

Un algorithme pour la génération de l'intonation dans la parole de synthèse

Piet MERTENS

Centrum voor Computerlinguïstiek, K.U.Leuven, BP 33, 3000 Leuven (Belgique)

Piet.Mertens@arts.kuleuven.ac.be

<http://bach.arts.kuleuven.ac.be/pmertens>

Résumé

L'article décrit l'implémentation d'un modèle d'intonation dans son application à la synthèse de la parole pour le français. Le modèle se caractérise par l'importance accordée à la syntaxe et par une approche analytique de l'intonation qui, en synthèse, permet une manipulation explicite et compositionnelle du sens intonatif. Le traitement proprement dit est précédé d'une analyse syntaxique identifiant les constituants, certains rapports de dépendance ou certaines constructions qui demandent une intonation particulière. Ces aspects intonatifs sont représentés par des marqueurs symboliques. À partir de l'arborescence sont constitués les groupes intonatifs, tout en tenant compte du rythme. Dans certaines conditions, des réajustements de la structure syntaxique seront effectués. Les tons mélodiques sont attribués aux groupes en fonction des marqueurs et des rapports syntaxiques.

1. Introduction

L'algorithme présenté ici vise le calcul de l'intonation d'une phrase ou d'un texte dans la parole de synthèse. La spécification de l'intonation en sortie consiste en une séquence de marques intonatives associées aux syllabes ou aux groupes de syllabes. Ce traitement symbolique constitue la première étape de la génération de la prosodie dans la synthèse à partir du texte, étape qui sera nécessairement suivie d'un traitement phonético-acoustique afin d'en déduire les valeurs concrètes de F_0 et de durée des phonèmes et des pauses. (Ce module acoustique ne sera cependant pas traité ici.)

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 (1991), Delais (1995), Boula de Mareüil (1997), Di Cristo *et al.* (1998), Gaudinat & Goldman (1998), Malfrère *et al.* (1998), Morlec *et al.* (1995), Véronis *et al.* (1998); dans la plupart des cas, il s'agit de méthodes automatiques.

L'approche présentée ici entre dans la classe des systèmes par règles. Elle se distingue des autres travaux par la priorité accordée à la syntaxe. Par là on entend non seulement la *structure de constituants*, mais aussi la *construction*. En effet, certaines constructions exigent des marques intonatives particulières, comme le montrent les exemples suivants.

- [1] Ils le chargent, le camion.
- [2] Le camion, ils le chargent.
- [3] C'est lui, qui l'avait trouvé.
- [4] C'est précisément là, que des problèmes d'emploi surgissent.
- [5] C'est l'homme (,) qui a pollué la mer.
- [6] Tu ne vas pas nous faire ça, franchement!

- [7] Tu vas travailler maintenant
 [8] L'herbe de la pelouse pousse lentement, décidément.
 [9] C'était un gars tellement aimable.
 [10] Il vend ça aux imbéciles
 [11] Nous envoyons ces documents aux clients.

En [1], le groupe nominal (GN) objet reçoit nécessairement l'*appendice* (c'est une séquence de syllabes atones au même niveau de hauteur, sans accent), la frontière intonative majeure (marquée par une chute au niveau infra-bas) étant située avant celui-ci. En revanche, si ce GN est antéposé [2], on n'a pas d'appendice. Le dispositif d'extraction en "*c'est ... que...*", tantôt demande l'appendice [3], tantôt l'admet [4]. Le dispositif et la relative [5], identiques au plan morphologique, sont pourtant différentes sur le plan intonatif et (donc) aussi sur le plan syntaxique. [7] peut être prononcé de plusieurs façons, avec ou sans appendice pour l'adverbe, ce qui donne lieu à des interprétations différentes. Par contre, [8] exige l'appendice pour "décidément", mais ne l'admet pas pour "lentement". L'accent d'insistance est très probable sur "tellement" en [9]. Dans [10] on est obligé d'inclure l'objet dans le groupe intonatif du verbe, alors que ce n'est pas le cas dans [11].

Ces exemples illustrent le rôle crucial de la structure syntaxique et de la construction. (Cet aspect a été peu étudié; citons toutefois: Mertens (1993, 1997), Eynde *et al.* (1998), Lonchamp (1998) et Rossi (1993).) Les approches automatiques actuelles, qui reposent sur un étiquetage grammatical, ne sauraient rendre compte des faits cités, parce que les catégories grammaticales ne permettent pas d'identifier la construction dans la plupart des cas.

La section 2 fournit une description sommaire du modèle. Des exemples permettront d'illustrer les phénomènes à expliquer. La section 3 présente l'algorithme lui-même. La dernière partie, enfin, parle des résultats, des limitations et des modifications envisagées.

2. Présentation sommaire du modèle intonatif

Le modèle intonatif implémenté est celui proposé dans Mertens (1997), enrichi d'observations nouvelles. On se limitera ici à une présentation très sommaire. Pour les détails, nous renvoyons aux ouvrages cités.

2.1 Les formes intonatives et leur fonction

Sur le plan formel, l'intonation d'un énoncé se manifeste comme une succession de groupes de syllabes, appelés *groupes intonatifs* (GI). À chaque GI est associé un contour mélodique. Ce contour se décompose à son tour en une suite de cibles ou de mouvements mélodiques, appelés *tons*, associés aux syllabes. La tâche du modèle intonatif (et donc de l'algorithme) consiste à constituer les GI et à leur attribuer les tons appropriés.

L'emploi des tons dépend de plusieurs facteurs. Les tons expriment des significations énonciatives précises et/ou imposent à la chaîne linéaire une certaine organisation hiérarchique (cf. *infra*, § 2.3, "regroupement intonatif"). De cette façon l'intonation contribue à la signification de l'énoncé, et constitue un facteur de structuration syntaxique. Dans [12], les tons HL-, L-L-, etc. associés aux syllabes donnent lieu à des interprétations différentes.

- [12] c'est impossible c'est impossible c'est impossible c'est impossible
 _____ $\overline{\text{HL-}}$ _____ $\overline{\text{h}}$ L-L- _____ $\overline{\text{H}}$ L-L- _____ $\overline{\text{H/H}}$

Les marques intonatives ne sont pas insérées librement. Leur emploi doit respecter certaines contraintes morphologiques, syntaxiques et rythmiques, que nous essaierons d'esquisser ci-dessous.

2.2 Formation des groupes intonatifs

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

Un même énoncé peut être prononcé de plusieurs façons, avec 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 aussi le terme “mot prosodique”). Dans les réalisations observées, chaque GI englobe un ou plusieurs GA. (Cette représentation à deux niveaux se justifie par les observations suivantes: 1. dans un énoncé avec peu 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.)

Chaque GA consiste d'un élément (“mot”) *tonique* (accentuable, “accentogène”, à accent lexical) 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 dépend, sauf s'il en est séparé par un autre élément tonique, auquel cas il s'intègre au groupe qu'il précède. Pour plus de détails, voir Mertens (1993, 1997: 34-5). Cela est 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. Les syllabes recevront les tons.

- [14] (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:

- [15] (lui avait-elle) (envoyé) (cette lettre)
 AL . . ○ . . . ○ . ○
 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 (1995: 125). Ensuite, les accents contigus sont évités. Le nombre de GI dépend en outre du rapport syntaxique entre les GA: le regroupement doit respecter l'emboîtement syntaxique. Enfin, il dépend de “l'unité de sens” (mais en synthèse à partir du texte, on n'a généralement pas accès à cette information).

2.3 Les marques énonciatives et le choix des tons

Pour toute phrase à prononcer, les *marques énonciatives* (modalité de la phrase, focus, appendice, pénultième haute, incise,...) seront déterminées à partir de plusieurs sources d'information: la ponctuation, la structure syntaxique, la construction syntaxique, la sémantique, ou manuellement. Certaines marques, comme l'accent d'instance ou le ton haut à la pénultième, pourraient même être introduites de façon aléatoire pour imiter un style de parole particulier, comme le style journalistique. Les marques sont associées à des endroits précis du squelette en fonction de l'entité prosodique qu'elles modifient. On distingue en effet plusieurs classes: certaines marques (la modalité phrastique (énonciative ou interrogative), le

changement de registre, l'appendice) s'associent à une séquence de GI, d'autres (la focalisation) à un GI, d'autres encore (l'accent initial, la pénultième) à une syllabe.

Une fois que les GI ont été constitués, on enrichit le squelette prosodique des marques énonciatives souhaitées (cf. les exemples dans Mertens (1997)). Il restera un certain nombre de positions non attribuées, pour lesquelles on choisit les tons appropriés en fonction de la structure syntaxique et de l'équilibre rythmique (eurythmie).

Le travail descriptif sur corpus a révélé le phénomène du *regroupement intonatif* lié aux tons placés aux syllabes finales de GI: le rapport hiérarchique entre ces tons donne lieu à un regroupement des GI en *paquets* intonatifs, qui ne sont autres que les constituants construits du plan intonatif. On constate également que la structuration intonative résultante doit être conforme à l'organisation syntaxique. Ceci suppose qu'entre les parties regroupées intonativement, il existe un rapport de dépendance, cf. Mertens (1997: 45-47). Dans l'exemple suivant, on regroupe les GI 2 et 3 (a), mais pas les GI 3 et 4 (b), parce qu'il n'y a pas de rapport syntaxique direct entre eux.

- [15]a (les citoyens) ((inquiets) (pour l'avenir)) ne consomment pas assez
 b * ((les citoyens) (inquiets)) ((pour l'avenir) (ne consomment pas assez))
 c ? ((les citoyens) (inquiets) (pour l'avenir)) (ne consomment pas assez)

Pour en tenir compte, la représentation syntaxique doit préciser pour chaque constituant le statut de tête ou de dépendant à l'intérieur du constituant père.

Vu le nombre restreint de tons mélodiques — et dès lors de rapports hiérarchiques possibles — le regroupement intonatif est limité à un maximum théorique de quatre niveaux d'emboîtement, mais cette complexité ne se rencontre qu'exceptionnellement. Lorsque la structure syntaxique compte davantage de "niveaux", on observe le *réajustement* de l'arborescence. L'énoncé [16] en fournit un exemple: le GN objet comporte une relative, donnant un arbre à plusieurs ramifications successives. L'intonation attestée comporte une frontière majeure *avant* la relative. C'est comme si la relative était juxtaposée à la phrase principale. Cette analogie justifie qu'on parle ici d'une rupture de l'arbre syntaxique, donnant une structure intonative plus plate.

- [16] les policiers ont retrouvé (les traces || que les ravisseurs avaient laissées)

Le phénomène inverse se rencontre également: le regroupement intonatif observé pour [16] après la rupture, est donné en [17]. Les GI 2 et 3 sont regroupés dans une même unité (intonative!), ce qui correspond à une structure syntaxique où ils seraient frères, et où le premier GI serait l'oncle.

- [17] ((les policiers) ((ont retrouvé) (les traces))) || ((que les ravisseurs) (avaient laissées))

Ces faits montrent le rôle crucial de la syntaxe pour l'intonation, à la fois au niveau de la formation des GA, de la formation des GI, du choix des tons et enfin de la constitution de la structure prosodique (et des frontières correspondantes).

3. Description de l'algorithme

Les paragraphes précédents visaient à situer la tâche dans toute sa complexité réelle, ce qui justifie en même temps la complexité de l'algorithme qui a été élaboré.

Peu de systèmes de génération de la prosodie de synthèse effectuent une analyse syntaxique, ne fût-ce partielle. Là où c'est le cas, on n'en retient que l'emboîtement des constituants — ou les degrés de frontière correspondants. L'entrée peut alors être représentée par

Un algorithme pour la génération de l'intonation

une simple chaîne, faite de mots et de frontières. Vu l'importance de la syntaxe, une approche radicalement différente a été adoptée qui prend comme structure de données l'arbre syntaxique lui-même et qui 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 GA et GI), 2. tantôt il s'agit de modifier des noeuds relativement élevés pour aplatir la structure syntaxique ou pour la réorganiser localement, 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 les noeuds de père en fils.

La suite de cette section précise les différentes étapes dans le calcul de l'intonation.

3.1 L'analyse syntaxique et l'identification de fonction énonciatives

Dans une telle perspective, la génération de l'intonation doit être précédée d'une analyse syntaxique identifiant les constituants, les relations entre eux, et même certaines constructions syntaxiques (parmi lesquelles la dislocation à droite, les adverbes "de phrase", l'extraction, l'incise, etc.). En l'état actuel des connaissances, une telle analyse est difficile à réaliser. Aussi le prototype utilise-t-il une analyse superficielle de type syntagmatique, mais enrichie d'informations sur le statut de dépendance (tête ou dépendant) de chaque constituant.

Le analyseur tabulaire ("chart") garde en mémoire les résultats intermédiaires. Même si l'analyse globale échoue, ces résultats restent utilisables. À partir de ces données, on construit l'arborescence (ou au besoin plusieurs arbres partiels juxtaposés) couvrant l'ensemble de l'énoncé, selon un critère heuristique.

```
+v2 [249]
  -n2 [130]
    +n1 [112]
      -det [2]: [les]
      +n [3]: [policiers]
  +v1 [123]
    -v1 [114]
      +v [6]: [ont]
    +v [7]: [retrouvé]
  -n2 [231]
    +n1 [115]
      -det [9]: [les]
      +n [10]: [traces]
    -rel [225]
      -prorel [13]: [que]
      +v2 [203]
        -n2 [137]
          +n1 [118]
            -det [16]: [les]
            +n [17]: [ravisseurs]
        +v1 [129]
          -v1 [120]
            +v [19]: [avaient]
          +v [20]: [laissées]
```

Dans l'exemple, l'arbre est couché sur le côté; le retrait sur l'axe horizontal indique le niveau de profondeur et les chiffres entre crochets les numéros des arcs dans le graphe; le + et le - indiquent une tête et un dépendant syntaxique, respectivement. (v2, v1, n2, n1, etc. sont des non-terminaux.)

Les marques énonciatives (cf. *supra*, § 2.3) sont représentées par des "marqueurs intonatifs" explicites, qui sont associés aux arcs du réseau (chart), donc à certaines parties du texte d'entrée, soit aux mots ou aux groupes de mots. Dans le prototype actuel, les marqueurs de modalité et de focalisation sont identifiés à partir la ponctuation. L'appendice est employé

lorsque, en position finale de phrase, on rencontre un adverbe “de phrase” (cf. § 3.5). L’extraction (“*c’est ... que...*”) est identifiée en tant que constituant dans l’analyse syntaxique.

3.2 La formation des groupes accentuels

Les GA sont constitués de façon récursive, en descendant l’arborescence. Pour chaque noeud on détermine la nature [\pm clitique]. Un noeud terminal est tonique (ou “non clitique”) s’il appartient à une catégorie grammaticale tonique (verbe, nom, adjectif, adverbe, pronom tonique, certains déterminants ou prépositions, ...). Tout autre élément terminal est classé clitique. Pour un noeud non terminal le caractère clitique dépend des noeuds fils. Chaque noeud non terminal regroupe des parties syntaxiquement liées (directement ou indirectement); elles peuvent être réunies dans un même GA à condition 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 le noeud frère qui suit immédiatement dans la chaîne, ou s’il n’y en a pas, avec celui qui précède. Il en résulte une réorganisation locale de l’arbre qui élimine les noeuds clitiques. 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.

```
+v2 ns=17
  -ga [2, 3] /le.pO.li.sje/ ns=4
  +v1 ns=3
    -ga [6] /O~/ ns=1
    +ga [7] /R@.tRu.ve/ ns=2
  -n2 ns=10
    +ga [9, 10] /le.tRas/ ns=2
    -v2 ns=8
      -ga [13, 16, 17] /k@.le.Ra.vi.s8R/ ns=4
      +v1 ns=4
        -ga [19] /a.vE/ ns=2
        +ga [20] /lE.se/ ns=2
```

3.3 La phonétisation et la syllabation

Afin de prendre en considération le rythme, il est nécessaire d’obtenir le nombre de syllabes (ou une mesure plus sophistiquée du poids sonore). La chaîne segmentale est fournie par la *phonétisation* qui est reprise au système LIPSS, cf. Dutoit (1997). Le *découpage syllabique* consiste à trouver, pour une chaîne donnée, une succession de syllabes telle qu’à chaque syllabe corresponde une voyelle pleine, précédée d’une attaque et suivie d’une coda. Au préalable on définit les attaques et les codas admises en français. En présence de plusieurs solutions, on maximalise l’attaque, ce qui donne l’enchaînement consonantique, caractéristique phonétique bien connue du français. La syllabation définitive ne sera effectuée qu’après la formation des GI (enchaînement consonantique et réalisation des schwas). L’information sur le nombre de syllabes est enregistrée dans chaque noeud, tant pour les noeuds GA que pour les noeuds supérieurs (voir “ns=N” dans les exemples).

3.4 La formation des groupes intonatifs

La formation des GI consiste à regrouper des GA courts, tout en respectant l’emboîtement syntaxique. Le modèle définit un paramètre indiquant la taille d’un GI long, égale à 6 syllabes par défaut. Ce seuil peut être modifié, p.ex., en fonction de la vitesse d’élocution.

D’abord, on obtient le nombre de syllabes pour chaque noeud (en allant des terminaux à la racine). Si ce nombre est inférieur ou égal au seuil, les GA aux noeuds fils sont regroupés

Un algorithme pour la génération de l'intonation

dans un même GI. Si la longueur maximum est dépassée, le noeud envisagé comportera plusieurs GI. La procédure est alors appliquée aux noeuds fils. L'emploi de l'arborescence garantit que seuls des constituants liés syntaxiquement sont regroupés.

Le regroupement crée une nouvelle entité (en l'occurrence un GI) qui 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.

```
+v2 ns=17
  -gi [2, 3] /le.pO.li.sje/ ns=4
  +v1 ns=3
    +gi [6, 7] /O~.R@.tRu.ve/ ns=3
  -n2 ns=10
    +gi [9, 10] /le.tRas/ ns=2
  -v2 ns=8
    -gi [13, 16, 17] /k@.le.Ra.vi.s8R/ ns=4
    +v1 ns=4
      +gi [19, 20] /a.vE.lE.se/ ns=4
```

3.5 L'incorporation des marqueurs intonatifs

Jusqu'ici les marqueurs intonatifs (cf. *supra*) étaient associés aux constituants syntaxiques. Maintenant que les GI ont été constitués, les marqueurs peuvent leur être attribués, afin de leur assigner plus tard les tons appropriés. Pour tout GI dans l'arbre, on recherche dans le "chart" les marqueurs associés, et on les ajoute au GI comme trait. Dans une phrase qui se termine par un appendice, les marqueurs de modalité et de focalisation (noté pmark(focus) dans les listings) éventuels seront transmis au GI qui précède l'appendice. Ces étapes sont illustrés sur l'exemple suivant pour l'énoncé "il nous ennuie, franchement".

```
Parse tree:
+v1 [41]
  -procl [1]: [il]
  -procl [3]: [nous]
  +v [5]: [ennuie]
-punc [6]: [(,)]
-adv [7]: [franchement]

Incorporate prosodic markers:
+gi [1, 3, 5] /il.nu.zA~.n9i/ ns=2
-gi [6, 7] /fRA~S.mA~/ ns=2 pmark(si(!)) pmark(focus) pmark(appendix)

Prosodic markers: handle appendix
+gi [1, 3, 5] /il.nu.zA~.n9i/ ns=2 pmark(si(!)) pmark(focus)
-gi [6, 7] /fRA~S.mA~/ ns=2 pmark(appendix)
```

3.6 Réajustements de la structure syntaxique — contraintes rythmiques.

Les exemples du § 2.3 montraient des regroupements intonatifs (observés) 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* syntaxique entre GI successifs. Ces données sont préparées à l'avance: pour chaque GI on détermine s'il entretient un lien de dépendance avec le GI suivant, quelle qu'en soit l'orientation: de tête à dépendant ou l'inverse. Dans les exemples, le trait l=+ indique la présence de ce lien.

1. L'exemple [16], avec la relative apposée à un groupe nominal, permet d'illustrer la *découpage*. Un constituant long sera découpé en deux parties juxtaposées s'il est structuré (il

comporte plusieurs GI), s'il entretient un lien de dépendance avec le GI précédant (en l'occurrence le verbe principal), et si le sous-constituant gauche est la tête syntaxique du constituant entier. Dans un cas pareil on coupera après le premier sous-constituant. Dans l'exemple donné on sépare la relative de son antécédent.

```
+v2 ns=17
  -gi [2, 3] /le.pO.li.sje/ l+= ns=4
  +gi [6, 7] /O~.R@.tRu.ve/ l+= ns=3
  -gi [9, 10] /le.tRas/ ns=2
-v2 ns=8
  -gi [13, 16, 17] /k@.le.Ra.vi.s8R/ l+= ns=4
  +gi [19, 20] /a.vE.lE.se/ ns=4 pmark(si(.))
```

2. Dans l'exemple “(les policiers) (ont retrouvé) (les traces)...” les GI 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 GI frères; si le nombre de syllabes cumulé est inférieur à une valeur critique, on reconstitue un constituant qui les englobe.

```
+v2 ns=17
  -gi [2, 3] /le.pO.li.sje/ l+= ns=4
  ?recons ns=5 l+=
    +gi [6, 7] /O~.R@.tRu.ve/ l+= ns=3
    -gi [9, 10] /le.tRas/ ns=2
-v2 ns=8
  -gi [13, 16, 17] /k@.le.Ra.vi.s8R/ l+= ns=4
  +gi [19, 20] /a.vE.lE.se/ ns=4 pmark(si(.))
```

3.7 L'attribution des frontières prosodiques

L'étape suivante associe aux noeuds (constituants et GI) des degrés de frontière prosodique qui permettront d'attribuer les tons appropriés. Le prototype 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: tel noeud qui reçoit un degré D , transmet ce même degré à son constituant fils droit, et un degré $D-1$ aux autres fils. L'entité majeure (soit l'énoncé entier) reçoit le degré de frontière maximal. L'appendice (forcément en fin d'énoncé, si présent) reçoit un traitement spécial où il passe la frontière D au constituant qui le précède. Les degrés de frontières sont notés comme “bound= D ” dans les listings.

```
+v2 bound=4 ns=17
  -gi [2, 3] /le.pO.li.sje/ bound=3 l+= ns=4
  ?recons bound=3 ns=5 l+=
    +gi [6, 7] /O~.R@.tRu.ve/ bound=2 l+= ns=3
    -gi [9, 10] /le.tRas/ bound=3 ns=2
-v2 bound=4 ns=8
  -gi [13, 16, 17] /k@.le.Ra.vi.s8R/ bound=3 l+= ns=4
  +gi [19, 20] /a.vE.lE.se/ bound=4 ns=4 pmark(si(.))
```

3.8 Assignation des tons

Toutes les données étant réunies, 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 la présence de focalisation.

Ce qui donne comme résultat final:

/le.pO.li.sje/	HH
/O~.R@.tRu.ve/	/LL
/le.tRas/	HH
/k@.le.Ra.vi.s8R/	HH
/a.vE.lE.se/	L-L-

ton	degré	modalité	focus
l..l-	4	(appendice bas)	-
h..h	4	(appendice haut)	-
L-L-	4	déclarative	-
H/H	4	interrogative	-
HL-	4	déclarative	+
HL	3	*	+
LH	3	*	-
HH	3	*	-
/LL	2	*	-
LL	1	*	-

4. Résultats et discussion

L'algorithme est intégré dans un système de synthèse de la parole, comportant plusieurs modules (lemmatisation, analyse syntaxique, phonétisation, syllabation, génération de l'intonation, et module acoustique), ce qui complique l'évaluation. En effet, telle erreur dans la sortie peut provenir de plusieurs endroits. Qui plus est, une évaluation objective, à partir d'un test d'écoute, suppose un module acoustique afin d'obtenir la forme acoustique correspondant à la transcription phonétique et intonative. Or, cet aspect n'était pas disponible dans le cadre de ce travail, pour la simple raison qu'une conversion des tons à la courbe F_0 n'existe pas encore. Pour cette raison, les données quantitatives manquent ici.

Mais l'évaluation qualitative est pour le moins aussi importante. Si l'approche proposée est bien plus onéreuse que les approches automatiques, cela se justifie par la nécessité de prendre en considération la syntaxe. Le résultat peut alors être évalué à partir d'un corpus de phrases qui poseraient des problèmes pour des approches sans information syntaxique. Voici, à titre d'exemple, le résultat obtenu pour quelques phrases de ce corpus.

Elle lui enverra une lettre?	el.l9i.A~.vE.Ra	HH	yn.lEtR	H/H		
Envoie-les moi.	A~.vwa.le.mwa	L-L-				
Devrions-nous lui proposer de se retirer ?	d@.vRjO~.nu	/LL	l9i.pRO.po.ze	HH d@.s@R.ti.Re	H/H	
Il vend ça aux imbéciles.	il.vA~.sa	HH	o.zE~.be.sil	L-L-		
Ils donnent celui-là à ceux-ci.	il.dOn	/LL	s@.l9i.la	HH a.s2.si	L-L-	
Je cherche la personne responsable pour les livres d'anthropologie.	Z@.SERS	/LL	la.pER.sOn.RE.spO~.sabl	HH puR.le.livR	HH dA~.tRO.pO.lO.Zi	L-L-
C'est le livre, qu'il donne.	sEl.livR	HL-	kil.dOn	l-l-		

Nous envisageons d'ores et déjà certaines modifications et extensions. Le système actuel identifie les groupes intonatifs et leur associe des tons finals. Les entités inférieures au GI sont ainsi délaissées. En vue d'introduire l'accent initial (ou accent d'insistance) et les tons de la pénultième, il est indispensable de préserver la structuration interne des GI, en mots et en syllabes. L'accent initial, par exemple, se place à la syllabe initiale, si celle-ci comporte une attaque consonantique, et sinon à la deuxième.

Le système proposé se distingue d'autres systèmes de génération de la prosodie par son caractère *syntaxique*, *détaillé* et *explicite*. La place importante accordée à la syntaxe se mani-

feste entre autres dans l'emploi de l'arbre syntaxique comme structure de données. (Il est vrai que cela nécessite une analyse syntaxique robuste.) Le caractère détaillé du modèle intonatif (et du traitement mis en oeuvre) permet de manipuler séparément plusieurs aspects de l'intonation, dont certains liés à la signification et d'autres à des aspects phono-stylistiques (e.g. l'emploi de certaines marques, la vitesse d'élocution). La différence cruciale entre notre système et un système à apprentissage automatique, réside dans son caractère explicite, inhérent aux systèmes par règles. Ceci permet de suivre les traitements successifs, de localiser les erreurs, et de vérifier ainsi le modèle intonatif. Ces propriétés réunies permettent au système d'exprimer un sens précis par le contrôle de son intonation, par l'insertion (automatique ou manuelle) de marqueurs, et de le faire d'une façon compositionnelle.

Remerciements

Je remercie Thierry Dutoit pour la mise à disposition de la phonétisation de LiPSS.

Références

- AUBERGÉ, V. (1991) *La synthèse de la parole, des règles aux lexiques*. Thèse de doctorat, Univ. P. Mendès-France, Grenoble, pp. 210
- DELAIS, E. (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
- 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 JEP XXII, pp. 135-138 (Martigny 15-19 juin 1998)
- DUTOIT, Thierry (1997), *An Introduction to Text-to-Speech Synthesis* (Dordrecht: Kluwer Academic Publ.) 312 pp.
- EYNDE, K. van den; MERTENS, P. & SWIGGERS, P. (1998) "Structuration segmentale et supra-segmentale 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
- GAUDINAT, Arnaud & GOLDMAN, Jean-Philippe (1998) "Le système de synthèse FIPSVox: syntaxe, phonétisation et prosodie". Actes JEP XXII, pp. 139-142 (Martigny 15-19 juin 1998)
- LONCHAMP, François (1998) "Prédire l'intonation d'une phrase affirmative: les facteurs accentuels, rythmiques, syntaxiques et énonciatifs." *Verbum* 98-1, pp. 35-47.
- MALFRÈRE, Fabrice; DUTOIT, Thierry & MERTENS, Piet (1998) "Un générateur de prosodie "tout automatique"". *Actes Journées d'Etude sur la Parole XXII*, 147-150 (Martigny 15-19 juin 1998)
- MALFRÈRE, Fabrice; DUTOIT, Thierry & MERTENS, Piet (1998) Fully Automatic Prosody Generator for Text-to-Speech Synthesis. *Proc. ICSLP*, 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
- MERTENS, P. (1993) Accentuation, intonation et morphosyntaxe, *Travaux de Linguistique* 26, 21-69.
- MERTENS, P. (1997) De la chaîne linéaire à la séquence de tons, *T.A.L.* Vol. 38, pp. 27-51
- MORLEC, Y.; AUBERGÉ, V.; BAILLY, G. (1995) Evaluation of automatic generation of prosody with a superposition model. *Proc. Int. Cong. Ph. Sc.* XIII, vol. 4, pp. 224-227 (Stockholm)
- ROSSI, M. (1993) A model for predicting the prosody of spontaneous speech (PPSS model), *Speech Communication* Vol. 13, pp. 87-107
- 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