

M. Klein et B. Laks

C.N.R.S. et Paris VIII

**ABSTRACT**

In this paper we present the first step of a research program. Our purpose is to evaluate and compare production systems with connexionist machines. The research focuses on phonic syllabation in relation with the graphic parsing of isolated French words. A first implementation of phonic/graphic hyphenation is presented.

**1. PROBLEMATIQUE DE LA SYLLABE**

On sait qu'il n'existe pas un indice phonétique clair et parfaitement individualisé de la discontinuité syllabique. Alors que la plupart des modèles phonologiques récents font, à différent niveaux d'analyse, centralement appel à une notion de découpe syllabique, la phonétique propose quant à elle une liste de paramètres partiellement redondants et partiellement contradictoires, dont aucun pris isolément ne livre cette discontinuité (degré d'aperture, tension, courbe de sonorité, *chest pulse*, implosion/explosion, etc.). Il reste que si les locuteurs peuvent faire montre d'une intuition de découpe syllabique d'une part, et que si les modèles d'explication phonologique démontrent la pertinence de la syllabe d'autre part, on peut se fixer pour objectif d'explicitier, tant en production qu'en réception, les liens entre indices phonétiques de la syllabation et structures phonologiques postulées. Les difficultés évoquées étant pour partie liées au caractère continu du signal et à la multiplicité d'indices contradictoires, les modèles connexionnistes constituent, à priori, de bons candidats pour expliciter le lien phonétique/phonologie s'agissant de la syllabe.

La mise en oeuvre de réseaux de neurones formels pour supporter cette analyse ne préjuge pas d'une modélisation de la syllabation en termes de courbes complexes ou en termes de frontières entre constituants. En phonologie, ces deux approches ont été proposées. On peut ainsi contraster, par exemple, le modèle à segmentation de Selkirk (1982, 1984) et le modèle à courbes de Goldsmith (1990). Les implémentations neuromimétiques quant à elles peuvent produire aussi bien une segmentation du signal (Elman 1990) qu'une intégration du signal sous forme de courbe (Goldsmith 1990). Le choix d'une implémentation connexionniste ne contraint donc pas le modèle phonologique, sous ce rapport au moins. La question de la segmentation a également un statut empirique. En effet, il est non seulement nécessaire de rendre compte de l'individualisation syllabique (à l'aide de frontières ou de zones d'inflection dans des courbes) mais aussi des phénomènes depuis longtemps relevés d'ambisyllabité et de multiplicité d'analyse pour un même mot. Soit, par exemple, les syllabations différentes du mot catastrophe :

- (1) [ka-tas-tRof]
- (2) [ka-ta-stRof]
- (3) [ka-tas-stRof] (ambisyllabité du [s])

Les systèmes à analyse unique et à coupe sont confrontés ici à des difficultés de traitement qui imposent un affaiblissement de la notion de coupe et/ou de discontinuité. Le concept de règle, central dans toutes les analyses de type parsing, doit être reformulé en termes de régularités incluant la possible existence de solutions multiples et contradictoires.

La reconnaissance de la notion syllabique ne suffit pas à définir son statut. En effet, aux analyses qui proposent une dérivabilité intégrale de l'architecture syllabique et qui postulent des algorithmes de syllabation (Anderson et Ewen 1987), on peut opposer les analyses qui font de l'architecture syllabique une partie de l'information lexicale et contestent ainsi l'existence de tels algorithmes (Encrevé 1988, Kaye, Lowenstamm et Vergnaud 1988). Les réseaux connexionnistes peuvent sans doute apporter une réponse à ces questions dans la mesure où la notion de dérivabilité y a un statut très différent. L'intégration dynamique de contraintes contradictoires et la capacité de réseaux à prendre en compte des phénomènes continus peut permettre de produire une analyse syllabique de la chaîne en termes de courbe ou de frontières sans pour autant postuler des algorithmes explicites de syllabation (Goldsmith 1990). Dans cette perspective, il reste à expliciter la nature de l'information sur laquelle les réseaux travaillent : information strictement segmentale et absence d'information syllabique vs. information segmentale et information syllabique minimale. Le codage initial sur la couche d'entrée incorpore nécessairement cette hypothèse. Ces deux conceptions du lexique sont toutes deux compatibles avec une approche connexionniste (approche *tabula rasa* de type Elman 1990 vs. approche phylogénétique de type Bienenstock 1990). Notre programme de recherche vise en particulier à tester ces deux hypothèses sur la syllabation des mots isolés en français.

**2. METHODOLOGIE.**

Avant même de s'interroger sur la richesse de l'information lexicale et du codage subséquent, une première difficulté doit être levée : sur quel matériel doit s'effectuer la modélisation : transcription orthographique ou phonémique, transcription phonétique, signal? A l'étape actuelle, une modélisation connexionniste prenant directement en entrée le signal digitalisé pose des problèmes pratiques et théoriques importants (constitution et taille des corpus, taille des fichiers de données, algorithmes de traitement du signal et de codage, etc.). Nous avons donc choisi, dans un premier temps de partir d'une transcription du signal.

Les transcriptions graphiques et phonétiques présentent pour notre projet les inconvénients de linéariser et de discrétiser le signal, et de filtrer les indices phonétiques dont on peut penser qu'ils constituent les bases de la syllabation. De plus, la transcription phonétique incorpore des décisions implicites de syllabation préjudiciables à l'analyse. Soit, par exemple, les transcriptions alternatives de lier :

- (4) [lje]
- (5) [lije]

La transcription graphique est un peu plus neutre sous ce rapport. Elle offre d'autre part l'avantage d'être constituée sous forme de corpus accessibles. Nous avons donc choisi comme point de départ d'appuyer nos modélisations sur un dictionnaire orthographique informatisé d'environ 90.000 formes. Deux lignes de recherche sont poursuivies en parallèle. La première, de type algorithmique, vise à expliciter des règles et à les organiser sous la forme d'un programme en C, la seconde vise à faire apprendre à un réseau de neurones formels des régularités par présentation des entrées et des sorties correspondantes.

Deux niveaux d'analyse peuvent être distingués.

- a. Formalisation, sur la graphie, de règles de coupe phonique. Le sous-ensemble des règles de coupe phonique inclus dans les prescriptions de coupe graphique constitue le point de départ. Ce module est progressivement enrichi par l'adjonction de règles de coupe phonique.
- b. Affaiblissement de la notion de règle et introduction des analyses multiples pour un même mot. Traitement de l'ambisyllabité. Introduction des régularités et des sous régularités.

La modélisation appuyée sur la graphie n'est, on l'a dit, qu'un choix provisoire. En effet, dès qu'on se propose de traiter des phénomènes phonologiques complexes mettant en jeu la continuité articulaire et/ou acoustique, les transcriptions discontinues de type orthographique voire phonétique/phonémique sont notablement insuffisantes. Dans une phase ultérieure, il sera donc nécessaire de prendre en compte directement le signal dans sa phase productive et réceptive. Ceci impliquera, en

retour des modifications considérables des architectures connexionnistes utilisées. L'architecture simple de réseaux à couches 'feed forward' de type PDP devra ainsi être abandonnée au profit d'architectures plus complexes incorporant en particulier des hypothèses cognitives sur l'intégration et la production du signal, à la manière de ce qui a été proposé pour la vision (cf. Bienenstock 1987). A ce niveau, la supériorité de modèles aptes à traiter le continu comme les neurones formels s'impose. Il en est ainsi, par exemple, du traitement de la syllabation dans des contextes de di-rèse/synrèse précédemment évoqués (cf. (4) et (5)).

### 3. PRINCIPES DE SYLLABATION

Comme nous l'avons dit, la modélisation d'un système expert de coupe syllabique et la comparaison avec les performances d'un réseau de neurones formels, travaillant tout deux sur une transcription graphique ne constitue que la première étape de notre projet de recherches. Ce sont les résultats de cette première étape que nous présentons à présent.

Au premier niveau, la syllabation est donc assurée par un système expert qui incorpore les principes suivants.

a. Les digraphes sont insécables (exemple : ph, th, rh, etc.). Les groupes graphiques occlusive-liquide et fr/vr/fl/vl sont insécables. Les géminées graphiques sont donc considérées comme sécables mais une règle tardive d'épellation assure leur unicité et leur syllabation à l'attaque.

b. Au niveau graphique, toute coupe syllabique laisse apparaître un noyau de syllabe.

c. Les groupes CV ne sont jamais hétérosyllabiques. Ceci implique ce qui a été nommé "priorité à l'attaque". La syllabation d'une ou plusieurs consonnes à la coda est une conséquence de l'impossibilité de les syllaber à l'attaque.

d. Les groupes C1C2 sont hétérosyllabiques ssi. ils ne forment pas un groupe insécable (priorité à l'attaque minimale).

e. Les groupes V1C2 sont hétérosyllabiques sauf si C2 ne peut être en position d'attaque (cf. b).

f. Les groupes V1V2 sont hétérosyllabiques si V1 et/ou V2 sont graphiquement accentués, à l'exception des quelques mots contenant certains groupes comme

eô, eû, oê, du groupe productif ai, et des finales de mot V1e, V1es.

g. Sauf exceptions traitées au deuxième niveau, notamment les très productives synrèses sur les groupes graphiques iV2, ainsi que les groupes eau qui ne coupent pas, les groupes V1V2 graphiquement inaccentués sont homosyllabiques ssi. V2 est un i ou un u graphiques, autrement ils sont hétérosyllabiques.

Au second niveau, les coupes prédites par le système expert sont testées manuellement et, en particulier, les ambisyllabicités et analyses multiples sont notées. On obtient ainsi la base minimale des entrées sorties fournies au réseau.

### 4. IMPLEMENTATIONS DES RESEAUX

La première implémentation est réalisée à l'aide d'un réseau à couches avec apprentissage par rétropropagation de type PDP (l'implémentation a été mise en oeuvre à l'ENST par C. Huynh dans le cadre de son mémoire de fin d'études). Sur la couche d'entrée on code des caractères. La fenêtre contextuelle étant une fenêtre glissante de 8 caractères avec test de coupe en position 3/4, et le nombre de caractères différents à prendre en compte étant de 43 (42 plus un symbole de fin de chaîne), la couche d'entrée comporte 344 unités. La couche de sortie sur laquelle est codée la possibilité d'une coupe en position 3/4 comprend une seule cellule dont le niveau d'activité est dans l'espace 0..1. Les niveaux de sorties supérieurs ou égaux à 0.5 sont considérés comme des réponses positives. L'optimum de résultat a été atteint avec une couche de 6 cellules cachées. La connectivité est strictement 'feed forward' sans inhibition bi-latérale.

L'apprentissage est réalisé sur un corpus de 1000 mots tirés au hasard, et les tests sont réalisés sur le dictionnaire complet de 90000 mots. Le corpus d'apprentissage est présenté 30 fois, et au total la convergence du réseau est atteinte après 45 mn de calcul sur une station SUN 3. Avec un corpus d'apprentissage au hasard, le pourcentage d'erreurs est de l'ordre de 2%. Une analyse du clustering sur la couche cachée fait apparaître que ses 6 cellules se divisent en 3 cellules excitatrices et 3 inhibitrices. Une des cellules excitatrices se comporte comme un détec-

teur typique de voyelles. On notera au passage que y graphique est, selon les cas, analysé comme voyelle et détecté par cette cellule ou bien analysé comme consonne. Le groupe des liquides est également détecté par une cellule spécialisée, mais il est de plus détecté, à un niveau faible d'activité, par la cellule spécialisée dans la détection des obstruantes. Enfin, il apparaît que les nasales et le h graphique sont détectés par une combinaison particulière des activités excitatrices et inhibitrices des cellules prenant en charge les voyelles, les liquides et les vrais consonnes. Une des cellules paraît spécialisée dans le traitement des groupes de voyelles graphiques, spécialement lorsqu'une de ces voyelles est accentuée.

L'analyse des erreurs et le caractère systématique et régulier d'un certain nombre d'entre elles nous ont conduit à modifier le protocole d'apprentissage. Le corpus d'apprentissage de 1000 mots a ainsi été divisé en 2 parties, les 500 premiers étant tirés au hasard, les 500 autres constituant une représentation statistique approchée des erreurs réalisées par le réseau dans un apprentissage strictement au hasard. On observe alors un comportement plus systématique et plus discret des cellules de la couche cachée. Le pourcentage d'erreurs est inférieur à 1%, soit de l'ordre de 600 mots, dont 1/4 environ est constitué de mots étrangers et 1/6 est constitué de mots à combinaison de caractères particulièrement rares (par exemple rhy, cz). Le reste des erreurs est constitué par quelques erreurs très systématiques portant sur des groupes lexicalement productifs de type préfixe dés + racine à h initial.

Bien que l'architecture de ce réseau n'incorpore aucune hypothèse cognitive ou phonologique sur la syllabation, on remarque qu'il couvre remarquablement le corpus des données avec un apprentissage faible. Ceci laisse supposer d'une part qu'une modification plus substantielle du protocole d'apprentissage incorporant des hypothèses sur la morphologie et la structure du lexique, d'autre part qu'une modification de l'architecture interne du réseau implémentant des hypothèses phonologiques devraient accroître encore les performances.

Ces performances pourront être comparées à celles d'un réseau traitant directement le signal en production et/ou en réception et assurant son apprentissage non seulement par une présentation simultanée des entrées et des sorties mais également par une réorganisation interne de type phylogénétique. L'architecture interne d'un tel type de réseau devra être riche et cognitivement pertinente. C'est dans cette orientation que se poursuit le travail présenté ici.

### OUVRAGES CITES

- Anderson, J. et Ewen, C., 1987 : *Principles of dependency phonology*, Cambridge University Press, Cambridge.
- Bienenstock, E., 1987 : Connexionist approaches to vision, in Imbert, ed., *Models of vision perception : from natural to artificial*, Oxford University Press, Oxford.
- Bienenstock, E. et Doursat, R., 1990 : Epigenetic development of spatio-temporal patterns in the brain, in Schuster, ed., *Nonlinear dynamics and neural networks*, VCH publisher.
- Elman, J. L., 1990 : Finding structure in time, *Cognitive Science* 14.
- Encrevé, P., 1988 : *La Liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Le Seuil, Paris.
- Goldsmith, J., 1990 : *Harmonic phonology*, ms., University of Chicago.
- Kaye, J. D., Lowenstamm, J. et Vergnaud, J.-R., 1988 : *Constituent structure and government in phonology*, ms.
- Selkirk, E. O., 1982 : The syllable, in van der Hulst and Smith, eds., *The Structure of phonological representations*, Foris, Dordrecht.
- Selkirk, E. O., 1984 : On the major class features and the syllable theory, in Aronoff and Oehrle, eds., *Language sound structure*, M.I.T. Press, Cambridge.