# Les théories linguistiques à l'épreuve des technologies vocales : l'exemple du synthétiseur "Kali"

Anne Lacheret-Dujour, Michel Morel

# Les théories linguistiques à l'épreuve des technologies vocales : l'exemple du synthétiseur « Kali »

*A. Lacheret-Dujour, M. Morel,*
*Laboratoire CRISCO,*
*Université de Caen, France*
{anne.lacheret;michel.morel}@crisco.unicaen.fr

Parmi les diverses technologies dites « émergentes », la synthèse de la parole apparaît comme un véritable moteur de développement social, au sens général du terme. Au même titre que d'autres outils informatiques avec lesquels elle collabore d'ailleurs souvent étroitement (image animée, interface tactile, réalité virtuelle par exemple), elle brouille les frontières entre l'écrit et d'autres formes de la communication humaine. A ce titre, elle implique une réflexion visant à situer la place de l'écrit dans l'ensemble des dispositifs communicatifs, en particulier par rapport à l'oral (transformation d'un texte écrit en un continuum sonore). Sous l'angle ergonomique, cette nouvelle forme de lecture oralisée – ou écriture sonore – a inévitablement des conséquences en termes d'usage et de charge cognitive (mobilisation auditive des capacités langagières). En amont la synthèse de la parole participe à l'élaboration des grandes orientations scientifiques contemporaines et à une nouvelle organisation de la recherche dans le domaine de l'informatique et de la linguistique, fondamentalement pluridisciplinaire et interactive.

Après un tour d'horizon rapide sur l'histoire de la parole artificielle en Europe, nous abordons l'univers actuel de la synthèse vocale dans le cadre général du traitement automatique des langues (TAL) et de la linguistique. Ce type de technologie, en effet, constitue un domaine tangible pour illustrer deux démarches fondamentales qui motivent les recherches en TAL, la première visant à la diffusion de produits d'ingénierie linguistique sur le marché des industries de la langue, la seconde concernant directement la recherche en linguistique théorique. Dans le premier cas, les sciences du langage sont au service de l'informatique dans la mesure où l'expertise linguistique contribue de manière active à résoudre des problèmes concrets de TAL. Dans le second, l'informatique se met à la disposition des linguistes pour, d'une part, proposer des outils robustes de traitement et d'analyse des données langagières, d'autre part interroger, tester et valider les théories linguistiques existantes, cela dans des domaines divers (sémantique, syntaxe, phonologie, phonétique, pour ne citer que les principaux). Ces deux volets font l'objet de notre communication, consacrée à la présentation de nos recherches en synthèse de la parole à partir du texte au laboratoire CRISCO depuis 1995. Si la nécessité de répondre à des problèmes applicatifs concrets (en l'occurrence : synthèse pour mal voyants), avec des outils particuliers, est à l'origine de nos travaux, la question récurrente qui se pose pour déterminer l'apport des modèles formels des syntacticiens et des phonologues, comme les approches substantielles des phonéticiens pour le traitement numérique des données sonores, constitue, bien sûr, pour nous une préoccupation majeure. C'est ce dialogue étroit entre visée modélisatrice et contraintes de traitement automatique dont nous souhaitons rendre compte ici. Pour ce faire, nous proposons de centrer notre exposé sur les recherches menées conjointement par les linguistes et les ingénieurs sur la prosodie – ou l'intonation – des langues, et sur les problèmes théoriques posés par la construction d'un modèle de congruence intonosyntaxique en vue de la génération automatique de la prosodie en synthèse de la parole. Une définition du domaine sera d'abord effectuée. Notre exposé nous amènera ensuite à préciser la pertinence et les limites de la notion de congruence. Cette thématique sera abordée à travers la présentation de deux générations de modèles, les premiers fondés sur la formalisation explicite de l'interaction entre les structures syntaxique et intonative, les seconds relatifs à la définition des contraintes rythmiques qui pèsent sur la dérivation de la structure intonative. Nous situerons enfin les choix théoriques retenus pour la modélisation de la prosodie dans notre système de synthèse Kali par rapport à ces perspectives théoriques. Ces différents points nous amèneront à conclure sur la nécessité d'envisager une troisième génération de modèles, fondée sur la formalisation explicite des traces que laissent les fonctions sémiotique et communicative dans le message sonore.

## 1. La synthèse de la parole : le domaine

Quelques jalons d'histoire permettront de comprendre les divers enjeux associés au développement de la parole artificielle. On verra, et c'est là toute la spécificité de la synthèse par rapport aux technologies dites nouvelles, qu'il s'agit en fait d'une science très ancienne.

### 1.1. Quelques mots d'histoire

Nouvelle technologie, avez-vous dit ? Erreur : jusqu'à la fin du XVIIIe siècle, les machines parlantes font l'objet de légendes et de fascinations diverses. Dès la Renaissance, les premiers modèles physiques de production du signal de parole sont développés. Un humaniste de l'époque déclare : « nous produisons à volonté des sons articulés et toutes les lettres de l'alphabet, soit les consonnes, soit les voyelles que nous imitons, ainsi que les différentes espèces de voix et de chants des animaux terrestres et des oiseaux ». Vers 1630, un projet de dispositif parlant proche de l'orgue à tuyaux est proposé : les voyelles proviennent de tuyaux à embouchure de flûte et divers dispositifs sont utilisés pour imiter les consonnes. Au Siècle des Lumières, cet engouement pour la parole artificielle atteint son apogée : dans les années 1780, apparaît en Russie un exemple de simulation du conduit vocal dont la réalisation est attribuée à un certain Kratzeinstein ; composée d'un ensemble de résonateurs acoustiques excités par une anche vibrante, la machine produit cinq voyelles. En 1791, une autre machine parlante est inventée par le baron Von Kempelen, gentilhomme de la cour d'Autriche-Hongrie, considéré aujourd'hui comme le véritable pionnier de la synthèse de la parole (Liénard 1977). Sa machine comprend un soufflet et une chambre à air comprimée munie d'une anche vibrante. Un résonateur constitué d'un cuir déformable à la main est utilisé pour produire les sons voisés. Les consonnes, quant à elles, sont créées par fermeture de certains orifices (production des plosives) ou par des sifflets actionnés par des leviers (génération des fricatives). La machine peut ainsi émettre une vingtaine de sons différents. Au XIXème siècle, d'autres systèmes sont construits sur les principes de la machine de Von Kempelen, mais ce n'est véritablement qu'à partir de 1939, aux Etats-Unis, que les progrès sont décisifs. D'abord le *Voder* (« voice Operation demonstrator »), premier codeur de voix électrique, a comme fonction première d'étudier le rôle relatif des différentes composantes du signal détecté par l'analyse acoustique. Ensuite, en 1950, à l'issue de nombreuses études sur la production et la transmission de la parole, le *Pattern Play-Back* est développé. Il joue un rôle essentiel pour la compréhension des caractéristiques articulatoires et perceptives de la parole. Puis, en 1953, une technique de synthèse, fondée sur la simulation articulatoire du conduit vocal, est présentée. Enfin, l'année 1959 est marquée par la présentation d'un nouveau type de synthèse aux Université de Stockholm et d'Edimbourg. L'enveloppe spectrale d'un signal de parole est décrite par ses composantes essentielles : les formants, qui résultent des fréquences de résonance des cavités du conduit vocal. Les années 1970 signent l'explosion de la synthèse de la parole en tant que technique attestée dans le domaine du traitement automatique des langues. En effet, tandis que les théories sur le traitement et la modélisation du signal deviennent de plus en plus pertinentes et opératoires, les ordinateurs se répandent sur le marché. Deux voies de recherche, déjà amorcées par le passé, font l'objet d'investigations rigoureuses : reproduire le signal de parole à partir de simulations fonctionnelles du conduit vocal humain ou bien simuler la propagation de l'onde sonore dans le conduit vocal à partir de connaissances physiologiques, articulatoires et mécaniques. Les enjeux scientifiques sont donc bien réels mais on est encore loin des objectifs industriels de commercialisation. Il faudra attendre le début des années 1980 pour que les synthétiseurs sortent des laboratoires, soit pour des applications grand public, à vocation essentiellement ludique, soit pour le handicap (synthèse pour aveugle ou handicap vocal par exemple).

### 1.2. Où les connaissances linguistiques s'avèrent nécessaires

L'objectif d'un système de synthèse de la parole est de produire un énoncé oral à partir d'une représentation phonétique de celui-ci. Deux méthodes peuvent être utilisées : la synthèse par concaténation d'éléments préenregistrés – il s'agit de codage – et la synthèse de vocabulaire illimité. Nous ne nous attarderons pas sur la première qui certes est encore très employée dans les applications grand public (jeux électroniques, appareils électroménagers, serveurs vocaux rudimentaires), mais présente peu d'intérêt pour la recherche fondamentale et ne concerne pas la linguistique. Plus précisément, un système de synthèse par mots remplit la même fonction qu'un magnétophone digital (enregistrement

d'un signal de parole prononcé par un locuteur humain, codage et compression de celui-ci, restitution du message). Lors de la restitution du signal, le vocabulaire est limité à celui qui a été prononcé et la voix est invariable (celle du locuteur enregistré). Cette technique présente l'avantage d'être économique et facile à mettre en œuvre, ce qui explique sa diffusion dans des applications bon marché. Néanmoins, elle a l'inconvénient de ne permettre la modification d'un message qu'en faisant appel à la personne qui a prononcé ce message. Si cette dernière n'est plus disponible, un ré-enregistrement complet est nécessaire ; ceci n'est évidemment pas envisageable pour des applications de grande envergure. D'où la nécessité de développer des systèmes de synthèse de vocabulaire illimité. Nous définirons cette technique comme la production par un ordinateur d'un énoncé oral de longueur quelconque qui n'a jamais été prononcé auparavant. Pour ce faire, deux étapes sont à distinguer (figure 1) :

- La première est de nature **linguistique**. Elle se décompose en deux volets : le traitement de la chaîne segmentale et celui de la chaîne suprasegmentale. Pour le premier aspect : il s'agit de faire correspondre à un texte écrit une suite de symboles phonémiques. On parle alors de **transcription graphème-phonème**. Le second nécessite le développement d'un **modèle intonatif pour la génération automatique de la prosodie** (génération de la durée, marquage des frontières entre groupes, allocation de pauses, positionnement des accents, génération de contours prosodiques qui traduisent les variations de la courbe mélodique). Pour effectuer ces deux tâches, une **analyse syntaxique automatique** préalable est nécessaire. Elle permet non seulement de détecter les ambiguïtés homographiques *(les poules du couvent couvent)*, mais également de distinguer les mots outils des mots lexicaux qui n'ont pas les mêmes caractéristiques accentuelles, de segmenter le texte à produire en constituants syntaxiques, étape nécessaire pour poser les accents associés à la structure intonative, enfin d'identifier les modalités des phrases à synthétiser marquées le cas échéant par des intonations variables[1].

- La deuxième étape est de nature **phonétique**. Elle consiste à transformer une suite de symboles phonétiques discrets en un signal de parole continue. Il s'agit donc d'effectuer la transition entre une représentation linguistique abstraite et la réalité acoustique.

[1] Voir par exemple, la différence prosodique entre les modalités assertive et interrogative dans les langues.

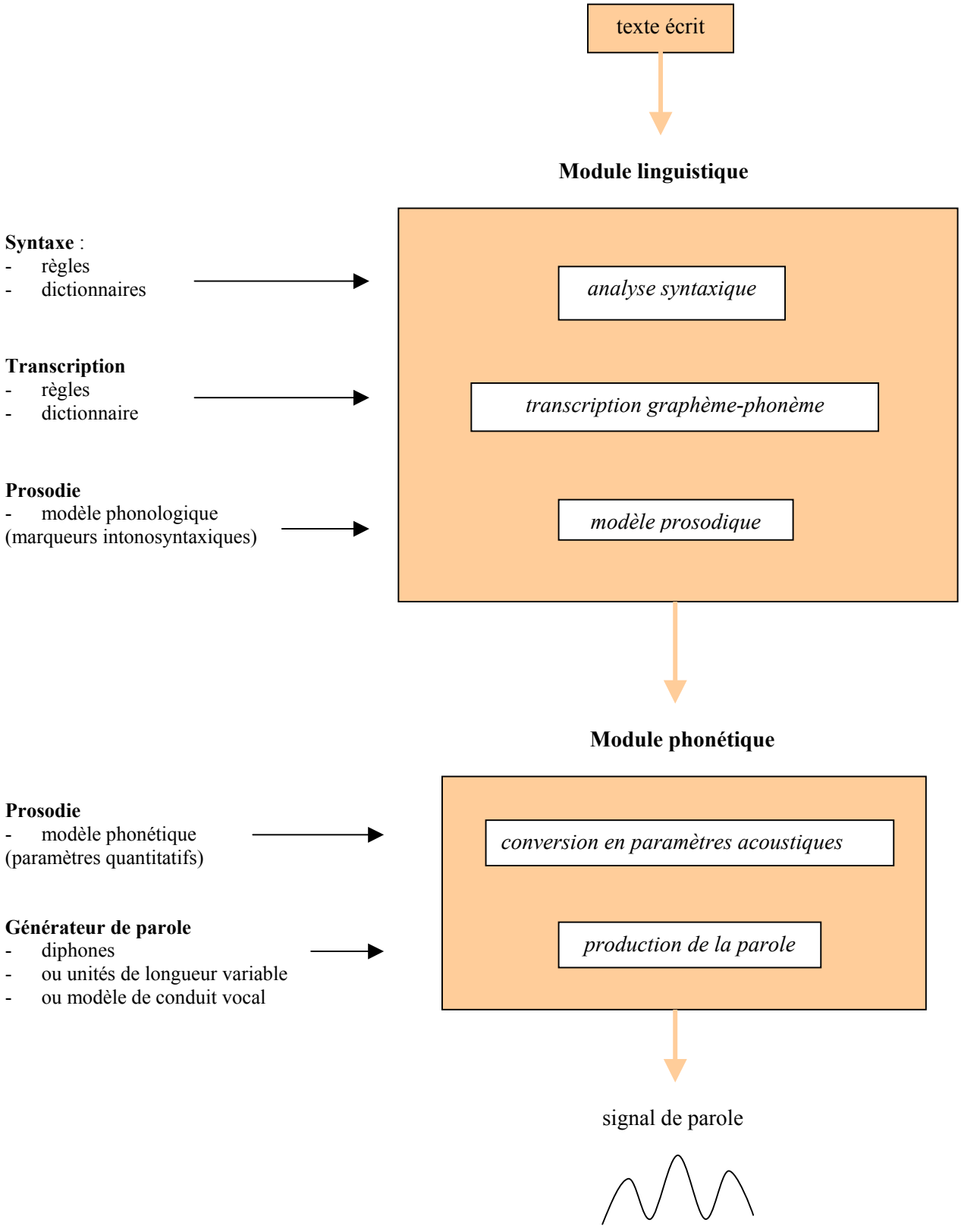



***Figure 1.*** *Les modules appelés dans un système de synthèse à partir du texte*

Seule la première étape nous intéressera ici. Si la transcription graphème-phonème ne pose plus de problèmes majeurs aujourd'hui pour le développement d'un système de synthèse de parole, en revanche la mise en place d'un générateur robuste de la prosodie reste une entreprise délicate. Essayons de voir pourquoi.

## 2. La prosodie : quelle définition ?

Le terme "prosodie" désigne une structure articulée autour de deux mécanismes imbriqués et à portée variable : tandis que **l'accentuation**, qui repose sur la proéminence d'objets syllabiques locaux, relève de la prosodie lexicale, **l'intonation** représente un processus postlexical qui se manifeste sur l'ensemble de l'énoncé et sur les groupes qui le constituent. C'est ainsi qu'on peut parler d'intonation de paragraphe, de phrase ou de groupe. Une fois cette définition posée, encore faut-il comprendre les corrélats physiques de la structure intonative tels qu'ils émergent du continuum acoustique.

### 2.1. Les paramètres mobilisés pour actualiser la structure prosodique dans la substance

Du point de vue de l'**interface** entre la **phonologie** et la **phonétique**, on peut proposer une définition cognitive de la prosodie : le niveau phonologique correspond à l'analyse du **système** prosodique, celui-là même qui donne lieu à l'apprentissage d'un jeu de règles par l'enfant en phase d'acquisition du langage, le niveau phonétique concerne l'étude des **paramètres physiques** mobilisés pour actualiser ce système dans la substance. Plus précisément, sous l'angle phonétique, le signal de parole est décomposable en deux niveaux d'analyse : le niveau segmental – ou phonématique – qui implique l'identification et la description des phonèmes produits dans le signal et le niveau suprasegmental – ou prosodique – marqué principalement par des variations de fréquence fondamentale (f0)[2], de durée[3] et d'intensité (figure 2)[4]. Parmi ces paramètres, représentés sur le plan perceptif par les termes "hauteur", "rythme" et "sonie", la f0 a pendant longtemps été considérée comme la plus significative, sans doute parce que la plus facile à percevoir. On a ainsi pu montrer que les variations de fréquence fondamentale, lorsqu'elles marquent des proéminences accentuelles, donnent lieu à la perception de hauteurs locales, le corrélat de l'intonation étant perçu comme une mélodie globale modélisable par une **ligne de déclinaison**[5]. Cette déclinaison – ou *downdrift* – serait le résultat de la tendance que connaît la fréquence fondamentale à décroître lentement du début à la fin d'une production[6]. L'hypothèse aujourd'hui admise consiste à poser que la déclinaison, processus physiologiquement déterminé, est également contrôlée par le locuteur pour délimiter des unités de taille et de statut linguistique variable (Ladd 1984). Elle s'accompagne notamment de phénomènes de **réinitialisation mélodique**[7] qui se situent à des frontières de groupes linguistiques.

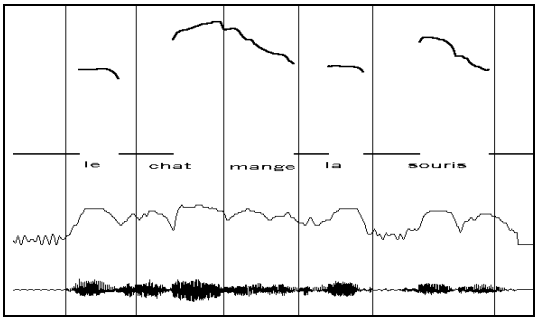



***Figure 2.*** *Visualisation d'un signal de parole avec, de bas en haut, respectivement le signal acoustique, l'intensité en décibels, et la fréquence fondamentale en hertz ; sur l'axe des abscisses : les segments prononcés au cours du temps (en millisecondes).*

---

2 Estimation du son laryngien à partir du signal acoustique à un instant donné.

3 Correspondant à la mesure d'un intervalle de temps nécessaire pour émettre le signal de parole, la durée concerne l'organisation temporelle du message et comprend le débit de parole (nombre de syllabes ou de phonèmes prononcés dans une unité de temps donnée), le tempo (accélération ou ralentissement du débit à l'intérieur d'un groupe prosodique) et les pauses.

4 L'intensité est relative à l'énergie contenue dans le signal de parole.

5 Voir Pike (1945) pour une première étude du phénomène.

6 L'effet du cycle de respiration se traduit par une tendance à réaliser une fréquence laryngée plus élevée au début du cycle et plus basse à la fin.

7 Dénommés également *resetting*.

Si les modélisations prosodiques se sont d'abord centrées sur les variations de la fréquence fondamentale, on ne peut pas, aujourd'hui, sous-estimer le rôle de la durée dans le marquage de proéminences accentuelles. Ces dernières, en effet, reposent non seulement sur des modulations de f0 mais également sur l'alternance de temps forts et de temps faibles. De même, les variations de durée jouent un rôle majeur dans la perception du rythme global de l'énoncé. L'intensité, quant à elle, reste un paramètre d'étude marginal, parce que souvent vue comme une simple co-variable de la fréquence fondamentale. Néanmoins, plusieurs travaux attestent de son importance dans la reconnaissance auditive des patrons intonatifs (Rossi 1979), dans la distinction des modalités énonciatives ou encore dans les stratégies interactives en situation de dialogue (Morel & Danon-Boileau 1998). Sur ce dernier point, l'intensité indiquerait la façon dont un locuteur entend gérer son tour de parole, s'il est prêt à laisser son interlocuteur intervenir ou si, au contraire, il souhaite poursuivre son discours. Enfin, on soulignera le rôle des caractéristiques spectrales des phonèmes[8] dans le marquage des variations temporelles et expressives. Une première constatation s'impose donc clairement : **la prosodie se manifeste dans la substance de manière pluriparamétrique**.

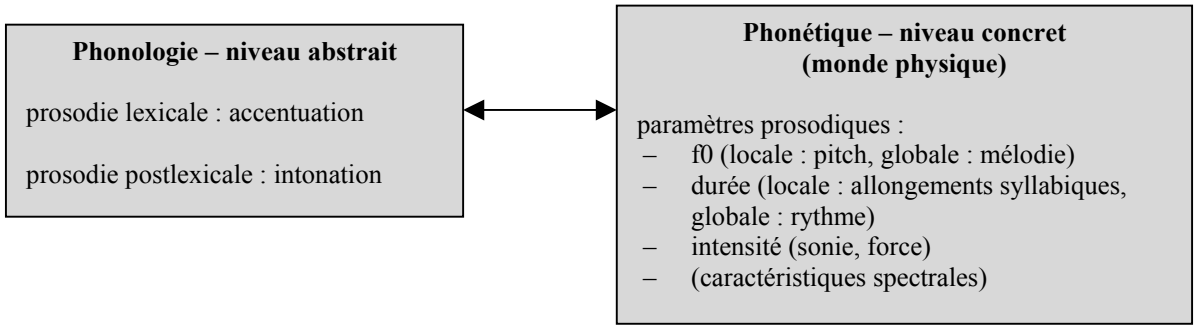



***Figure 2.*** *Phonologie prosodique et corrélats phonétiques*

### 2.2. Un champ d'étude complexe, des fonctions linguistiques variées

La complexité du domaine présenté ici, tant pour les linguistes que pour les spécialistes de traitement automatique de la parole, découle tout d'abord de la nature intrinsèque du champ à explorer : représentant à la fois ce qu'il y a de plus universel (il n'y a pas de langue sans intonation, sans accentuation ou sans rythme), les constructions prosodiques sont également spécifiques à chaque langue. Dès les premiers balbutiements, avant les premiers mots, le signe prosodique apparaît comme moyen de communiquer et de négocier sa prise de parole ; c'est aussi lui qui perdure lors de l'apprentissage de langues secondes, celui, enfin, qui nous ramène vers ce langage primitif, d'où, dit-on, la double articulation est absente. Cette irréductibilité de la prosodie est bien sûr associée à sa charge significative : contrairement au seul rôle qu'on a bien voulu lui attribuer pendant longtemps, la prosodie n'est pas une simple musique vocale qui accompagne de manière facultative les modulations de la pensée et qui, de ce fait, doit rester cantonnée dans la sphère du paralinguistique (de l'émotion ou de l'expressivité), elle contribue fondamentalement à structurer le niveau référentiel et assure ainsi la cohésion sémantique de l'énoncé. En conséquence, lorsqu'on propose une définition linguistique du domaine, on ne peut que mettre d'emblée l'accent sur la polyvalence fonctionnelle des unités prosodiques. La prosodie assume d'abord une **fonction démarcative** puisqu'elle permet à l'auditeur de segmenter le continuum sonore en isolant des unités de taille variable (énoncé, paragraphe, phrase, constituant, voire mot). Elle répond ensuite à une **fonction de hiérarchisation** et marque les rapports d'inclusion syntactico-sémantique ou, au contraire, d'autonomie qui peuvent exister entre les unités détectées. Enfin, par sa **fonction énonciative,** elle a pour objet de rendre saillantes certaines parties du discours qui, par-là même, deviennent des centres d'attention perceptive.

[8] Liées à la distribution de l'énergie dans le spectre.

## 3. Les modèles prosodiques

En premier lieu, le rôle actif de la synthèse de la parole dans le développement des théories et modèles prosodiques doit être souligné. Certes, elle intervient en aval pour valider les approches conceptuelles, mais elle participe dès le départ activement à la construction des modèles. Dans ce contexte de collaboration étroite entre recherche fondamentale et recherche appliquée, les travaux sur la prosodie des langues ont été menés très tôt pour préciser les relations qui peuvent exister entre la structuration syntaxique et l'organisation intonative d'une phrase. Trois hypothèses ont été explorées : (i) Les structures intonative et syntaxique sont parfaitement isomorphes, (ii) la structure intonative est totalement indépendante de la syntaxe, (iii) ni complètement congruente à la structure syntaxique, ni totalement indépendante, la structure intonative est associée à la syntaxe dans la mesure où elle respecte par ailleurs un jeu de contraintes rythmiques de nature universelle.

### 3.1. Modèles de la première génération : la structure intonative dérive de points d'ancrage syntactico-sémantiques

Envisager une congruence étroite entre les structures intonative et syntaxique revient à considérer que l'intonation remplit une fonction d'actualisation et de hiérarchisation des constituants syntaxiques. Cette fonction est d'autant plus marquée que l'intonation peut lever tout ou partie des ambiguïtés potentiellement porteuses des problèmes de segmentation en mots ou en composants syntaxiques (Di Cristo 1978). Soit les exemples suivants :

1. *Mais oui mon cher, réellement*
2. *Mais oui mon cher Rey, elle ment*
3. *La sœur de Pierre-Olivier et son cousin*
4. *La sœur de Pierre, Olivier et son cousin*
5. *L'assassin a tué l'homme avec un revolver*
6. *La petite fille a vu l'homme avec un revolver*

Autrement dit, l'intonation intègre, délimite ou segmente suivant les contraintes syntactico-sémantiques et, donc, oriente le récepteur vers le découpage pertinent. Néanmoins, il est vite apparu, notamment en français, qu'étant donné les phénomènes de désaccentuation nombreux dans cette langue, plusieurs sorties intonatives pouvaient être engendrées pour une structure syntaxique donnée. Ainsi, la phrase *7. Paul a rencontré un marchand de tableaux d'origine auvergnate* ne fera sans doute jamais l'objet du découpage intonatif suivant :

*(Paul) (a rencontré) (un marchand) (de tableaux) (d'origine) (auvergnate)*.

En revanche, diverses segmentations sont possibles :

*(Paul) (a rencontré) (un marchand de tableaux) (d'origine auvergnate)*
*(Paul a rencontré) (un marchand de tableaux) (d'origine auvergnate)*
*(Paul) (a rencontré un marchand de tableaux) (d'origine auvergnate)*
*(Paul a rencontré) (un marchand de tableaux d'origine auvergnate)*

### 3.2. Modèles de la seconde génération : les contraintes rythmiques

Après avoir exploré la congruence intonosyntaxique et posé ses limites, les chercheurs ont cherché à mieux préciser les contraintes rythmiques pesant sur la mise en place de la structure prosodique, c'est en cela qu'on peut parler de **modèles de la seconde génération**. Ces modèles reposent sur le **principe de régulation accentuelle** et **d'équilibre rythmique** (Martin 1987, Pasdeloup 1990). Suivre ce principe implique d'abord de limiter à quatre le nombre de syllabes inaccentuées, ce qui explique la formation des pieds accentuels et donc l'apparition d'accents secondaires non finals de mots. Par ailleurs, la répétition régulière de groupes de taille équivalente satisfait au principe de récurrence et de périodicité qui constitue une des caractéristiques fondamentales des rythmes moteurs. Les groupes intonatifs tendent donc à être ni trop longs, ni trop courts et se construisent autour d'un nombre de syllabes proche, la moyenne étant de sept selon Wioland (1985) ; d'où des processus compensatoires de suraccentuation ou, à l'inverse, de désaccentuation. Ainsi, le syntagme nominal <déterminant + adjectif + nom> est représenté, conformément à la structure syntaxique, comme un seul groupe intona-

tif *(ex. un beau ca**nard**)*, mais la structure syllabique peut imposer la formation de deux groupes *(ex. (d'intermi**nables**) (escala**tors**))*. A l'inverse, deux constituants syntaxiques distincts, comme le sujet et le verbe, peuvent donner lieu à un regroupement intonatif s'ils sont de petite taille *(ex. Paul **lit**).)*. Les modèles convoquent enfin un principe de degré accentuel (Dell 1984), de nature perceptive : il existe une hiérarchie accentuelle dans la perception des accents qui s'aligne sur la hiérarchie syntaxique. Ainsi, dans la phrase : *c'est les poils du rat d'eau,* si *poil* est accentué, son degré accentuel est moins fort que celui porté par *eau*.

Si, comme nous venons de le voir, les contraintes rythmiques jouent un rôle important dans la bonne formation de la structure intonative, les tentatives de représenter cette dernière indépendamment de la syntaxe ont bien évidemment échoué. Bref, la seconde hypothèse s'avère totalement irréaliste et c'est bien la troisième qui fait l'objet des modélisations actuelles. Nous la résumerons comme suit : dire que, pour une entrée syntaxique donnée, plusieurs découpages s'avèrent possibles, c'est préciser également que les plus **eurythmiques** – ou équilibrés rythmiquement – sont sélectionnés, à condition toutefois qu'ils respectent la règle de **non-collision syntaxique** (Martin 1987). Selon cette dernière, le regroupement d'unités accentuables au sein d'un même groupe intonatif est bloqué lorsqu'elles sont dominées par des têtes syntaxiques différentes. Par exemple :

7. *Pour la première fois de sa vie le général a décoré son meilleur chameau*

qui fait l'objet du découpage syntaxique :

*((Pour la première fois) (de sa vie)) (le général) (a décoré) (son meilleur chameau)*

peut donner lieu aux découpages intonatifs suivants (Martin 1999) :

*(pour la première fois de sa vie) (le général) (a décoré son meilleur chameau)*
*(pour la première fois de sa vie) (le général a décoré)(son meilleur chameau)*

le regroupement ci-dessous étant en revanche interdit :

*(pour la première fois) (de sa vie le général) (a décoré son meilleur chameau)*

## 4. Le modèle intonosyntaxique implémenté dans le système de synthèse vocale Kali

Le système de synthèse à partir du texte que nous avons développé au laboratoire CRISCO à l'université de Caen se situe dans la mouvance des modèles de la seconde génération : il repose essentiellement sur la modélisation interactive des contraintes rythmiques et syntaxiques, à laquelle s'ajoute la prise en compte de la dimension textuelle du message à synthétiser (identifications de titres et de paragraphes).

### 4.1. Le modèle syntaxique

La méthode retenue pour ce travail repose sur le calcul de mise en relations syntaxiques, c'est-à-dire sur la mise en évidence de processus en nombre fini à l'origine de la production et de l'interprétation des structures syntaxiques rencontrées (Vergne 1999). Dans cette approche, la phrase est considérée comme le codage linéaire d'une représentation dépendantielle abstraite, qui, selon l'hypothèse formulée par Vergne (1999), doit obéir à deux contraintes cognitives fondamentales : (i) la contrainte de minimisation de l'effort mémoriel implique la minimisation des distances entre deux nœuds syntaxiquement liés dans l'ordre linéaire, (ii) l'information contenue dans la représentation profonde ne doit pas être perdue au cours de la linéarisation. Or, étant donné les contraintes de propriétés géométriques du matériau observable (caractère unidimensionnel de l'énoncé), une éventuelle relation de dépendance syntaxique peut être rendue opaque lors de la production des séquences (cf. ex. 5 vs. 6 en *supra*). Il paraît bon de rappeler ici que : « syntaxiquement, la vraie phrase, c'est la phrase structurale dont la phrase linéaire n'est que l'image projetée tant bien que mal, et avec tous les inconvénients d'aplatissement que comporte cette projection dans la chaîne » (Tesnière[9] 1959, 20). Nous

[9] Voir aussi Robert (1997, 30) : « Dans l'énonciation, le locuteur doit projeter sur un axe ***linéaire*** une pensée ***multidimensionnelle*** et la ***discrétiser*** en unités séquentielles. Du fait des propriétés physiques du langage qui est un matériau sonore produit séquentiellement dans le temps, la verbalisation suppose donc de faire passer la pensée par un code particulier qui constitue un goulet d'étranglement (…) ». Et aussi Berrendonner

montrerons dans la section 4.2. que dans de tels contextes, c'est bien en définitive la structure prosodique (temporelle et tonale) qui permet de résoudre ce hiatus : la variation dans le marquage et la durée des pauses, ainsi que dans le calcul des proéminences accentuelles permet d'évaluer la distance entre deux constituants linéairement adjacents et ainsi de récupérer l'information nécessaire à l'auditeur pour reconstruire l'arbre de dépendance associé au matériau linéaire produit.

En pratique, le modèle syntaxique (Vannier 1999) prend en charge trois tâches : (i) la **segmentation** du texte à synthétiser **en paragraphes, phrases et mots**, (ii) **l'étiquetage** des mots en parties du discours qui s'appuie conjointement sur la **consultation de bases de données lexicales** et **morphologiques**, et sur l'utilisation de **règles de déduction contextuelles**, enfin (iii) la **segmentation en constituants syntaxiques** et leur **mise en relation**.

Concernant le premier point, si la segmentation, en paragraphes et en phrases ne présente pas de difficulté particulière, il en va autrement pour l'unité « mot ». D'une façon très générale, on peut affirmer que le mot se définit comme une chaîne de caractères compris entre deux blancs ou entre un blanc et un signe de ponctuation. Notre principe de segmentation est donc le suivant : les caractères de ponctuation, les tirets, parenthèses, guillemets et leurs variantes servent de séparateurs de mots et sont eux-mêmes considérés comme des mots[10]. L'apostrophe (dans le cas du français) est rattachée au mot qui la précède. A première vue donc, le repérage des mots qui composent une phrase semble simple. Néanmoins, on ne peut pas négliger l'ambiguïté réelle de certains caractères comme l'apostrophe et le tiret (la chaîne *aujourd'hui* correspond à un mot, *j'arrive* est formé de deux mots, ibid. pour *porte-monnaie vs. voulez-vous*). En conséquence, les critères typographiques sont nécessaires mais non suffisants. Ce point explique en partie l'utilisation de dictionnaires.

L'étiquetage des mots ensuite, dénommé aussi « catégorisation » est activé essentiellement pour traiter les formes graphémiques ambiguës syntaxiquement (ex. *a priori* : adverbe ou nom) et traiter à part les verbes et les mots grammaticaux. Les premiers, en effet, servent ici de points d'ancrage à la segmentation en constituants : toute séquence qui ne se construit pas autour d'une forme verbale est considérée par défaut comme nominale. D'où la nécessité d'avoir repéré correctement les verbes au préalable. En outre, les formes verbales, beaucoup moins nombreuses que les formes nominales et plus stables (moins de néologismes et d'emprunts), justifient ce choix : les parties du discours sujettes à de fortes variations ne peuvent bien évidemment pas faire l'objet d'un codage lexical. Pour autant, il est impossible de lister toutes les formes verbales dans le dictionnaire, étant donné le nombre non négligeable de terminaisons qui nous amènerait à multiplier les entrées (quelques dizaines de milliers). Une solution plus judicieuse consiste à explorer simultanément une base de terminaisons (59 groupes différents, étant donné les nombreux verbes irréguliers) et une base de racines verbales (2500 racines). La recherche s'effectue sur chaque mot non encore catégorisé, jusqu'à ce qu'une correspondance puisse être établie entre une terminaison et un radical. Pour le mot *intéresse* par exemple, la terminaison *se* est identifiée (groupes *coudre* et *cuire*), mais le radical *intéres-* n'existe pas pour ces groupes. En revanche, la terminaison *-e* est répertoriée dans le groupe 1, qui contient le radical *intéress-* :

| |
|---|
| e = (verbe groupe 1) (singulier 3ème ou 1ère personne)<br>intéress = (verbe groupe 1) |

Le mot est alors étiqueté verbe, singulier, 3ème ou 1ère personne.

Quant aux mots grammaticaux, c'est fondamentalement autour d'eux que s'articulent les relations entre les constituants syntaxiques, ils permettent en outre de lever un grand nombre d'ambiguïtés. Soit la phrase *: Le musicien soufflait sans trêve dans ses bazouks et ses strapons*. (B. Vian), les déterminants permettent ici de catégoriser comme noms les néologismes *bazouks* et *strapons*. Précisons que la nature souvent polycatégorielle de ces entrées nous amène à proposer plusieurs étiquettes grammaticales, filtrées ultérieurement par l'étape de déduction contextuelle. Enfin, les mots grammaticaux sont

(2002) : « Il importe de bien distinguer entre un discours et un texte. Celui-ci n'est qu'une simple trace instrumentale de celui-là, dans laquelle toutes les relations combinatoires qui structurent le discours se trouvent écrasées sur une seule dimension, celle des successivités. ».

[10] Au sens typographique du terme.

porteurs de patrons prosodiques spécifiques car souvent associés à des compressions syllabiques marquées.

A ce stade de l'analyse, certains mots ne sont pas encore étiquetés et d'autres le sont dans plusieurs catégories. C'est là qu'interviennent les **règles de déduction contextuelle**. Au nombre de 60, celles-ci utilisent les informations apportées par l'étiquetage déjà réalisé et les propagent par déduction. Par exemple, si un pronom sujet est suivi d'un mot étiqueté déterminant ou pronom objet, ce dernier est ré-étiqueté pronom objet. Ainsi, dans la phrase *il l'entraîna dehors*, la catégorie du *l'* est précisée par la règle de déduction suivante :

| |
|---|
| (pronom sujet) + ((déterminant) ou (pronom objet))<br>→ (pronom sujet) + (pronom objet) |

Les catégories nouvellement attribuées peuvent à leur tour être utilisées pour d'autres déductions. Mais les règles de déduction contextuelle ne suffisent pas à terminer l'étiquetage, de nombreux mots restant inconnus. Un complément est alors apporté par une base de **suffixes** (600 entrées) fondée sur des indices statistiques. Un suffixe, en effet, est souvent porteur d'une information syntaxique (Boula de Mareüil 1997), y compris pour les néologismes (*foultitude*, *trompage*, etc.), par exemple, la finale *-age* permet d'identifier un nom masculin singulier[11].

A l'issue de ces traitements, les mots non étiquetés sont considérés par défaut comme nominaux (voir *supra* le dictionnaire de verbes).

Concernant la dernière étape, consacrée à la segmentation en constituants syntaxiques et à leur mise en relation, quelques déductions simples effectuées directement par le programme conduisent à regrouper les mots fortement liés syntaxiquement en tronçons (appelés également « chunks », Abney 1992) – ex : déterminant + nom (+ adjectif) ; auxiliaire + verbe, etc. Ainsi la phrase :

*8. L'a.d.n. des hommes célèbres intéresse le président*

est segmentée en 4 tronçons :

*(l'a.d.n.) (des hommes célèbres) (intéresse) (le président)*

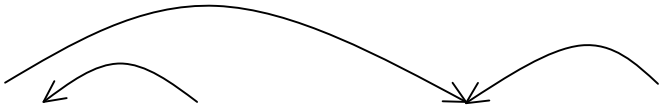
La mise en relation des tronçons entre eux, très importante pour hiérarchiser les frontières prosodiques (cf. *infra* 4.2.), est une opération beaucoup plus complexe, qui fait appel à un jeu de 42 règles. Notre système dispose de différentes mémoires dédiées chacune à une relation syntaxique déterminée (ex : relation sujet-verbe, nom-complément de nom, etc.). Chacune de ces mémoires est gérée comme une pile : de gauche à droite de la chaîne à traiter, chaque nouveau tronçon inséré dans une mémoire est empilé et devient donc le premier candidat pour une éventuelle mise en relation. Un groupe qui n'est plus en attente d'autres groupes (parce que déjà relié) est effacé de la mémoire[12]. Ainsi pour la phrase 9, le premier tronçon nominal (*l'a.d.n.*) est mémorisé comme sujet possible, donc en attente d'un tronçon verbal, le deuxième (*des hommes célèbres*), est relié au premier (relation nom-complément de nom) et retiré de la mémoire. Le tronçon verbal (*intéresse*) valide *l'a.d.n.* comme sujet, celui-ci est donc effacé de la mémoire. Pour le quatrième tronçon (*le président*), construit autour d'un homographe hétérophone, la règle appliquée est la suivante :

*(verbal) + (homographe nom/verbe) → (verbal) + (nominal) (relié -1) {N}*

Autrement dit, le tronçon homographe est redéfini comme nominal (adjonction du marqueur {N} à destination du phonétiseur) et relié au tronçon précédent. A l'issue de la mise en relation, la phrase 8 est représentée de la façon suivante :

[11] Les finales adverbiales en *-ment* sont répertoriées dans cette base.

[12] Cette notion d'attente peut être rapprochée du concept de saturation valancielle chez Tesnière (1959).



*(l'a.d.n) (des hommes célèbres) (intéresse) (le {N} président)*

## 4.2. Le modèle prosodique

Des constituants générés par l'analyse syntaxique, dérive une structure phonologique abstraite formée d'une succession de groupes accentuels : hormis le pronom sujet toujours atone, tout constituant syntaxique donne lieu à la formation d'un groupe accentuel virtuel, défini comme une chaîne de syllabes dont la dernière est frappée par un accent démarcatif :

*9. (Le prési**dent**) (parler**a**) (de**main**) vs. (il parler**a**) (de**main**)*

La question qui se pose est alors la suivante : comment dériver une structure intonative hiérarchisée et rythmiquement bien formée à partir de cette représentation accentuelle de base ? En d'autres termes, quels sont les groupes intonatifs (ici appelés tronçons) effectivement actualisés dans le signal de parole ? En suivant l'hypothèse selon laquelle, pour rappel, la structure intonative s'articule autour de trois types de contraintes fondamentales : textuelles, syntaxiques et rythmiques, l'application de ces contraintes fait émerger six degrés de frontières intonatives.

Dans le détail, les **contraintes textuelles** nous amènent à manipuler trois unités de traitement : le paragraphe, la phrase et le groupe de souffle, ce dernier étant basé principalement sur la ponctuation interne à la phrase[13]. Ces trois unités se caractérisent par une déclinaison et une pause terminale de durée variable (la pause la plus forte est attribuée à l'unité 'paragraphe'). D'où un premier jeu de frontières, hiérarchisées de la façon suivante :

| | |
|---|---|
| *Niveau **1 : FTPg*** | *Frontière Terminale de Paragraphe* |
| *Niveau **2 : FTPh*** | *Frontière Terminale de Phrase* |
| *Niveau **3 : FCGS*** | *Frontière Continuative de groupe de Souffle* |

Les **contraintes d'alignement syntaxique** dérivent du calcul des dépendances syntaxiques (contiguës ou à distance). En reprenant les notations utilisées dans la section précédente, nous illustrons la dépendance contiguë par l'exemple suivant :

*10. (les enfants) (mangent) (leur soupe)*

Dans ce contexte, la relation de contiguïté syntaxique entre un tronçon 'a' et un tronçon 'b' linéairement adjacents s'exprime sur le plan intonatif par une proéminence accentuelle associée à un allongement de la dernière syllabe pleine du tronçon 'a'.

Dans les contextes de dépendance à distance, l'élément régi et l'unité régissante n'entrent pas dans une relation linéaire de contiguïté, comme dans l'exemple suivant :

*11. (les étudiants) ((avaient appris) (en arrivant)) (la triste nouvelle)*

où le circonstanciel *en arrivant* isole le complément d'objet de son régissant. Dans ce contexte, la relation de non-contiguïté syntaxique entre un tronçon 'a' et un tronçon 'b' est marquée intonativement par une proéminence accentuelle associée à un allongement plus prononcé de la dernière syllabe pleine du tronçon 'a'. Autrement dit, les deux degrés accentuels générés sont associés à un **principe**

[13] Unité démarquée à sa droite par une virgule.

**de dominance intonative** qui traduit explicitement les deux types de dépendances syntaxiques (contiguës ou à distance). Deux nouvelles frontières sont ainsi définies :

| | |
|---|---|
| *Niveau **4 : FCGI*** | *Frontière Continuative Majeure de Groupe Intonatif (relation de dépendance à distance)* |
| *Niveau **5 : FcGI*** | *Frontière Continuative Mineure de Groupe Intonatif (relation de contiguïté)* |

En dernier ressort, la prise en compte de l'effort de mémorisation nécessaire pour relier un constituant à distance conduit à insérer une pause dont la durée est proportionnelle au nombre de syllabes à parcourir dans la phrase pour relier l'unité régie à son régissant. Dans la pratique, cette pause n'est effective que si au moins 8 syllabes séparent les groupes qui contractent cette relation, par exemple :

12. *(il promène) (ses enfants) (dans le jardin)*
13. *(il promène) (les enfants) (de Nathalie) (et Vincent) # (dans le jardin)*

Dans *13* l'allongement de la dernière syllabe de *ses en**fants*** est important (FCGI), mais aucune pause n'est insérée, contrairement à *14*, où la distance de 10 syllabes impose un effort de mémorisation plus grand. Nous émettons alors l'hypothèse que lorsqu'une pause est insérée, il n'y a pas de différence fondamentale entre ce type de frontière et celui qui serait généré par l'existence d'une virgule typographique. Nous regroupons ainsi dans la même catégorie (FCGS) tous les groupes de souffle de taille inférieure à la phrase. Les règles suivantes récapitulent les conditions de détermination des niveaux hiérarchiques 3, 4 et 5 :

| | |
|---|---|
| **((tronçon) et (virgule)) → FCGS** | (niveau 3) |
| **(tronçon) + ((tronçon) et (distance[14] >= 8)) → FCGS** | (niveau 3) |
| **(tronçon) + ((tronçon) et (0 < distance < 8)) → FCGI** | (niveau 4) |
| **(tronçon) + ((tronçon) et (distance[15] = 0)) → FcGI** | (niveau 5) |

Considérons maintenant la **contrainte de régulation accentuelle** selon laquelle un groupe constitué d'un nombre de syllabes inaccentuées trop important (4 syllabes ou plus) est accentué sur la première syllabe à attaque consonantique de son premier mot lexical *(u**ne ac**tivité valorisante)*, et donne ainsi lieu à la formation d'un pied métrique[16]. Une dernière frontière est donc enfin définie :

| | |
|---|---|
| *Niveau **6 : FPM*** | *Frontière de Pied Métrique* |

Une synthèse de la hiérarchie intonative ainsi définie est proposée dans le tableau 1 :

[14] Où la distance correspond au nombre de syllabes qui séparent deux tronçons contractant une relation de dépendance syntaxique.

[15] La distance est nulle quand les tronçons en relation sont linéairement adjacents (paramètres de relation +1 ou –1, cf. *supra*, phrase 9).

[16] Pour rappel, on désigne par *pied métrique* une unité prosodique caractérisée par un accent rythmique non terminal.

***Tableau 1.*** *Hiérarchie des frontières intonatives et paramètres phonétiques associés (F0 = fréquence fondamentale)*

| Niveau hiérarchique | Déclinaison (F0) | Pause | Allongement | Proéminence (F0, intensité) |
|---|---|---|---|---|
| **1** | **FTPg** | **FTPg** | **FTPg** | |
| **2** | **FTPh** | **FTPh** | **FTPh** | |
| **3** | **FCGS** | **FCGS** | **FCGS** | **FCGS** |
| **4** | | | **FCGI** | **FCGI** |
| **5** | | | **FcGI** | **FcGI** |
| **6** | | | | **FPM** |

où :

- Les pauses résultent de contraintes typographiques et syntaxiques.
- Une ligne de déclinaison est associée aux groupes intonatifs terminés par des pauses.
- L'allongement caractérise toutes les frontières sauf le pied métrique.
- Les proéminences accentuelles dérivent de principes syntactico-rythmiques.

L'exemple suivant illustre le marquage effectué par les modèles syntaxique et prosodique (la réduction vocalique des mots grammaticaux est symbolisée par une police plus petite) :

*14. Dimanche matin, un autre vol pour la même ville a été annulé.*

*→ **Di**(FPM)manche ma**tin**(FCGS), un autre **vol**(FcGI) pour la même **ville**(FCGI) a été annu**lé**(FTPh).*

Le modèle ainsi construit, les marqueurs abstraits peuvent être transformés en paramètres acoustico-phonétiques adéquats pour la génération du signal synthétique. A chaque marqueur sont associées des variations spécifiques de fréquence fondamentale, d'intensité ou d'allongement vocalique, ainsi que l'insertion de pauses plus ou moins longues.

## Conclusion

Puisque les travaux que nous avons présentés se situent dans le cadre de la recherche appliquée, où les critères d'efficacité, de fiabilité, d'évolutivité et de maintenance sont décisifs, nous ne saurions conclure cette communication sans quelques mots sur l'évaluation qui présente un double avantage : si elle permet de situer les performances du système sous l'angle ergonomique, les erreurs détectées peuvent, en outre, avoir un impact direct sur les attentes du TAL vis à vis de la linguistique et, donc, sur certaines orientations de recherche que pourrait prendre cette dernière. Une évaluation "maison" a pu ainsi être menée, mettant principalement en lumière des erreurs provenant de mauvais regroupements syntaxiques ou de mauvaises mises en relation, ces deux types d'erreurs se situant essentiellement dans des contextes liés à des figements lexicaux non connus par le système. Les exemples suivants montrent quelques-unes des nombreuses erreurs détectées, par ordre de difficulté croissante[17] :

*15. Je dis **ce** que je pense.*
(locution *ce que*)
*16. Il traversa La Rochelle d'un bout # à l'autre.*
(expression figée *d'un bout à l'autre*)
*17. Il n'a pas voulu annoncer sa décision définitive, par égard # pour le Parlement.*
(expression figée *par égard pour*)
*18. La rencontre s'est soldée par un succès # à l'arraché de Kafelnikov.*
(expression figée *succès à l'arraché*)
*19. Plus tard, il apprendra qu'il s'agissait de projectiles # à l'uranium appauvri.*
(expression en voie de figement *projectiles à l'uranium appauvri* ?)

[17] La notion de figement devenant de plus en plus discutable, il s'agit bien là d'un continuum de formes plus ou moins lexicalisées.

20. *Il a réclamé les 10 milliards de dollars d'investissements nécessaires # au développement.*
    (expression *nécessaires au développement* ?)
21. *L'énergie est fournie par une batterie rechargeable # à travers la peau*
    (expression *rechargeable à travers* ?)

Si les locutions oubliées peuvent facilement être placées dans le dictionnaire adéquat (phrase 18), les expressions dites figées sont beaucoup plus nombreuses, plus complexes (phrases 21 à 23) et plus sujettes à discussion (phrase 24, dans laquelle l'expression *rechargeable à travers* n'est pas vraiment courante). Par leurs répercutions sur les liaisons et le positionnement des accents démarcatifs, ces erreurs nuisent à la compréhension et à l'agrément d'écoute. Sur le plan de la recherche linguistique, l'observation de telles erreurs fait apparaître une première piste de recherche : la nécessité de concevoir une interaction étroite entre le lexique et la syntaxe et une meilleure modélisation des phénomènes de figement et de grammaticalisation dans les langues (Mejri 1998).

Lorsque de tels problèmes seront résolus avec une marge d'erreurs acceptable pour l'utilisateur, on pourra affirmer que, parmi les trois fonctions essentielles de la prosodie, posées dans la section 2.2 (fonction démarcative, hiérarchisante et énonciative), la modélisation des deux premières aboutit à un résultat satisfaisant sous l'angle de la synthèse : la génération d'une prosodie parfaitement intelligible. Mais peut-on se contenter d'un tel résultat ? On reste, en effet, dans un contexte minimaliste : sans parler des fonctions expressive et émotive, reconnues depuis longtemps comme faisant partie intégrante de la prosodie (Fonagy 1983) mais qui font appel à la modélisation de paramètres ectolinguistiques que l'on est encore loin de maîtriser, la fonction énonciative liée à la matérialisation linguistique de la structure communicative, demeure la grande absente. Une troisième génération de modèle s'impose donc : celle qui traite les instructions projetées par la structure communicative sur le formatage des objets prosodiques. Dans de tels modèles, l'identification automatique du thème et du rhème, le premier étant défini comme le point de départ psychologique de l'énoncé (ce dont l'énoncé parle), le second désignant la partie informative du message, implique nécessairement d'aborder non seulement les aspects formels, mais également communicatifs et cognitifs – en terme de représentation mentale – qui déterminent de manière cruciale les variations prosodiques. Dans cette perspective, il semble totalement irréaliste d'envisager, à l'instar de la Grammaire Générative, l'autonomie et le primat de la syntaxe pour dériver les unités prosodiques. Autrement dit, ces dernières ne peuvent pas être vues comme des éléments stables et déterminés syntaxiquement une bonne fois pour toutes, qui se combineraient de manière préétablie et immuable au sein des relations linéaires ou séquentielles. Elles sont tout au contraire considérées comme rentrant dans le processus général de construction progressive du sens et, par là même, sont vues comme des entités dynamiques qui ont leur part d'indétermination et se structurent en fonction des relations conceptuelles dans lesquelles elles entrent. Cette approche est par ailleurs conforme à l'objectif de la phonologie contemporaine : « il s'agit toujours de comprendre et d'expliquer les systèmes phonologiques (…) en construisant des dispositifs formels cognitivement pertinents permettant, en production comme en réception, de relier des états mentaux à des productions sonores attestées » (Laks 1997, 9). A la platitude des énoncés encore inhérente aux systèmes de synthèse aujourd'hui, sans relief, ni profondeur et, de ce fait, encore très artificiels, pourront se substituer des messages plus acceptables pour l'utilisateur, parce que tendant vers une voix proche de la réalité humaine. De tels modèles doivent tenir compte à part entière des mécanismes variables de mise en saillance qui indiquent clairement les parties focalisées ou, à l'inverse, parenthétiques de l'énoncé. En particulier, une modélisation fine des paramètres prosodiques permettant d'identifier les divers phénomènes de détachement (extraction topicale : *moi, mon frère, sa maison, il l'a vendue*, clivage focalisant : *c'est Charlotte qui a mangé tout le chocolat*) s'avère dorénavant nécessaire pour atteindre cet objectif. Elle suppose une collaboration soutenue entre la recherche fondamentale à différents niveaux de l'analyse linguistique (sémantique, pragmatique, syntaxe, phonétique), la modélisation informatique et le traitement du signal. Cette troisième approche constitue évidemment une remise en question radicale des paradigmes computationnels classiques et s'inscrit directement dans les nouvelles représentations du langage, ce dernier étant abordé non pas comme un ou plusieurs modules spécifiques et autonomes, mais comme une propriété émergente qui entre dans les mécanismes généraux de la cognition.

## Références


**Abney S. (1992) :** « Prosodic structure, performance structure and phrase structure », *proceedings, Speech and Natural Language Workshop,* Morgan Kaufmann Publishers, San Mateo, CA, 425-428.

**Berrendonner A. (2002) :** « Morpho-syntaxe, pragma-syntaxe, et ambivalences sémantiques », *Macro-syntaxe et macro-sémantique,* H. NØlke (éd.), Bernes, Peter. Lang, à paraître.

**Boula de Mareüil P**. (1997) : *Étude linguistique appliquée à la synthèse de la parole à partir du texte*, Thèse de Doctorat, Université de Paris XI, Orsay.

**Carré R. *& al.* (1991) :** *Langage humain et machine,* Presses du CNRS, Paris.

**Dell F., Hirst D *&* Vergnaud J.R. (1984) :** *Forme sonore du langage, structure des représentation en phonologie,* Paris, Hermann.

**Dell F. (1984) : «** L'accentuation dans les phrases françaises », in F. Dell *&* *al.* (1984 éd.), pp. 65-122.

**Di Cristo A. (1978) :** *De la microprosodie à l'intonosyntaxe*, Thèse d'Etat, Université de Provence, Diffusion Jeanne Laffite, 1985.

**Fonagy Y. (1983) :** *La vive voix*, Paris, Payot.

**Fuchs C. *&* Robert S. (1997) :** *Diversité des langues et représentations cognitives,* Paris, Ophrys.

**Gaudin T (1990) :** *2100, récit du prochain siècle,* Paris, Payot.

**Lacheret A. *&* Beaugendre F. (1999) :** *La prosodie du français,* paris, éditions du CNRS.

**Ladd D.R. (1984)** : « Declination : a Review and some Hypotheses », *Phonology Yearbook,* 1, C.J. Ewen *&* J.M. Anderson (éd.), Londres, pp. 54-61.

**Laks B. (1997) :** *Phonologie accentuelle – Métrique, autosegmentalité et constituance* – Paris, éditions du CNRS.

**Liénard J.S. (1977) :** *Les processus de la communication parlée*, Masson.

**Martin Ph. (1987) :** « Prosodic and rhythmic structures in French », *Linguistics,* vol. 25-5, pp. 925-949.

**Martin Ph. (1999) :** « Prosodie des langues romanes : analyse phonétique et phonologie », *Recherches sur le français parlé,* Publications de l'Université de Provence, vol. 15, pp. 233-253.

**Mejri S. (1998) :** *Le figement lexical,* RLM, Tunis.

**Morel, M.A. *&* Danon-Boileau L. (1998) :** *Grammaire de l'intonation – l'exemple du français,* Paris, Ophrys.

**Morel M. & Lacheret-Dujour A. (2002),** « Le logiciel de synthèse vocale Kali : de la conception à la mise en œuvre », *Traitement Automatique des Langues,* Ch. D'Alessandro (éd.), Paris, Hermès, vol. 42, pp. 193-221.

**Pasdeloup V. (1990) :** *Modèle de règles rythmiques du français appliqué à la synthèse de la parole,* Thèse de Doctorat, Université de Provence 1, Institut de Phonétique d'Aix-en-Provence.

**Pike K. (1945) :** *The Intonation of American English,* Ann Arbor, University of Michigan Press.

**Robert S. (1997) :** « Variation des représentations linguistiques : des unités à l'énoncé », in C. Fuchs *&* S. Robert (éd.), Paris, Ophrys, pp. 25-37.

**Rossi M. *& al.* (1981) :** *L'intonation : de l'acoustique à la sémantique*, Paris, Klincksieck.

**Rossi M. (1979) : «** Les configurations et l'interaction des pentes de F0 et de i », *Travaux de l'Institut de Phonétique d'Aix-en-Provence,* vol. 8, pp. 51-72.

**Rossi M. (1999) :** *L'intonation, le système du français – description et modélisation,* Paris, Ophrys.

**Rossi M. (2002) :** « L'intonation : prémisses, théories, statut », actes des *Journées Prosodie 2001*, V. Aubergé *&* A. Lacheret (éd.), à paraître.

**Tesnière L. (1959) :** *Eléments de syntaxe structurale,* Paris, Klincksieck, édition de 1988.

**Vannier G. (1999) :** *Etude des contributions des structures textuelles et syntaxiques pour la prosodie : application à un système de synthèse vocale à partir du texte,* Thèse de Doctorat, Université de Caen.

**Vergne J. (1999) :** *Etude et modélisation de la syntaxe des langues à l'aide de l'ordinateur, analyse syntaxique automatique non combinatoire,* dossier d'habilitation, Université de Caen.

**Wioland F. (1985) :** *Les structures rythmiques du français,* Paris, Champion.