	ɥ : 1,51	j : 1,39	ɳ : 1,12
	r : 0,72	ʀ : 0,70	l : 0,60	ʃ : -0,01	ʒ : -0,06
	z : -0,31	n : -0,57	g : -0,59	f : -0,81	v : -0,86
	k : -0,87	m : -1,02	b : -1,15	s : -1,28	d : -1,37
	p : -1,50	t : -2,13			

Bien que les valeurs numériques soient différentes, la hiérarchie de sonorité apprise est très semblable à celle obtenue en (7), mais sans aucun présupposé quant à la sonorité relative des unités. Comme pour le modèle biaisé, la voyelle /a/ émerge comme la plus sonore et, en bas de l'échelle, nous rencontrons les plosives bilabiales et alvéolaires ainsi que le /s/. L'appendice nasal a une sonorité proche de celle des glissantes, comme nous l'avions déjà observé pour le modèle biaisé. Le fait que ce regroupement soit dans ce cas complètement émergent nous conforte dans l'idée qu'il est correct, d'un point de vue phonologique, de traiter cet appendice comme une glissante. Sa sonorité est en tout cas très supérieure aux autres nasales, en particulier /n/ et /m/. Le fait le plus remarquable qui ressort de ce modèle moyen est qu'il y a une séparation nette entre les obstruantes et les sonantes (y compris les voyelles), la ligne de départ étant 0 : les sonantes ont toutes une sonorité positive, alors que les obstruantes ont toutes une sonorité négative, fût-elle quasi-nulle comme dans le cas du /ʃ/.

#### 5.4 Analyse qualitative des erreurs

Comme nous l'avons dit, les modèles non biaisé et biaisé n'échouent respectivement que sur 6 et 7 des 497 formes du corpus de test. De manière intéressante, les 6 formes erronément prédites sont partagées par les deux modèles, ce qui n'a rien de surprenant dans la mesure où les paramètres et l'échelle de sonorité sur lesquels ils convergent sont comparables. Nous donnons en (9) les 7 formes, la dernière n'étant incorrectement prédite que par le modèle biaisé.

(9)	<b>mot</b>	<b>syllabation réelle</b>	<b>syllabation prédite</b>
	<i>islamique</i>	is.la.mi.kə	i.sla.mi.kə

<i>patchwork</i>	patʃ.wœrk	pat.ʃwœrk
<i>spécialité</i>	spe.sja.li.te	s.pe.sja.li.te
<i>spécialisé</i>	spe.sja.li.ze	s.pe.sja.li.ze
<i>station</i>	sta.sjɔ̃N	s.ta.sjɔ̃N
<i>stage</i>	sta.ʒə	s.ta.ʒə
<i>technologique</i>	tek.no.lo.ʒi.kə	te.kno.lo.ʒi.kə

Ces formes appellent plusieurs remarques. Tout d'abord, l'erreur de prédiction pour le mot *islamique* illustre le cas d'un mot (non natif) où un groupe consonantique dont l'initiale est /s/ a été mal séquencé. Or, il s'agit précisément d'un site où le rattachement syllabique du /s/ est loin d'être clair et son statut continue de faire débat, et il n'est pas du tout certain que ce soit systématiquement le cas étant donnée la tendance générale à l'enchaînement (et donc à la maximisation des attaques) en français (Encrevé 1988). On relèvera par ailleurs qu'il n'y avait qu'un seul autre mot contenant un groupe /sl/ dans le corpus d'entraînement (et dans tout le corpus), à savoir *slip*, et le groupe est dans ce cas initial.

Dans le cas du mot *patchwork*, la fricative<sup>xv</sup>, qui est suivie d'une glissante, est reconnue comme un creux local de sonorité. Il faut néanmoins souligner qu'il s'agit là d'un mot d'emprunt et d'un hapax dans le corpus : le fait que le modèle échoue à le syllabifier doit donc plutôt être interprété comme un signe de ce que ce mot diverge des régularités phonotactiques de la langue (telles qu'elles sont manifestées dans le corpus d'apprentissage).

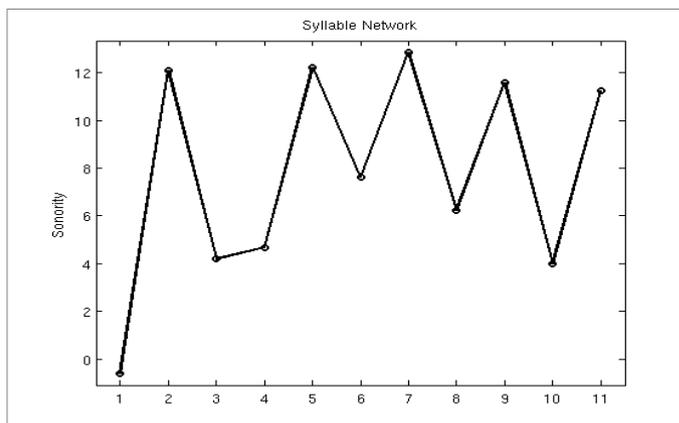
Les mots *spécialisé*, *spécialité*, *station* et *stage* présentent tous le même problème de prédiction, à savoir que le /s/ d'un groupe /sC/ initial a été identifié comme un sommet syllabique. On notera que le mot *scolarité*, présent dans le corpus de test, a été correctement prédit par les deux modèles. La différence s'explique par la sonorité relative de la consonne qui suit le /s/ : /p/ et /t/ sont les segments les moins sonores dans les deux modèles, alors que la sonorité inhérente de /k/ est plus élevée que celle de /s/. Signalons que la prédiction des deux modèles se rapproche de certaines analyses phonologiques : par exemple, en phonologie du gouvernement, ce segment est considéré comme la coda d'une syllabe dont le noyau est vide, ce noyau étant licencié « par magie » (Kaye 1992). La prosthèse romane dans les mots latins commençant par un groupe /sC/ (*schola* > *\*iscola* > fr. *école*) est un argument souvent avancé à l'appui d'une telle analyse, et la prédiction du DCN pourrait également être interprétée comme allant dans ce sens.

Enfin, le mot *technologique* illustre un cas où un groupe « obstruante non voisée + nasale » a été reconnu comme un groupe de sonorité dérivée croissante, correspondant encore une fois à la maximisation de l'attaque. L'analyse « hétérosyllabique » de ces groupes consonantiques est justifiée par le fait que la voyelle moyenne précédant le /k/ est mi-ouverte ([teknoloʒikə] et non \*[teknoloʒikə] ce qui, en vertu de la loi de position, suggère que le /k/ est en coda), et par le fait qu'il n'existe pas en français de groupe d'attaque \*/kn/ en début de mot. Il faut noter que les groupes « obstruante non voisée + nasale » non dérivés sont très rares en français du Midi et proviennent de mots savants généralement issus du grecs (cf. *rythme*, également présent dans le corpus, ou encore *acmé*, *apnée*). De manière intéressante, nous avons observé à plusieurs reprises le mot *technique* prononcé [teʒnikə] par des locuteurs ayant un système méridional conservateur. Cela pourrait suggérer que la prononciation avec un groupe hétérosyllabique (la plus répandue, et celle qui est attestée dans notre corpus) n'est pas la seule possible et qu'il y a donc variation pour ces groupes consonantiques. Quoi qu'il en soit, l'erreur de prédiction, qui est propre au modèle biaisé, s'explique aisément : avec une valeur de  $\alpha$  relativement élevée ( $\approx 0.21$ ) et une valeur de  $\beta$  quasi-nulle, il tend à laisser la sonorité du segment à gauche du sommet inchangée et à augmenter la sonorité du segment à sa droite, le rendant plus susceptible d'apparaître en « coda ». Le modèle prédit la sonorité dérivée  $\mathbf{d} = (-0.58, 12.07, 4.20, 4.69, 12.20, 7.61, 12.83, 6.22, 11.58, 3.99, 11.25)$ , illustrée à la

Fig. 4. Bien que le sommet [ɛ] « attire » à lui le [k], cela ne suffit pas à combler la différence de sonorité sous-jacente entre /k/ et /n/ : il en résulte que ces deux segments ont une sonorité dérivée très proche mais que le [k] est identifié comme un creux local de sonorité.

On mentionnera ici qu'il est possible, et sans doute nécessaire, de permettre au modèle d'identifier des plateaux de sonorité. Une stratégie possible serait de considérer que des segments successifs qui excèdent ou qui se situent en dessous d'un certain seuil de sonorité doivent être considérés comme un plateau (Larson 1993 : 108). C'est la stratégie adoptée par Laks (1995) pour l'identification des séquences de sommets VV dans des mots du type *aorte*. Le même raisonnement peut être appliqué à l'identification des creux, auquel cas deux segments successifs ayant une sonorité inférieure à un seuil minimal seraient considérés comme un plateau. Une stratégie alternative serait de considérer qu'un segment est considéré comme un sommet ou un creux si et seulement si la valeur absolue de la différence entre sa sonorité dérivée et celle de ses voisins immédiats excède un certain seuil. Ce seuil est nécessairement dépendant des échelles numériques sur lesquelles sont exprimées la sonorité inhérente d'une part, et les coefficients  $\alpha$  et  $\beta$  d'autre part. Nous laissons cette question ouverte pour des recherches ultérieures.

Figure 4. Sonorité dérivée de *technologique*



## 6 Conclusion

Dans cet article, nous avons défendu une conception connexionniste dynamique de la syllabe, dans la lignée des travaux de Goldsmith, Larson et Laks. Après avoir exposé le modèle « standard » développé par Goldsmith et Larson et le modèle « élargi » de Laks, nous avons montré que le modèle standard, plus restrictif, est tout à fait capable d'apprendre la structure syllabique du français, étant donné un algorithme d'apprentissage et une paramétrisation adéquats. Sur la base d'un corpus de 2 486 formes de français méridional, construit à partir d'une enquête en Languedoc issue de la base de données PFC, nous avons mesuré et analysé la performance de deux modèles : un dispositif d'acquisition biaisé, qui dispose d'une échelle de sonorité préétablie comparable à celles qui sont utilisées dans la littérature, et un dispositif non biaisé, qui ne fait aucune hypothèse quant à la sonorité sous-jacente des segments. Nous avons montré que les deux modèles ont tous les deux une très bonne capacité de généralisation, puisqu'ils atteignent respectivement 98,59% et 98,79% sur des données non vues. Cette performance est très supérieure à celle d'un modèle statique reposant sur une simple échelle de sonorité, lequel ne prédit que 93,56% de formes correctes sur le corpus de test. Nous avons également montré que la performance du modèle non biaisé était significativement supérieure à celle du modèle biaisé à partir de 1000 simulations sur le corpus de validation. Nous en avons conclu que l'échelle de sonorité n'est en rien nécessaire pour prédire la structure syllabique dans un RCD.

L'ensemble de ces résultats nous semble offrir des perspectives intéressantes : tout d'abord, et bien que nos résultats ne soient pas directement comparables à ceux de Laks (1995), ils montrent qu'un RCD

standard, qui n'est contrôlé que par les paramètres  $\alpha$  et  $\beta$ , offre des performances très satisfaisantes pour l'analyse de la syllabation, et ce sur un corpus authentique construit à partir d'un échantillon de parole spontanée. Étant un modèle plus restrictif que sa variante généralisée, il nous semble qu'il serait intéressant d'explorer les propriétés du RCD standard plus avant. Deuxièmement, ces résultats font écho aux travaux critiques vis-à-vis d'une analyse syntagmatique et statique de la syllabe et montrent que la structure syllabique peut s'analyser comme un phénomène émergent de la dynamique d'un réseau connexionniste. Enfin, la performance du dispositif d'acquisition non biaisé peut être recontextualisée dans le cadre du débat sur la richesse de l'état initial : il existe un nombre croissant de travaux qui suggèrent que l'apprenant n'est pas doté d'un « instinct du langage » (Pinker 1994), et que la langue est le produit de capacités cognitives très générales telles que les capacités de catégorisation, d'inférence et de symbolisation (Langacker 2008 ; Tomasello 2003). Les résultats obtenus fournissent un point de repère utile sur ce qu'il est possible d'apprendre sur la syllabation sans aucune information sur la sonorité. Les RCD offrent ici un cadre formellement simple et élégant, à forte valeur explicative et d'une grande parcimonie conceptuelle.

Pour autant, il est clair qu'il s'agit d'un programme de recherche ouvert et nous voudrions mentionner les aspects qui nous semblent les plus cruciaux pour des travaux futurs. Parmi les questions à résoudre, la question de l'analyse des plateaux de sonorité, qui n'a pas été traitée ici, figure en ligne de front. C'est un problème particulièrement délicat, surtout pour un modèle non biaisé, mais il est nécessaire de le résoudre et, dans la mesure du possible, d'une manière raisonnée et non *ad hoc*. La mise au point d'algorithmes d'apprentissage non paramétriques est un autre domaine d'investigation qui nous semble important. Le modèle de recuit simulé que nous avons utilisé est très performant en pratique, mais il est d'un réalisme cognitif douteux. Enfin, si la performance du modèle non biaisé est très stimulante, il serait faux de dire que le modèle ne dispose d'aucune information linguistique. En effet, chaque nœud d'un RCD syllabique encode un segment, ce qui est déjà un découpage de la chaîne parlée en unités et *ipso facto* une hypothèse linguistique forte sur la structure phonologique de la langue. Ici, les systèmes dynamiques (linéaires et non linéaires) offrent sans doute les outils mathématiques et conceptuels adéquats pour véritablement passer du domaine du discret au continu.

## Références bibliographiques

- Angoujard, J.-P., (1997). *Théorie de la syllabe. Rythme et qualité*, Paris: CNRS Editions.
- Angoujard, J.-P., (2006). *Phonologie déclarative*, Paris: CNRS Editions.
- Blevins, J. (1995). The Syllable in Phonological Theory. In J. Goldsmith (éd.). *The Handbook of Phonological Theory*, Cambridge : Blackwell Publishing, 206-244
- Bechtel, W. et A. Abrahamsen (2002). *Connectionism and the Mind. Parallel Processing, Dynamics, and Evolution in Networks*, Malden/Oxford : Blackwell Publishing.
- Chomsky, N. et M. Halle (1968). *The Sound Pattern of English*, Cambridge, Mass : MIT Press.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston et M. Beckman (éds), *Papers in Laboratory Phonology I*, Cambridge : Cambridge University Press, 283–333.
- Côté, M.-H., (2000). *Consonant Cluster Phonotactics: a perceptual approach*. Thèse de doctorat non publiée, MIT.
- 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 et Ch. Lyche (2002). La phonologie du français contemporain: usages, variétés et structure. In: C. Pusch & W. Raible (éds.) *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, 93–106.
- Durand, J., B. Laks et Ch. Lyche (2009). Le projet PFC : une source de données primaires structurées. In J. Durand, B. Laks et Ch. Lyche (éds) *Phonologie, variation et accents du français*. Paris: Hermès, 19–61.

- Durand, J., C. Slater et H. Wise, (1987). Observations on schwa in Southern French. *Linguistics*, 25(2), 983–1004.
- Encrevé, P. (1988). *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Paris : Éditions du Seuil.
- Eychenne, J. (2006). Aspects de la phonologie du schwa dans le français contemporain. Optimalité, visibilité prosodique, gradience. Thèse de doctorat non publiée, Université de Toulouse-Le Mirail.
- Goldsmith, J. (1990). *Autosegmental & Metrical Phonology*. Oxford : Blackwell.
- Goldsmith, J. (1992). Local Modeling in Phonology. In S. Davis (éd.) *Connectionism: Theory and Practice*, Oxford: Oxford University Press, 229–246.
- Goldsmith, J. (1993). Harmonic Phonology. In J. Goldsmith (éd.) *The Last Phonological Rule*, Chicago et Londres : The University of Chicago Press, 21–60.
- Goldsmith, J. et G. Larson. 1990. Local modeling and syllabification. *Papers from the 26th Annual Meeting of the Chicago Linguistic Society: Parasession on the syllable in phonetics and phonology*.
- Goldsmith, J. et G. Larson. 1993. Using Networks in a Harmonic Phonology. With Gary Larson. In *Papers from the 28th Annual Meeting of the Chicago Linguistic Society*, vol. 2. Chicago: Chicago Linguistics Society, 94–125.
- Goldwater, Sh. (2006). Nonparametric Bayesian Models of Lexical Acquisition, thèse de doctorat non publiée, Brown University.
- Hannahs, S., (1995). *Prosodic Structure and French Morphophonology*, Tübingen : Niemeyer.
- Hayes, B. et C. Wilson (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning, *Linguistic Inquiry* 39(3), 379–440.
- Kaye, Jonathan (1992). Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics* 2, 293–3013.
- Klein, M., (1993). La syllabe comme interface de la production et de la réception phoniques. In B. Laks, B. et M. Plénat (éds.). *De natura sonorum. Essais de phonologie*, Saint-Denis: Presses Universitaires de Vincennes, 101–142.
- Laks, B. (1995). A Connectionist Account of French Syllabification, *Lingua* 95, 51–76.
- Laks, B. (1996). *Langage et cognition : l'approche connexioniste*. Paris : Hermès.
- Laks, B. (1997). *Phonologie accentuelle. Métrique, autosegmentalité, constituance*. Paris : CNRS Éditions.
- Langacker, R. (2008). *Cognitive Grammar. A Basic Introduction*. New York : Oxford University Press.
- 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, thèse de doctorat non publiée, University of Chicago.
- Marr, D. (1982). *Vision: A Computational Approach*. San Francisco : Freeman & Co.
- McClelland J., D. Rumelhart (éds.) (1986). *Parallel Distributed Processing. Explorations in the microstructure of Cognition. Volume 1 : Foundations*. Cambridge et Londres : The MIT Press.
- McClelland J., D. Rumelhart (éds.) (1986). *Parallel Distributed Processing. Explorations in the microstructure of Cognition. Volume 2 : Psychological and Biological Models*. Cambridge et Londres : The MIT Press.
- McLeod, P., K. Plunkett et E. Rolls (1998). *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford : Oxford University Press.
- Pinker, S. (1994). *The Language Instinct. How the Mind Creates Language*. Harper Perennial Modern Classics.
- 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, Rapport technique n°1, Rutgers University Center for Cognitive Science.

- Prince, A. et 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. Thèse de doctorat non publiée, Université de Nice-Sophia Antipolis.
- Scheer, T. (2008). Why the Prosodic Hierarchy is a Diacritic and Why the Interface must be Direct, *Sounds of Silence. Empty Elements in Syntax and Phonology*, JM Hartmann, V. Hegedüs & H. van Riemsdijk, Elsevier, Amsterdam.
- Smolensky, P. (1988) On the Proper Treatment of Connectionism. *Brain and Behavioral Sciences* 11, 1–74.
- Tomasello, M. (2003). *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, Mass & Londres : Harvard University Press.

---

<sup>i</sup> Le connexionnisme a connu dans les années 80 un nouvel élan avec la publication du monumental *Parallel Distributed Processing* (McClelland, Rumerhart 1986a,b). Pour une introduction au connexionnisme, voir par exemple McLeod, Plunkett et Rolls (1998) ainsi que Betchel et Abrahamsen (2002).

<sup>ii</sup> Smolensky (1988) appelle ce niveau le niveau « sous-conceptuel » (*subconceptual*).

<sup>iii</sup> Cité par Goldwater (2006 : 8).

<sup>iv</sup> cf. *The Sound Pattern of English* (Chosmsky & Halle 1968)

<sup>v</sup> Notons que des formes de rétropropagation entre les niveaux sont tout à fait possibles, selon le schéma A  $\leftrightarrow S \leftrightarrow M$  (Larson 1993 : 39-40). De telles rétroactions introduisent toutefois des non-linéarités dans le modèle et leur étude dépasse le cadre de cet article.

<sup>vi</sup> cf. les contraintes du type \*Clash « pénaliser les syllabes accentuées successives », \*VV « pénaliser les voyelles (syllabiques) successives » en théorie de l'optimalité.

<sup>vii</sup> Rien n'exclut en théorie de spécialiser les paramètres plus avant, par exemple en distinguant davantage de classes (comme nous le faisons pour la sonorité inhérente) voire en les individualisant complètement, de sorte que chaque segment  $s$  ait ses propres paramètres  $\alpha_s \beta_s$ .

<sup>viii</sup> La matrice identité est une matrice de dimension  $n \times n$  contenant des 1 sur la diagonale principale et des 0 sur toutes les autres diagonales. La matrice identité est aux matrices ce que 1 est aux nombres réels, toutes choses égales par ailleurs.

<sup>ix</sup> Le modèle a été implanté dans le Python, à l'aide de la bibliothèque de calcul numérique Numpy (<http://numpy.scipy.org>).

<sup>x</sup> Ce terme est emprunté au domaine de la métallurgie.

<sup>xi</sup> D'un point de vue informationnel, la distribution uniforme est celle qui a l'entropie la plus élevée, c'est-à-dire le plus haut degré d'incertitude.

<sup>xii</sup> Voir <http://www.projet-pfc.net>

<sup>xiii</sup> Dans tous les modèles ayant une bonne performance que nous avons examinés,  $\alpha$  a toujours une valeur positive comprise entre 0,20 et 0,35 et  $\beta$  une valeur (positive ou négative) très proche de zéro.

<sup>xiv</sup> On se souviendra à cet égard du « licenciement magique » de Kaye (1992) destiné à rendre compte du comportement de cette consonne dans le cadre de la phonologie du gouvernement.

<sup>xv</sup> Eychenne (2006) présente des arguments pour une analyse de [tʃ] en une séquence bi-phonématique en français méridional.