

LA SYNTHÈSE DE L'INTONATION À PARTIR DE STRUCTURES SYNTAXIQUES RICHES

Piet Mertens*, Jean-Philippe Goldman**, Éric Wehrli**, Arnaud Gaudinat**

RÉSUMÉ - ABSTRACT

Le développement d'un système complet de synthèse de la parole à partir du texte de bonne qualité nécessite beaucoup de ressources de domaines très variés. Dans cet article, une approche modulaire intégrant plusieurs composants spécifiques est présentée. On décrira principalement deux systèmes existants de traitement automatique du langage naturel: 1. l'analyseur syntaxique *Fips*, qui effectue en premier un traitement linguistique détaillé (prétraitement, lemmatisation, analyse syntaxique et phonétisation) et 2. le système *Mingus* qui génère une intonation appropriée à partir de l'arborescence syntaxique. Dans un premier temps, la génération associe une séquence de tons à des groupes de terminaux, puis chaque ton est transformé en une suite de cibles acoustiques associées aux sons. Comme cette procédure de génération suppose une étude théorique de la relation entre syntaxe et intonation, cette étude est présentée également. Enfin, l'article décrit les ajustements nécessaires pour assurer une intégration optimale de ces composants tout en respectant les choix théoriques de chacun des systèmes.

The development of a high quality text-to-speech system requires many resources covering various aspects of linguistic processing. This paper describes a modular approach integrating components initially designed for specific purposes. We basically describe two existing NLP systems: 1. the *Fips* parser, which provides a detailed linguistic analysis of the input text (performing preprocessing, lemmatization, syntactic analysis, and grapheme-to-phoneme conversion), and 2. the *Mingus* system for intonation generation on the basis of the syntactic structure. Intonation generation first assigns a tone sequence to groups of terminals and then converts these contour specifications into a sequence of acoustic pitch targets synchronized with (time instants within) phonetic segments. As the generation procedure relies on a theory about the relation between syntax and intonation, this theory is described as well. Finally, the study describes the necessary interfacing step ensuring the optimal integration of these components, while respecting theoretical choices of both NLP systems.

MOTS-CLEFS - KEYWORDS

Synthèse de la parole à partir du texte. Analyse syntaxique. Intonation. Prosodie.
Text-to-speech synthesis. Parsing. Prosody generation.

* CCL, K.U.Leuven, BP 33, 3000 Leuven (Belgique) Piet.Mertens@arts.kuleuven.ac.be

** LATL, Université de Genève, 2 rue de Candolle, Genève {goldman,wehrli,gaudinat}@latl.unige.ch

INTRODUCTION

Plusieurs études récentes ont permis de mettre en lumière l'interaction entre syntaxe et intonation (Martin Ph. 1981, Eynde K. & Dooren K. 1983, Mertens P. 1993, 1997, Eynde K. *et al.* 1998, Rossi M. 1987, 1999, Lacheret A. & Beaugendre, F. 1999). L'application de ces observations en synthèse de la parole (Gaudinat A. & Goldman J.-P. 1998, Mertens P. 1999) est loin d'être triviale; elle suppose en effet plusieurs traitements de complexité élevée: d'une part, une analyse syntaxique sophistiquée, à la fois riche, complète, et robuste, et d'autre part, un algorithme de génération de l'intonation, qui associe à la structure syntaxique des formes intonatives.

Or, il est extrêmement rare que tous ces modules soient réunis dans un seul système de traitement du langage naturel ou dans un même système de synthèse à partir du texte (*text-to-speech*). Néanmoins pour réaliser l'objectif, on est amené à combiner des systèmes existants, traitant chacun un ou plusieurs aspects du problème global. Il s'avère que les systèmes combinés, tout en étant adaptés à leur tâche, sont rarement compatibles. Pour régler ces problèmes d'interfaçage, on se voit obligé de rajouter des traitements supplémentaires.

Nous décrivons ici une étude visant à rapprocher (et connecter) deux systèmes existants: 1. l'analyseur syntaxique *Fips* élaboré au LATL qui effectue en amont un traitement linguistique et 2. le système *Mingus* dans une version étendue de l'algorithme décrit dans Mertens P. 1999 qui utilise le résultat syntaxique de *Fips* pour générer l'intonation.

Cet article présente d'abord l'analyseur *Fips* (section 1). Ensuite on résume le cadre descriptif pour la représentation linguistique de l'intonation française (section 2). La section 3 aborde la question de la relation entre intonation et syntaxe: comment rendre compte de leur interaction. La section 4 décrit la génération de l'intonation dans le système *Mingus*. On explicite le type d'analyse syntaxique attendu par ce système. Les deux sections suivantes (5 et 6) présentent le modèle mélodique et le modèle de durée, qui permettent de passer de la transcription phonétique enrichie de marqueurs intonatifs aux commandes phonético-acoustiques utilisées par le synthétiseur. Le module de phonétisation est décrit en section 7. La comparaison des représentations syntaxiques dans *Fips* et *Mingus* mène à la formulation d'une procédure de conversion visant à modifier la sortie fournie par *Fips* en vue de son utilisation par *Mingus* (section 8). Les résultats obtenus sont commentés dans la section 9.

La génération de la prosodie décrite ici se limite essentiellement à la congruence des structures syntaxique et intonative et au phénomène des constructions syntaxiques exigeant une organisation intonative particulière (comme les clivées, incisives, extrapositions,...). La plupart des aspects pragmatiques et émotionnels sont reportés à une version ultérieure, bien que tout soit prévu pour les intégrer rapidement à partir d'informations supplémentaires pragmatiques (non textuelles) telle qu'une annotation manuelle à l'aide de balises, ou leur insertion automatique à partir d'autres sources ou à partir de paramètres phonostylistiques ajustables.

Toute mise en rapport des organisations intonative et syntaxique suppose l'adoption d'un cadre descriptif particulier des données intonatives et de la structure syntaxique. On précise donc les choix effectués.

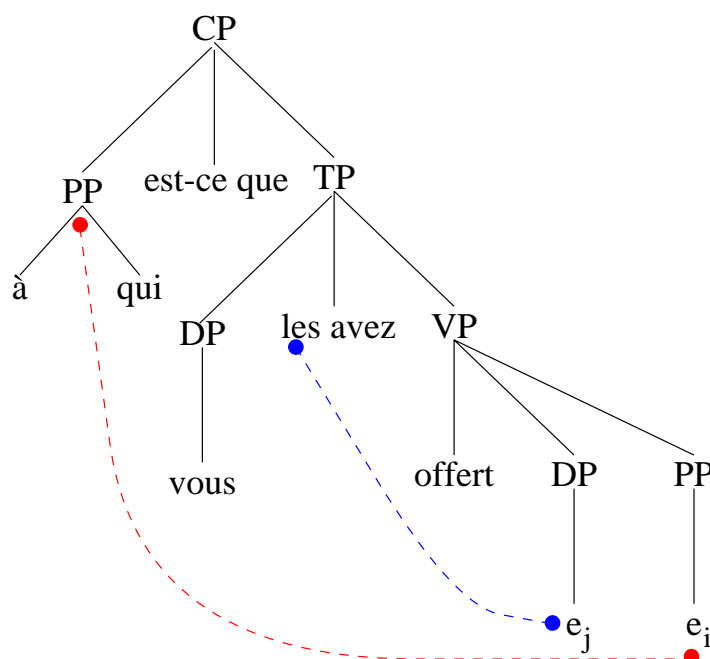
1. L'ANALYSEUR SYNTAXIQUE FIPS

Fips est un analyseur syntaxique robuste basé sur la théorie chomskyenne des principes et paramètres (cf. Chomsky N. 1995). Il prend en entrée un texte brut et tente de le découper en phrases, et d'associer à chacune d'elles une structure syntaxique. Cette structure prend la forme d'un arbre syntagmatique correspondant aux structures de surface enrichies du modèle chomskyen. Rappelons brièvement que les structures de surface reflètent l'ordre et l'agencement des constituants tels qu'ils apparaissent dans la phrase, par opposition à des structures plus abstraites qui ne respecteraient pas forcément l'ordre linéaire des mots, comme les structures profondes des théories transformationnelles. On appelle *structures de surface enrichies* des structures de surface augmentées de catégories vides, éléments abstraits correspondant à des syntagmes dépourvus de réalisation lexicale. Parmi ces catégories vides, les *traces* d'éléments extraposés jouent un rôle particulièrement important, puisqu'elles signalent la position d'origine des éléments déplacés.

Un exemple simple permettra de mieux illustrer ces concepts¹.

À qui est-ce que vous les avez offerts ?

[_{CP} [_{PP} à qui]_i est-ce que [_{TP} [_{DP} vous] les_j avez [_{VP} offerts [_{DP} e_j] [_{PP} e_i]]]]



¹ Les étiquettes utilisées dans les représentations syntaxiques sont : CP (phrase avec complémenteur), TP (phrase), DP (syntagme nominal), PP (syntagme prépositionnel).

Figure 1. Analyse syntaxique de l'exemple ci-dessus selon *Fips*.

Dans cet exemple, le syntagme prépositionnel interrogatif à *qui* est lié à sa trace, la catégorie vide [_{PP} e] complément du verbe *offrir*. Ce lien, qui est représenté par l'indice *i* dans la figure ci-dessus, exprime le fait que le syntagme interrogatif fonctionne comme complément du verbe *offrir*. De même, un autre lien, indicé *j*, représente la relation entre le pronom clitique *les* et le syntagme nominal [_{DP} e_j] en position d'objet direct du verbe *offrir*.

L'usage de ces catégories vides et de leur association avec des éléments extraposés permet une grande simplification dans l'assignation des fonctions grammaticales : un objet direct est un constituant de catégorie DP, gouverné par un verbe transitif.

L'algorithme d'analyse développé pour *Fips* repose sur quelques principes relativement simples. Parcourant les mots d'une phrase de gauche à droite, le processus d'analyse commence par une recherche lexicale, chargée d'établir la liste des items lexicaux correspondant aux mots orthographiques. Cette étape incorpore les tâches de lemmatisation, d'identification des mots composés, etc.

Le traitement syntaxique comprend 3 types d'opérations : projection, combinaison et formation de chaînes.

La projection est une opération syntaxique qui crée un constituant syntaxique sur la base soit d'un item lexical, soit d'un autre constituant syntaxique. Ainsi, par exemple, les items lexicaux de catégorie substantif, comme *maison* ou *pommes de terre*, donnent lieu à des projections de syntagmes nominaux, respectivement [_{DP} maison] et [_{DP} pommes de terre]. Comme exemple de projection à partir d'un constituant syntaxique, mentionnons la projection d'un constituant adverbial à partir d'un syntagme nominal à valeur temporelle ([_{AdvP}[_{DP} l'autre jour]], [_{AdvP}[_{DP} l'année dernière]]).

La combinaison de deux constituants adjacents A et B est l'opération fondamentale du processus d'analyse. Elle permet soit d'intégrer A comme sous-constituant gauche de B, soit B comme sous-constituant droit de A. La combinaison de constituants A=[_{DP} le chat] et B=[_{TP} a mangé] donne un nouveau constituant du type [_{TP}[_{DP} le chat [_{T'} a mangé]] où A est le sujet de B. La combinaison de ce nouveau constituant avec le syntagme nominal [_{DP} une petite souris] donnera la structure complète :

Le chat a mangé une petite souris.

[_{TP}[_{DP} le chat [_{T'} a mangé]_{DP} une petite souris]]

Le processus de formation de chaînes est déclenché par l'attachement d'un constituant dans une position d'extraposition, position pour laquelle aucune fonction syntaxique n'est déterminée. C'est le cas typiquement de la position initiale de phrase (spécificateur de CP), ou des positions de pronoms clitiques. La formation de chaîne cherche à établir un lien entre ces éléments et des positions canoniques de constituants syntaxiques, de manière à pouvoir assigner une fonction syntaxique aux éléments déplacés (extrapositions, clitiques, etc.).

Les trois mécanismes que nous venons de présenter sont soumis à des contraintes, parfois complexes, et souvent liées à des propriétés lexicales des items concernés, comme par exemple, les propriétés de valence et autres propriétés sélectionnelles des verbes, les règles d'accord, etc.

Les processus d'analyse, principalement la combinaison et la formation de chaînes, peuvent donner lieu à des résultats multiples. Un ensemble réduit d'heuristiques, basées en partie sur des principes psycholinguistiques, est utilisé pour ordonner les analyses, voire en réduire le nombre, lorsque la prolifération d'analyses s'avérerait par trop pénalisante.

2. THÉORIE DE L'INTONATION EN FRANÇAIS

Pour la description des courbes intonatives du français, leur notation et leur interprétation linguistique, nous adoptons le modèle proposé par P. Mertens (1987-2000). On se limitera ici à une présentation sommaire. Pour les détails, nous renvoyons aux publications citées.

L'intonation d'un énoncé se présente comme une suite de contours mélodiques portés par des groupes de syllabes. Chaque contour peut à son tour être analysé comme une configuration de cibles. Chaque cible correspond à un niveau de hauteur localisé à un endroit précis de la chaîne syllabique, endroit défini en fonction de l'accentuation.

2.1. Les niveaux de hauteur et les tons.

Les *niveaux de hauteur* sont les *paliers relatifs* sur l'échelle de hauteur mélodique. Ils se définissent les uns par rapport aux autres, à partir des intervalles mélodiques observés entre syllabes successives ou à l'intérieur d'une même syllabe. Les niveaux extrêmes se définissent en outre en fonction du registre du locuteur individuel, c'est-à-dire de sa plage de hauteur usuelle.

Les quatre *niveaux de base* sont le bas, le haut, l'infra-bas et le suraigu; ils sont notés *v*, *h*, *v-* et *h+*, respectivement, lorsqu'ils apparaissent en syllabe accentuée, et par les minuscules correspondantes en syllabe atone. L'interprétation des variations mélodiques en termes de niveaux de hauteur s'établit progressivement et localement, en fonction de la présence d'intervalles mélodiques dans l'énoncé.

On distingue deux catégories d'*intervalles*: majeur et mineur; typiquement, ils correspondent à des écarts de 5 et 3 demi-tons respectivement, mais ces valeurs varient en fonction du locuteur et en fonction de facteurs expressifs, phonostylistiques, ou paralinguistiques (tel que l'état émotif du locuteur). Les intervalles mineurs donnent lieu au rehaussement ou au rabaissement du niveau de base déjà atteint. Ils sont notés par les signes / et \ placé devant le niveau de base.

Le *ton* indique l'ensemble de niveaux de hauteur associés à une syllabe. L'intonation d'un énoncé se présente alors comme une succession de tons.

2.2. L'accentuation

L'accentuation d'une syllabe peut se manifester de plusieurs façons. La syllabe peut être prononcée avec une force phonatoire particulière (dont témoignent la forme de l'onde glottale, le spectre résultant, et enfin le timbre perçu). Elle peut être allongée. Ensemble ces deux facteurs produisent l'impression d'un niveau sonore plus élevé.

Le français connaît deux types d'accent: l'accent final et l'accent initial. L'accent final (AF) s'accompagne généralement d'un allongement de la syllabe; il peut être suivi d'une pause et accepte des contours dynamiques (montée ou chute). L'accent initial (AI), également appelé accent d'insistance, frappe le plus souvent la syllabe initiale d'un mot sans pourtant l'allonger. Il se caractérise par la force dans l'articulation de l'attaque syllabique; il peut être précédé d'une pause. Si le mot commence par une voyelle, l'AI soit passe à la deuxième syllabe, soit entraîne la prononciation d'un coup de glotte. Dans les exemples suivants, la syllabe portant l'AI sera soulignée. L'astérisque précède une forme agrammaticale.

la circulation
*la circulation
absolument
absolument

La présence éventuelle d'un accent et sa nature permettent donc de distinguer trois *classes de tons*. Ils correspondent aux paradigmes² de tons de certaines positions: la position de l'*accent final*, celle de l'*accent initial* et celle d'une séquence de *syllabes non accentuées* (NA). Les paradigmes de tons sont énumérés plus loin. Cette typologie des classes est basée sur leurs propriétés distributionnelles. Par exemple: tel ton, qui apparaît en position AF, ne s'utilise pas en position AI ni en position NA.

2.3. Le groupe intonatif

L'unité descriptive du *groupe intonatif* (GI) correspond à une suite de syllabes comportant un seul accent final et éventuellement d'autres syllabes de type NA ou AI. Il apparaît que les séquences attestées obéissent toutes à un schéma structural unique³. Ce schéma définit l'ordre des classes de tons dans le groupe. Les parties facultatives sont données entre crochets. AF et AI représentent une syllabe; NA représente une ou plusieurs syllabes atones.

$$GI \rightarrow [[NA] AI] [NA] AF [NA]$$

² La notion linguistique de "paradigme" indique l'ensemble des éléments qui peuvent figurer en un point de la chaîne parlée. Ainsi dans la chaîne "je le veux", on peut remplacer le deuxième élément par "les" ou par "la", mais non par "il", "où", "de". On en déduit que "le, la, les" appartiennent au même paradigme, où ne figurent pas "il", "où" et "de".

³ Le raisonnement distributionnel qui sous-tend la définition du GI est donné dans (Mertens P. 1987: 93-101). Comme le schéma unique comporte plusieurs parties facultatives, il regroupe en fait plusieurs séquences différentes.

L'intonation d'un énoncé se présente ainsi comme une succession de GI , où chaque GI se décompose à son tour en une suite de tons appartenant à des classes précises.

Le *ton final* est celui qui dans un GI donné occupe la position AF . Le ton final tombe sur la dernière syllabe pleine du GI (sauf si celui-ci contient un appendice, cf. *infra*). Par syllabe pleine on entend ici une syllabe qui comporte une voyelle autre que le schwa.

Le tableau suivant illustre les paradigmes des tons admis pour chacune des localisations dans le GI . La *notation* suit quelques conventions qui explicitent le statut du ton, soit sa position dans le GI . 1. L'emploi des majuscules et des minuscules reflète le caractère accentué ou atone du ton. 2. Les tons à l' AF comportent systématiquement deux niveaux de hauteur, même pour les tons statiques. Ceci permet de distinguer les tons homophones des positions AI et AF , par exemple H et HH. 3. Une suite atone (NA) est représentée par deux localisations, indiquant le niveau de hauteur au début et à la fin de la séquence. Quand une partie atone ne comporte qu'une seule syllabe, les deux localisations coïncident dans la même syllabe, et cette syllabe reçoit alors deux niveaux de hauteur, forcément identiques.

Pour les positions atones, le tableau ne fournit que les niveaux de hauteur primaires (b, h), qui peuvent être rehaussés (/b, /h) ou abaissés (\b, \h). La combinaison de tons des différents paradigmes doit respecter le principe de définition des niveaux rehaussés et abaissés: ces niveaux supposent un intervalle mineur qui s'applique au niveau de base précédent; pour cette raison des suites telles que "b..b /HH", "b../h BB", etc. sont exclues.

	NA	AI	NA	AF	NA
b	b	H	b	B-B-	b-...b-
h	h	B	h	H+H+	
				HB-	
				H/H	h..h
				/HH	
				\HH	
				HB	
				HH	
				/BB	
				BB	
				\BB	

Tableau 1. Distribution des tons dans le groupe intonatif maximal (i.e. comportant un nombre maximal de paradigmes). Les lignes horizontales indiquent des contraintes sur la combinaison d'éléments appartenant aux paradigmes traversés par cette ligne. Par exemple: les tons b- (appendice) ne se combinent qu'avec les tons B-B- HB- et H+H+ du paradigme AF.

L'*appendice* correspond à la séquence de syllabes atones (NA) après l' AF ; il s'agit d'une partie facultative dans la structure maximale du GI . Cette séquence a des propriétés distributionnelles propres et présente un contour plat, cf. *infra*.

Le schéma structural du GI génère les séquences suivantes, ainsi que ces mêmes chaînes suivies d'un appendice.

séquence de localisations	exemple
AF	oui HB-
NA AF	sûrement bb HB-
AI AF	assez H B-B-
AI NA AF	précisément H h.b B-B-
NA AI NA AF	dans ma situation b...b H bb HH
NA AI AF	pour ce projet bb H hh B-B-

Figure 2. Exemples de séquences de localisations générées par la structure générale du GI.

2.4. Le groupe accentuel

- (a) [nous avons visité] [son nouvel appartement]
 (b) [nous avons][visité] [son nouvel][appartement]

Une même suite de mots peut être prononcée de plusieurs façons, avec un nombre variable de syllabes accentuées finales et un nombre variable de GI. Dans l'exemple ci-dessus, les GI sont délimités par des crochets; l'énoncé (b) compte plus de GI que (a). Pour rendre compte des variantes observées, on suppose un niveau de représentation intermédiaire (entre les GI et les mots), constitué de *groupes accentuels* (GA), qui, eux, sont prévisibles. (On rencontre couramment le terme "mot prosodique" pour désigner ces GA.) Dans les réalisations observées, chaque GI englobe un ou plusieurs GA⁴.

Comment déterminer les GA d'un énoncé ? La règle renvoie à la présence ou l'absence d'un *accent lexical*, aussi appelé *accent de mot*. Cette propriété oppose deux classes de mots: d'une part les *clitiques*, dépourvus d'accent de mot, et d'autre part les *non clitiques*, portant l'accent lexical. Pour la deuxième classe on parle aussi de mot *tonique*, *accentuable*, ou *accentogène*⁵. La liste de catégories clitiques et toniques sera donnée plus loin (§ 2.5.).

Chaque GA consiste d'un élément ("mot") *tonique* et des mots *clitiques* dépendants contigus. Plus exactement: chaque élément tonique constitue le centre d'un GA et tout élément clitique s'intègre au groupe formé par l'élément tonique dont il

⁴ Cette représentation à deux niveaux se justifie par les observations suivantes: 1. dans un énoncé avec moins de GI, les syllabes accentuées finales de groupe coïncident avec celles de la variante avec le maximum de GI; 2. les frontières de GI coïncident avec celles des GA englobés.

⁵ Dans la terminologie de Garde (1968).

dépend syntaxiquement (le clitique précède généralement le mot lexical, sauf dans l'inversion: "disait-il"). Précisons que si le clitique est séparé de l'élément dont il dépend par un autre élément tonique, il s'intègre au groupe qu'il précède: "ce magnifique bateau". Pour plus de détails, voir Mertens P. (1993: 30-40, 1997: 34-5). Cela peut être représenté dans le *squelette prosodique* de l'énoncé: sur la couche AL (pour "accent lexical") un rond indique la dernière syllabe pleine d'un mot tonique; chaque syllabe est représentée par un tiret ou un point selon qu'elle est accentuable ou non; les parenthèses indiquent les frontières de GA résultantes. Les syllabes porteront les tons.

(nous) (on y cherchait) (un nouvel) (appartement)
 AL ○ . . . ○ . . ○ . . . ○
 GA (-) (. . . -) (. . -) (. . . -)

Déjà à ce niveau très bas (celui des préterminaux), la syntaxe intervient, puisque les éléments clitiques (pronoms clitiques, articles,...) se rangent du côté de l'élément dont ils dépendent, qui se situe tantôt à gauche, tantôt à droite:

(lui avait-elle) (envoyé) (cette lettre)
 AL . . ○ . . . ○ . ○
 GA (. . . -) (. . -) (. -)

L'accent final frappe la dernière syllabe pleine du GI. Si cette syllabe appartient à un élément clitique (par exemple à un pronom enclitique: "disait-il"), c'est cet élément clitique qui porte l'accent⁶. On indiquera l'accentuation effective sur la couche des GI:

(lui avait-elle) (envoyé) (cette lettre)
 AL . . ○ . . . ○ . ○
 GA (. . . -) (. . -) (. -)
 GI (. . . +) (. . +) (. +)

Un GI consiste donc d'un ou plusieurs GA. Le choix du nombre de GI, soit la façon de regrouper les GA en GI, dépend de plusieurs facteurs. D'abord, de la longueur du GI résultant: la longueur moyenne observée par GI est de 3.8 syllabes (cf. Delais E. 1995: 125). Ensuite, les accents contigus sont évités (Mertens P. 1992). Le nombre de GI dépend en outre du rapport syntaxique entre les GA: le regroupement doit respecter l'emboîtement syntaxique (§ 2.7.). Il dépend également de la sémantique, des "unités de sens"; mais en synthèse à partir du texte, on n'a généralement pas accès à cette information (cf. Mertens P. 1987:111, 1993: 49-50). Finalement il faut mentionner d'autres facteurs comme le débit de parole, le style (soigné ou non), ou le type de parole (texte lu, parole spontanée).

2.5. Accent de mot et catégorie grammaticale

La *nature clitique*⁷ ou non d'un mot peut être déterminée à partir de sa catégorie grammaticale⁸. Les catégories classées non clitiques sont: le verbe, le nom,

⁶ Notons cependant le comportement irrégulier de certains pronoms. Dans "dis-le", le schwa est prononcé soit [ə], [œ] ou [ø], et accentué; en revanche dans "puis-je" on ne prononce pas de schwa et c'est la syllabe "puis" qui reçoit l'accent.

⁷ Par "clitique" on entend toujours: prosodiquement clitique, c'est-à-dire dépourvu d'accent de mot.

l'adjectif, l'adverbe (sauf la particule "ne"), les numéraux, et les "mots-phrases"⁹ ("non, bonjour, salut, parbleu, soit, allez, voyons, ..."). Pour les pronoms, on distingue selon la sous-catégorie.

Certains *pronoms personnels* sont strictement clitiques: "je, tu, il, ils, on, le, la, les, me, te, se, ..." appartiennent à un paradigme clitique. D'autres, comme "moi, toi, eux,..." appartiennent à un paradigme non clitique. Un petit groupe de pronoms personnels, à savoir "nous, vous, elle, elles, lui, leur", sont ambigus quant au trait [\pm clitique]: il s'agit en fait de formes homonymes, l'une appartenant à un paradigme clitique, l'autre à un paradigme non clitique. Dans "nous, on leur demandera", "nous" est tonique alors qu'il est clitique dans "nous demanderons le leur". Le caractère clitique dépend donc de la position syntaxique (tantôt proclitique ou enclitique, tantôt forme d'emphase).

Les autres pronoms sont non-clitiques: il s'agit de pronoms indéfinis ("aucun, chacun, personne, certains, tous, tout, rien,..."), interrogatifs et relatifs ("qui, quoi,..., lequel, auxquels,...", à l'exception de "que"), et démonstratifs ("ceux-ci, celles-là,...", sauf "ce, c").

Les déterminants sont clitiques, sauf les indéfinis ("*certain*s enfants") et les pré-déterminants ("*tous* mes livres").

Les prépositions et conjonctions se comportent comme des clitiques dans la plupart des cas¹⁰. Les locutions prépositives ("à la *suite* de") ou conjonctives ("à *condition* que") se comportent comme des suites d'éléments toniques ou clitiques selon leur catégorie grammaticale¹¹.

⁸ Alternativement, elle peut être définie à partir de la position syntaxique; mais cela supposerait bien entendu une grammaire complète.

⁹ Selon la terminologie de Tesnière (1959: 94-96)

¹⁰ Pour les exceptions, cf. Mertens (1992)

¹¹ Contrairement aux pronoms enclitiques, les prépositions ou conjonctions finales appartenant à une locution prépositive ou conjonctive ne s'intègrent pas au GA qui précède: dans "à condition que tu partes", l'accent frappe la syllabe "tion" plutôt que la syllabe "que".

catégorie	[±clitique]	exemples
verbe	-	<i>calculer</i>
nom	-	<i>maison, casserole</i>
adjectif	-	<i>machinal</i>
adverbe (sauf "ne")	-	<i>rapidement, très, assez,...</i>
numéral	-	<i>cinq, sixième,...</i>
mot-phrase	-	<i>bonsoir, zut,...</i>
pronom		
• personnel	+	<i>je, tu, il, ils, on, le, la, les, me, te, se,...</i>
	-	<i>moi, toi, eux, soi</i>
• (selon position:)	±	<i>nous, vous, elle, elles, lui, leur</i>
• possessif	-	<i>le mien, les siennes, les leurs,...</i>
• indéfini	-	<i>aucun, chacun, certains, personne, rien,...</i>
• interrogatif	-	<i>combien, quand, où, qui, ...</i>
• relatif	+	<i>que</i>
	-	<i>qui</i>
• démonstratif	+	<i>ce, c'</i>
	-	<i>ceci, ça, ceux, ceux-là,...</i>
déterminant		
• article	+	<i>le, la, les, l', un, une, des</i>
• interrogatif	+	<i>quel, quelle, quels,...</i>
• possessif	+	<i>mon, ta, ses,...</i>
• démonstratif	+	<i>ce, cette, ces</i>
• indéfini	-	<i>aucun, chaque, plusieurs, certains,...</i>
• prédéterminant	-	<i>tout</i>
conjonction	±	<i>et, ou, mais, puisque, comme, bien que,...</i>
préposition	±	<i>à, de, sur, avec, depuis, après,...</i>

Tableau 2. Catégorie grammaticale et trait [±clitique]

2.6. La fonction des unités intonatives

Le choix des tons n'est pas arbitraire. Pour une définition de la valeur sémantique de chacun des tons, nous renvoyons à Mertens P. (1997: 35-37).

2.7. Le regroupement intonatif

Tout enchaînement de contours résulte en un regroupement particulier des GI consécutifs:

((d'abord)(la viande)) ((les légumes) (ensuite)) ((laisser) (mijoter))
 \BB HH \BB HH /BB B-B-

((d'abord)(la viande)) (les légumes)) ((ensuite) ((laisser) (mijoter)))
 \BB HH H/H HH /BB B-B-

Selon l'intonation employée, l'interprétation de la chaîne "ils regardent les voitures et les filles les pilotes" peut basculer de "eux, ils regardent les voitures, et, elles, elles regardent les pilotes" en "les pilotes, ils regardent les voitures et les filles".

ils regardent les voitures et les filles les pilotes
 /BB HH /BB B-B-

ils regardent les voitures et les filles les pilotes
 /BB HH B-B- b-.....b-

Ces exemples peuvent s'expliquer en termes de regroupement d'entités. Toute entité implique deux frontières: son début et sa fin. Il en est de même du σ . À chaque ton *final* est associé un degré de frontière. L'enchaînement de σ implique alors une suite de frontières d'importance variée, ce qui revient à un emboîtement de constituants ou – graphiquement – à une arborescence.

Pour indiquer le mécanisme de l'organisation de la chaîne linéaire grâce à l'intonation, Mertens P. (1987: 112-119, 1990: 169-172) propose le terme de *regroupement intonatif*. Pour chaque paire de σ consécutifs on détermine les degrés de frontière liés aux deux tons finals. Le rapport de force entre ces deux frontières résulte soit en la *juxtaposition* des deux groupes, soit en l'*inclusion* du premier dans le deuxième. La hiérarchie des tons, soit l'association ton – degré de frontière, suit la hauteur relative (.. B < .. < H < /H < ...), à l'exception du niveau infra-bas qui marque la frontière maximale (cf. Mertens P. 1987:112, 1990:170-171). Les unités emboîtées résultantes sont appelées *paquets intonatifs*. Soulignons que le regroupement est dicté par les tons finals; les tons de l'AI n'y participent pas.

La hiérarchie résultant du regroupement intonatif peut être représentée sous forme d'arborescence.

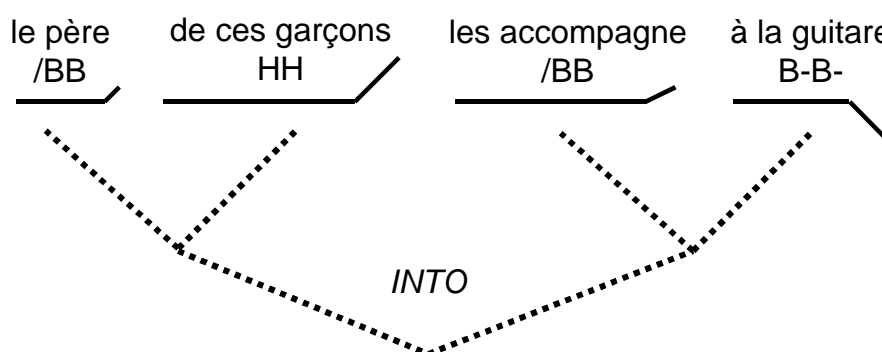


Figure 3. Regroupement intonatif pour la phrase "le père de ces garçons les accompagne à la guitare" prononcée avec le contour intonatif indiqué.

La représentation suivante est équivalente, mais elle visualise mieux la mise en place progressive et incrémentale de l'organisation prosodique, dans le temps, "de gauche à droite". Comme c'est la frontière droite qui est à l'origine de la structuration, la place d'une entité dans l'ensemble ne devient claire que quand cette entité est achevée. De même, le rapport entre deux paquets successifs n'est donné qu'au terme du deuxième paquet.

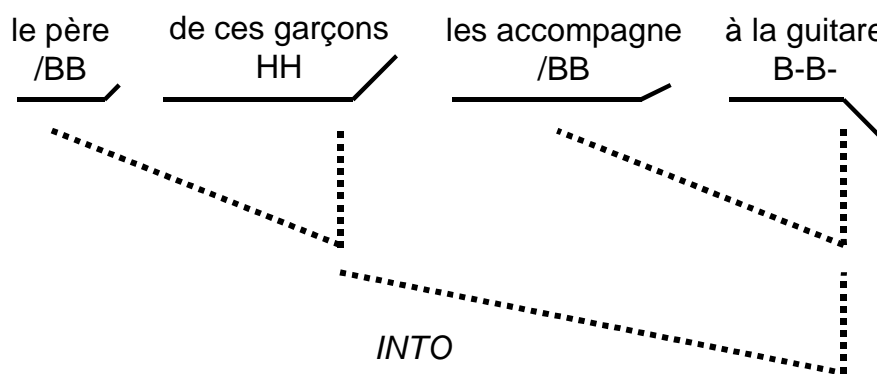


Figure 4. Représentation alternative du regroupement intonatif.

3. INTONATION ET SYNTAXE, CONGRUENCE OU NON-CONGRUENCE ?

La question de la congruence entre intonation et syntaxe est de savoir si les entités du plan prosodique sont conformes à celles du plan syntaxique, ou si en revanche la structuration intonative possède une certaine autonomie. Dans le premier cas, l'intonation serait prévisible et redondante; dans l'autre, les marques intonatives remplissent des fonctions que l'organisation syntaxique ne permet pas d'exprimer. La question est très débattue: dans Rossi *et al.* (1981), par exemple, les deux points de vue opposés s'affrontent. Pour y répondre, il faut d'abord préciser quelles sont les entités des deux plans et en quoi consisterait une conformité entre elles.

Comme on vient de le voir dans la section précédente, les entités du plan intonatif sont les groupes accentuels, les groupes intonatifs et les paquets intonatifs résultant du mécanisme de regroupement intonatif. Les entités syntaxiques sont bien sûr les constituants emboîtés (quoique l'organisation interne des constituants puisse varier considérablement d'une théorie syntaxique à l'autre, ce qui complique le raisonnement).

3.1. Congruence et constituance

Le critère de congruence couramment adopté est formulé le plus clairement par Dell F. (1984): pour qu'il y ait congruence entre les structurations syntaxique et intonative, il faut que les frontières prosodiques soient proportionnelles aux frontières syntaxiques. Ceci impliquerait donc que le regroupement intonatif respecte toujours l'emboîtement des constituants.

Les rapports syntaxiques et intonatifs, ainsi que leur similarité ou dissimilitude, peuvent être explicités de la façon suivante. Pour la clarté, les deux plans, syntaxique et intonatif, sont séparés; il faut cependant garder à l'esprit leur caractère simultané.

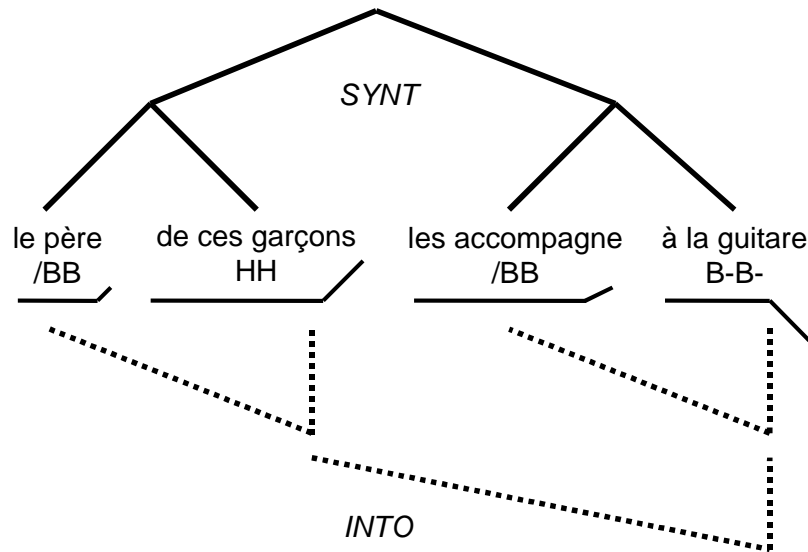


Figure 5. Structurations syntaxique et intonative pour la phrase "le père de ces garçons les accompagne à la guitare" prononcée avec le contour intonatif indiqué.

Le caractère agrammatical de la réalisation (a) ci-dessous s'explique alors par le regroupement intonatif qui réunit des sous-constituants appartenant à des constituants différents. Dans la variante (b), en revanche, l'enchâssement des constituants est respecté.

- le frère de Françoise chantait une mélodie compliquée
- (a) *(HH) (/BB HH) (HH B-B-)
- (b) (/BB HH) (HH) (/BB B-B-)

La représentation graphique de l'organisation résultant d'une intonation erronée met en évidence la dissymétrie entre la structure syntaxique sous-jacente et l'organisation imposée par le contour intonatif.

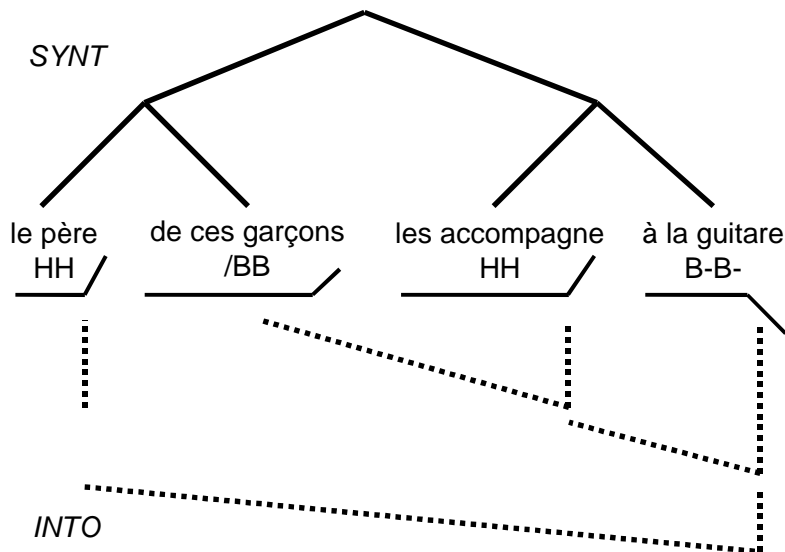


Figure 6. Conflit entre les organisations syntaxique et intonative.

Le critère de congruence basée sur la conformité entre les structures des deux plans rencontre pourtant certaines *difficultés*. (Celles-ci sont détaillées dans Mertens P. (1993, 1997), où l'on trouvera également de nombreux exemples tirés de corpus parlés.)

1. Rappelons d'abord que l'entité du GA a été introduite pour rendre compte du *nombre variable* de σ_1 pour une même structure syntaxique. Le nombre d'entités intonatives effectivement employées n'est donc pas strictement égal au nombre de syntagmes. L'énoncé suivant comptera quatre, trois, ou éventuellement deux σ_1 .

les amis de Pierre déménageront dans une semaine
 () () () ()
 () () () ()
 () () () ()

2. Ensuite, il faut noter que aussi bien le GA que le σ_1 peut intégrer plusieurs éléments qui, sur le plan syntaxique, constituent autant de constituants indépendants: c'est le cas des constituants *clitiques*: pronoms clitiques, articles, déterminants, etc. Pour écarter ce problème, il faudra réserver un statut particulier à ces constituants clitiques.

ils déménageront dans une semaine
 *() () ()
 () () ()

3. Un autre problème, c'est que le *nombre de degrés de frontière marqués intonativement* est bien inférieur au nombre d'emboîtements de constituants dans des phrases d'une certaine longueur.

Pour maintenir l'idée de conformité entre les entités des plans syntaxique et intonatif, il serait donc nécessaire d'apporter les précisions suivantes. 1. Les éléments clitiques ne comptent pas comme des constituants prosodiques; c'est ce qui est explicité par l'entité du GA. 2. Un constituant prosodique peut perdre son autonomie et s'intégrer à un constituant plus large (frontière basse non explicitée); la règle sur le regroupement des GA en σ_1 explicite ce comportement. 3. Si la hiérarchie de frontières dépasse une certaine complexité, l'intonation ne reproduit qu'un sous-ensemble des emboîtements (frontière élevée non exprimée). On modifiera le critère de la façon suivante: il faut que les frontières prosodiques soient proportionnelles à *un sous-ensemble des frontières syntaxiques*.

Cette précision est pourtant loin de résoudre tous les problèmes.

4. L'unité d'un constituant syntaxique construit peut être dissolue dans sa réalisation orale, et ses sous-constituants peuvent être soudés intonativement à un constituant extérieur au même syntagme. C'est le cas de l'objet dans la phrase suivante (a), où la relative est séparée de son antécédent. Pourtant le phénomène ne s'applique pas à toutes les relatives (b).

(a) les policiers ont retrouvé les traces || que les voleurs ont laissées
 (b) les policiers qui étaient là || ont retrouvé les traces

5. Enfin, une même chaîne linéaire, syntaxiquement univoque, peut être réalisée avec des regroupements intonatifs différents – même pour un nombre égal de σ_1 . Des exemples de ce type sont fournis par Mertens (1997: 44-47).

Ces deux dernières observations semblent irréconciliables avec un critère de congruence basée sur la constituance.

3.2. Congruence et dépendance

Mertens P. (1993, 1997) avance un critère de congruence alternatif qui permet de rendre compte de la plupart des problèmes cités. Ce critère est basé sur le *rapport de dépendance*¹² entre les parties regroupées dans une même entité prosodique. S'il existe un lien de dépendance direct entre deux σ consécutifs, leur regroupement intonatif est admis, comme dans l'exemple ci-dessous.

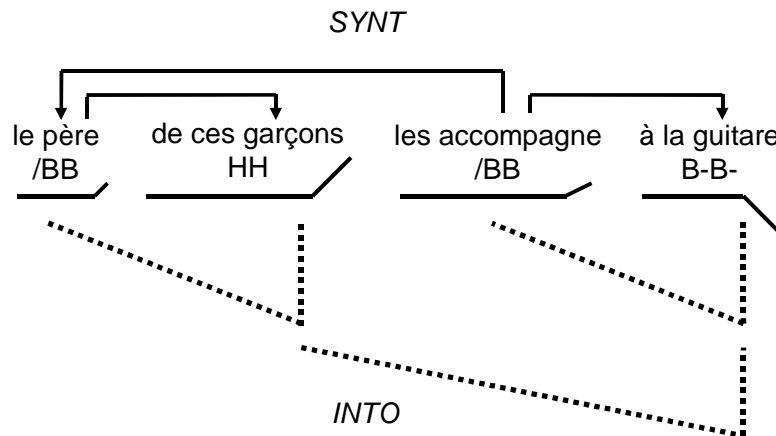


Figure 7. Relations de dépendance et regroupement intonatif pour la phrase "le père de ces garçons les accompagne à la guitare" prononcée avec le contour intonatif indiqué. La flèche (allant de la tête du constituant subordonnant à la tête du constituant dépendant) représente un lien de dépendance.

En revanche, si le paquet intonatif réunit deux σ sans lien direct, alors qu'un troisième σ pour lequel ce lien existe n'est pas inclus dans ce même paquet, alors l'intonation devient erronée. C'est le cas dans l'exemple suivant.

¹² Pour une comparaison des notions de dépendance, de valence, et de constituance: voir par exemple le chapitre 4 de P.H. Matthews (1981,1992).

SYNT

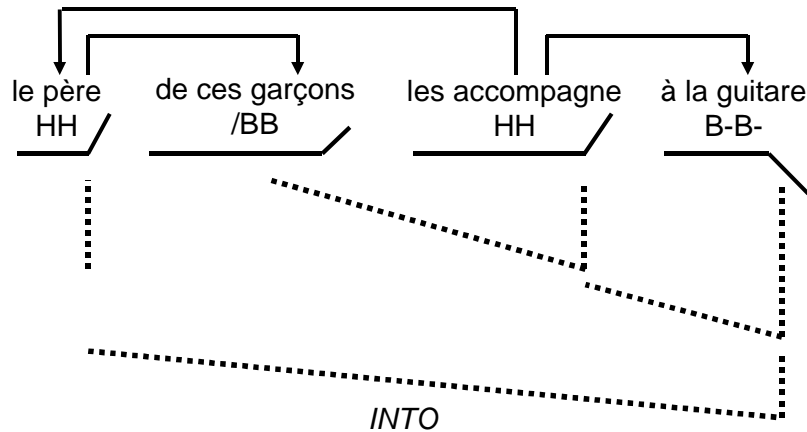


Figure 8. Conflit entre les rapports de dépendance et le regroupement intonatif pour la même phrase prononcée avec le contour intonatif indiqué.

Le critère peut alors être formulé comme suit: *Il y a congruence si le regroupement intonatif (des α_I en paquets) ne va pas à l'encontre des rapports de dépendance entre les (têtes locales des) α_I consécutifs.*

Il permet de rendre compte de certains cas problématiques pour le critère initial, basé sur la constituance. Dans l'exemple avec la relative ("les policiers ont retrouvé les traces que ..."), il existe en effet un lien direct entre l'antécédent ("traces") et le verbe ("ont retrouvé"), ce qui autorise leur regroupement par l'intonation.

Ce critère, qui était d'abord appliqué aux α_I , peut être généralisé aux entités prosodiques inférieures: soit aux mots regroupés dans le même α_A , soit aux α_A regroupés dans un même α_I :

1. Les clitiques sont attachés tantôt au α_A de gauche, tantôt à celui de droite, suivant qu'ils sont régis par la tête de gauche ou de droite.

(devrais-je) (le lui donner)

2. Plusieurs α_A peuvent être réunis dans un seul α_I , si l'un dépend de l'autre, ou en tout cas, si ce regroupement ne va pas à l'encontre des liens de dépendance qu'entretiennent ces α_A avec les α_A voisins. L'exemple suivant accepte plusieurs intonations.

((je cherche) (la personne)) ((responsable) (pour les livres))
 /BB HH /BB B-B-

(je cherche) ((la personne) (responsable)) (pour les livres)
 HH /BB HH B-B-

Si l'organisation due aux frontières intonatives ne rend pas tous les liens syntaxiques présents, elle reproduit toujours des liens locaux et ne va pas non plus à l'encontre des liens non exprimés.

On peut donc conclure que, telle qu'elle était formulée au début de cette section, la question de la congruence était en fait mal posée.

3.3. Congruence et construction syntaxique

Il reste des cas que le critère alternatif lié à la dépendance ne permet pas d'expliquer. Il s'agit de certaines *constructions* ou de certains *types de constituants*¹³: les dislocations, les phrases clivées, les adverbes de phrase et les incises.

1. Quand il est *postposé*, le constituant *disloqué* porte nécessairement un contour d'appendice, et le σ qui le précède reçoit un contour terminal ou conclusif (frontière intonative maximale, marquée par une chute au niveau infra-bas) (a). Il serait impossible de considérer ces deux σ comme deux entités de même statut, auxquelles on attribuerait des tons qui en feraient un seul paquet (constituant) construit (b).

- (a) Elles les avaient envoyés, les livres
 b.....h BB- b....b-
- (b) * Elles les avaient envoyés, les livres
 b.....b HH bb BB-

Si l'élément disloqué apparaît en premier (a), il reçoit un ton à frontière majeure (tel que HH, BH, ou H/H). Il est impossible de le regrouper avec un σ suivant autre que la phrase complète (b): l'élément disloqué forme un îlot; son regroupement est bloqué¹⁴.

- (a) Les livres, elles les avaient envoyés.
 bb HH b.....b HH b..b BB-
- (b) *Les livres, elles les avaient envoyés.
 bb /BB b.....b HH b..b BB-

2. Dans une *clivée*, l'élément extrait en "c'est ... qu" reçoit un ton à frontière majeure ou terminale (a). La suite de la phrase reçoit l'appendice si elle est brève (a); si elle est longue, on trouve soit une suite de σ indépendante (b), soit l'appendice (c).

- (a) C'est lui, qui l'avait trouvé.
 bb HB- b.....b-
- (b) C'est précisément là, que des problèmes d'emploi surgissent
 b.....b HH b.....b HH bb B-B-
- (c) C'est précisément là, que des problèmes d'emploi surgissent
 b.....b HB- b.....b-

Dans certains cas, l'intonation permet à elle seule d'imposer une interprétation de clivée, grâce à l'appendice (d), alors que sans appendice, la construction reste

¹³ Ces termes sont employés ici avec leur définition très générale venant de l'analyse en constituants immédiats (cf. Gleason H.A.1967:132): le constituant est une entité comportant plusieurs parties et qui peut entrer dans une configuration plus large. La construction est une configuration de constituants. "A constituent is any word or construction (or morpheme) which enters into some larger construction."

¹⁴ Rossi (1999: 64) note en outre qu'il est impossible d'appliquer à la suite de la phrase un contour d'appendice (qu'il appelle "parenthèse basse", reprenant la terminologie de P. Delattre):

ce livre, je sais pourquoi Jean l'a acheté
 * bb HB- b.....b-

ambiguë (e). La clivée et la relative, identiques au plan morphologique, sont donc différentes sur le plan intonatif.

- (d) C'est l'homme qui a pollué la mer. (clivée)
 bb HB- b-.....b-
- (e) C'est l'homme qui a pollué la mer. (relative)
 bb HH b.....b HH bb BB-

3. Les *adverbes de phrase* (ainsi que les locutions adverbiales apparentées) portent l'appendice lorsqu'ils apparaissent en finale d'énoncé (a). En cela ils se distinguent d'autres catégories d'adverbes, qui n'accepteraient pas l'appendice (b). Si la phrase comporte plusieurs adverbes, seul l'adverbe de phrase reçoit l'appendice (c).

- (a) Tu ne vas pas nous faire ça, franchement!
- (b) Il marchait lentement.
- (c) L'herbe de la pelouse pousse lentement, décidément.

Certains adverbes sont ambigus: tantôt adverbe de phrase (d), tantôt adverbe de temps (e); l'intonation lève l'ambiguïté.

- (d) Tu vas travailler maintenant (= tu vas travailler, enfin ?)
 HB- b-....b-
- (e) Tu vas travailler maintenant (=c'est maintenant que tu dois travailler)
 HH B-B-

En position initiale, l'adverbe de phrase constitue un α autonome (a); il ne se regroupe pas dans un même α avec le constituant suivant, même s'il en dépend syntaxiquement (b). C'est comme si cette construction introduisait une barrière qui empêche le regroupement intonatif¹⁵.

- (a) Décidément, ça s'améliore.
 b...b HH b.....b B-B-
- (b) Décidément, ça s'améliore.
 * b.....b B-B-

4. Les α à l'intérieur d'une incise prosodique (marquée par un changement de registre) peuvent se regrouper entre eux, mais pas avec ceux en dehors de l'incise.

Ces exemples illustrent le rôle crucial de la construction syntaxique. (Cet aspect a été peu étudié; citons toutefois: Mertens P. 1993, 1997; Eynde K. *et al.* 1998; Rossi M. 1993, 1999. La liste de constructions citées n'est d'ailleurs pas exhaustive.) Les approches automatiques actuelles en génération de l'intonation ne sauraient rendre compte des faits cités: ces méthodes reposent sur un étiquetage grammatical; or, dans la plupart des cas, les catégories grammaticales ne permettent pas à elles seules d'identifier la construction.

Plutôt que de se limiter à une simple énumération des constructions concernées, il serait préférable de fournir un critère unique qui rende compte des différents cas. La première solution qui vient à l'esprit consiste à dire que ces constructions introduisent une frontière infranchissable bloquant le regroupement intonatif habituel.

¹⁵ Rossi (1999: 71-72) fait la même observation.

Alternativement, on pourrait envisager l'hypothèse que le rapport syntaxique entre ces constituants îlots et le reste de l'énoncé ne correspond pas à un lien de dépendance, mais à un type de lien fondamentalement différent. Dans cette vue, les cas avec les constituants îlots ne violeraient pas la règle de regroupement, puisque celle-ci ne s'appliquerait pas à eux.

4. DESCRIPTION DE L'ALGORITHME DE GÉNÉRATION *MINGUS*

Mingus est un système qui génère des courbes mélodiques pour la synthèse à partir du texte, et à partir d'une représentation de la structure syntaxique des phrases. Dutoit (1997: 147-160) classe les méthodes existantes en trois catégories selon qu'elles se servent d'heuristiques simples, de règles morpho-syntaxiques, ou d'un apprentissage automatique à partir d'un corpus d'entraînement. Parmi les travaux récents pour le français, on peut citer Aubergé V. 1991, Beaugendre, F. 1994, Delais E. 1995, Boula de Mareüil Ph. 1997, Di Cristo *et al.* 1998, Gaudinat A. & Goldman J.-P. 1998, Malfrère F. *et al.* (1998), Morlec Y. *et al.* (1995), Véronis J. *et al.* (1998); dans la plupart des cas, il s'agit de méthodes automatiques.

Alors que la plupart des systèmes récents sont de type automatique, *Mingus* est à situer parmi les approches par règles morpho-syntaxiques; il se caractérise par le grand nombre d'informations syntaxiques prises en compte.

La génération de l'intonation comporte un nombre assez élevé d'étapes. Pour des raisons de clarté, elles seront présentées en deux temps, allant du général au particulier. On décrira d'abord les mécanismes de base, qui s'appliquent aux constituants enchâssés (§ 4.2. - 4.6.). Leur application produit une intonation dite "neutre", où le regroupement intonatif respecte les liens syntaxiques. Ensuite seront traitées les extensions requises pour rendre compte de constructions syntaxiques particulières, ainsi que des marqueurs prosodiques particuliers, tels que le focus et l'appendice (§ 4.7.). Enfin, on précisera l'ordre des étapes dans le système complet (§ 4.9.).

4.1. Choix de la structure de données

Peu de systèmes de génération de la prosodie partent d'une analyse syntaxique, même partielle. Certains (Mareüil Ph. 1997, Vannier G. 1999) effectuent une analyse locale ou "superficielle" (*shallow parsing*), identifiant des constituants dits "non récursifs" (sans emboîtement) ou même des entités ad hoc (tels que les *chunks*), sans indication des liens syntaxiques entre eux. L'entrée peut alors être représentée par une simple chaîne, faite de mots et de frontières.

Vu l'importance de la syntaxe, une approche radicalement différente est adoptée par *Mingus*. Elle prend comme structure de données l'arbre syntaxique lui-même et conçoit la génération de la prosodie comme des manipulations successives de l'arborescence. Ces opérations sont d'ailleurs assez variées: 1. tantôt il s'agit de remplacer des noeuds relativement bas par des entités correspondant à des unités prosodiques (tels les α et Γ); 2. tantôt il s'agit de modifier des noeuds relativement élevés pour aplatir la structure syntaxique ou pour la réorganiser localement (§ 4.8.); 3. enfin, les noeuds sont enrichis en leur associant des marques intonatives. Pour

chacune des opérations, on part de la racine de l'arborescence et on descend vers les noeuds terminaux, de façon récursive.

Étant donné que les liens de dépendance syntaxique interviennent dans plusieurs étapes du traitement, l'arborescence doit expliciter cette information. Jusqu'à ce jour il n'existe pas, à notre connaissance, d'analyseur spécialement conçu pour une grammaire de dépendance *pure*. C'est pourquoi on utilise un formalisme hybride qui enrichit la grammaire hors-contexte classique d'une marque indiquant la tête du constituant (cf. aussi la proposition de Matthews P. 1981.). Bien sûr, une grammaire d'arbres adjoints (Abeillé A. 1993) ou une grammaire syntagmatique guidée par les têtes (Pollard C.J. & Sag I. 1987), plus riches que la grammaire hors contexte puisqu'elles intègrent la notion de tête, conviendraient tout aussi bien.

Pour l'illustrer, prenons l'exemple d'une règle de réécriture comportant trois noeuds fils, où le noeud du milieu serait le noeud tête, sans préciser le caractère terminal ou non des noeuds. La notation en grammaire hors-contexte (CFG) serait: $X \rightarrow A B C$. L'arbre correspondant n'indique pas lequel des noeuds est la tête. La grammaire de dépendance de Lucien Tesnière (1959), représenterait cette information sous la forme d'un *stemma*; il s'agit d'un graphe où l'élément subordonnant (le régissant, la tête) est placé au-dessus des éléments subordonnés (régis, dépendants). Le stemma ne comporte que des terminaux. Le format hybride est identique au format CFG, à l'exception du signe + indiquant la tête. Comme il n'y a pas de contrainte sur le nombre de fils, les arborescences sont généralement plus plates que celles produites par une grammaire régulière, ou par une grammaire de type X-bar.

CFG	Stemma	Forme hybride
$X \rightarrow A B C$		$X \rightarrow A +B C$
<pre> X / \ A B C </pre>	<pre> B / \ A C </pre>	<pre> X / \ A +B C </pre>

Figure 9. Configuration de l'arborescence syntaxique et format des règles de réécriture dans une grammaire hors contexte (CFG), dans la grammaire de dépendance de L. Tesnière et dans une forme hybride.

Passons aux différentes étapes dans le calcul de l'intonation.

4.2. Constitution des groupes accentuels

Un noeud terminal sera étiqueté *tonique* (ou "*non clitique*") s'il appartient à une des catégories grammaticales toniques énumérées dans le § 2.4. Tout autre élément terminal est classé *clitique*. Pour les noeuds non terminaux, le caractère clitique dépend des noeuds fils: s'il comporte un fils tonique, le noeud père sera lui aussi tonique. Chaque noeud non terminal regroupe des parties syntaxiquement liées (directement ou indirectement); elles peuvent être réunies dans un même **GA**

pourvu qu'elles ne comportent qu'un seul élément tonique. Pour tout noeud non terminal, on examine donc les noeuds fils. Un fils clitique est regroupé avec son frère droit tonique (qui le suit immédiatement dans la chaîne), ou s'il n'y en a pas, avec le frère gauche. Pour traiter les séquences de clitics, comme dans "que tu ne les lui prennes pas", l'association est répétée jusqu'à ce que tout clitique soit associé. Il en résulte une réorganisation locale de l'arbre qui insère les noeuds clitiques dans le noeud dont le non-clitique est la tête. Grâce à l'emploi de l'arborescence, les parties clitiques sont associées correctement à l'élément tonique dont elles dépendent, sans qu'on ait à préciser la nature du lien syntaxique ou le type de constituant. (L'arborescence est ensuite simplifiée par l'élimination des noeuds à un seul fils.)

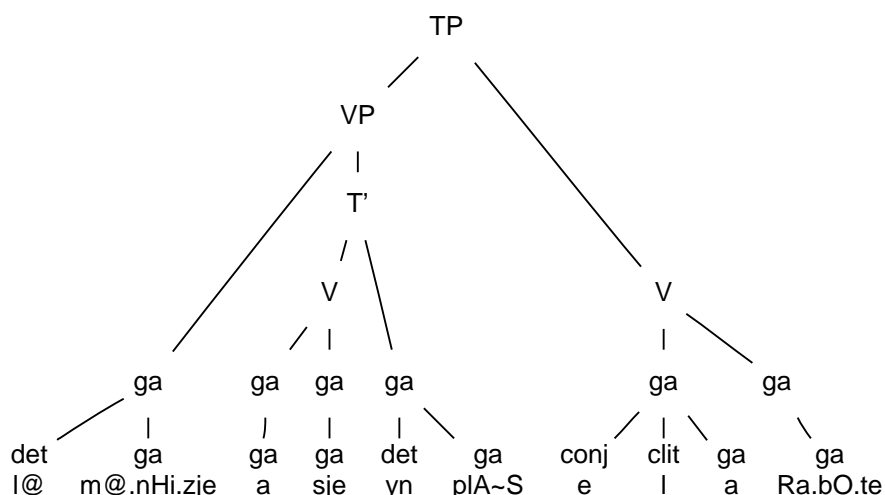


Figure 10. Arborescence après la constitution des \mathfrak{GA} , pour la phrase: "le menuisier a scié une planche et l'a rabotée". Transcription phonétique selon la norme SAMPA. La conjonction clitique "et", qui dans la structure syntaxique constitue une branche indépendante, a été intégrée au \mathfrak{GA} suivant (cf. *infra*, § 4.3.). L'emboîtement "((le menuisier) ((a scié) (une planche)))" s'explique à partir du modèle syntaxique adopté par *Fips*.

4.3. La constitution des groupes intonatifs à partir des GA

La formation des \mathfrak{GI} consiste à regrouper des \mathfrak{GA} courts, tout en respectant l'emboîtement syntaxique. Seuls des constituants liés syntaxiquement sont regroupés, grâce à l'arborescence qui fournit la constituance. Ce regroupement dépend d'un paramètre indiquant la taille d'un \mathfrak{GI} long, égale à 6 syllabes par défaut. Ce seuil peut être modifié, par exemple en fonction de la vitesse d'élocution, ou en fonction de la longueur de la phrase.

La procédure parcourt l'arbre à partir du noeud racine. Si le nombre de syllabes à un noeud donné est inférieur ou égal au seuil, les \mathfrak{GA} aux noeuds fils sont fusionnés dans un même \mathfrak{GI} : le noeud en question reçoit le trait 'GI'. Dans le cas contraire, le noeud envisagé comportera plusieurs \mathfrak{GI} ; la procédure est alors appliquée aux noeuds fils.

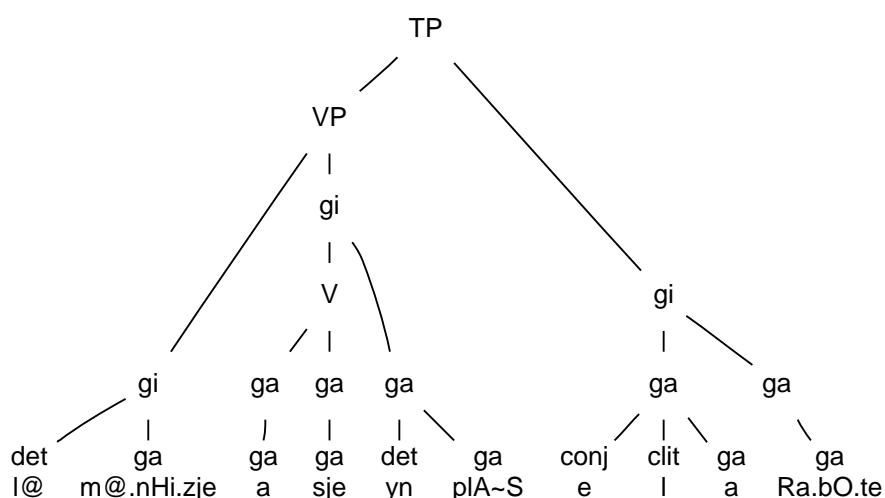


Figure 11. Arborescence de la phrase de la figure précédente, après la constitution des σ_I .

Il arrive qu'un constituant prosodique long comporte des noeuds fils de longueur très inégale, comme dans “((c'est) ((la maison) (de mon collègue)))” (1 (3 4)). Pour éviter les σ_I monosyllabiques, la procédure est modifiée de sorte qu'un σ_A monosyllabique sera d'abord intégré au σ_A suivant avant d'appliquer le critère de regroupement: on obtient “(c'est la maison) (de mon collègue)” (4 4). Ce traitement est réservé aux cas où il existe un lien de dépendance entre le σ_A monosyllabique et le σ_A suivant.

L'entité résultante après l'insertion du σ_A court doit être positionnée syntaxiquement comme tête ou dépendant par rapport aux noeuds frères. Cette information s'obtient par l'inspection (du statut de dépendance) des parties regroupées: si une des parties était la tête syntaxique, l'entité englobante le sera également. “($X_{(+A)}$ c'est) ($Y_{(+B)}$ la maison) (C de mon collègue))” correspond à [X +A [Y +B C]]. Le réarrangement donne [X [+Y +A B] C] .

Dans l'exemple suivant, le σ_A monosyllabique “prend” sera intégré au σ_A qui le suit.

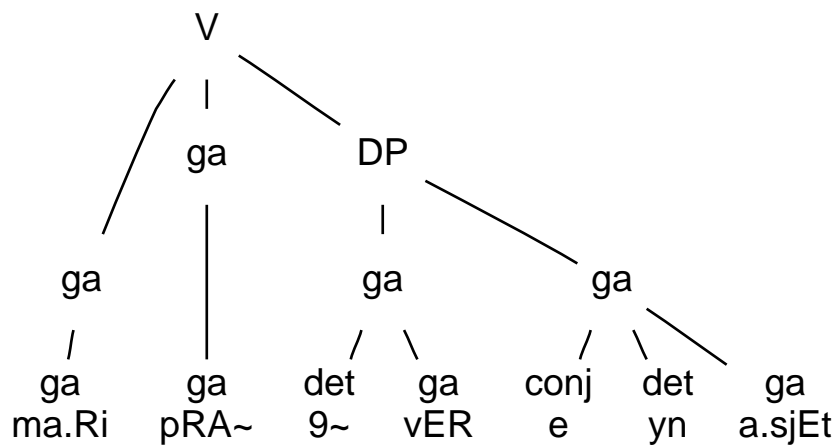


Figure 12 a. Arborescence de la phrase “Marie prend un verre et une assiette”, après la formation des $\mathcal{G}\mathcal{A}$.

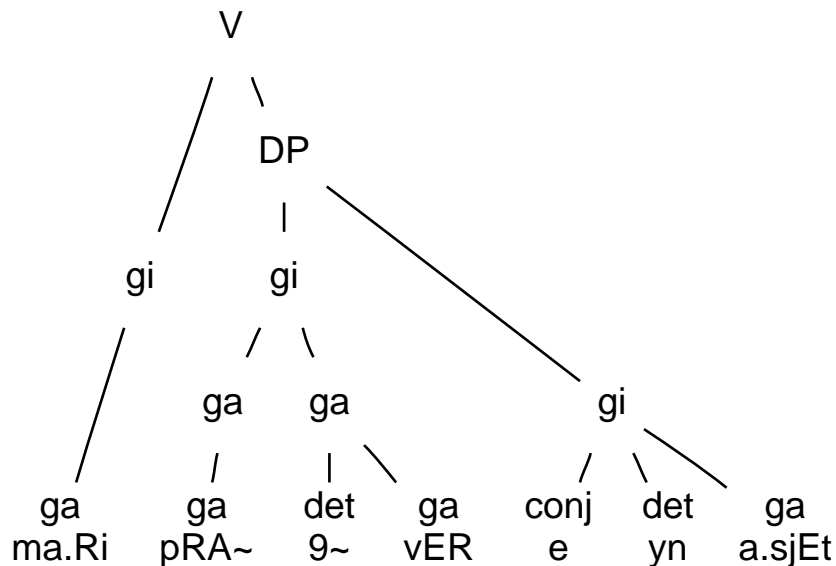


Figure 12 b. Arborescence de la phrase “Marie prend un verre et une assiette”, après après la constitution des $\mathcal{G}\mathcal{I}$.

4.5. L'attribution des frontières prosodiques

Maintenant que les $\mathcal{G}\mathcal{I}$ sont formés, on peut leur attribuer des degrés de frontière prosodique. Ceux-ci détermineront à leur tour le choix des tons à l'accent final, de telle sorte que le mécanisme de regroupement intonatif soit respecté.

Cette étape associe à tout noeud (constituants et $\mathcal{G}\mathcal{I}$) un degré de frontière. La version actuelle utilise quatre degrés de frontière, où la valeur 4 correspond à la frontière maximale, soit la fin de la phrase. Les degrés de frontière sont attribués de façon récursive, en partant de la racine qui reçoit le degré maximal: tel noeud qui reçoit un degré D , transmet ce même degré à son fils droit, et un degré $D-1$ aux fils précédents.

L'appendice (forcément en fin d'énoncé, donc de degré $D = 4$) reçoit un traitement spécial: il passe la frontière D au constituant qui le précède (cf. *infra*)¹⁶.

4.6. Attribution des tons de l'accent final.

Le choix des tons se limite à la recherche du ton dans un tableau en fonction du degré de frontière, de la modalité (interrogative ou énonciative) et de certains marqueurs prosodiques (focus, appendice) dont il sera question dans les paragraphes suivants.

ton	degré	modalité	focus
b-..b-	4	appendice bas	-
h..h	4	appendice haut	-
B-B-	4	déclarative	-
H/H	4	interrogative	-
HB-	4	déclarative	+
HB	3	*	+
BH	3	*	-
HH	3	*	-
/BB	2	*	-
BB	1	*	-

Tableau 3. Tableau permettant la sélection du ton AF en fonction des propriétés associées: degré de frontière, modalité, focus.

Le graphe ci-dessous indique les GI et les tons attribués.

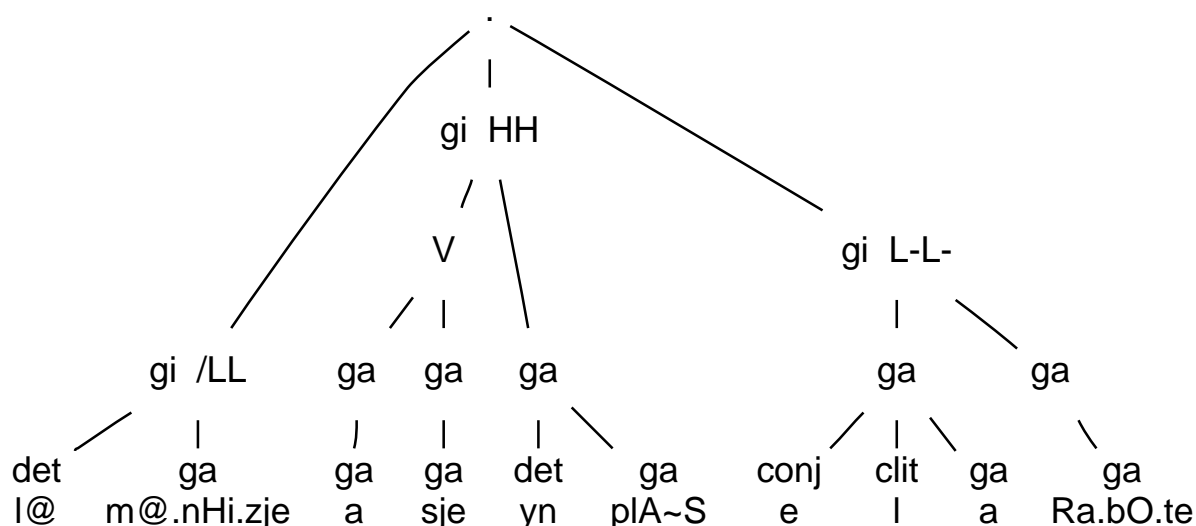


Figure 13. Arborescence de la phrase “Le menuisier a scié une planche et l’a rabotée” après l’attribution des tons finals aux groupes intonatifs.

¹⁶ Le lecteur attentif aura remarqué que l'appendice est traité comme un GI autonome par *Mingus*, alors que la théorie intonative les considère comme une partie de GI.

Les paragraphes suivants décrivent les ajouts nécessaires pour les marqueurs prosodiques.

4.7. Traitement des marqueurs prosodiques.

Le terme *marqueur prosodique* désigne une annotation d'un segment de la phrase à analyser (ou une de ses formes dérivées: constituant, \mathcal{G}_A , \mathcal{G}_I). La présence du marqueur aura des répercussions sur l'intonation produite: un marqueur peut entraîner l'emploi d'un ton précis, ou le blocage d'une opération telle que le regroupement. Les marqueurs employés actuellement sont: la modalité, le focus, l'appendice, la clivée. Une version future pourrait ajouter l'incise, la délimitation gauche (l'accent initial), la pénultième haute, d'autres commandes pragmatiques ou rythmiques.

Dans le prototype de *Mingus* de 1999, la modalité phrastique et la focalisation sont identifiées à partir des signes de ponctuation: '.', '?' ou '!'. Le point d'exclamation entraîne l'insertion du marqueur "focus" qui a son tour engendra le ton HB- ou HB, comme l'indique le tableau 3. (En l'absence de ponctuation, la modalité déclarative est attribuée par défaut.) L'appendice est associé à tout adverbe "de phrase" placé en position finale de phrase. L'extraction ("c'est ... que...") est identifiée en tant que constituant dans l'analyse syntaxique. Tous ces éléments sont représentés par des marqueurs explicites, qui sont associés aux arcs du *chart*, donc à certaines parties (ou positions) du texte d'entrée, soit aux mots ou à des groupes de mots. Il aurait été possible de les associer aux noeuds de l'arborescence, à condition de les déplacer à chaque fois que celle-ci est modifiée par l'algorithme.

Dans la construction *clivée*, la partie placée en extraction reçoit un marqueur qui bloque son regroupement éventuel avec des \mathcal{G}_A à sa droite.

Après la constitution des \mathcal{G}_I , les marqueurs peuvent leur être attribués, afin de sélectionner plus tard les tons appropriés. Pour tout \mathcal{G}_I dans l'arbre, on recherche dans le "chart" les marqueurs associés, et on les ajoute au \mathcal{G}_I comme des traits.

Dans une phrase qui se termine par un *appendice*, les marqueurs de modalité et de focalisation éventuels passent au \mathcal{G}_I qui le précède.

4.8. Réajustements de la structure syntaxique.

On rencontre certains regroupements intonatifs qui reviennent à des réajustements de l'arbre syntaxique. Il peut s'agir du découpage d'un constituant long (déliage d'un emboîtement), ou de la réunion de deux parties dans une entité reconstituée. Ces opérations supposent que soit connu le *lien de dépendance* entre \mathcal{G}_I successifs. Entre deux groupes A et B, il existe un lien de dépendance si la tête syntaxique à l'intérieur de A régit la tête de B, ou vice versa. Ces données sont préparées à l'avance: pour chaque \mathcal{G}_I on détermine s'il entretient un lien de dépendance avec le \mathcal{G}_I suivant, quelle qu'en soit l'orientation: de tête en dépendant, ou l'inverse.

1. L'exemple "les policiers ont retrouvé (les traces || que les ravisseurs avaient laissées)" (§ 3.1), avec la relative apposée à un groupe nominal, permet d'illustrer le *découpage*. Un constituant long sera coupé en deux parties s'il est structuré (il comporte plusieurs \mathcal{G}_I), s'il entretient un lien de dépendance avec le \mathcal{G}_I précédant

(en l'occurrence le verbe principal), et si le sous-constituant gauche est la tête syntaxique du constituant entier. On coupera alors après le premier sous-constituant. Dans l'exemple donné on sépare la relative de son antécédent.

2. Dans l'exemple "(les policiers HH) (ont retrouvé /BB) (les traces HH)..." les σ_2 et 3, qui sont juxtaposés dans l'arbre syntaxique, peuvent être emboîtés intonativement s'il existe entre eux un lien de dépendance: "(les policiers) ((ont retrouvé) (les traces))...". Afin d'obtenir ce *regroupement*, on examine les paires de σ_1 frères; si le nombre de syllabes cumulé est inférieur à une valeur critique, on reconstitue un constituant qui les englobe.

4.9. Ordonnement des étapes de traitement

Afin de tenir compte de l'interaction des différents facteurs, les étapes sont ordonnées de la façon suivante.

1. Identification des marqueurs intonatifs.
2. Calcul du nombre de syllabes pour chaque symbole terminal.
3. Constitution des σ_A .
4. Constitution des σ_1 .
5. Indication du lien de dépendance direct entre σ_1 consécutifs.
6. Réajustement de structure en vue du découpage éventuel de constituants
7. Réajustement de structure en vue du regroupement éventuel de constituants
8. Attribution des degrés de frontière
9. Attribution des tons (en fonction des frontières et des marqueurs)

5. DE LA REPRÉSENTATION TONALE AU CONTOUR ACOUSTIQUE

Pour réaliser une courbe mélodique, le synthétiseur (MBROLA¹⁷) utilise des valeurs de fréquence fondamentale (F_0) à des instants temporels précis, appelées *cibles*. La courbe obtenue correspondra à une suite de segments de droite reliant les cibles successives par interpolation linéaire sur l'axe de fréquence. (C'est le fonctionnement adopté par le synthétiseur MBROLA.) Il nous faut donc un module "mélodique" qui transforme la représentation tonale obtenue en une suite de cibles acoustiques associées aux sons successifs de l'énoncé. Comment y parvenir? Ce paragraphe décrit l'approche suivie dans *Mingus*.

Rappelons d'abord que le ton n'est autre qu'une suite de *niveaux de hauteur*. Les niveaux (B-, B, H, H+) se représentent naturellement par des *lignes droites superposées*¹⁸. Leur hauteur (donc la fréquence fondamentale) s'ajuste à la tessiture du locuteur, à son état émotif,... Leur pente peut être choisie pour reproduire la "ligne de déclinaison", phénomène de la diminution progressive de la fréquence moyenne au cours du cycle respiratoire (le groupe de souffle). À partir de la position des sons sur l'axe du temps, on peut donc déterminer la fréquence correspondant à tel niveau

¹⁷ cf. Dutoit (1997)

¹⁸ L'utilisation de lignes de déclinaison en synthèse de la mélodie est très répandue quelle que soit l'approche: le "grid" de l'école de Lund, le *gabarit* de Beaugendre F. 1994, chez Goldman J.-P. 1997, etc.

de hauteur. Pour un son situé à l'instant t et portant le ton T , on choisit la ligne du niveau T et on détermine la hauteur de la ligne à cet instant t .

Pour chaque ton, on spécifie la liste des cibles. Dans ce modèle simple les cibles s'associent aux voyelles (plutôt qu'aux syllabes, par exemple). Chaque cible est une coordonnée (x,y) ; elle comporte deux éléments: une position sur l'axe temporel et une valeur indiquant le niveau de hauteur atteint à cette position. La position temporelle s'exprime comme une fraction de la durée de la voyelle. Par exemple, pour le ton HB- on a trois cibles: le niveau H est atteint après 33% de la durée de la voyelle, et maintenu jusqu'à 40% de la durée, puis on atteint le niveau B- à 90%.

ton	liste des cibles
b-	(33, B-), (66, B-)
b	(33, B), (66, B)
h	(33, H), (66, H)
B-B-	(80, B-), (90, B-)
H/H	(33, H), (50, H), (100, /H)
HB-	(33, H), (40, H), (90, B-)
HB	(33, H), (40, H), (90, B)
BH	(33, B), (40, B), (90, H)
HH	(33, H), (66, H)
/BB	(33, /B), (66, /B)
BB	(33, B), (66, B)

Tableau 4. Liste de cibles pour chaque ton employé. Chaque paire (x,y) représente les coordonnées (*fraction de la durée du son, niveau de hauteur*) d'une cible.

Le *nombre* de cibles nécessaire pour réaliser un ton donné dépend bien sûr de la complexité de sa courbe: un ton dynamique demande plus de cibles qu'un ton statique. À première vue, on se contenterait d'un minimum de cibles. Avec une seule cible pour le ton statique (soit une cible par syllabe), on obtient cependant des courbes différentes selon que le niveau de hauteur à la syllabe précédente est identique ou non. C'est pourquoi on utilisera un minimum de deux cibles par ton, l'une pour le "début" (33%) du ton, l'autre pour la "fin" (66%).

Si l'on veut obtenir des courbes mélodiques naturelles, la *position* des cibles, c'est-à-dire le moment dans le son où le niveau de hauteur est atteint, est absolument cruciale.

L'intérêt de cette approche est qu'elle permet un nombre variable de cibles, et donc des formes variées. Ceci permet de choisir une forme particulière de la courbe pour un même ton abstrait; ce choix peut être déclenché par plusieurs facteurs, comme le contexte phonétique, le débit, la durée syllabique, ou des critères phonostylistiques (style de prononciation, accent régional, dialecte,...).

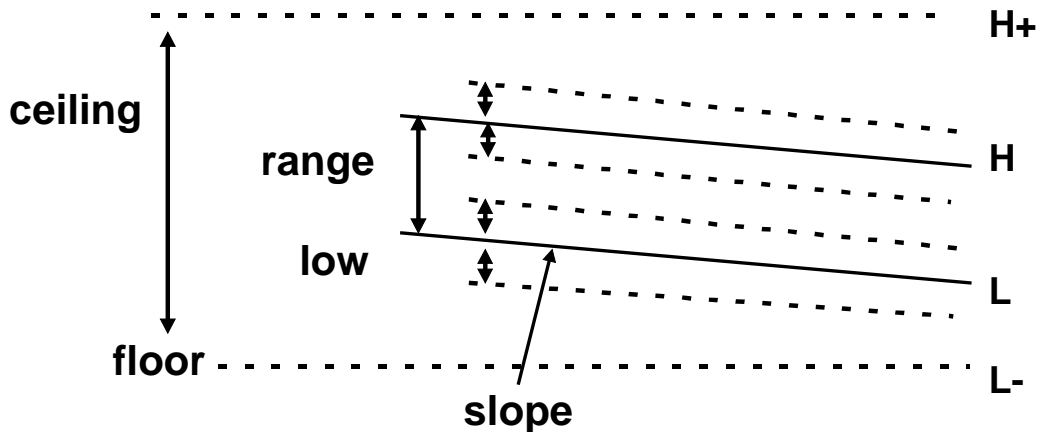


Figure 14. Ensemble de lignes représentant l'évolution des niveaux de hauteur au cours de l'énoncé (échelle logarithmique en demi-tons). *Floor*, *low*, *range*, etc. sont les paramètres qui définissent cet ensemble de lignes de repère. Les signes L- et L remplacent ici B- et B.

Les propriétés générales de la "portée mélodique" sont ajustées à l'aide des paramètres suivants:

1. Le *plancher* (ligne inférieure, *floor*): la valeur (absolue) de fréquence correspondant au niveau infra-bas (B-), en Hz.
2. Le *plafond* (ligne supérieure, *ceiling*): c'est le niveau suraigu (H+); il est spécifié comme un intervalle mélodique au-dessus du plancher, en demi-tons (ST).
3. Le *bas* (*low*), soit la fréquence correspondant au niveau de hauteur bas (B), au début de l'énoncé.
4. La taille de l'intervalle majeur (*range*): c'est l'intervalle mélodique (en demi-tons) qui sépare le niveau haut (H) du niveau bas (B).
5. L'intervalle mineur: la distance (en demi-tons) qui sépare les niveaux de hauteur rehaussé ou abaissé du niveau primaire (B ou H) correspondant.
6. La *pente* (*slope*): c'est l'inclinaison donnée aux lignes des niveaux H et B (et indirectement aux valeurs abaissées ou rehaussées), soit l'intervalle mélodique entre ses valeurs au début et à la fin de l'énoncé.

Notons que les lignes des niveaux B- et H+ ne sont pas affectées par ce paramètre de pente. Ils sont définis par rapport à la tessiture du locuteur.

Le système des niveaux est défini de façon relative et en demi-tons, soit sur une échelle logarithmique. Il en est de même des valeurs interpolées, par exemple la fréquence fondamentale qui correspond à un ton T à l'instant t . Les seules exceptions sont le plancher et le bas, définies de façon absolue en Hertz. Il faut en effet un point de référence absolu.

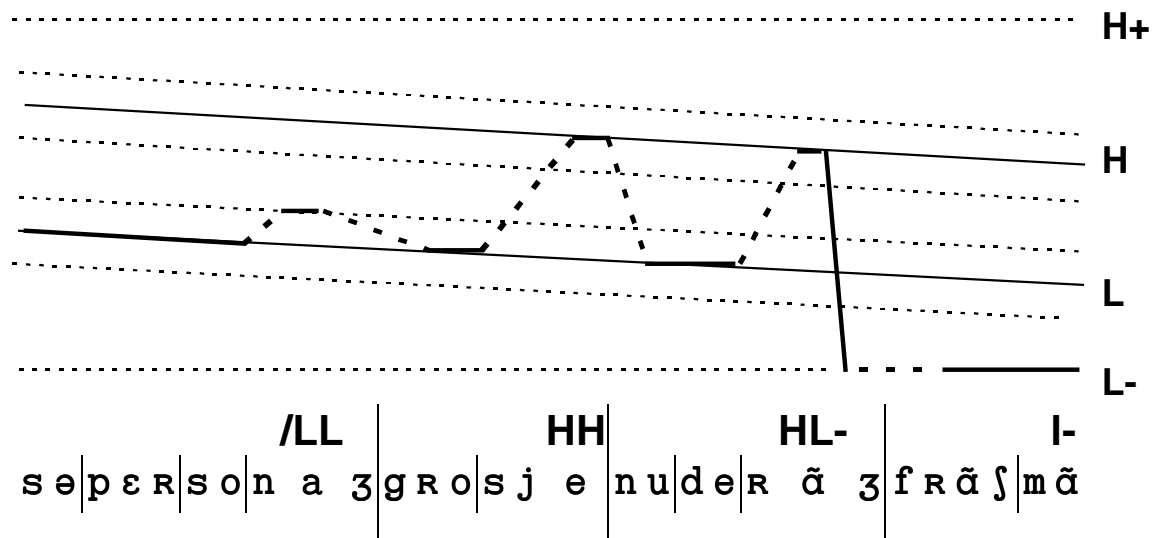


Figure 15. Calcul de la courbe de fréquence fondamentale à partir des durées segmentales et des tons associés aux segments. Exemple d'une phrase concrète avec les phonèmes, les tons, les cibles et la courbe interpolée (sur une échelle logarithmique en demi-tons). Les signes L- et L remplacent ici B- et B.

Ce modèle mélodique paramétrisé permet d'adapter les propriétés globales de la courbe, indépendamment de la représentation tonale "abstraite", en fonction registre du locuteur individuel (c'est-à-dire de la base de données de diphones utilisée par le synthétiseur), en fonction de facteurs paralinguistiques (caractéristiques émotionnelles ou expressives), etc.

6. PRÉDICTION DE LA DURÉE DES SONS

L'unité rythmique choisie étant la syllabe, la prédiction de la durée de chaque phonème d'une phrase à synthétiser se fait en deux étapes: la première détermine la durée totale de chaque syllabe à partir des moyennes issues de mesures dans des corpus; puis la seconde ajuste la durée de chaque phonème au sein de cette syllabe. Nous décrivons succinctement les aspects généraux de ces deux étapes.

6.1 Prédiction de la durée syllabique et des pauses

Les travaux de référence dans le domaine de la prédiction de la durée syllabique sont les deux études Padeloup V. (1992a, 1992b) et Kruckenberg A. & Fant G.(1995). L'une prend en compte le degré d'accentuation de la syllabe, sa position dans le groupe intonatif ainsi que la longueur de ce dernier, l'autre ne considère que le caractère accentué de la syllabe et son nombre de phonèmes.

Les études de Padeloup n'ont pas pour objectif initial de faire une prédiction sur la durée syllabique mais simplement d'étudier celle-ci en fonction de trois facteurs : le type d'accent (primaire, secondaire¹⁹, non-accentué), la position syllabique par

¹⁹ Padeloup (1992a: 531) distingue 3 types d'accent: 1. l'accent primaire, qui correspond à notre AF; 2. l'accent secondaire qui apparaît "à l' initiale du mot, sur l'antépénultième d'un mot lexical, à la

rapport à la prochaine syllabe accentuée si elle n'est pas accentuée, enfin le nombre de syllabes du groupe intonatif. Le principal résultat montre "un ralentissement progressif du débit qui s'accroît sur la syllabe pénultième et culmine sur la syllabe accentuée", quelles que soient les valeurs de l'accent et du nombre de syllabes dans le groupe accentuel.

Nous confrontons à ces mesures les études de Kruckenberg A. et Fant G. (1995) qui comparent les durées syllabiques en français et en suédois et donnent des résultats sous forme de régression linéaire de la *durée moyenne* des syllabes accentuées et non-accentuées en fonction du nombre de phonèmes n , variable non prise en compte par Padeloup.

	Non-accentuée	Accentuée
Durée syllabique (en ms)	$d = -10,5 + 64,5 \cdot n$	$d = 82,5 + 49 \cdot n$
Longueur syllabique moyenne (en phonemes par syllabe)	2.1	2.3

Tableau 5. Régression linéaire des durées syllabiques en fonction du nombre de phonèmes n et longueur syllabique moyenne

Une mise en parallèle de ces deux références montre que les deux études concordent (voir tableau suivant) et nous avons intégré les avantages des deux en un seul modèle de prédiction de la durée.

	Padeloup (1992)		Kruckenberg & Fant (1995)
	accent primaire (AF)	accent secondaire (AI)	
syllabe pénultième	152 ms	120 ms	124 ms
autres syllabes inaccentuées	131 ms	120 ms	124 ms
syllabe accentuée	250 ms	154 ms	195 ms

Tableau 6. Durées syllabiques moyennes dans deux études

Le modèle proposé prédit en trois étapes la durée syllabique en fonction du degré d'accentuation, du nombre de phonèmes et du degré de frontière prosodique :

1. la durée syllabique initiale est obtenue selon les données de Padeloup (1992) en fonction de la position syllabique et du degré d'accentuation (colonnes 1 et 2)
2. un ajustement selon le nombre de phonèmes est effectué par les équations suivantes :

$$\begin{aligned}
 dur_{FINALE} &= dur_{PASDELOUP} + 49(n - 2.3) && \text{pour les syllabes accentuées} \\
 dur_{FINALE} &= dur_{PASDELOUP} + 64.5(n - 2.1) && \text{pour les syllabes inaccentuées}
 \end{aligned}$$

finale d'un morphème dans un mot polymorphémique ("anticonstitutionnel") et qui correspond à notre AI dans la plupart des cas; 3. l'accent "monosyllabique", réalisé dans un mot monosyllabique.

Nous sommes conscients du fait que les coefficients d'ajustements des études de Kruckenberg & Fant (1995) ne tiennent pas compte du degré d'accentuation du groupe accentuel contrairement à Padeloup (1992). De plus, cette fusion suppose que les nombres moyens de phonèmes par syllabe du corpus utilisé par cette dernière soient équivalents à ceux de Kruckenberg & Fant. Néanmoins leurs corpus respectifs sont de taille suffisante pour tendre vers les valeurs standards du français.

3. Finalement nous appliquons un coefficient d'allongement aux syllabes portant un accent final en fonction du ton et du degré de frontière qui le suit. Cette technique issue de Beaugendre (1994) a été simplifiée et les coefficients corrigés.

ton	allongement
B-B-	1.5
HB-	1.6
H/H	1.4
HB	1.4
HH	1.1
/LL	1.1
autres	1

Tableau 7. Allongement syllabique selon le type de frontière

6.2. Prédiction de la durée phonémique

Nous adoptons pour cela la technique de modèle de répartition Campbell W. (1992) en prenant les données de Barbosa P. (1994) sur les durées intrinsèques du français et de leurs écarts-types.

Ainsi pour le calcul des durées segmentales le **modèle de répartition** part de deux hypothèses : 1. chaque phonème i possède une durée intrinsèque moyenne (μ_i) mesurable dans des corpus et dont l'écart-type (σ_i) peut être calculé, 2. la syllabe confère à chacun de ses phonèmes un même score d'étirement appelé z-score. Ainsi la somme des durées des phonèmes étirés (ou compressés) selon ce z-score donne la durée de la syllabe.

En pratique, à partir de la durée de la syllabe et de la durée intrinsèque de chacun de ses phonèmes, on calcule d'abord le z-score optimal, c'est-à-dire un degré d'étirement à appliquer à chaque phonème, en résolvant l'équation suivante :

$$dur_{\text{syllabe}} = \sum_{i=1}^n \exp(\mu_i + z\sigma_i)$$

où n représente le nombre de phonèmes dans la syllabe et (μ_i, σ_i) respectivement le logarithme de la moyenne et de l'écart type du phonème i . L'utilisation de la fonction exponentielle est justifiée par le fait que la distribution du logarithme de la durée des phonèmes est sensiblement normale et peut donc se modéliser par ces deux paramètres.

Puis nous 'répartissons' la durée syllabique selon ce z-score optimal pour calculer la durée de chaque phonème selon :

$$dur_{phoneme_i} = \exp(\mu_i + z_{syllabe} \sigma_i)$$

En conclusion, la durée de chaque phonème est donc calculée par l'intermédiaire de la durée syllabique, déterminée dans un premier temps à partir des propriétés de la syllabe, que l'on distribue dans tous les phonèmes la composant au moyen d'un coefficient d'étirement global à toute la syllabe, le z-score.

7. LA PHONÉTISATION

7.1. Méthodes utilisées

L'étape de phonétisation ou conversion graphème-phonème consiste à créer la chaîne phonétique à partir du texte orthographique. Elle fait appel à deux méthodes et à deux sources d'informations:

1. le lexique utilisé par l'analyseur syntaxique *Fips*, qui spécifie aussi la forme phonétisée de chaque forme lexicale ainsi que la consonne latente de liaison propre à chaque mot (par exemple, "nous": [nu] [z])
2. un système expert qui contient un jeu de règles de conversion graphème-phonème propre au français. Ces règles possèdent plusieurs champs qui contraignent leur utilisation lors de la conversion d'un mot. Non seulement la séquence graphémique à convertir doit concorder, mais également les contextes graphémiques gauche et droit, ainsi que le contexte syntaxique. Ces champs de contexte sont optionnels. Par exemple, dans la phrase *Les amis du président président*, la séquence "ent" en fin de nom se prononce [ã] (i.e. le contexte droit est une fin de mot, il n'y a pas de contexte gauche et le contexte syntaxique impose la catégorie grammaticale *nom*). Ainsi le nom commun *président* se prononce bien [prezidã] et non [prezid] comme l'est le verbe *présider* conjugué à la 3^{ème} personne du pluriel de l'indicatif présent.

La méthode lexicale, plus directe, est la méthode par défaut, mais elle doit parfois laisser la place au système expert pour la phonétisation de mots inconnus (c'est-à-dire les mots hors-lexique). Malgré ces deux moyens complémentaires de déterminer la séquence de phonèmes, on est confronté à certaines difficultés que l'on résout par des traitements locaux et particuliers mais qui puisent toujours leur information dans la richesse de l'analyse syntaxique.

Les problèmes les plus couramment rencontrés lors de la phonétisation sont les **homographes hétérophones** ("Les poules du couvent couvent", dans laquelle le mot "couvent" est prononcé de deux manières différentes) et les **phénomènes phonologiques** tels que l'élision du schwa et la liaison (Gaudinat A. *et al.* 1998). Finalement, les **mots inconnus** peuvent se manifester sous plusieurs formes: noms propres, néologismes, mots inusités, adresses web ou e-mail de plus en plus rencontrés dans les textes, sigles, abréviations et toute sorte de nombres et expressions numériques...

7.2. Les homographes et les phénomènes phonologiques

Pour tenter de résoudre ces problèmes, l'approche classique recourt à des heuristiques de surface, entraînant un traitement local, voire linéaire, des données (pour un état de l'art du domaine, voir Boula de Mareuil Ph. 1997). Notre approche diffère de ces systèmes classiques par le fait qu'elle utilise l'analyse syntaxique complète déterminée en amont par *Fips*. Cette analyse fournit au module de phonétisation les informations syntaxiques et grammaticales indispensables, donc une description complète et très détaillée de l'organisation de toute la phrase. Dans le lexique, on accède ainsi à la forme phonétique adéquate, sinon on fournit au système expert les informations contextuelles graphémiques et syntaxique pour une bonne phonétisation.

La liaison est un bon exemple de phénomène phonologique pour lequel l'utilisation d'une analyse syntaxique se justifie. De même, le traitement des homographes hétérophones devient alors aisé si la catégorie grammaticale du mot phonétiquement ambigu est identifiée correctement comme *président*. Il faut noter que d'autres ambiguïtés ne pourront être levées sans approche sémantique ("les fils de couture" et "les fils de mon oncle"). De plus, il existe un certain nombre de mots courants pour lesquels la réalisation phonétique peut varier en fonction du contexte sémantique et syntaxique. C'est le cas de *plus* (*j'en veux plus / je n'en veux plus*), *tous* (*ils sont tous partis / tous les enfants*) ainsi que de certains nombres comme *cinq*, *six*, *huit*, *neuf*, *dix*, *vingt* (*Il y en a cinq. Cinq enfants. Cinq gamins. Cinq cents bambins. Neuf années, neuf ans*). On peut noter que, dans les différents exemples, la consonne finale se fait entendre ou non ([ply] / [plys]). De plus, la liaison peut provoquer une troisième variante phonétique du même mot (*Je ne veux plus avoir d'ennuis.*) dans laquelle la consonne finale est réalisée /z/ : [plyz].

Enfin, il faut signaler que le module de phonétisation traite aussi les phénomènes résultant de la liaison tels que la **dénasalisation** (*un bon ami* se prononce [œbɔnami] au lieu de [œbɔ̃nami], de même que pour *divin enfant* [divinãfã]) ou l'**aperture du /e/ final** dans des adjectifs pré-nominaux (*léger accroc* se prononce [lezɛrakro] et non [lezɛrakrɔ]).

Cette vision globale de la phrase ou de l'extrait à prononcer permet surtout d'éviter des erreurs de homographie, de liaison (1), de prononciation de la consonne finale comme dans *tous*(2) ou *plus*, mais aussi les erreurs dues aux nombres (3).

1. L'homme qui est avec **vous** | **a donné son avis** / L'homme **vous**, **a donné son avis**.
2. Nous regardons **tous** les jeunes journalistes récemment engagées.
3. Les **31** filles. (prononcer "les trente-et-une filles")

7.3. Les mots inconnus: noms propres, sigles, nombres

Les mots inconnus ou hors lexique sont, par définition, traités par le système expert. Deux pré-traitements filtrent le texte pour repérer ceux qui pourraient être considérés comme des noms propres ou comme des sigles.

Pour les *noms propres*, on tente de déterminer son origine linguistique, c'est-à-dire de lui attribuer une langue ou un groupe de langues parmi les huit que l'on considère. Pour ce faire, on combine trois approches: par règles, par détection de préfixes ou d'affixes et par un calcul statistique de combinaisons de graphèmes (ou n-gram). Le cas échéant, on dote le système expert d'un jeu de règles supplémentaires propres à cette langue, provoquant une prononciation "sonnant" dans la langue détectée. Par exemple, le préfixe "Pfeiffen-" dans "Pfeiffenberger" confère au nom propre une connotation germanophone et force la prononciation du suffixe *-ger* comme [bɛrgœʀ] et non [bɛʀʒe].

En ce qui concerne les *sigles*, une méthode heuristique détermine par règles leur mode de prononciation : lu ("ONU", "OTAN") ou épilé ("TSR", "TSOL"). Il existe en outre une liste d'abréviations qui comporte leur signification complète.

Les *nombres* et *expressions numériques* (heure, date, monnaie) sont traités et une lecture étendue est générée de manière interne (*123456789, cent vingt trois millions quatre cent cinquante six mille sept cent quatre vingt neuf; 12h30, douze heure trente ; \$12.50, douze dollars cinquante cents*). Cette représentation permet une bonne prononciation de l'exemple (3) ci-dessus. Finalement, les prononciations spécifiques comme les régionalismes sont respectées (*90, nonante*).

8 CONVERSION DES ARBORESCENCES *FIPS*

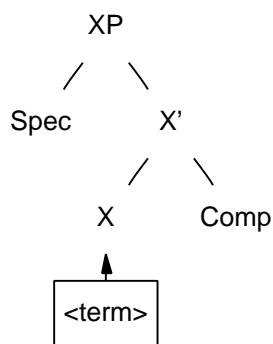
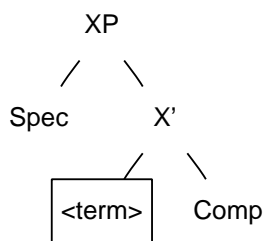
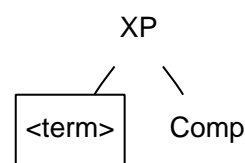
Afin d'utiliser les arborescences générées par *Fips* comme entrée de *Mingus*, plusieurs opérations ont été effectuées.

8.1. Identification des têtes de constituant

Comme la possibilité de regroupement intonatif dépend de la présence d'un lien de dépendance entre les parties à regrouper, on cherche à déterminer, pour chaque constituant, son statut de tête ou de dépendant.

Dans une grammaire de type X-bar, cette information n'est donnée qu'indirectement, à travers la configuration de l'arbre (si on maintient systématiquement le schéma de base à 5 noeuds: XP, Spec, X', X, Comp; la figure ci-dessous) ou à travers des conventions de dénomination des (symboles auxiliaires) non terminaux: XP désigne la projection maximale (le constituant complet), et X le préterminal tête. Le format *Fips* comporte en outre des formes simplifiées pour les XP sans spécifieur ou sans complément.

schéma de base X-bar

format simplifié *Fips*format simplifié *Fips* d'un XP sans Spec**Figure 16.** Configurations d'arbres dans *Fips*.

À partir de l'étiquette du noeud mère et de celles des fils immédiats (correspondant aux parties gauche et droit de la règle de réécriture correspondante, respectivement), on détermine le fils tête, à l'aide de règles du genre:

<i>noeud mère</i>	<i>liste des noeuds fils</i>	<i>noeud tête dans la partie droite</i>
NP	celui étiqueté N
DP	celui étiqueté N
VP	celui étiqueté V
TP	celui étiqueté T'
AP	celui étiqueté A'
VP	AdvP V'	V'
DP	D X , où $X \in \{ NP, AP, DP, PP \}$	X
PP	P X	X
PP	D DP	DP
CP	Conj TP	TP
N	N PP	N
T'	Aux VP	VP

Tableau 8. Identification des têtes syntaxiques dans les syntagmes de *Fips*.

8.2. Détermination du statut clitique des terminaux

L'accent de mot (soit le trait [\pm clitique]) est déterminé à partir de la catégorie grammaticale des terminaux, cf. § 2.4.

8.3. Réajustement de l'arborescence: les verbes auxiliaires

L'illustration suivante illustre la structure après le passage de *Fips* à une CFG avec indication de la dépendance. Pour les formes composées du verbe, la grammaire de *Fips* considère l'auxiliaire comme le X' (soit la tête) et le verbe principal comme le Comp (soit le dépendant).

```

?TP
  -DP
    -D un : û
    +N loup : lu
  +T'
    -clit s' : s
    +V est : è
    -VP
      +V jeté : j%té
      -A immédiatement : imédiatmâ
    -PP
      -P sur : syr
      +DP
        -D la : la
        +NP
          -A petite : p%tit
          +N chèvre : /èvr

```

Figure 17. Représentation textuelle de l'arbre syntaxique à la sortie de *Fips* après l'insertion des signes + et – indiquant les têtes et les dépendants. Les noeuds terminaux mentionnent la forme orthographique et sa transcription phonétique selon la convention de *Fips*: û = [œ], è = [ɛ], % = [ə], é = [e], r = [R], / = [ʃ].

Cependant, du point de vue de l'intonation, l'auxiliaire se comporte comme un dépendant du verbe principal. Pour obtenir l'effet souhaité, il est nécessaire (1) de distinguer l'auxiliaire des autres verbes, (2) de lui attribuer le statut de dépendant (cf. la dernière règle dans le tableau 8), (3) et enfin de modifier l'arbre de sorte que l'auxiliaire et le verbe principal (tête locale), ainsi que leurs clitiques, soient réunis dans un même constituant. Ce dernier sera à son tour la tête locale à laquelle les autres compléments du verbe principal sont juxtaposés. Après ces modifications, l'arbre prend la forme suivante. (Notons que l'interprétation des étiquettes T' et VP s'écartera de leur sémantique dans *Fips*. C'est sans importance pour *Mingus*, qui ne se sert que des étiquettes des terminaux.)

```

?TP
  -DP
    -D un : û
    +N loup : lu
  +VP
    +T'
      -clit s' : s
      -Aux est : è
      +V jeté : j%té
    -adv immédiatement : imédiatmâ
  -PP
    -P sur : syr
    +DP
      -D la : la
      +NP
        -A petite : p%tit
        +N chèvre : /èvr

```

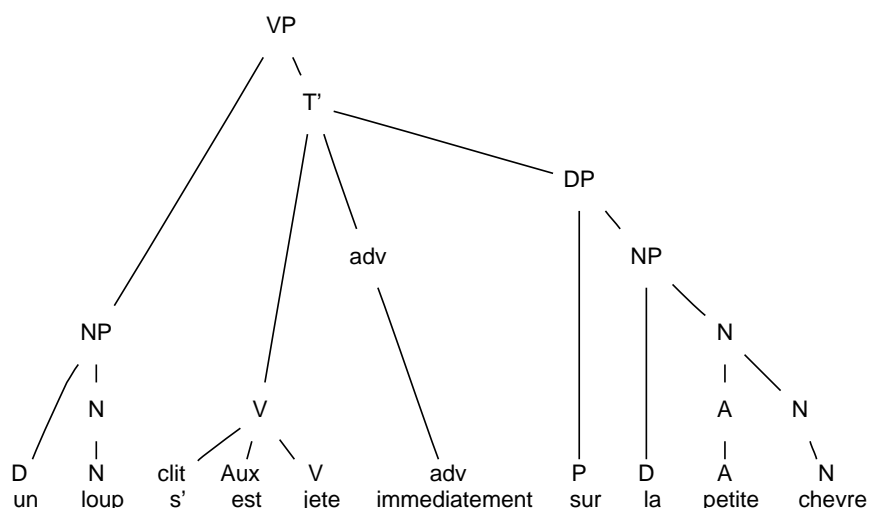


Figure 18. Représentations textuelle et graphique de l'arbre syntaxique après le réajustement pour les auxiliaires appliqué à l'arborescence de la figure précédente. Pour la notation, voir la figure précédente.

À la suite de ces modifications, les symboles des noeuds peuvent perdre leur interprétation logique dans *Fips*, ce qui prête à confusion. Pour tout noeud non terminal, on remplace donc l'étiquette par celle du noeud tête local, ce qui au moins fournit des données cohérentes.

8.5. Réorganisation de l'arborescence: les mots clitiques

L'arborescence comporte toujours un nombre de branchements qui pourront être simplifiés (aplatis). Tout élément dépendant qui précède l'élément dont il dépend est déplacé sous la tête en question.

```

+T'
  -N
    -D un : û
    +N loup : lu
  +V
    -clit s' : s
    -Aux est : è
    +V jeté : j%té
  -adv immédiatement : imédiatmâ
  +N
    -P sur : syr
    -D la : la
    -A petite : p%tit
    +N chèvre : /èvr
  
```

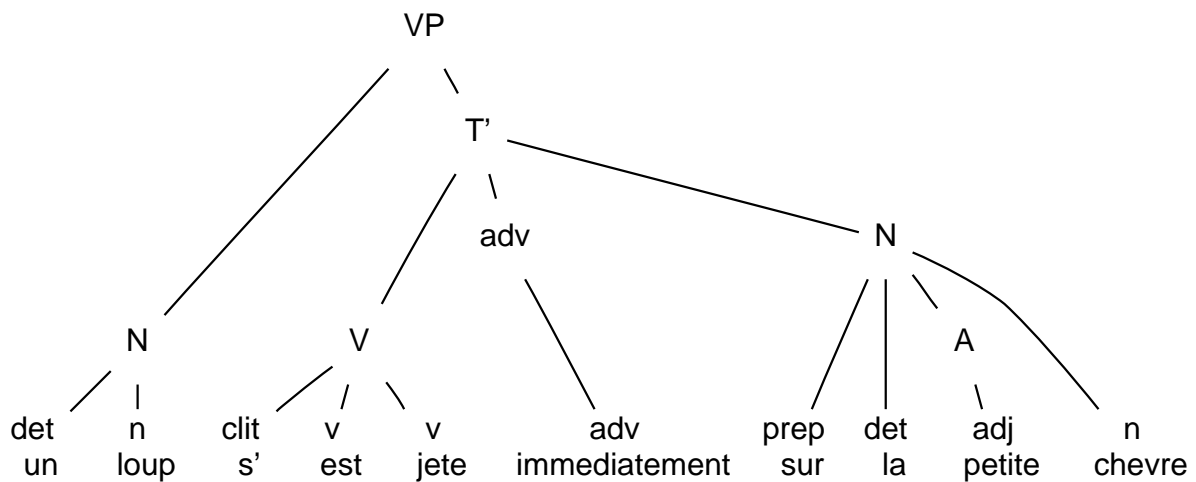


Figure 19. Arborescence après le réajustement pour les clitiques appliqué à l'arborescence de la figure précédente. Pour les conventions de notation, voir la figure précédente.

On arrive ainsi au type d'arbre qui peut être interprété correctement par *Mingus*.

9. RÉSULTATS ET DISCUSSION

Le système complet a été testé sur deux corpus de nature différente. Le premier corpus est tiré de la base de données BDSONS; il comporte des phrases assez courtes (10 à 15 syllabes) syntaxiquement simples. Le deuxième est un extrait du récit "Babar en ballon", qui comporte des phrases de longueur et de complexité variées.

La plupart des sujets participant à ces tests informels ont du mal à séparer, dans leur jugement, les différents aspects de l'intonation: la forme du contour, la durée des syllabes, le rythme ou les pauses. Qui plus est, le moindre écart d'ordre segmental, comme la qualité de la voix synthétique, la phonétisation ou le traitement du schwa, les frappe et influence leur jugement sur l'intonation. Par exemple, certains ont remarqué que le traitement des schwas, ajusté dans *Fips* à un style plutôt familier, ne colle pas avec l'intonation neutre que les auditeurs associent à un style plus soutenu, ou encore à la lecture oralisée. Pour cette raison, une évaluation objective et expérimentale se révèle assez délicate.

L'écoute informelle de l'intonation générée conduit aux observations suivantes.

1. D'une manière générale, l'intonation peut être caractérisée comme étant celle d'un texte lu de façon *neutre*, c'est-à-dire *sans expressivité*. Cela s'explique par le fait que l'algorithme se limite à attribuer les tons de fin de groupe intonatif, qui reproduisent la structuration syntactico-prosodique. Les syllabes autres que la syllabe finale sont toutes considérées atones et reçoivent le ton bas (c'est-à-dire non marqué). Or, ce sont précisément le ton de la syllabe pénultième et l'accent initial qui produisent des effets expressifs. La conception de l'algorithme permettrait facilement d'attribuer des tons à ces localisations. Seulement, leur insertion

effective supposerait la présence de marqueurs (morphèmes) intonatifs supplémentaires (de délimitation pour l'AI, de consensualité pour la pénultième). Les indices sémantiques, pragmatiques ou phonostylistiques qui permettraient de générer ces marqueurs manquent dans la synthèse à partir du texte. Il est en de même d'autres informations telle que la topicalisation, le focus, la relation sémantique avec des éléments du contexte gauche,... Par exemple: l'énoncé à prononcer constitue-t-il une réponse à quelle question? Où est le rhème?

Pour augmenter le naturel dans la synthèse à partir du texte, on pourrait partir d'observations probabilistes. Par exemple, pour les adjectifs antéposés ("le petit livre", "un immense fardeau",...) et les adverbes de degré modifiant des adjectif ou des adverbes ("très important", "assez difficile", "assez lourd", "peu lentement",...), l'accent initial est très fréquent. Alternativement ces marques "expressives" pourraient être introduites de façon aléatoire, pour introduire de la variation.

2. Certains problèmes observés sont liés à des constituants ou constructions syntaxiques qui introduisent des frontières intonatives fixes. Il s'agit des adverbes de phrase en début de phrase, des éléments disloqués, etc. Parfois, ces éléments ne peuvent pas être identifiés à partir de l'arborescence, parfois cette information n'est pas encore exploitée dans *Mingus*.

3. D'autres erreurs trouvent leur origine dans un déséquilibre entre les représentations syntaxiques adoptées par *Fips* et *Mingus*. Prenons le cas des prédicateurs adjectivaux (un adjectif précédé d'un verbe support: "être originaire de") ou des prédicateurs complexes ("faire preuve de, faire face à, perdre pied, faire le point (de)"). Dans "je suis resté sourd à ses cris", l'analyse syntaxique de *Fips* donne l'emboîtement (a) où le verbe régit l'adjectif, alors que du point de vue de l'organisation intonative, on préférerait (b), avec l'adjectif prédicateur qui fait unité avec le verbe support (éventuellement composé). Les parenthèses indiquent les constituants et les flèches les liens de dépendance.

(a) (suis) -> ((resté) -> (sourd) (à ses cris))

(b) (((suis) <- resté) + sourd) -> (à ses cris)

Une situation analogue s'observe pour le verbe auxiliaire ou ses formes composées: (a) est dérivé de *Fips* alors que (b) serait plus adapté du point de vue intonatif.

(a) (avait -> (été -> (condamné)))

(b) ((avait <- été) <- condamné)

4. Dans les phrases testées analysées à l'aide de *Fips* il ne reste pas de trace de la ponctuation du texte original, alors que la génération de la prosodie pourrait s'en servir pour déterminer la modalité énonciative ou interrogative, la présence d'un focus, ou la présence de pauses (cf. Vannier G. *et al.* 1999). La prise en compte des signes de ponctuation n'entraînera qu'une amélioration du résultat, les aspects syntaxiques ayant été réglés dans la plupart de cas.

10. CONCLUSION

L'objectif initial était de générer l'intonation dans la parole de synthèse, et cela en partant de la structure syntaxique. Cet objectif a nécessité la réalisation d'un système complet de synthèse à partir du texte. Vu la complexité d'une telle application, nous avons préféré intégrer deux systèmes existants: un analyseur syntaxique (avec tous les composants inférieurs) et un système de génération de l'intonation. D'autres modules viennent compléter le système: la phonétisation, le modèle de durée.

La combinaison de ces parties entraîne une couche supplémentaire d'interface, pour compenser certaines différences d'approche syntaxique. Cette conversion d'une représentation syntagmatique à une représentation dépendancielle comporte l'identification des têtes de constituant, la modification de certains rapports de dépendance implicites, le traitement particulier des constituants clitiques, etc.

La combinaison de systèmes existants reliés par un module d'interface a permis de réaliser un système de synthèse dans lequel l'intonation est gérée de façon explicite tant au niveau de son interaction avec la syntaxe, qu'au niveau du contrôle de propriétés prosodiques globales, indépendantes de la structuration linguistique: le registre du locuteur, la dynamique des mouvements mélodiques, la pente de déclinaison, le débit, la structure temporelle des syllabes, et ainsi de suite. Les résultats obtenus permettent de vérifier les règles à propos de l'interaction entre intonation et syntaxe, et d'identifier des écarts, permettant ainsi d'affiner ces règles. L'application de la synthèse à partir du texte constitue ainsi un outil extrêmement utile pour l'étude des rapports entre intonation et syntaxe.

Le recours à une analyse syntaxique détaillée fournit une solution à la fois élégante, systématique et unifiée à la plupart des problèmes liés au rôle de la syntaxe dans le calcul de la forme sonore. Pensons à la phonétisation, où la syntaxe est indispensable pour régler les phénomènes de liaison, d'élision, du schwa, des homographes hétérophones, et ainsi de suite. Pensons également à la génération de la prosodie proprement dite: à la formation des unités accentuelles qui suppose l'information sur la catégorie grammaticale des éléments. Une simple lemmatisation ne suffirait pas ici, ni la désambiguïsation locale; certains cas demandent une analyse syntaxique globale. Pensons ensuite à tous les aspects où interviennent les liens de dépendance: la formation des groupes accentuels et des groupes intonatifs, le regroupement intonatif, l'attribution des degrés de frontière, ... Et nous n'avons certainement pas épuisé tous le cas.

Ce travail appliqué nous offre ainsi un panorama de phénomènes syntaxiques liés à l'intonation. Il permet ainsi de dresser un bilan (provisoire) des propriétés syntaxiques que l'analyseur idéal devrait identifier, en dehors la constituance et la dépendance, bien entendu. Il s'agit de certaines constructions telles que la clivée (ou autres procédés d'extraction), la dislocation, les adjoints de phrase, les incises, la focalisation, la topicalisation,... Il s'agit aussi du statut de certains éléments comme les adverbes de degré, les prédicateurs (verbaux, adjectivaux, nominaux) complexes, les locutions de tous genres, les mots composés, et ainsi de suite. On peut espérer que l'étude de l'intonation permette d'éclaircir certaines questions en syntaxe ou d'orienter celle-ci vers de nouvelles questions.

À ce jour il n'existe point de théorie syntaxique, ni *a fortiori* d'analyseur syntaxique, qui réponde à tous les besoins de la génération de la prosodie. Ce serait en effet difficile puisque ces besoins ne s'identifient que progressivement. Le système *Fips* peut se prévaloir d'atouts tels que l'analyse détaillée, la large couverture syntaxique et lexicale, et la phonétisation robuste et sophistiquée.

RÉFÉRENCES

- ABEILLÉ, Anne (1993) : *Les Nouvelles Syntaxes. Grammaires d'unification et analyse du français*. Paris, Armand Colin, Collection Linguistique.
- AUBERGÉ, Véronique (1991) : La synthèse de la parole, des règles aux lexiques. Thèse de doctorat, Université P. Mendès-France, Grenoble, pp. 210.
- BAILLY, Gérard (1989) : "Integration of rhythmic and syntactic constraints in a model of generation of French prosody", *Speech Communication*, 8, 137-146.
- BARBOSA, Plinio (1994) : Caractérisation et génération automatique de la structuration rythmique du français. Thèse de doctorat. Institut National Polytechnique de Grenoble, Grenoble, France.
- BARBOSA, Plinio & BAILLY, Gérard (1994) : "Characterisation of rhythmic patterns for text-to-speech synthesis", *Speech Communication*, 15, 127-139.
- BEAUGENDRE, Frédéric (1994) : Une étude perceptive de l'intonation du français: Développement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte. Thèse de doctorat en Sciences de l'Université de Paris XI.
- CAMPBELL, W.N. (1992) "Syllable-based segmental durations" in BAILLY, G. & BENOÎT, C. (ed) (1992), *Talking Machines: Theories, Models and Designs*. Elsevier Science Publ. North Holland. pp. 211-224.
- CHOMSKY, N. & LASNIK, H.: "The Theory of Principles and Parameters", in CHOMSKY, N. (1995) *The Minimalist Program*. Cambridge, MIT Press, pp. 13-127.
- DELAIS, Elisabeth (1995) : Pour une approche parallèle de la structure prosodique. Étude de l'organisation prosodique et rythmique de la phrase française. Thèse de doctorat, Univ. Toulouse - Le Mirail
- DELL, François (1984) : "L'accentuation dans les phrases en français", in DELL, F.; HIRST, D.J. & VERGNAUD, J.R. (eds.) (1984) *La forme sonore du langage. Structure des représentations en phonologie*. Paris: Hermann. pp. 65-122.
- DI CRISTO, Albert; DI CRISTO, Philippe & VÉRONIS, Jean (1998) : "Optimisation d'un modèle prosodique pour la synthèse par règles à partir du texte en français". *Actes XXII Journées d'Etude sur la Parole*, 135-138 (Martigny 15-19 juin 1998).
- DUTOIT, Thierry (1997) : *An Introduction to Text-to-Speech Synthesis*, Dordrecht: Kluwer Academic Publ. 312 pp.
- EYNDE, Karel van den & DOOREN, K. van (1983) : "Intonation and syntactic structure in Dutch", *ITL-Review of Applied Linguistics*, vol. 60/61, 27-42.
- EYNDE, Karel van den; MERTENS, Piet & SWIGGERS, Pierre (1998) : "Structuration segmentale et suprasegmentale en syntaxe: Vers un modèle intégrationniste de l'écrit et de l'oral." in Bilger, M.; van den Eynde, K. & Gadet, F. *Analyse linguistique et approches de l'oral*. Leuven-Paris: Peeters, pp. 33-57.
- GARDE, Paul (1968) : *L'accent*. Paris: Presses Universitaires de France, 171 pp.
- GAUDINAT, Arnaud; GOLDMAN, Jean-Philippe & WEHRLI, Eric (1998) : "Le système de synthèse FipsVox: syntaxe, phonétisation et prosodie". *Actes XXII Journées d'Etude sur la Parole*, 139-142 (Martigny 15-19 juin 1998).

- GAUDINAT, Arnaud & WEHRLI, Eric (1997) : "Analyse syntaxique et synthèse de la parole: le système FipsVox" *Traitement Automatique des Langues*, n° 38, 121-130.
- GLEASON, H.A. (1955, 1961, 1967) : *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston, 503 pp. (Revised edition: 1961)
- GOLDMAN, Jean-Philippe (1997) : *Génération automatique de l'intonation en français*. Mémoire de DES. Université de Genève. Novembre 1997.
- KRUCKENBERG, A. & FANT, G. (1995) : "Notes on syllable duration in French and Swedish.", *Proceedings International Congress of Phonetic Sciences 1995* (Stockholm), 158-161.
- LACHERET-DUJOUR, Anne & BEAUGENDRE, Frédéric (1999) : *La prosodie du français* Paris, Editions du CNRS, pp. 354.
- LAENZLINGER, C. & WEHRLI, E. (1991) FipsIPS : "Un analyseur interactif pour le français", *TA Informations*, 32:2, pp. 35-49.
- MALFRÈRE, Fabrice; DUTOIT, Thierry & MERTENS, Piet (1998) : "Un générateur de prosodie "tout automatique"". *Actes XXII Journées d'Etude sur la Parole*, 147-150 (Martigny 15-19 juin 1998)
- MALFRÈRE, Fabrice; DUTOIT, Thierry & MERTENS, Piet (1998) : "Fully Automatic Prosody Generator for Text-to-Speech Synthesis", *Proc. International Conference on Spoken Language Processing*. Sidney, Australia, pp. 1395-1398.
- MAREÛIL, Ph. Boula de (1997) : *Étude linguistique appliquée à la synthèse de la parole à partir du texte*. Thèse de doctorat, Université de Paris X, Orsay, France
- MARTIN, Philippe (1981) : "Pour une théorie de l'intonation. L'intonation est-elle une structure congruente à la syntaxe ?" in Rossi M. *et al.* (1981), 234-271.
- MARTIN, Philippe (1998) : "Association prosodie-syntaxe: validation par synthèse". *Actes XXII Journées d'Etude sur la Parole*, 119-122 (Martigny, 15-19 Juin 1998).
- MATTHEWS, Peter Hugoe (1981, 1992) : *Syntax*. Cambridge: Cambridge University Press..
- MERTENS, Piet (1987) : *L'intonation du français. De la description linguistique à la reconnaissance automatique*. Thèse de doctorat. Université de Leuven.
- MERTENS, Piet. (1990) : "Chap. IV. L'intonation", in BLANCHE-BENVENISTE, C.; BILGER, M.; ROUGET, C. & EYNDE, K. VAN DEN (1990), *Le français parlé. Etudes grammaticales*. Paris: Editions du CNRS. Collection "Sciences du Langage", pp. 159-176
- MERTENS, Piet (1991) : "Local prominence of acoustic and psychoacoustic functions and perceived stress in French." *Actes du Congrès International de Phonétique*, vol. 3, 218-221 (Aix-en-Provence)
- MERTENS, Piet (1992) : "L'accentuation de syllabes contiguës" *ITL-Review of Applied Linguistics*, 95/96, 145-165
- MERTENS, Piet (1993) : "Accentuation, intonation et morphosyntaxe", *Travaux de Linguistique*, 26, 21-69.
- MERTENS, Piet (1997) : "De la chaîne linéaire à la séquence de tons", *Traitement Automatique des Langues*, . vol. 38, pp. 27-51.
- MERTENS, Piet (1999) : "Un algorithme pour la génération de l'intonation dans la parole de synthèse", *Actes Conférence Traitement Automatique du Langage Naturel 1999 (Cargèse, 12-17 juillet 1999)*, pp. 233-242.
- MORLEC, Y.; AUBERGÉ, V.; BAILLY, G. (1995) : "Evaluation of automatic generation of prosody with a superposition model." *Proc. International Congress Phonetic Sciences*. XIII, vol. 4, pp. 224-227 (Stockholm)
- PASDELOUP, Valérie (1990) : *Modèles de règles rythmiques du français appliqué à la synthèse de la parole*. Thèse de doctorat, Université de Provence.
- PASDELOUP, Valérie (1992) : "Durée intersyllabique dans le groupe accentuel en français", *Actes 19e Journées d'Etude sur la Parole*, Bruxelles (mai 1992), pp 531-534.
- PASDELOUP, Valérie (1992). : "A prosodic model for French text-to-speech synthesis: a psycholinguistic approach" in Bailly, G. & Benoît, C. (ed) (1992), *Talking Machines: Theories, Models and Designs*. Elsevier Science Publ. North Holland. pp. 335-348

- POLLARD, C.J. & SAG, I.A. (1987) : *An Information-based Approach to Syntax and Semantics. Vol 1. Fundamentals*. CSLI Lecture Notes, 13. Chicago University Press.
- ROSSI, Mario (1987) : "Peut-on prédire l'organisation prosodique du langage spontané?" *Etudes de Linguistique Appliquée*, 66, 20-48.
- ROSSI, Mario (1993) : "A model for predicting the prosody of spontaneous speech (PPSS model)", *Speech Communication* Vol. 13, pp. 87-107
- ROSSI, Mario (1999) : *L'intonation, le système du français: description et modélisation*. Gap - Paris : Ophrys. Coll. L'essentiel français
- ROSSI, M.; DI CRISTO, A.; HIRST, D.; MARTIN, Ph. & NISHINUMA, Y. (1981) : *L'intonation. De l'acoustique à la sémantique*. Paris: Klincksieck, 364 pp.
- TESNIÈRE, Lucien (1959, 1969, 1988) : *Éléments de syntaxe structurale*. Paris: Klincksieck, pp. 674.
- VANNIER, Gérald (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, Juillet 1999.
- VANNIER, Gérald; LACHERET-DUJOUR, Anne; VERGNE, Jacques (1999) : "Pause location and duration calculated with syntactic dependencies and textual considerations for TTS system." *Proceedings International Congress of Phonetic Sciences*,
- VÉRONIS, J.; DI CRISTO, Ph.; COURTOIS, F.; CHAUMETTE, C. (1998) : "A stochastic model of intonation for text-to-speech synthesis", *Speech Communication*, 26(4), 233-244
- WEHRLI, Eric (1997) : *L'analyse syntaxique des langues naturelles : problèmes et méthodes*, Paris, Masson.