Une grammaire en tronçons appliquée à la génération de la prosodie

Philippe Boula de Mareüil, Christophe d'Alessandro, Frédéric Beaugendre, Anne Lacheret-Dujour

Une grammaire en tronçons appliquée à la génération de la prosodie

Philippe Boula de Mareüil – Christophe d'Alessandro, Frédéric Beaugendre – Anne Lacheret-Dujour

LIMSI-CNRS
Bâtiment 508 - Université Paris XI, BP 133 F91403 Orsay
{mareuil,cda}@limsi.fr,
frederic.beaugendre@lhs.be, anne.lacheret@info.unicaen.fr

RÉSUMÉ. *Dans cet article, nous décrivons une grammaire en tronçons appliquée au parenthésage prosodique et à la génération de la prosodie en français. Nous présentons un analyseur rapide, robuste et déterministe, qui utilise les informations attachées aux parties du discours et un jeu de règles, pour assigner des frontières et des mouvements prosodiques en synthèse de la parole à partir du texte. L'interface syntaxe-prosodie est exposée : les séquences définies permettent de placer des frontières prosodiques potentielles (mineures, majeures et majeures intermédiaires). Des accents sont ensuite attribués aux mots lexicaux et des règles phonotactiques sont appliquées. Enfin, la description accentuelle est traduite en la réalisation de 9 contours mélodiques (synchronisés avec la structure de surface), de pauses dans certains cas, et d'allongements. Des mesures quantitatives effectuées sur les « tronçons » délimités par les frontières prosodiques ont montré l'avantage de cette grammaire sur une approche plus simple, fondée uniquement sur les mots outils et les signes de ponctuation.*

ABSTRACT. *In this paper, we describe experiments in text chunking for prosodic phrasing and generation in French: we present a quick, robust and deterministic parser which uses part-of-speech information and a set of 20 rules, to consistently assign prosodic boundaries and movements in Text-To-Speech synthesis. The syntax-prosody interface is presented: the sequences enable the location of potential prosodic boundaries (minor, major or mid-major). Stresses are then assigned to lexical words, and stress deletion rules based on phonotactic constraints are applied. Eventually, the accentual description is linked to the realisation of 9 melodic contours (synchronised with the surface structure), pauses in certain cases and lengthening. Quantitative measurements computed on the so-called "chunks" delimited by prosodic boundaries showed the advantage of our chunk grammar over a simpler approach, only based on function words and punctuation.*

MOTS-CLÉS : *grammaire en tronçons, parenthésage prosodique, génération de la prosodie, synthèse de la parole à partir du texte.*

KEYWORDS: *chunk grammar, prosodic phrasing, generation of prosody, text-to-speech synthesis.*

1. Introduction

Il est généralement reconnu que le parenthésage prosodique est lié à la syntaxe : une analyse syntaxique est donc nécessaire, pour un système de synthèse de la parole à partir du texte. Dans cet article, nous décrivons dans ses grands traits un analyseur et son application à la génération de la prosodie en français.

La correspondance entre syntaxe et prosodie a été débattue par de nombreux auteurs (par exemple [VAI 80, MAR 80, CAE 91, MER 00] pour le français). En la matière en effet, il y a interaction entre tous les niveaux de l'analyse linguistique, de la phonétique à la sémantique et à la pragmatique (interprétation dans une situation d'échange d'information). La syntaxe est reflétée dans les fonctions démarcative et modale de la prosodie, mais celle-ci assure également une fonction expressive, et dépend aussi du nombre de syllabes ainsi que de la vitesse d'élocution. Donner à un même énoncé un grand nombre de variantes de contours mélodiques, c'est précisément tout l'art de l'acteur. La prosodie est un mécanisme multidimensionnel, à plusieurs entrées.

La synthèse de parole a joué un rôle fondamental pour les recherches sur la prosodie, car elle a révélé à quel point les fonctions prosodiques étaient variées et fondamentales dans de la parole véritable. Ce domaine, qui avait été plutôt négligé en linguistique sauf dans le cas des langues à tons, réapparu dans toute sa complexité : ainsi, de nombreuses études sur la phonologie et la phonétique de l'intonation se sont constituées pour diverses langues, à partir des années 70, entre autres sous la pression de la piètre qualité d'une parole synthétique qui faisait l'économie de la prosodie.

On peut arguer que la prosodie véhicule la « substance » plus que la « forme ». Cependant, dans l'état actuel de nos connaissances, la machine n'a pas accès au sens, pour le tout-venant des textes. Au contraire, la composante prosodique d'un système de synthèse peut recevoir des informations utiles d'un analyseur syntaxique robuste, rapide et déterministe – une seule façon de lire une phrase est prévue. Pour un système de synthèse, l'enjeu est de rendre compte d'un grand nombre de faits, et non de sélectionner les phrases grammaticales d'une langue, définies comme des candidats éligibles, et de rejeter les autres. Dans ce cadre, la syntaxe (prise ici au sens large) est apparue très vite comme une donnée essentielle pour définir la prosodie d'un énoncé, ou du moins son découpage en constituants prosodiques. Dès 1975, dans [CHO 75, LIE 77] la succession des mots et de leurs catégories est utilisée afin de synthétiser la prosodie dans un système automatique de synthèse de la parole à partir du texte. Il s'agissait alors de prendre en compte la distinction entre mots pleins et mots vides (ou « mots outils »), pour définir les suites de mots écrits qui forment un même groupe prosodique : une analyse superficielle en constituants syntaxiques est ainsi réalisée. Ce procédé s'est révélé très rentable, puisqu'une analyse rudimentaire (une simple liste de mots outils), de complexité très faible, fournit des constituants qui sont souvent tout à fait acceptables. C'est ce type d'approche qui sera utilisé par la suite dans de nombreux systèmes jusqu'à aujourd'hui, mais de façon plus approfondie, et mieux justifiée linguistiquement, dans le cadre des grammaires de dépendance.

Issue des travaux de Tesnière [TES 59], une grammaire de dépendance est bien adaptée pour des applications à grande échelle. Des exemples en sont, pour la synthèse du français, [LAR 89, BAI 89, VER 90, CON 91], ainsi que le système du LIMSI présenté ici. Comme le rappelle Ejerhed [EJE 88], même les analyseurs à large couverture sont d'un intérêt à la fois pratique et scientifique. C'est pourquoi nous voyons aujourd'hui resurgir les techniques empiriques et statistiques en vogue dans les années 50 [CHU 88]. L'apprentissage automatique du parenthésage prosodique a été rendu possible par le fait qu'on dispose maintenant d'importantes quantités de données (par exemple [OST 94, SHA 96, BLA 97, VER 97]).

En comparaison avec d'autres systèmes récents qui utilisent également des analyse syntaxiques pour la prosodie en français, comme [VAN 99, VER 97, DIC 98], nous avons plutôt privilégié les critères structurels par rapport aux probabilités, et une approche intensive, fondée sur des règles plutôt que sur le lexique. Dans ce qui suit, nous proposons une grammaire en tronçons du français, inspirée des grammaires de dépendance (notamment dans le rôle pivot accordé au verbe, siège de la prédication). Nous avons également tenté de décrire de façon pratique et explicite les heuristiques et les règles utilisées pour l'analyse syntaxique et la génération de la prosodie : ainsi, ce travail devrait pouvoir être aisément dupliqué et amélioré par d'autres chercheurs.

Cet article est organisé comme suit. Section 2, l'analyseur « superficiel » (*shallow parser*) est présenté : nous ne discuterons pas la *tokenisation* (segmentation en phrases et en mots) ; des méthodes non lexicalistes sont proposées pour l'étiquetage morpho-syntaxique (ou *tagging*), utilisant un dictionnaire partiel de mots outils, adjectifs antéposables et formes verbales, ainsi que des informations sur les suffixes et des règles de désambiguïsation. Le parenthésage prosodique est ensuite abordé : il consiste à segmenter les phrases en séquences non récursives, définies en termes de catégories possibles.

Section 3, l'interface syntaxe-prosodie est exposée : des règles sont présentées pour l'accentuation et la génération de la prosodie. La méthode préconisée consiste à simplifier la courbe originale d'intonation par des segments de droite élémentaires (sur une échelle temps-fréquence semi-logarithmique) et à classifier ces segments en un nombre restreint de mouvements standard. Cette procédure, initialement proposée pour le néerlandais [HAR 91] (et depuis appliquée à l'allemand, à l'anglais britannique, au russe et à l'arabe), élimine ainsi les détails les moins pertinents. Elle a été conduite en deux étapes (*stylisation* et *standardisation*), sur la base de critères perceptifs : une série de tests a prouvé la validité de cette schématisation [BEA 94]. Pour les durées, le modèle est en comparaison relativement simple.

La section 4 est consacrée à des expériences : des mesures quantitatives sur les tronçons définis par les frontières prosodiques sont présentées, de même qu'une comparaison de notre approche avec une approche uniquement fondée sur les mots outils et la ponctuation. La section 5 discute les résultats et conclut.

2. Analyse syntaxique

2.1. *Étiquetage morpho-syntaxique*

Un analyseur syntaxique robuste, capable de traiter les néologismes et les erreurs d'orthographe ou d'accord, a été proposé dans [VER 90] et repris dans [VAN 99] : il utilise uniquement un dictionnaire partiel. Comme dans une grammaire de contraintes [KAR 90], les règles morpho-syntaxiques résultent de corpus observés. Un parenthésage prosodique utilisant un petit dictionnaire avec des règles sur les suffixes [OSH 87] et/ou identifiant les mots outils [QUE 92, QUA 89] a également été exploré dans la communauté du traitement automatique de la parole, pour l'anglais, le néerlandais et l'italien.

Dans notre cas, le dictionnaire contient :

1. des pronoms,
2. des déterminants,
3. des prépositions,
4. des conjonctions,
5. des adverbes (un millier) auxquels des adverbes en *-ment* ont été ajoutés,
6. des formes verbales (environ 60 000, issues de BDLEX [PER 92]),
7. des adjectifs antéposables (un millier) car les adjectifs en français apparaissent en majorité après le nom, mais 1/3 de cas d'antéposition peut être observé).

Ce dictionnaire est complété par une liste de 340 terminaisons qui permettent de déduire la catégorie grammaticale : par exemple, le suffixe *-ieuse* indique toujours un adjectif féminin singulier. Les mots non identifiés se voient attribuer l'étiquette par défaut *nom* – les noms propres et les sigles notamment.

Si une et une seule catégorie est affectée aux mots, le problème majeur est bien sûr celui de la polycatégorie. Tous les verbes, par exemple, ont la même forme à la 1^{re} et à la 2^{e} personne du singulier, au conditionnel présent ou à l'imparfait de l'indicatif. Des classes mixtes ont donc été introduites. Une centaine d'homonymies (outre celles avec les noms) est également notable, entre adjectifs antéposables, mots outils et formes verbales.

Dans l'ensemble, notre dictionnaire privilégie les adjectifs antéposables par rapport aux mots outils, et les mots outils par rapport aux formes verbales. Par exemple, *célèbre*, qui peut être une forme du verbe *célébrer*, est plutôt considéré comme un adjectif, sur la base d'importants corpus du journal *Le Monde*. De cette manière, un certain équilibre est rétabli par rapport à notre dictionnaire, qui donne un grand poids aux formes verbales.

À l'intérieur de la classe des mots outils, les cas d'homonymie tels que *ce, leur, en, s'* doivent être désambiguïsés. Les étiquettes les plus fréquentes sontt d'abord assignées, dans une phase d'amorce (*bootstrapping*), toujours à partir d'importants corpus

du journal *Le Monde* ; puis d'autres étiquettes possibles sont analysées, en fonction d'ensembles d'étiquettes pour les mots suivants. Ces étiquettes les plus fréquentes sont, dans un ordre décroissant de préférence :

préposition > conjonction > adverbe > déterminant > pronom.

Cette contrainte n'est pas très éloignée de l'heuristique suggérée par J. Vergne dans le cadre de l'action GRACE [ADD 99]. Par exemple, *en* reçoit l'étiquette par défaut *préposition*, et est considéré comme un pronom si le mot suivant est un verbe conjugué : nous entendons par là un verbe à l'indicatif, au subjonctif, au conditionnel ou à l'impératif. Exemple :

elle n'en veut pas.

Examinons plus en détail les cas d'homonymie que représentent *le, la, les, leur, l'*, qui concernent près d'un mot sur dix en discours, et qui est un obstacle notoire pour toute analyse automatique du français. Ils reçoivent l'étiquette par défaut *déterminant*, mais peuvent aussi être des pronoms – normalement placés avant le verbe en français. La désambiguïsation de ces mots (désormais désignés par det/P) suit le principe de l'ensemble des catégories possibles (comme le parenthésage syntaxique que nous verrons ci-dessous), avec un regard en avant. Si le mot suivant est un verbe transitif ou auxiliaire, un pronom personnel complément ou *leur* lui-même, suivi par un verbe transitif ou auxiliaire, le det/P ambigu est considéré comme un pronom. Bien sûr, cette contrainte n'est pas systématique : dans *le manger*, par exemple, *le* peut être un déterminant (*cf.* 4.2.).

Si elle dérange par sa fréquence, cette homonymie est moins grave que celle entre verbe et non-verbe, qui concerne plus de 2 000 entrées différentes de BDLEX avec formes fléchies. Cette ambiguïté a été encodée dans le dictionnaire, de même que l'information *intransitif* à partir de [BES 90]. Ainsi six heuristiques, négatives et à caractère distributionnel, ont-elles été déployées, pour faire basculer un mot d'abord reconnu comme verbe conjugué dans la catégorie *nom*. Provenant d'une analyse de corpus et de recoupements avec des études précédentes [CON 91], elles sont du type utilisé dans les grammaires de contraintes [KAR 90].

Heuristique 1 : après une préposition, il ne peut y avoir un verbe conjugué qui en soit séparé par rien, par le mot *en* ou par une séquence nominale sans nom et sans pronom possessif – on comprendra, dans la sous-section suivante, que l'étiquetage morpho-syntaxique ne présuppose pas le parenthésage en séquences. Exemples :

sans le mauvais sort

avec en poche

Heuristique 2 : après un verbe, il ne peut y avoir un verbe conjugué qui en soit séparé par un det/P, par le mot *en* ou par une séquence nominale sans nom. Exemples :

il voit mal la petite marche
il n'est pas en mesure

Heuristique 3 : immédiatement après un det/P ou le mot *en* en début de phrase, il ne peut y avoir un verbe conjugué, si on n'a pas après un pronom personnel sujet (inversé). Exemples :

La porte étroite vs *la porte-t-il*
En cours, vs *en cours-tu le risque*

Heuristique 4 : immédiatement après un det/P, il ne peut y avoir un verbe intransitif conjugué. Exemple :

et le voyage

Heuristique 5 : immédiatement après un déterminant autre qu'un det/P, il ne peut y avoir un verbe conjugué. Exemple :

mais un avantage indéniable

Heuristique 6 : immédiatement après un adjectif antéposable au pluriel (resp. singulier), il ne peut y avoir un verbe conjugué à la 2^e personne du singulier (resp. à la 3^e personne du pluriel). Exemple :

les petites brises

Les heuristiques 1, 2 et 3 ont la priorité sur la désambiguïsation des det/P et du mot *en*. Exemples :

aimer la danse pour la danse
être en demeure en la demeure

D'autres exemples, qui ne sont pas acceptables, sont fournis dans la section 4.2 (*Évaluation de l'analyseur syntaxique*).

2.2. *Parenthésage syntaxique*

Comme l'étiquetage morpho-syntaxique, le parenthésage syntaxique tire son inspiration des travaux de Vergne [VER 90], repris dans [VAN 99]. Les phrases sont découpées en séquences nominales, verbales et « transjonctives ». Le terme générique de « transjonctif », que nous introduisons en référence à la *translation* et à la *jonction* de Tesnière, englobe les prépositions, les conjonctions, les pronoms relatifs, certains adverbes et signes de ponctuation comme la virgule et les parenthèses. Rappelons que la *connection* (*i.e.* le lien qui existe entre deux mots), la *jonction* (juxtaposition ou coordination) et la *translation* (éclairant les compléments du nom et les propositions relatives) sont les structures syntaxiques fondamentales de Tesnière.

Les séquences sont faites de mots contigus, et ne sont pas récursives. Par exemple, « une belle vue de Paris » est décomposé en trois séquences : « une belle vue » (séquence nominale), « de » (séquence transjonctive) et « Paris » (séquence nominale). De cette non récursivité, nous voulons comme justification (psycho)linguistique le fait que l'enchâssement est limité dans la langue. Les propriétés récursives du langage sont d'ailleurs sujettes à caution : comme l'écrit P. Mertens : « Souvent un locuteur entame une phrase sans savoir comment elle finira et dès lors sans avoir à l'esprit sa structure syntaxique entière. » [MER 97]. En outre, les dépendances entre les séquences représentent un problème complexe, pouvant demander un accès au contenu lexical ou à la sémantique : on peut avoir des dépendances lointaines ; on peut coordonner des sujets, des verbes, des objets et des phrases. Notre choix est donc aussi et surtout guidé par des raisons de simplification du calcul.

Nous nous sommes cantonnés aux dépendances entre les mots à l'intérieur des séquences, ce qui nous rapproche des *chunk grammars* [ABN 91], grammaires en tronçons qui aboutissent à un *partial parsing*. La grammaire en tronçons consiste simplement à diviser la phrase en segments. Elle est en partie inspirée d'études en psychologie sur la durée des pauses, en lecture, et sur la structuration « naïve » de phrases. Fondée sur une analyse assez superficielle, non exhaustive, sa motivation est également procédurale. Si elle diffère les difficiles décisions d'attachement à une étape ultérieure, cette grammaire peut servir à la découverte d'unités de traduction, à l'extraction d'information ou à la génération automatique d'index : dans ce domaine, la plupart des efforts se sont concentrés sur l'identification des groupes nominaux de base [RAL 95, ANB 96]. Semblables aux techniques utilisées dans [RAL 95] (issues de Brill [BRI 93]), des arbres de classification et de régression (CART) ont également été appliqués dans [HIR 96], pour positionner des frontières intonatives.

Les grammaires en tronçons proposées dans la littérature, qu'elles soient probabilistes ou par règles, intègrent des termes coordonnés ou certains syntagmes prépositionnels, prenant ainsi une décision de rattachement. Pour la synthèse de la parole, il semble illusoire de désirer énumérer la totalité des séquences possibles. Celles-ci peuvent être assez longues (lors d'une construction disloquée avec un verbe « à montée » notamment), et, inévitablement, certaines nous échappent. De surcroît, nous pouvons avoir une approche plus tolérante qu'en génération, suivant en cela [CON 91], qui ne parle pas de « séquences » mais de « bandes généralisées » nominales et verbales. La bande nominale généralisée (BNG) se définit comme « suite de mots comprise entre deux mots du type jonctif, translatif ou bien verbe » ; et « la bande verbale représente le verbe et les différents éléments qu'il gouverne localement » [CON 91].

Dans notre cas, les séquences sont définies par des ensembles de catégories possibles (*cf*. tableau 1). Ceci peut certes être représenté par des règles de réécriture. Cependant, exprimer les séquences en termes d'ensembles de catégories possibles est beaucoup plus simple et plus concis, puisqu'elles ne correspondent qu'à un niveau de parenthésage.

Les ensembles utilisés dans la définition des séquences ne sont pas disjoints : la plupart des adverbes, par exemple, peuvent appartenir aux trois types de séquences.

<table>
<tr><th>séquence nominale</th><th>séquence verbale</th><th>séquence transjonctive</th></tr>
<tr><td>nom
adjectif
(pré)déterminant
pronom possessif
adverbe d'adjectif</td><td>verbe conjugué
infinitif
négation
pronom personnel
pronom adverbial
pronom indéfini
pronom démonstratif</td><td>préposition
conjonction
pronom relatif
ponctuation (, -)</td></tr>
<tr><td colspan="2">participe</td></tr>
<tr><td colspan="3">adverbe (non de négation ni d'adjectif)</td></tr>
</table>

Tableau 1. *Définition des catégories possibles dans les séquences nominales, verbales et transjonctives.*

C'est le premier mot de la séquence qui décide, par propagation gauche-droite, les séquencess étant examinées dans l'ordre transjonctive-nominale-verbale. Ainsi, si un adverbe (non de négation ni d'adjectif) est en début de phrase, il ouvre une séquence transjonctive. Un adverbe (autre que *pas* et *point*) est un adverbe d'adjectif si le mot suivant immédiatement est un adjectif.

Une table indiquant qu'entre deux catégories successives (dont la première peut être « début de phrase »), on passe d'un type de séquence à un autre, ne peut générer cette analyse, pas plus qu'un algorithme tel que *chinks 'n chunks* [LIB 92], qui n'est qu'une détection modifiée de mots outils.

On note que la classe traditionnelle des pronoms a été subdivisée en plusieurs catégories : les pronoms possessifs (dans les séquences nominales), les pronoms personnels, adverbiaux (*en* et *y*), indéfinis ou démonstratifs (dans les séquences verbales) et les pronoms relatifs (dans les séquences transjonctives). Outre le fait que dans certaines langues (telles que l'espagnol, l'italien ou l'arabe), le pronom personnel sujet est facultatif, et qu'en français le sujet n'est qu'apparent dans des phrases comme « il pleut (des cordes) », deux arguments nous ont semblé militer en faveur d'un rattachement du pronom personnel, adverbial, indéfini ou démonstratif à la séquence verbale. D'une part le pronom (aussi bien sujet que complément) peut s'insérer au milieu d'un groupe verbal (ex. *Paul ne t'a pas vu, as-tu vu Paul ?*). D'autre part, il n'est souvent que la reprise anaphorique du sujet (ex. *mon père, il a vu Paul...*).

Pour **chaque** phrase en entrée, l'analyse peut fournir **une** partition de la chaîne écrite en trois types de séquences (nominales, verbales et transjonctives) qui ne se chevauchent pas, ainsi qu'un alignement de mots et de parties du discours. Cette opération dirigée par les données (*data-driven*) utilise des contraintes locales, faciles à implémenter dans un automate d'états finis. L'algorithme procède phrase par phrase, et est de complexité linéaire par rapport au nombre de mots.

3. Interface syntaxe-prosodie

3.1. *Une methodologie ascendante*

La sortie de l'analyseur syntaxique, qui comprend une suite de séquences représentant la phrase donnée en entrée ainsi que la catégorie grammaticale de chaque mot et la modalité (assertive ou interrogative), est connectée à des règles prosodiques, comme décrit figure 1. Les règles pour la génération automatique de la prosodie sont organisées en trois modules :

– un module syntaxique :

- pour délimiter des unités prosodiques virtuelles, de taille variable, et leur associer une frontière spécifique,

- pour fournir les catégories morpho-syntaxiques qui serviront à générer les accents ;

– un module phonotactique, pour prendre en compte les contraintes rythmiques et les phénomènes de désaccentuation (voir en particulier les règles 10 et 11) ;

– un module phonético-acoustique, permettant de lier la structure prosodique de surface aux paramètres de mouvements mélodiques, pauses dans certains cas, et allongement.

Cet ensemble de règles a été élaboré en suivant une méthodologie inductive (*bottom-up*), à partir de l'inspection d'un corpus d'apprentissage de 220 phrases isolées. La construction de ce corpus a pris en considération des contraintes syntaxiques (modalité, inversion, dislocation, nature et fonction des groupes), morphologiques (structure des mots), distributionnelles (position des mots), phonotactiques (nombre de syllabes) et phonétiques (il a été évité de faire commencer un mot par une occlusive sourde, afin de clairement distinguer les pauses). Ce corpus a été lu par un locuteur parisien, à un débit d'élocution « normal », et avec une intonation « neutre » (sans emphase, qui ne véhicule pas d'émotion) : ainsi le style de prosodie est-il simple, et en relative adéquation avec la syntaxe. Des mesures de moyenne au sens statistique ont été effectuées, et différentes expériences ont été conduites, où l'on demandait à des sujets de comparer des stimuli, ou de transcrire les proéminences accentuelles perçues.

Cette section décrit les contraintes syntaxiques et phonotactiques utilisées pour enrichir la chaîne phonématique de marqueurs pour le parenthésage prosodique et le calcul de la structure accentuelle. La stratégie considérée se situe entre les propositions extrêmes qui prônent l'une une relation bijective totale entre structures syntaxiques et prosodiques, l'autre l'indépendance complète des deux, supposant la prosodie entièrement dirigée par les contraintes rythmiques.

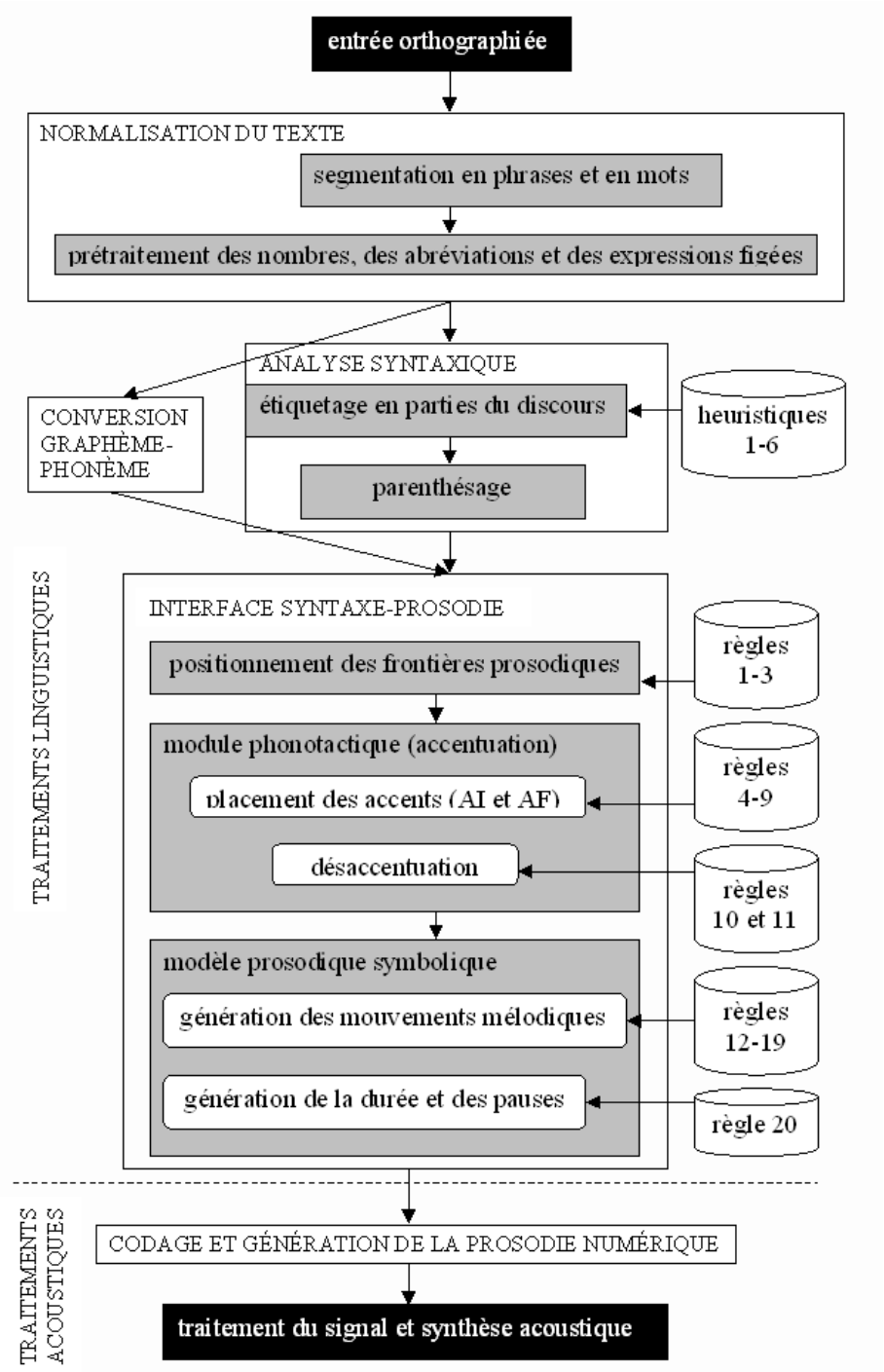

Figure 1. *Diagramme bloc du système de synthèse du LIMSI*

3.2. *Des séquences aux tronçons définis par des frontières prosodiques*

Quatre types de frontières prosodiques sont définis :

– des frontières mineures, #fm, ou intermédiaires, #FI, à la fin des sujets nominaux, des syntagmes verbaux et des compléments ;

– des frontières majeures, #FM (continuation majeure + pause), après un signe de ponctuation faible ;

– des frontières terminales, #FT, qui peuvent être réalisées comme montantes (questions) ou descendantes, en fin de phrase – c'est l'intonème conclusif, dans la terminologie de Rossi [ROS 93].

Les règles syntactico-prosodiques pour le parenthésage prosodique distribuent ces frontières de la façon suivante :

Règle 1 : une frontière de continuation mineure (#fm) est placée à la fin des séquences nominales et verbales, quand celles-ci ne sont pas suivies d'un signe de ponctuation.

Règle 2 : des marqueurs prosodiques sont associés aux signes de ponctuation faible (#FM) et fort (#FT).

Règle 3 : une frontière majeure intermédiaire est définie (#FI), qui atteint le même niveau de hauteur que #FM, mais sans la pause qui lui est associée.

Ces frontières intermédiaires ne semblent pas obéir à des lois syntaxiques fixes, du moins dans le corpus utilisé dans ce travail [BEA 94]. Cependant, les sept types de tronçon suivants sont presque toujours suivis par #FI :

préposition + nom,
conjonction de coordination + nom,
début de phrase + déterminant + adjectif + nom,
déterminant + nom + adjectif,
préposition + pronom personnel,
conjonction de coordination + déterminant + nom,
préposition + nom + nom (ex. *de François Mitterrand*).

Les exemples ci-dessous illustrent le parenthésage prosodique obtenu avec ces règles (SN = séquence nominale ; ST = séquence transjonctive ; SV = séquence verbale) :

(Ils chantent)$_{SV}$ #fm
(une cantate)$_{SN}$ #fm
(de)$_{ST}$ (Jean Sébastien Bach.)$_{SN}$ #FT
(Il est bien arrivé)$_{SV}$ #fm

(à)$_{ST}$ (Paris)$_{NS}$. #FT

(Un hommage international)$_{SN}$ #fm
(de)$_{ST}$ (première grandeur)$_{SN}$ #fm
(doit être rendu)$_{SV}$ #fm
(à)$_{ST}$ (la nouvelle Roumanie)$_{SN}$ #fm
(et à)$_{ST}$ (son peuple)$_{SN}$. #FT

(Personne ne leur a enseigné)$_{SV}$ #fm
(la manière)$_{SN}$ #fm
(de)$_{ST}$(se présenter)$_{SV}$(,)$_{ST}$ #FM
(et)$_{ST}$(ils n'ont généralement pas)$_{SV}$ #fm
(les moyens)$_{NS}$ #fm
(d')$_{ST}$(être correctement habillés)$_{SV}$. #FT

S'il existe à l'intérieur de chaque ligne (*i.e.* chaque tronçon) une grande cohésion, la lecture de phrases ainsi découpées ressemble à celle qu'en ferait un enfant encore hésitant. Les tronçons définis par ces frontières, obtenus sur des critères purement syntaxiques (donc intermédiaires dans les différents niveaux de l'analyse linguistique), ne sauraient définir des groupes de souffle séparés par des pauses. Pour les rattacher à la prosodie, il faut notamment les ajuster avec le nombre de syllabes, selon [BEA 94].

3.3. *Accentuation*

Après la segmentation primaire en constituants syntaxiques, le traitement prosodique requiert trois étapes :

– comptage du nombre de mots accentuables par groupe, identification des mots « pleins » ;

– activation des règles de placement et d'effacement d'accents initiaux (AI) et finaux (AF) ;

– passage de la structure accentuelle schématisée à la réalisation de contours mélodiques (enchaînement des mouvements).

L'accentuation est calculée sur la base des catégories des mots. Les mots pleins, qui peuvent être frappés par un accent, sont : les verbes principaux, les noms, les adjectifs, les adverbes (non de négation) et les pronoms personnels suivis d'une virgule (voir la règle 5 ci-dessous). Les autres catégories, c'est-à-dire les déterminants, les conjonctions, les prépositions, les négations, les verbes auxiliaires et les autres pronoms sont considérés comme des clitiques : ils n'ont pas d'autonomie suprasegmentale et ne peuvent être accentués.

Les règles suivantes gouvernent l'assignation de l'accent, lequel est à relier avec l'organisation linguistique spécifique de la phrase. Le jeu de règles prosodiques développé par [BEA 94], pour un débit de parole normal, a été adapté. Les unités considérées sont les groupes prosodiques, caractérisés en termes d'AI et AF, comme dans

[MER 93]. En français en effet, l'accent a une place fixe dans les mots lexicaux, mais ceux-ci peuvent être inaccentués : c'est un accent de groupe plus qu'un accent de mot. Un exemple d'accentuation pour une courte phrase est :

Les enfants discutaient de leurs vacances.

le z $\underbrace{[\ \tilde{a}]}_{AI}$ $\underbrace{[\ f\tilde{a}]}_{AF}$ dis ky $\underbrace{[\ t\varepsilon]}_{AF}$ də lœʀ va $\underbrace{[\ k\tilde{a}s]}_{AF}$

Règle 4 : par défaut, les syllabes finales des mots pleins sont accentuées.

Règle 5 : un mot clitique avant une frontière prosodique étant aussi accentuable (ex. *LUI, il...*), un AF est placé sur la dernière finale ferme de chaque groupe prosodique.

Règle 6 : un AI est placé sur la syllabe initiale du premier mot plein du groupe. Si le premier phonème est une voyelle, l'accent est déplacé sur la syllabe suivante. Cependant, trois exceptions sont ajoutées (règles 7, 8 et 9).

Règle 7 : en début de phrase, une première syllabe commençant par une voyelle peut être accentuée (comme si elle suivait un coup de glotte).

Règle 8 : la première syllabe d'un mot plein commençant par une voyelle peut être accentuée, si elle est précédée d'une liaison.

Règle 9 : un mot plein de plus de 4 syllabes reçoit un AI, même s'il ne se trouve pas au début d'un groupe prosodique. Cet AI correspond à l'*ictus* mélodique décrit par Rossi [ROS 93] et Pasdeloup [PAS 90].

Règle 10 : lorsqu'un AF est suivi d'un AI sur des syllabes consécutives, l'AI est désaccentué, sauf si une pause les sépare – dans ce cas, les deux accents sont réalisés.

Règle 11 : quand on est en présence de deux accents adjacents, séparés par des frontières :

– de même type, le second accent est supprimé ;

– de nature différente, l'accent précédant la frontière la plus forte est maintenu, l'autre est éliminé.

3.4. *Mélodie*

De nombreuses méthodes ont été proposées pour la génération de la mélodie, dans le cadre de la synthèse de la parole à partir du texte (pour une revue détaillée de ces méthodes, on pourra consulter [LAC 99]). Bien qu'il soit difficile de classer l'ensemble de ces méthodes, on peut tenter de les regrouper en trois grandes classes :

– **les modèles phonologiques**. Ces modèles sont souvent réalisés du point de vue phonétique par l'intermédiaire de points cibles, qui décrivent des événements linguis-

tiques, comme ToBI (*Tones and Break Indices*) [SIL 92] ou des configurations acoustiques, comme le modèle MOMEL associé au système de desription phonologique INTSINT (*INternational Transcription System for INTonation*) [HIR 98, DIC 98] ;

– **les modèles de superposition**. Dans ces modèles, des commandes associées à divers niveaux de structuration prosodique se superposent [FUJ 84] pour former les contours. Ceux-ci peuvent aussi être obtenus par concaténation/superposition d'unités stockées en mémoire, dans un « lexique prosodique » contenant des contours prototypes extraits d'un corpus [AUB 91, MOR 97] ;

– **les modèles phonétiques**. Les contours de fréquence fondamentale (F_0) sont construits au moyen de prototypes mélodiques, en prélevant des valeurs prosodiques à partir d'un corpus qu'il s'agit de préparer soigneusement, par des règles ou des réseaux de neurones [LAR 89, TRA92]. Notre approche entre dans ce cadre.

Dans cette étude, une description de la mélodie en termes de mouvements de hauteur a été adoptée. Ce sont les unités intonatives de base qui doivent être reliées directement à l'analyse syntactico-prosodique. Il s'agit de l'application au français de la méthode décrite dans [HAR 91], qui a été appliquée à plusieurs autres langues, pour la synthèse de la parole en particulier. Elle est acoustique plutôt que phonologique : nous pensons que cela est plus approprié pour notre application pratique.

L'approche consiste à négliger toute variation de F_0 inférieure à un certain seuil. Les patrons intonatifs peuvent être réduits à des segments de droite : des *close copies* (*i.e.* des copies sans dégradation audible de l'original) sont obtenues en interpolant avec un minimum de segments de droite (sur une échelle logarithmique), entre un minimum de points perceptivement pertinents de la courbe de F_0 [HAR 91]. Cette stylisation ne peut pas, à l'oreille, être distinguée des patrons originaux. L'étape suivante ramène les segments de droite à un petit nombre de mouvements *standard*, qui ont un rôle fonctionnel : des différences entre l'original et les contours de hauteur faits de mouvements standard sont susceptibles d'être identifiées, mais la structure intonative est inchangée.

L'application de cette méthodologie au français [BEA 94] a conduit à une classification en neuf mouvements (cinq montées, trois descentes, un plateau), illustrés figure 2. Ces mouvements sont synchronisés avec la structure segmentale ; quant au registre mélodique global, il est défini par un gabarit constitué d'une ligne haute et d'une ligne basse déclinantes, avec une pente plus forte pour la première.

Un ensemble de règles (12 à 19) a été défini pour relier l'accentuation aux mouvements mélodiques, comme suit :

Règle 12 : si un AI est séparé d'un AF par moins de deux syllabes, il correspond à un mouvement R_5. Dans tous les autres cas, il est réalisé par R_2.

Règle 13 : un AF suivi par une frontière mineure est réalisé par R_4 dans le cas où il est séparé d'un AI par moins de deux syllabes. Il est réalisé par R_3 dans les autres cas.

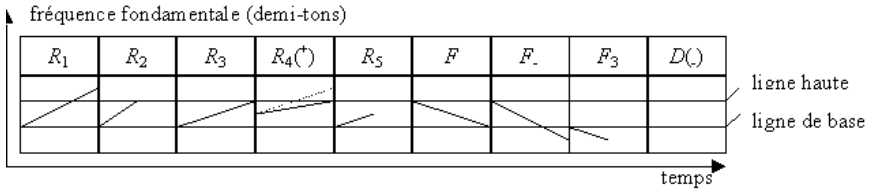

Figure 2. *Illustration des neuf mouvements standard pour la description mélodique du français. On peut voir dans leurs points d'arrivée cinq niveaux de hauteur (sur, entre, en dessus ou en dessous des lignes de gabarit)*

Règle 14 : un AF suivi par une frontière majeure est réalisé par R_4^+ dans le cas où il est séparé d'un AI par moins de deux syllabes. Il est réalisé par R_1 dans les autres cas.

Règle 15 : un mouvement descendant F est utilisé pour revenir au registre mélodique de base. Il est situé après chaque mouvement montant, sauf R_5, qui précède toujours soit R_4 soit R_4^+. Il se poursuit jusqu'au mouvement montant suivant, ou jusqu'au plateau D suivant (voir règle 19).

Règle 16 : une fin de phrase déclarative reçoit un mouvement F_3 sur le groupe consonantique de la dernière syllabe, à condition que celle-ci soit séparée par au moins deux syllabes inaccentuées de la dernière montée mélodique ; dans le cas contraire, le dernier mouvement est F_-, directement lié à la dernière montée mélodique (*cf.* figure 2).

Règle 17 : une fin de phrase interrogative reçoit un mouvement R_1 sur la dernière syllabe.

Règle 18 : si le dernier AI d'une phrase déclarative est réalisé, R_2 est remplacé par R_5.

Règle 19 : à l'intérieur d'une phrase, si deux montées mélodiques sont séparées par plus de deux syllabes inaccentuées, un plateau D est intercalé sur la syllabe précédant le second mouvement.

Le tableau 2 résume les types de mouvements qui peuvent être associés aux AI et AF, ainsi que la correspondance entre les types de frontière, les accents et les mouvements.

Le tableau 3, lui, met en évidence un exemple d'analyse : les frontières prosodiques sont indiquées.

L'exemple suivant illustre la transcription phonétique et les mouvements mélodiques correspondants, comme ils apparaissent en sortie des modules de traitement

AI	AF		
	#FM (ou #FI)	#fm	#FT
R_2	R_1	R_3	$F\_$
R_5	R_4^+	R_4	F_3
	F	F	R_1

Tableau 2. *Mouvements standard associés aux AI et AF – avant une frontière majeure (#FM ou #FI), une frontière mineure (#fm) ou une frontière terminale (#FT)*

(*Ils sont*	*partis*) #fm	(*pour*	*Paris.*) #FT
il sɔ̃	paʀti	puʀ	paʀi
D	$R_5\ R_4$	F	$R_2\ F\_$

Tableau 3. *Mouvements générés pour la phrase* ils sont partis pour Paris.

linguistique du système de synthèse. Les mouvements sont séparés du dernier phonème couvert par /, et le signe # après R_1 désigne une pause :

Les alcooliques de ma ville ont suivi une cure de désintoxication qui, d'après eux, a été très douloureuse.

début de phrase + det. + nom #fm prep. + det. + nom #fm verbe auxiliaire + participe #fm det. + nom #fm prep. + nom #FI pronom relatif + signe de ponctuation faible #FM prep. + pronom personnel + signe de ponctuation faible #FM verbe auxiliaire + participe #fm adv. + adj. + signe de ponctuation forte #FT

le/Dz/R_5alkɔli/R_4kdə/Fma/Dvi/R_3lɔ̃/F sɥ/R_5ivi/R_4y/Fnky/R_3ʀdə/Fd/R_2

ezɛ̃ tɔksikasjɔ̃ki#da/Fpʀɛ/Dzø/R_1# ae/Ft/R_5e/R_4tʀɛdulu/Fʀ/F_3øz./$D\_$

3.5. *Durée*

La génération des durées des phonèmes est liée aux aspects segmentaux intrinsèques, à des facteurs co-intrinsèques (influence d'un phonème adjacent) et à des faits prosodiques comme l'allongement syllabique. Les systèmes de prédiction automatique de la durée des phonèmes peuvent être classifiés comme suit (restreints aux méthodes éprouvées pour le français) :

– **méthodes fondées sur les phonèmes**. Dans un modèle multiplicatif, la durée d'un phonème spécifique est obtenue à partir de sa durée intrinsèque mesurée sur un corpus de logatomes (CVC et VCV), d'un certain nombre de coefficients prenant en compte les effets contextuels (allongement des fins de mots ou avant une consonne allongeante, réduction des durées de groupes consonantiques, par exemple) et d'autres informations concernant la structure prosodique de l'énoncé [BAR 87, TZO 95].

– **méthodes fondées sur les syllabes** [BAR 94]. La durée relative des unités rythmiques (dans ce cas le **G**roupe **I**nter-**C**entre-**P**erceptif) augmente tout au long du

frontière	nombre de syllabes	nombre de syllabes
	> 10	< 10
#fm (par defaut)	$\times$ 1,5	$\times$ 1,15
#FI	$\times$ 1,9	$\times$ 1,5
#FM	$\times$ 1,9 + pause de 250 ms	$\times$ 1,5 + pause de 150 ms
#FT	$\times$ 2,1 + pause de 500 ms	

Tableau 4. *Allongement et pauses potentielles, pour un type de frontière donné, que le nombre de syllabes dépasse 10 syllabes ou non*

groupe prosodique, pour atteindre un maximum à la fin. La durée de chaque phonème dans le GICP est alors calculée selon l'algorithme proposé par [CAM 92].

– **méthodes statistiques** [KEL 97].

Le modèle de durée développé dans cette étude est plutôt sommaire et vise simplement à générer des allongements et des pauses en fonction de la structure prosodique globale. Une étude plus récente [VAN 99], observant qu'un tiers des pauses en lecture n'est pas marqué par la ponctuation, inclut des pauses en prenant appui sur les relations de dépendance entre éléments linéairement adjacents et sur la force des frontières syntaxiques.

Règle 20 : chaque syllabe accentuée et chaque syllabe précédant une frontière prosodique sont allongées – cet allongement peut parfois être le seul corrélat acoustique de l'accent.

La durée de la syllabe est multipliée par un coefficient qui dépend de la nature de la frontière et de la longueur de la phrase (supérieure ou non à dix syllabes). Une pause est placée après un signe de ponctuation : les valeurs numériques sont reportées dans le tableau 4.

4. Expériences

L'analyseur syntaxique et les règles prosodiques ont été implémentés dans le système de synthèse de la parole du LIMSI, avec lequel d'importants corpus ont été synthétisés. Les paragraphes suivants fournissent des données quantitatives et des évaluations.

4.1. *Mesures quantitatives sur les tronçons*

Même si l'étiqueteur morpho-syntaxique lui-même prévoit un jeu de 41 catégories (venant de ce que les mots de la langue appartiennent souvent à plusieurs parties du discours), les règles syntactico-prosodiques utilisent un ensemble d'étiquettes plus

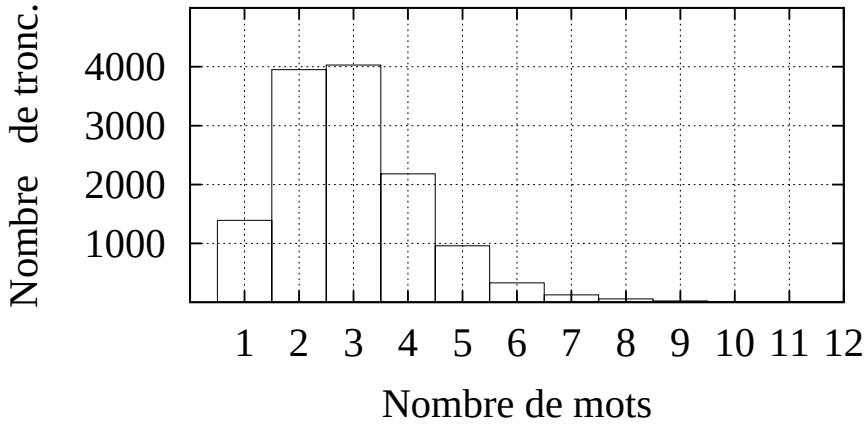

Figure 3. *Nombre de tronçons d'une taille donnée, exprimée en nombre de mots*

restreint, qui est résumé dans le tableau 1 : *nom* (incluant les mots *uns*, *unes* et les pronoms possessifs), *adjectif* (antéposable ou déduit par la terminaison + « -ci », « -là », « -même »), *(pré)déterminant, verbe principal conjugué, verbe auxiliaire conjugué, infinitif, participe, pronom de couleur verbale, adverbe (non de négation), négation, conjonction de coordination, préposition, pronom relatif ou conjonction de subordination, signe de ponctuation faible, signe de ponctuation forte*.

Un corpus de 192 000 mots a été analysé (et synthétisé) en utilisant ces quinze classes (et un marqueur supplémentaire de début de phrase). Nous avons obtenu 65 000 tronçons, dont 5 000 différents. Rappelons que ce sont les frontières prosodiques qui délimitent et par là même définissent les tronçons. Le nombre de mots dans chaque tronçon est reporté dans la figure 3. Il apparaît que la taille moyenne d'un tronçon est de trois mots, ce qui correspond grosso modo à la taille moyenne d'un « mot prosodique » [DEL 95], souvent de l'ordre de quatre ou cinq syllabes. On obtient un grand nombre de types de tronçons différents : la combinatoire théorique est importante, même si certaines étiquettes ne peuvent se trouver qu'à certaines positions dans le tronçon, ou à l'exclusion d'autres étiquettes : un nom et un verbe, par exemple, ne peuvent appartenir à un même tronçon.

Dans un texte dont on a extrait les mots différents, les mots fréquents sont en petit nombre : ce sont les mots outils. De même ici, les tronçons fréquents sont peu nombreux et, au contraire, on a beaucoup de tronçons rares – pouvant être considérés comme du bruit, provenant en partie d' « accidents », ou comme le témoin d'une richesse de structures. La figure 4 montre ce fait : la plupart des types de tronçon ne se rencontrent qu'une seule fois, et peu de types de tronçon se rencontrent très souvent.

Les types de tronçon les plus couramment observés sont dans l'ordre :

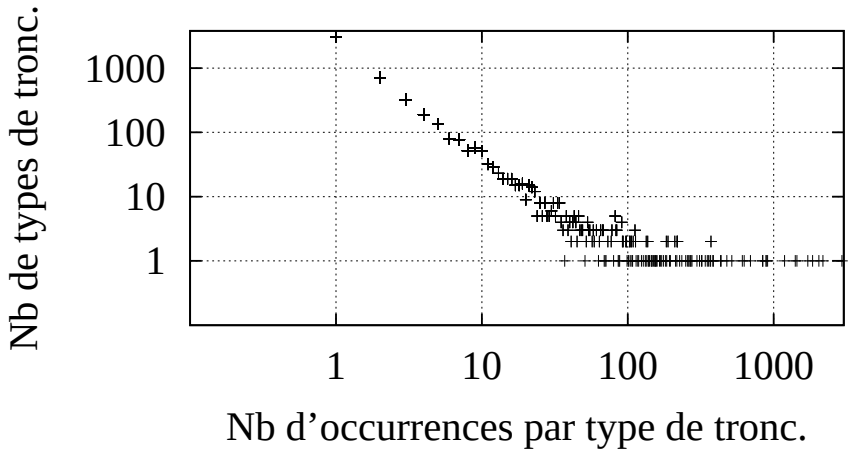

Figure 4. *Répartition des tronçons, selon leur fréquence. En abscisse est représenté le nombre N d'occurrences d'un type de tronçon donné. En ordonnée est représenté le nombre de types de tronçon qui apparaissent N fois*

préposition + nom
déterminant + nom
verbe principal conjugué
préposition + déterminant + nom
préposition + nom + signe de ponctuation faible
préposition + nom + signe de ponctuation forte
nom

Les chiffres sont bien entendu tributaires du nombre de ces étiquettes et de la tokenisation des mots (par exemple, *29 UDF* est compté comme cinq mots). Précisons aussi que les amalgames *du, des, au, aux*, etc. sont étiquetés comme prépositions. Cependant, le résultat suivant nous semble remarquable : les 18 types de tronçon différents les plus fréquents couvrent 38 % des occurrences. De cette façon, on peut se focaliser sur eux pour perfectionner les règles syntactico-prosodiques.

4.2. *Évaluation de l'analyseur syntaxique*

Pour évaluer l'analyseur en n'utilisant que des critères syntaxiques, un premier test a été conduit sur le corpus GRACE [ADD 99], qui est constitué de 302 phrases : 90 phrases du journal *Le Monde* et 212 phrases construites pour concentrer des difficultés syntaxiques. Les 4 547 mots de ce corpus ont été étiquetés par un expert.

Le critère d'évaluation que nous avons retenu est particulièrement sévère. Il est en un sens plus défavorable que le paradigme originel de GRACE, où le silence est

c_i / c'_i	**T**	**N**	**V**
T	”	T T N V # T N	T T V N # T V
N	T N N V # N N	”	T N # T
V	T V V N # V V	T V # T	”

Tableau 5. *Cas où le remplacement de c_i par c'_i – parmi les valeurs T (transjonctif), N (nominal), V (verbal) – n'a aucune incidence sur les frontières de tronçons (#) dans des suites $c_{i-1}c_ic_{i+1}$*

permis, où les expressions figées peuvent être éclatées, et où une phase d'*adjudication* est prévue. Ce n'est pas le cas ici.

Nous avons étudié les répercussions des erreurs de catégorie sur les séquences et les tronçons. Le taux d'erreur d'étiquetage est de 10 %. Toutefois, le taux d'erreurs engendré sur les séquences n'est que de 3 %. Le parenthésage syntaxique a été estimé en termes de rappel (*recall*) et de précision, par rapport à une segmentation « naïve ». Le rappel représente le rapport entre le nombre de séquences correctement prédites et du nombre total de séquences attendues, tandis que la précision représente le rapport entre le nombre de séquences correctement prédites et le nombre total de séquences prédites.

Ces deux mesures, bien connues dans le domaine de la recherche d'information, pénalisent respectivement l'oubli de séquences et l'extraction de séquences incorrectes (le bruit). Ces deux taux tournent ici autour de 97 %. On s'aperçoit en fait que les erreurs de parenthésage proviennent moins des erreurs d'étiquetage que des expressions figées et de l'inclusion de participes dans les séquences nominales (ex. **(jusqu'à)$_{ST}$ (présent conçues)$_{SN}$*). Des erreurs entre nom, adjectif et déterminant, notamment, n'ont aucune incidence sur le type de la séquence, pas plus que celles entre préposition, conjonction et pronom relatif. Des erreurs sur les adverbes et les participes peuvent aussi être sans impact sur le type de séquence, qui dépend du contexte gauche.

Dans certaines configurations, même, une erreur sur le type de séquence n'a aucun effet sur la frontière entre tronçons. Des exemples de telles situations sont donnés dans le tableau 5, où c_i représente le type de séquence produit par l'analyse, pour un mot donné w_i, et c'_i est le type de séquence correspondant à la catégorie grammaticale de référence. Les valeurs que peuvent prendre c_i et c'_i sont T (transjunctif), N (nominal) et V (verbal). Soit c_{i-1} (resp. c_{i+1}) le type de séquence du mot précédent (resp. suivant). C_i et c'_i étant reportés respectivement en ligne et en colonnne, le corps de la matrice (moins la diagonale) est rempli par les suites $c_{i-1}c_ic_{i+1}$ telles que les frontières de tronçons (#) ne changent pas si l'on remplace c_i par c'_i. Le début de phrase et les signes de ponctuation peuvent ici être considérés comme un transjonctif T.

Ces suites vérifient les systèmes suivants :

$$\begin{cases} c_i &= c_{i+1} \neq c_{i-1} \\ c'_i &= T \end{cases} \quad [1]$$

$$\begin{cases} c'_i &= c_{i+1} \neq c_{i-1} \\ c_i &= T \end{cases} \quad [2]$$

$$c_{i+1} = c_{i-1} = T \quad [3]$$

On note la symétrie entre c_i et c'_i dans [1] et [2], où, contrairement à [3], $c_{i+1} \neq c_{i-1}$. Si l'on compte les frontières qui apparaissent ou disparaissent en remplaçant, en un point de la chaîne parlée, la catégorie de m_i produite par l'analyse, par la catégorie de référence, on arrive à un changement pour 1 % des mots. Des erreurs consécutives peuvent de plus se compenser, et ne pas modifier l'emplacement des frontières : par exemple, *sauvegarder # le pouvoir*, peut être analysé comme V # N N ou N # V V (si *sauvegarder* n'est pas connu comme verbe), sans conséquence fâcheuse pour la prosodie.

4.3. *Comparaison avec une approche à base de mots outils*

Les tronçons donnés par l'analyse syntaxique ont aussi été comparés avec ceux que fournit le positionnement d'une frontière après tout mot non reconnu comme mot outil, et après les signes de ponctuation. Comme tout dépend de ce qu'on met sous cette étiquette de « mot outil », nous avons repris la liste de GRAPHON+ (le programme de conversion graphème-phonème du LIMSI [BOU 97]), constituée d'une centaine de mots outils. Précisons que lorsque deux « mots » sont séparés par un trait d'union et/ou une consonne de liaison, GRAPHON+ place un marqueur prosodique et un seul après le second « mot » (ex. *attrape-nigaud, parlons-en*).

Ce test a été conduit sur 90 phrases extraites du journal *Le Monde*, sur lequel l'étiquetage morpho-syntaxique était correct à 90 % sur les mots. Par rapport aux 868 frontières mineures ou intermédiaires prédites par notre analyse, l'approche à base de mots outils en supprime 2 et insère 314 nouvelles frontières. L'analyse syntaxique apporte une amélioration sensible notamment pour le traitement des tronçons à tête nominale : des suites d'adjectifs et de noms telles que des dates – pas de frontière au milieu de *jeudi* 1^{er} *mai 1997*, ni entre un prénom et un nom (ex. *le président Jacques Chirac*), ni après un adjectif antéposé (ex. *une belle fille*). Et on peut aller plus loin dans la hiérarchisation, ce que ne permet pas un découpage sur la seule base des mots outils.

Pour évaluer l'opportunité des frontières intermédiaires et la performance de notre analyse syntaxique, nous les avons enfin comparées par un test d'écoute avec les résultats d'une approche à base de mots outils. Ce test d'acceptabilité de la prosodie a été conduit sur un ensemble de 26 paires de phrases du journal *Le Monde*, parmi les 90 du paragraphe précédent. Deux versions étaient présentées pour chaque phrase, dans un ordre aléatoire, de sorte qu'autant de stimuli commençaient avec ou sans analyse syntaxique. Les autres modules (conversion graphème-phonème, modification quantitative de la prosodie, etc.) étaient sinon les mêmes. Il était demandé à 9 auditeurs d'écouter successivement chaque paire de phrases, et d'indiquer quelle version ils préféraient. Une synthèse par concaténation d'unités sonores a été utilisée, avec les diphones et le traitement du signal MBROLA [DUT 96]. Ce test nous paraît plus satisfaisant qu'un jugement sur une échelle absolue, qui dépend trop de l'humeur des auditeurs. D'autres chercheurs [BLA 97] qui se sont attaqués à l'évaluation de la prosodie se sont heurtés à ce problème, et ont choisi d'effectuer un tel test différentiel. Dans notre cas, la tendance qui se dégage est que les auditeurs préfèrent à 70 % ce qui sort de notre analyse. Pour tirer plus de leçons de ce protocole, il faudrait plus de données, car de nombreux paramètres s'entremêlent, du nombre de syllabes à la charge sémantique des mots, en passant par la position dans la phrase. Cependant, on peut affirmer que la cohérence et le côté plus ou moins répétitif sont de grande importance, perceptivement.

5. Remaques finales et perspectives

5.1. *Discussion*

Dans une certaine mesure, notre modèle rejoint un cadre théorique proposé par Jun et Fougeron pour le français [JUN 95], inspiré de ToBI. En effet, tout comme dans ce cadre, l'analyse présente une hiérarchie à trois niveaux : le groupe intonatif (généralement suivi d'une pause), le groupe intermédiaire et le groupe accentuel (le niveau le plus bas de représentation, qui est caractérisé par un faible allongement). Elle distingue aussi un accent primaire final, qui délimite le groupe accentuel, et un accent secondaire initial, sur la première ou la deuxième syllabe. Au reproche souvent adressé à ce type de description répond un certain nombre d'heuristiques, notamment concernant les lignes de déclinaison. Cependant, ce système de notation ne définit pas en lui-même de mécanisme pour passer des étiquettes à un contour de F_0, pour combiner les accents successifs et contrôler la hauteur : les transitions entre les cibles sont décrites par des règles de réalisation *phonétiques*. En outre, ce standard *de facto* pour la représentation de l'intonation, en accord avec le rôle général de la phonologie, repose sur des intuitions et des jugements d'experts (sur les types d'accents et de *boundary tones* en frontière de la hiérarchie prosodique) difficiles à automatiser – d'autant que le signifié est ici pragmatique. L'utilisation du diacritique « ? » pour indiquer l'incertitude sur l'accentuation en est symptomatique.

Notre modèle peut également être comparé à l'algorithme de Mounin et Grosjean [MOU 93]. Les « structures de performance » de Grosjean (utilisées dans [BAC 90],

[ZEL 94] et [KEL 98]) sont des regroupements psycholinguistiques de mots mettant en évidence une cohésion syntactico-sémantique (entre un adjectif et un nom, par exemple), qui peuvent être identifiés par diverses tâches expérimentales comme la segmentation subjective. Elles reflètent un certain équilibre de « poids » – nombre, taille et hiérarchie des constituants. Cette tendance pourrait trouver ses origines dans les aspects moteurs de la parole humaine [KEL 98], comme le confirment des investigation en neurophysiologie. Par exemple, un syntagme nominal particulièrement long suivi d'un syntagme verbal plutôt court a tendance à être subdivisé. Mais une de nos règles (la règle 9), pour les mots de plus de quatre syllabes, rend particulièrement compte de ce fait. Les propriétés d' « eurythmie » et la « symétrie » – voulant que les unités soient équilibrées – invoquées par Grosjean, ont pour point de départ l'analyse de la structure temporelle de l'anglais, donc d'une langue à chronométrage accentuel (*stress-timed*) qui possède une organisation prosodique très différente du français. Le rattachement prédit au verbe plutôt qu'à l'adjectif, dans *possède un magnifique*, de surcroît, est sujet à discussion, de même que le parenthésage suivant.

(La fille)(s'est déguisée)(en une jolie)(petite fée espiègle).

par rapport à notre grammaire (en outre plus concise), qui regroupe *une jolie petite fille espiègle*. L'algorithme de Mounin et Grosjean a été conçu sur un corpus de neuf phrases, et n'a pas été testé sur un autre matériel.

Nos séquences sont des groupes syntaxiques au sens de Le Goffic [LEG 93]. Abney a également mis en évidence la pertinence psychologique des tronçons [ABN 91]. Mentionnons toutefois les limitations de notre analyseur : elles concernent essentiellement la détection des ellipses, des incises et des phénomènes d'extraction (phrases clivées). Comme les frontières de propositions ne sont pas détectées, le système ne va pas marquer de frontière prosodique entre *vous* et *aime*, dans *l'homme qui est avec vous aime les femmes* et fera même la liaison. Enfin il ne permet d'associer que des montées d'intonation aux virgules.

5.2. *Conclusion*

Bien qu'il n'existe pas une congruence parfaite entre syntaxe et prosodie (un principe d'eurythmie, en particulier, entre en ligne de compte), une approche syntactico-prosodique a été adoptée dans le système de synthèse de la parole du LIMSI. Elle repose sur l'hypothèse que des accents et des frontières prosodiques peuvent être affectées sans interprétation sémantique.

Une grammaire en tronçons inspirée des grammaires de dépendance a été implémentée pour le français, avec vingt règles syntactico-prosodiques écrites à la main. Le parenthésage consiste à segmenter la phrase en séquences non récursives, qui sont définies comme des ensembles de catégories possibles (ex. pronoms personnels, verbes... pour les séquences verbales). Une frontière prosodique mineure, majeure ou intermédiaire est ensuite placée après les séquences nominales et verbales. Cette segmenta-

tion, réalisée en insérant des labels dans la chaîne de mots, reflète bien le parenthésage prosodique, comme l'ont démontré des tests systématiques.

Ainsi, les résultats que donnent la synthèse vocale, avec une représentation simplifiée de la syntaxe et de la prosodie (à l'aide de neuf mouvements) peuvent-ils contribuer à l'étude des corrélations entre ces deux pôles. Ils tendent à montrer que la structure prosodique est plus plate que l'arbre syntaxique. Si les aspects cognitifs n'ont pas été au centre de nos préoccupations, il y aurait certainement des conséquences à en tirer. Il faudrait les tempérer par le fait même que l'objet manipulé est de la parole lue. Mais c'est ici la tâche et là, comme en discours spontané, l'humain n'est pas sensé planifier tout ce qu'il va dire : il est assez improbable qu'il le programme de façon récursive (comme cela est fait dans [BAI 89]).

5.3. *Perspectives*

Nous avons privilégié une approche structurelle par intension. On pourra néanmoins concéder que l'intégration de critères quantitatifs devrait affiner l'étiquetage en plus de la gestion des homographes hétérophones : par exemple, *notions* n'est que rarement un verbe. De même, un lexique de locutions figées, comme *de temps en temps*, devrait affiner le parenthésage plus la gestion des liaisons. La définition de telles expressions constitue un sujet de recherche en soi. Plus généralement, le recours au lexique doit pouvoir résoudre des problèmes quand une analyse purement syntaxique est mise en défaut.

En reprenant l'étiquetage en parties du discours et les tronçons définis par les frontières – à tête nominale ou verbale –, plusieurs perspectives s'offrent à nous :

– perfectionner les règles syntactico-prosodiques, notamment sur les durées, pour mieux exploiter l'information disponible et restituer une voix moins facile à caricaturer, moins stéréotypée et génératrice de fatigue (avec par exemple un comportement différent des noms, porteurs d'information, et des adjectifs) ;

– ou envisager un apprentissage automatique, à partir de l'identification de la performance d'un locuteur humain.

Dans la première perspective, il serait souhaitable d'aller au-delà du tronçon : après une séquence déterminant + nom, on ne trouvera pas nécessairement le même type de frontière, selon que ce qui suit est un verbe ou une préposition. On aurait donc besoin de plus de données ; ce qui suggère, comme pour la seconde perspective, de se diriger vers une description tonale de la prosodie, pour permettre un étiquetage automatique – la mise en place de cette représentation, en effet, nécessite un moindre investissement en temps pour l'expert.

Comme souvent dans une démarche scientifique qui vise à discrétiser les problèmes, le souci de simplification a ici présidé. Réduire le nombre de symboles, tel était l'objectif. Mais il faut tenir compte du fait que la synthèse vocale ne peut sortir

que des éléments fournis en entrée, pour pallier le côté un peu monotone de la voix produite.

Pour un apprentissage automatique, que l'on peut adapter à différents styles ou registres de langue, et à divers locuteurs, une idée est de construire une base de données de types de tronçons, avec des représentants pour chacun, fournissant des échantillons du nombre correspondant de syllabes et de patrons mélodiques observés. Les positions initiales et finales dans la phrase pourront également être utilisées. Ce qu'il faut, c'est trouver un compromis entre :

– d'un côté, multiplier les contraintes phonotactiques et syntaxiques (positionnelles et catégorielles) afin de trouver le meilleur candidat pour chaque tronçon ;

– de l'autre, avoir une large couverture.

Ces impératifs, que l'on retrouve pour la construction du corpus de référence, posent l'alternative suivante : assignation automatique ou manuelle des niveaux de hauteur dans ce corpus de parole naturelle. Les règles syntactico-prosodiques déjà existantes ont l'avantage certain d'être automatisées : elles permettent de passer du texte à une suite théorique de niveaux de hauteur, que l'on peut aligner directement avec les valeurs de fréquence fondamentale réalisées ; mais elles ont l'inconvénient de ne pas rendre compte des idiosyncrasies. Des outils développés au LIMSI peuvent permettre de remédier à ce fait : on utilisera la reconnaissance de la parole, couplée avec un modèle qui simule une perception tonale de l'intonation [DAL 95].

Les types de tronçons fréquents étant peu nombreux et recouvrant un bon pourcentage des occurrences, une piste consiste également à se concentrer sur eux. Quoi qu'il en soit, les lacunes et les conflits dans le corpus d'apprentissage doivent être prévus. Une méthode d'apprentissage automatique de règles, inspirée par E. Brill [BRI 93] est envisagée. Il s'agit alors de définir avec précaution des patrons de règles, prenant en compte la taille des constituants – ce qui n'est pas prévu par l'auteur.

Par ailleurs, il est intéressant d'étudier dans quelle mesure l'approche que nous avons proposée est portable à d'autres langues. Un travail sur le créole et sur l'espagnol a déjà été ébauché.

Remerciements

Les auteurs tiennent à remercier ici Philippe Blache et Véronique Aubergé pour leurs remarques et critiques sur la première version de cet article.

6. Bibliographie

[ABN 91] ABNEY S., « Parsing by chunks » *in* Berwick R.,Abney S. & Tenny C. (eds.), *Principle-based parsing*, Kluwer Academic Publishers, Dordrecht, p.257-278, 1991.

[ANB 96] ABNEY S., « Part-of-speech tagging and partial parsing », *in* Church K., Young S. & Bloothooft G. (eds.), *Corpus-based methods in language and speech. An ELSNET book*, Kluwer Academic Publishers, Dordrecht, 1996.

[ADD 99] ADDA G., MARIANI J., PAROUBEK P., RAJMAN M., LECOMTE J., « Métrique et premiers résultats de l'évaluation GRACE des étiqueteurs morphosyntaxiques pour le français », *Actes de TALN'99*, Cargèse, p. 15-24, 1999.

[AUB 91] AUBERGÉ V., La synthèse de la parole : « des règles aux lexiques », Thèse de doctorat, Université Pierre Mendès-France, Grenoble, 1991.

[BAC 90] BACHENKO J. & FITZPATRICK E., « Computational Grammar of Discourse-Neutral Prosodic Phrasing in English », *Computational Linguistics*, 16(3), p. 155-170, 1990.

[BAI 89] BAILLY G., « Integration of rhythmic and syntactic constraints in a model of generation of French prosody », *Speech Communication*, 8(1), p. 137-146, 1989.

[BAR 94] BARBOSA P. & BAILLY G., « Characterization of rhythmic patterns for text-to-speech synthesis », *Speech Communication*, 15, p. 127-137, 1994.

[BAR 87] BARTKOVA K. & SORIN C., « A model of segmental duration for speech synthesis in French », *Speech Communication*, 6, p. 245-260, 1987.

[BEA 94] BEAUGENDRE F., Une étude perceptive de l'intonation du français, d'eveloppement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte, Thèse de doctorat, Université Paris XI (Orsay), 1994.

[BES 90] BESCHERELLE, *La conjugaison, 12000 verbes*, Hatier, Paris, 1990.

[BLA 97] BLACK A.W. & TAYLOR P., « Assigning phrase breaks from part-of-speech sequences », *actes de Eurospeech'97*, Rhodes, p. 995-998, 1997.

[BOU 97] BOULA DE MAREÜIL P., Étude linguistique appliquée à la synthèse de la parole à partir du texte, Thèse de doctorat, Université Paris XI (Orsay), 1997.

[BRI 93] BRILL E., « Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging », *Computational Linguistics*, 21(4), 1993, p. 545-565, 1993.

[CAE 91] CAELEN-HAUMONT G., Stratégies des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques, Thèse d'État, Université de Provence, Aix-en-Provence, 1991.

[CAM 92] CAMPBELL N., « Syllable-based segmental duration », *in* Bailly G. & Benoît C. (eds.), *Talking Machines: theories, models and designs*, Elsevier Science Publishers, North Holland, p. 211-224, 1992.

[CHO 75] CHOPPY C., LIÉNARD J.S., TEIL D., « Un algorithme de prosodie automatique sans analyse syntaxique», *actes des* 6^{e} *Journée d'Étude sur la Parole*, Toulouse, 1975.

[CHU 88] CHURCH K.W., « A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text », *Second Conference on Applied Natural Language Processing*, Austin, p. 136-143, 1988.

[CON 91] CONSTANT P., Analyse syntaxique par couche, Thèse de doctorat, ENST (Paris), 1991.

[DAL 95] D'ALESSANDRO C. & MERTENS P., « Automatic pitch contour stylization using a model of tonal perception », *Computers Speech and Language*, 9(3), p. 257-288, 1995.

[DEL 95] DELAIS E., Pour une approche parallèle de la structure prosodique française, Thèse de doctorat, Université de Toulouse-le-Mirail, Toulouse, 1995.

[DIC 98] DI CRISTO P., Génération automatique de la prosodie pour la synthèse à partir du texte, Thèse de doctorat, Université de Provence, Aix-en-Provence, 1998.

[DUT 96] DUTOIT T., PAGEL V., PIERRET N., BATAILLE F., VAN DER VRECKEN O., « The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes », *Actes de ICSLP'96*, Philadelphia, p. 1393-1396, 1996.

[EJE 88] EJERHED E., « Finding clauses in unrestricted text by finitary and stochastic methods », *Second Conference on Applied Natural Language Processing*, Austin, p. 129-227, 1988.

[FUJ 84] FUJISAKI H. & KEIKICHI H., « Analysis of voice fundamental frequency contours for declarative sentences of Japanese », *Journal of the Acoustical Society of Japan*, vol. 5, p. 233-241, 1984.

[HIR 96] HIRSCHBERG J. & PRIETO P., « Training intonational phrasing rules automatically for English and Spanish text-to-speech », *Speech Communication*, 18, p. 281-290, 1996.

[HIR 98] HIRST D. & DI CRISTO A. (EDS.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge University Press, Cambridge, 1998.

[JUN 95] JUN S.-A. & FOUGERON C., « The accentual phrase and the prosodic structure of French », *Actes de ICPhS'95*, Stockholm, p. 722-725, 1995.

[KAR 90] KARLSSON F., « Constraint grammar as a parsing framework for parsing running text », *Actes de COLING*, Helsinki, p. 168-173, 1990.

[KEL 97] KELLER E., ZELLNER B. & WERNER S., « Improvements in prosodic processing for speech synthesis », *Actes de Speech Technology in the Public Telephone Network*, Rhodes, 1997.

[KEL 98] KELLER E. & ZELLNER B., « Motivations for the prosodic predictive chain », *Actes du Third ESCA COCOSDA International Workshop on Speech Synthesis*, Jenolan Caves, 1998, p. 137-141, 1998.

[LAC 99] LACHERET-DUJOUR A. & BEAUGENDRE F., *La prosodie du français*, CNRS Éditions, Paris, 1999.

[LAR 89] LARREUR D., EMERARD F. & MARTY F., « Linguistic and prosodic processing for a text-to-speech synthesis system », *Actes de Eurospeech'89*, Paris, p. 510-513, 1989.

[LIB 92] LIBERMAN M.Y. & CHURCH K.W., « Text Analysis and Word Pronunciation in Text-to-Speech Synthesis », *in* Furui S. & Sondhi M.M. (eds.), *Advances in Signal Processing*, Dekker, p. 791-831, 1992.

[LEG 93] LE GOFFIC P., *Grammaire de la Phrase Française*, Hachette Supérieur, Paris, 1993.

[LIE 77] LIÉNARD J.S., CHOPPY C., TEIL D., RENARD G., SAPALY J., « Diphone synthesis of French: vocal response unit and automatic prosody from the text », *Actes de IEEE-ICASSP'77*, Hartford, 1977.

[MAR 80] MARTIN P., « Pour une théorie de l'intonation », *in* Rossi M. Di Cristo A., Hirst D., Martin P., Nishinuma Y., *L'intonation : de l'Acoustique à la Sémantique*, Éditions Klincksieck, Paris, p. 231-271, 1980.

[MER 93] MERTENS P., « Accentuation, intonation et morphosyntaxe », *Travaux de linguistique*, vol. 26, Gent, 1993.

[MER 97] MERTENS P., « De la chaîne linéaire à la séquence de tons », *TAL*, 38(1), 1997.

[MER 00] MERTENS P., « Intonation and syntax: natural language processing for speech synthesis », *Actes de TALN*, Lausanne, 2000.

[MOR 97] MORLEC M., BAILLY G. & AUBERGÉ V., « Synthesising attitudes with global rhythmic and intonation contours », *Actes de Eurospeech'97*, Rhodes, p. 219-223, 1997.

[MOU 93] MOUNIN P. & GROSJEAN F., « Les structures de performance en français : caractérisation et prédiction », *L'année psychologique*, 93, p. 9-30, 1993.

[OSH 87] O'SHAUGHNESSY D., « Specifying Intonation in a Text-to-Speech System using only a Small Dictionary », *Actes de IEEE-ICASSP'87*, Dallas, p. 1430-1433, 1987.

[OST 94] OSTENDORF M. & VEILLEUX N., « A Hierarchical Stochastic Model for Automatic Prediction of Prosodic Boundary Location », *Computational Linguistics*, 20(1), p. 27-54, 1994.

[PAS 90] PASDELOUP V., Modèle de règles rythmiques du français appliqué à la synthèse de la parole, Thèse de doctorat, Université de Provence, Aix-Marseille I, 1990.

[PER 92] PÉRENNOU G., DE CALMÈS M., FERRANÉ I., PÉCATTE J.-M., « Le projet BDLEX de bases de données lexicales du français écrit et parlé », *Séminaire lexique*, Toulouse, p. 41-56, 1992.

[QUA 89] QUAZZA S., VARESE G. & VIVALDA E., « Contextual syntactic analysis for text-to-speech conversion », *Actes de European Conference on Speech Communication and Technology*, Paris, p. 388-392, 1987.

[QUE 92] QUENÉ H. & KAGER R., « The derivation of prosody for text-to-speech from prosodic sentence structure », *Computer Speech and Language*, 6(1), pp 77-98, 1992.

[RAL 95] RAMSHAW L.A. & MARCUS M.P., « Text Chunking using Transformation-Based Learning », *actes du ACL Third Workshop on Very Large Corpora*, Cambridge, p. 82-94, 1995.

[ROS 93] ROSSI M., « A model for predicting the prosody of spontaneous speech (PPSS model) », *Speech Communication*, 13, p. 87-107, 1993.

[SHA 96] SHARMAN R.A. & WRIGHT J.H., « A fast stochastic parser for determining phrase boundaries for text-to-speech synthesis », *Actes de IEEE-ICASSP*, Atlanta, p. 357-360, 1996.

[SIL 92] SILVERMAN K., BECKMAN M., PITRELLI J., OSTENDORF M., WIGHTMAN C., PRICE P., PIERREHUMBERT J., HIRSCHBERG J. « ToBI: a standard for labeling English prosody », *Actes de ICSLP'1992*, Banff, p. 867-870, 1992.

[TES 59] TESNIÈRE L., *Éléments de syntaxe structurale*, Éditions Klincksieck, Paris, 1959.

[HAR 91] HART J.'T, COLLIER R. & COHEN A., *A perceptual study of intonation: an experimental-phonetic approach to speech melody*, Cambridge University Press, Cambridge, 1991.

[TRA92] TRABER C., « F_o generation with a database of natural F_o patterns and with a neural network », *in* Bailly G. & Benoît C. (eds.), *Talking Machines: theories, models and designs*, Elsevier Science Publishers, North Holland, p. 287-304, 1992.

[TZO 95] TZOUKERMANN E. & SOURNOY O., « Segmental Duration in French Text-to-Speech Synthesis », *Actes de Eurospeech'95*, Madrid, p. 607-610, 1995.

[VAI 80] VAISSIÈRE J., « La structuration acoustique de la phrase française », *Annali della Scuola Normale Superiore di Pisa*, p. 530-560, 1980.

[VAN 99] VANNIER G., LACHERET-DUJOUR A., VERGNE J., « Pauses location and duration calculated with syntactic dependencies and textual considerations for T.T.S. system », *Actes de ICPhS'99*, San Francisco, p. 1569-1572, 1999.

[VER 90] VERGNE J., « A parser without a dictionary as a tool for research into French syntax », *Actes de COLING*, Helsinki, p. 70-72, 1990.

[VER 97] VÉRONIS J., DI CRISTO P., COURTOIS F., LAGRUE B., « A stochastic model of intonation for French text-to-speech », *Actes de Eurospeech'97*, Rhodes, p. 2643-2647, 1997.

[ZEL 94] ZELLNER B., « Pauses and the Temporal Structure of Speech », *in* Keller E., Wiley J. & Sons (eds.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art and Future Challenges*, Chichester, p. 41-62, 1994.