

## GABARITS PROSODIQUES, EMPATHIE(S) ET ATTITUDES

**Antoine AUCHLIN<sup>o</sup> & Anne Catherine SIMON<sup>o\*</sup>**  
**Université de Genève<sup>o</sup> – FNRS-Université de Louvain\***

### 1. INTRODUCTION

Qu'est-ce que la prosodie apporte à la construction du sens, dans le discours ? Cette question, dont la redoutable simplicité est due à Piet Mertens, nous paraît totalement d'actualité. Sans prétendre à l'exhaustivité et la simplicité de la réponse idéale, nous voulons plutôt établir ici des conditions pour y répondre. Pour cela, nous nous penchons sur quelques petits faits discursifs tirés de notre corpus, petits faits à la fois parfaitement « ordinaires », banals, mais dont la description permet d'établir un certain nombre de propositions – de thèses :

- (a) la prosodie contribue au sens comme donnée d'expérience, pas seulement comme donnée d'interprétation (= *nature des faits à considérer*, voir 2.1) ;
- (b) les différentes dimensions qui constituent globalement et solidairement la prosodie sont à la fois relativement autonomes (conformées selon des principes et des règles propres), non exclusivement fonctionnellement attribuées (susceptibles de coopérer avec différentes sphères de l'organisation langagière) et opportunistes, dans le choix des plans auxquels elles contribuent et dans la manière d'y contribuer, qui va de la redondance à la complémentarité (« enjambement ») – aussi bien qu'à la cacophonie (= *principes généraux de fonctionnement*, voir 2.2) ;
- (c) les données prosodiques interviennent toujours comme médiation entre deux systèmes expérimenteurs (deux sujets parlants), plus ou moins synchronisés l'un à l'autre, plus ou moins sensibles et adaptés à cette synchronisation –

plus ou moins **empathiques** (= *condition interactionnelle de la prosodie*, voir 2.3) ; la perception de la synchronisation, qui constitue une condition à son établissement, s'exerce également dans les interactions dissymétriques (interaction en différé et sans rétroaction, par exemple dans l'observation de données discursives).

Par ces propositions, on tente d'adopter (de simuler) le point de vue panoramique et syncrétique dont dispose toute personne engagée dans une activité de discours. Ce point de vue d'utilisateur ordinaire, non professionnel, peut être partiel ; celui que nous adoptons ici, « scientifique », ne l'est pas : certes, pour présenter et analyser les exemples, nous nous fondons sur une certaine interprétation-appréhension de ceux-ci, mais nous ne tenons pas à celle-ci plutôt qu'à telle autre (a priori en tout cas). Ce qui nous intéresse leur serait commun : ce sont les conditions générales dans lesquelles s'établit l'interprétation - appréhension du discours et les modalités d'intervention de la prosodie dans ce processus.

Si l'analyse du discours veut prendre au sérieux la *granularité prosodique* ; si l'étude de la prosodie veut prendre au sérieux la *complexité des faits discursifs*, elles doivent constater qu'on ne peut pas relier formellement, et terme à terme, une suite d'événements prosodiques complexes et une séquence verbale complexe, sans faire intervenir d'une façon ou d'une autre des « êtres d'expérience », entités parlantes et donneuses de sens, susceptibles de disposer d'attitudes et d'émotions, soit par expérience intérieure subjective, soit, à défaut, parce qu'on leur en attribue (si ce sont des machines)<sup>1</sup>.

## 2. PROPOSITIONS THÉORIQUES

### 2.1. Nature des faits à considérer

Si la prosodie contribue au sens dans la parole spontanée, elle s'intègre « naturellement » dans la masse du discours en tant que donnée d'expérience, et pas seulement comme donnée d'interprétation<sup>2</sup>. Une part des données prosodiques ne s'intègre en effet que dans l'*expérience* du discours, n'a de raison d'être ou de sens qu'à opérer dans ce qui est *vécu* par les participants : cela va d'une influence

<sup>1</sup> Ce qui, pour les interactions homme-machine, serait déjà un progrès considérable.

<sup>2</sup> Ceci concerne à la fois la notion d'interprétation telle qu'établie par la théorie de la pertinence (SPERBER & WILSON 1989) et les conceptions moins définies ayant cours généralement en analyse du discours ; cela concerne également la notion d'indice de contextualisation bien que, jusqu'à un certain point, cette dernière possède quelques points communs avec notre idée du caractère opportuniste et coopératif du fonctionnement des phénomènes prosodiques – traitée ci-dessous.

hédonique, comme la contagion du rire, au partage d'une pulsion rythmique, simple partage de formes mais dont la valeur est celle d'un vécu commun, de la construction d'un événement de parole partagé.

Dans le volume global des informations acoustiques présentes (à supposer qu'il soit quantifiable), quelle est la proportion des données prosodiques expérimentiellement adressées ? Si elle est difficile à quantifier, cette part est qualitativement nécessaire – ce que démontrent *a contrario* les machines de synthèse, qui n'arrivent à susciter ni un réel engagement de l'auditeur dans le discours qu'elles produisent, ni un intérêt et une attention de longue durée. Cela montre que si les données prosodiques expérimentielles ne sont pas nécessaires à l'*interprétation* des différentes parties constituant le discours (en effet, pour ces unités produites hors co-texte et hors contexte, les machines font relativement bien leur travail ; voir SEEBER ici-même), elles le sont en revanche pour que l'interprétation advienne dans le cadre d'un vécu qui nous paraisse vivant.

Cette « attribution de vivance », dans la réalité du discours, n'est pas un léger supplément, cosmétique, qui s'ajoute après-coup, si nécessaire (même si c'est l'un des pas qui reste à franchir dans l'histoire de la synthèse de la parole). Dans la phylogenèse, dans l'ontogenèse, et vraisemblablement dans la plus simple des interprétations, cette attribution de vivance est une condition nécessaire de premier plan. On s'en convainc facilement en écoutant les différentes versions synthétisées de notre extrait de corpus. Les informaticiens s'occupant de synthèse de la parole n'utilisent pas d'autre terme que celui-ci pour définir le cahier des charges d'une collaboration avec les spécialistes du discours : il faut « plus de naturel », autrement dit il faut que la synthèse « ait l'air plus vivante ».

La prosodie du discours, dans la parole naturelle plus ou moins « spontanée », nous met nécessairement en contact avec la matière extrêmement complexe qui fait la texture du « vivant », y compris le fait que ce « vivant » offre une certaine résistance à l'appréhension, une volatilité, qui pourrait être ce qu'on cherche parfois à désigner comme « âme<sup>3</sup> ». Le « vivant » du discours authentique, c'est qu'il consiste en une constante variation de l'intérêt engagé (que reflète chaque variation d'intensité, chaque différenciation mélodique), en une constante variation d'attitudes associées aux foyers d'intérêt successifs, d'attitudes à l'égard du fait de parler (envie ou non), dans un certain climat d'interaction (positions d'interaction, proximité-intimité, pouvoir, etc.). C'est ainsi que nous vivons le discours, et il importe pour le vivre ainsi de pouvoir imputer un tel vécu à toute source de parole. La parole des machines est beaucoup plus acceptable lorsqu'elle présente des caractéristiques attitudinelles, des

---

3 « La quantité des mots est bornée ; celle des accents est infinie. (...) [L'intonation] c'est l'image même de l'âme rendue par les inflexions de la voix, les nuances successives » (Diderot, Salon de 1767 : 136, 268 ; cité in FÓNAGY 1983 : 121). De même : « Quand j'entends parler quelqu'un, au travers de la mélodie du mot, je vois bien plus profondément dans son âme » (Janos Janacek, cité par MENAHEM 1983).

traits prosodiques qui à proprement parler ne relèvent pas d'un traitement interprétatif, mais d'un traitement perceptif : leur objet est de faire sentir l'adoption d'une attitude, et leur fin est l'accréditation, par l'interprète-expérienceur, d'un autre organisme avec lequel il entre en interaction.

## 2.2. Principes généraux de fonctionnement

Les données prosodiques sont organisées en différents « sous-systèmes<sup>4</sup> », plus ou moins autonomes, dotés de principes et de règles propres ; ils se conforment pour une part selon ceux-ci, et pour une part selon les contraintes provenant de leurs interactions avec les différentes sphères de l'organisation de la parole (de la morpho-syntaxe aux constructions discursives). Ces différentes formes prosodiques s'articulent avec les autres plans de façon « coopérative » : la coopération va de la redondance, cas obligé des machines de synthèse *text-to-speech* qui travaillent uniquement à partir (des structures syntaxiques) du texte, à la complémentarité de l'« enjambement » prosodique, telles les frontières prosodiques majeures associées à des frontières textuelles mineures, ou les pauses après les déterminants.

La contribution des formes prosodiques à la construction de l'expérience discursive est plutôt « opportuniste » : dans une large mesure, ces formes manifestées se montrent aptes à coopérer avec différentes sphères de l'organisation verbale, de la délimitation des actes de parole (dans la dimension hiérarchique-textuelle) au système des tours de parole (dimension interactionnelle), en passant par l'organisation énonciative, polyphonique, informationnelle – topicale, etc. (voir SIMON, GROBET & AUCHLIN ici-même, pour illustration).

Enfin, on suppose que les différentes manifestations prosodiques concomitantes ne font pas nécessairement l'objet d'un traitement « centralisé » au plan interprétatif, *contra* une doxa ambiante plus ou moins explicite ou affirmée<sup>5</sup>. Ainsi, être « pris » dans une structure rythmique (patterns de listes, voir 3.1.3), dans un *decrecendo* d'intensité ou dans une chute mélodique qui suggère le rapprochement de celui (celle) qui parle (voir 3.2), sont des processus qui se déroulent de façon largement indépendante les uns des autres. Et ils contribuent à la construction du sens interprétatif à la mesure de leur degré de mise en phase (ou de coopération par

---

<sup>4</sup> Accentuation, rythme, contours intonatif, effets de registre, etc.

<sup>5</sup> Soutenue indirectement par WICHMANN : « Speaker meaning, in so far as it deviates from the surface meaning of an utterance, is generated in a systematic way, usually when there is a perceived mismatch between the content of an utterance and the context in which it is conveyed. I assume that at least *some* intonationally conveyed attitudes are also conveyed by some kind of mismatch, for example between the intonation and the message, or between the intonation and the context » (2000 : 145-146) ; cette doxa est globalement en phase avec la modularité centralisée de Fodor et les formes de cognitivisme qui s'y rattachent.

opposition de phase) avec les données d'un autre plan, d'une autre sphère de l'organisation verbale.

### 2.3. Condition interactionnelle de la prosodie

Nous supposons donc que le discours est une chose qui existe pour un organisme susceptible non seulement d'interpréter (décodage et inférences) mais aussi d'expérimenter ; nous supposons également que la prosodie intervient à la fois comme reflet ou symptôme d'aspects divers de l'expérience du parleur<sup>6</sup>, et comme déclencheur d'expérience (ou plus exactement d'expérimenter) pour l'interprète ; la première de celles-ci est celle de sa mise en synchronisation, de sa mise à disposition à partager du vécu discursif (son « ouverture expérientielle »).

Les aspects de l'expérience reflétés par les données prosodiques peuvent être schématiquement représentés comme un repère à trois axes – un espace à trois dimensions – mobile dans le temps :

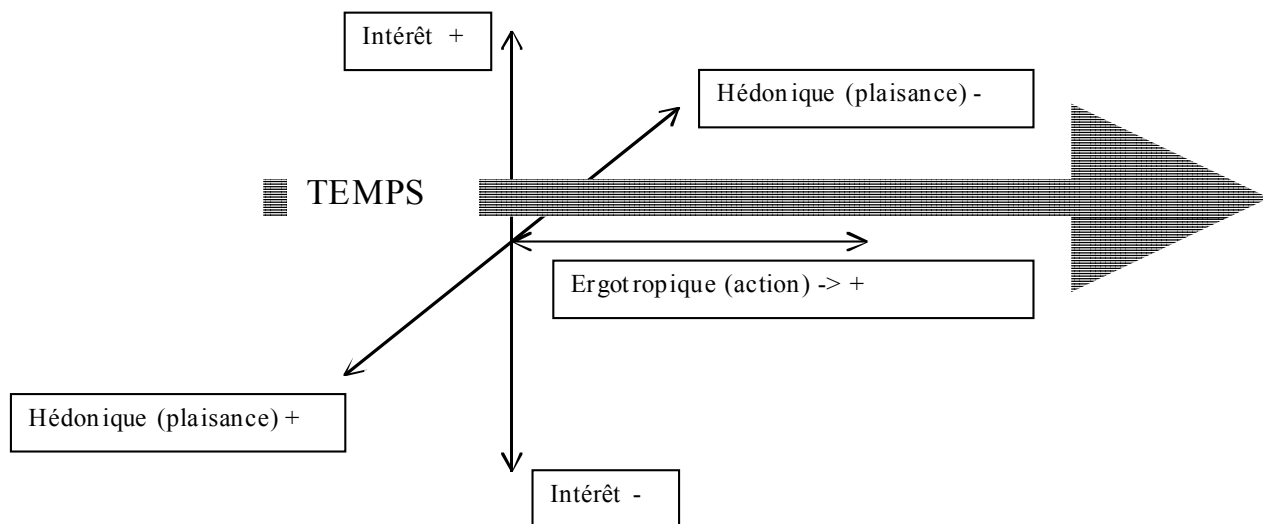


Figure 1. Dimensions de l'expérience prosodiquement réglées

Cette figure appelle quelques commentaires.

- (a) L'axe de l'**intérêt** note les variations pour l'expérimenter dans ce dans quoi il est engagé ; cet axe est relié de façon privilégiée à la quantité d'énergie investie ou engagée dans le signal<sup>7</sup>. Celle-ci peut se manifester (être dissipée) de différentes façons, en fonction, notamment, des valeurs sur l'axe

<sup>6</sup> Voir BOLINGER (1986) pour qui l'intonation, avant tout, reflète symptomatiquement la manière dont le parleur se sent vis-à-vis de ce dont il parle, de ce qu'il dit.

<sup>7</sup> Et inversement : l'énergie investie dans le signal est le vecteur des « inférences émotionnelles » relatives à l'intérêt investi dans l'événement – indépendamment de la valeur hédonique.

hédonique : sous forme de volume sonore (en dB) aussi bien qu'en précision articulatoire, serrement de mâchoires, etc. On peut investir davantage d'énergie à chuchoter fort qu'à parler normalement ; l'énergie peut être dissipée par la mobilisation de mouvements antagonistes, produisant un rapport signal/bruit où le bruit prend une place significative ; ces modalités de dissipation de l'énergie sont l'objet de l'axe ergotropique.

- (b) L'axe **ergotropique** décrit, globalement, l'activité déployée - mobilisée par l'expérienceur ; il contrôle l'activité générale, et le mode de dissipation de l'énergie (le degré d'activité) ; le degré d'activité se manifeste autant dans la précision/dispersion articulatoire, dans le débit (en syllabes par seconde), le volume sonore (en dB), que dans le rythme (réalisation des accents secondaires ; ratio de syllabes accentuées par groupe intonatif, etc.). L'activité déployée et ses caractéristiques sont saisies dans leurs rapports avec l'évolution de l'intérêt et de la valeur hédonique, et dans ce jeu s'établissent des traits essentiels du sens de l'événement de parole.
- (c) L'axe **hédonique** décrit les variations de plaisir - déplaisir associées à ce qui a lieu ; celles-ci se manifestent de différentes manières, à différents niveaux : macro (caractéristiques globales d'un accès de colère vs une parole enjouée), et micro – effets de timbre, liés au mode phono-articulatoire (coins de la bouche vers le haut/le bas, type de respiration, etc.<sup>8</sup>). Bien entendu ces modifications interviennent de façon notablement différente selon le degré d'intérêt et la valeur ergotropique (à valeur hédonique égale, différence entre un éclat de rire et un sourire vocal).

Chacun des axes fait l'objet d'un traitement plus ou moins distinct ou cloisonné ; le signal prosodique contient en fait **plusieurs systèmes de signaux prosodiques** (FIRTH 1948, cité par OGDEN 2001<sup>9</sup>). Et nous adoptons la notion de fractalité (au moins comme métaphore, voir point 4) pour rendre compte de ce fonctionnement complexe.

De nombreux signaux prosodiques ont une première valeur proprioceptive : ils interviennent dans l'autorégulation de la production et on peut supposer que c'est à ce titre-là qu'ils prennent une « fonction communicative ». La fonction communicative des signaux prosodiques s'établit également comme une autorégulation non plus au niveau du système de la personne, mais au niveau du macro-

<sup>8</sup> Voir dans le corpus les manifestations de « sourire vocal » (AUBERGÉ & LEMAÎTRE 2000) après la réplique *C'est qui, on ?* À la fois ténues et parfaitement perceptibles, ces manifestations sont incontournables dans l'interprétation. Voir aussi D'ALESSANDRO *et al.*, ici même.

<sup>9</sup> Parmi les éléments de l'approche de FIRTH retenus par OGDEN, il y a l'idée que les observations linguistiques doivent être faites à plusieurs niveaux tous mutuellement compatibles – autrement dit, que les analyses menées à différents niveaux doivent être réintégrées, ce qui fonctionne à un niveau devant pouvoir être relié à un autre (strictement linguistique ou « contextuel ») (OGDEN 2001 : 3).

système qu'est l'interaction en cours (notamment en termes de synchronisation interactionnelle – d'empathie – voir les *supra-individual biological processes* décrits par NÚÑEZ 1990). Le degré, la qualité de la synchronisation interactionnelle (adaptation mutuelle des systèmes en contact) est une donnée de fond des interactions verbales à laquelle participent potentiellement tous les aspects de chaque contribution : alternance et passation des tours de parole, alignement rythmico-temporel et débit, distance proxémique vocalement évoquée, partage des alternances hédoniques (je souris quand tu souris...), de l'intérêt investi, etc.

Lorsque deux systèmes interagissent, leur rapport est localisé sur deux axes distincts (KERBRAT-ORECCHIONI 1990 ; WICHMANN 2000 : 147) : l'un horizontal sur lequel se mesure la proximité ou la distance entre les systèmes, et l'autre, vertical, où se mesure le « pouvoir » de l'un sur l'autre<sup>10</sup>. Les positions sur ces deux axes : (i) ne sont pas forcément partagées par les systèmes en présence ; (ii) sont virtuellement modifiables ; (iii) sont constamment reflétées par le comportement – entre autre vocal.

Bien entendu, tout cela ne revient pas à dénier l'existence, plus ou moins conventionnelle et *language dependant*, de différentes formes intonatives – des traits ou des séquences de traits voco-mélodiques – dotées plus ou moins conventionnellement d'un certain contenu, soit de dénier l'existence de morphèmes intonatifs. Leur contenu peut être de type instructionnel, comme pour le regroupement intonatif ou la focalisation ; il peut être de type plus « modal », voire « représentationnel », iconique. Quel que soit l'inventaire de ces formes (et de leurs types), l'occurrence d'une forme dans la chaîne a lieu *d'une certaine façon*. Et le traitement, l'identification, l'accès à la forme intonative présuppose le traitement de cette « façon » : par exemple, un morphème HB (ton dynamique en position d'accent final qui marque en français un groupe intonatif comme focalisé) sera nécessairement attaqué à une certaine hauteur, réalisé avec une certaine pente et une certaine vitesse de descente, et jusqu'à un point d'arrivée, réglé au dixième de ton. Le traitement (interprétatif) du morphème et l'expérienciation (perceptuelle) de la façon dont il est réalisé, de son occurrence dans ce contexte-là, se font conjointement.

### 3. ÉTUDE DE CAS

Les descriptions qui suivent mettent essentiellement l'accent sur les aspects qui étayent les positions développées ci-dessus, et ne prétendent pas à une quelconque

---

<sup>10</sup> Il faut nuancer, car la position haute n'implique pas l'imposition de pouvoir ; de surcroît, ni position ni pouvoir n'indiquent les particularités de la place d'où cela se fait, les co-déterminations d'identité (FLAHAULT 1978).

« exhaustivité » descriptive<sup>11</sup>. Il s'agit de listes, d'une part, et de phénomènes de synchronisation « d'humeur » ou de « disposition » (vis-à-vis de l'interlocuteur), qui se manifestent entre autres aux changements de tour de parole, d'autre part. D'un point de vue monologique (pour les listes) ou dialogal (pour les (dé)synchronisations aux changements de tour), nous nous sommes penchés sur ces effets de « gabarits » ou d'empathie, par lesquels les interactants (ou l'analyste) se laissent parfois entraîner. À ces moments plus qu'à d'autres, on est localement sollicité à entrer dans (participer à) l'interaction, par la perception-contagion très physique (expérientielle) des patterns qui sont mis en œuvre et des attitudes qui y sont temporellement déployées, de manière complexe et entremêlée.

### 3.1. Les listes ou énumérations<sup>12</sup>

Une transcription prosodique basée sur un modèle morpho-phonologique de l'intonation du français (comme celle proposée par MERTENS 1987) ne capte que partiellement la complexité de l'organisation prosodique des listes. En effet, la construction et la reconnaissance des schémas prosodiques de liste s'appuient en grande partie sur des phénomènes **rythmiques** – régularité du retour des temps accentués, espacement régulier des attaques, accents initiaux par élément mis en liste – et de « **modulation** » – dynamisme accru dans la réalisation des tons, élargissement du registre tonal, allongement (ou réduction) de la durée des syllabes accentuées, etc. Il en résulte que les schémas de liste sont perceptibles, voire très manifestes, lorsqu'on écoute (participe à) une interaction verbale, mais qu'ils sont en revanche beaucoup plus difficiles à repérer quand on n'a accès qu'à la transcription prosodique de cette interaction.

<sup>11</sup> Voir SIMON, GROBET & AUCLIN (ici-même) pour une « mise à plat » des différentes dimensions en jeu dans un extrait plus réduit du corpus.

<sup>12</sup> « The notion of 'itemization' is introduced by Jayyusi (1984 : chap. 3), who distinguishes 'lists' – which have gestalt properties and constituent elements displaying that they are 'co-selected to stand together' (78) – from 'itemizations', enumerative paradigms where the items are singled out to stand as individuals. The procedure seems to correspond to a notion in intensification theory (*Stapelung*; see van Os 1989). For example, in an emphatic version of Germ. *Ganz besonders herzlichen Dank*, 'very many special thanks', the stresses tend to be aligned on a par in a serial array and are detached from the hierarchical constituency of (((*ganz*) (*besonders*))) (*herzlichen*) (*Dank*). Seen more generally, 'itemizations' as a prosodic resource would seem to run counter to – or rather create tension with – the stress hierarchies as they are described in metrical trees and grids in phonology » (AUER, COUPER-KUHLEN & MÜLLER 1999 : n. 17, p. 217). La tension repérée par AUER *et al.* s'exerce également avec d'autres dimensions, comme nous le montrons ; de sorte que, en suivant leur raisonnement, la distinction due à JAYYUSI entre « listes » et « itemisation » présente l'inconvénient de « figer », dans ces effets, certes peut-être prototypiques, le jeu qui s'exerce entre les dimensions engagées, produisant tensions et accords, effets de phase, déphasages, hors phases – et qui n'exécute que rarement la figure idéalisée du prototype. Voir §4.

L'organisation interne d'une séquence énumérative typique peut se décomposer en trois positions : la position occupée par le premier élément de la séquence (qui met ou non en jeu son pouvoir cataphorique d'annonce), la position médiane (occupée par un certain nombre d'items – au moins deux<sup>13</sup>), et la coda. Nous aimerions montrer que les « patterns » ou les organisations prosodiques qui sont utilisés par les locuteurs dans chacune de ces trois positions (i) peuvent fonctionner de manière autonome ; (ii) peuvent être interprétés à différents niveaux de l'organisation du discours ; (iii) ne font pas nécessairement l'objet d'un traitement centralisé et orienté vers un seul et unique foyer attentionnel de l'interaction verbale en cours.

### 3.1.1. Le marquage prosodique des listes

Tantôt le **premier élément de la liste** s'annonce comme tel, en affichant un pouvoir de projection (qui annonce l'énumération à venir), tantôt il est non marqué – et n'acquiert le statut de « premier élément d'une liste » que rétrospectivement, par l'apparition d'un second élément présentant une similitude syntaxique, prosodique et/ou sémantique avec lui. Le pouvoir cataphorique du premier élément peut se manifester à au moins deux niveaux, syntaxique et prosodique.

Certaines constructions syntaxiques appellent au moins deux éléments (ou plus) pour être considérées comme complètes<sup>14</sup>. C'est le cas de constructions comme *qu'il s'agisse de ...*, qui projettent automatiquement une suite : *ou de...* ; ou de la préposition *ni*, qui appelle au moins deux éléments coordonnés de polarité négative (*on voit qu'aucune [catégorie] n'est épargnée, ni... ni...*). Dans d'autres cas, la structure syntaxique ne projette aucune liste, mais un contour prosodique particulier présente un élément comme le premier d'une énumération.

(1) alors que là ça a eu une grande gravité  
 parce que c'était → les grands chefs religieux  
 h.....b/B\B H bb BB b..b HH  
 les hommes de science  
 H b.../b HH  
 les hommes de loi euh  
 (H) b.../b HH x  
 les hommes de lettres  
 b.....b HH x  
 vraiment toutes les professions  
 b.....b H b.....b H/H

Dans (1), la construction syntaxique *parce que c'était* implique certes la présence d'un complément, mais pas nécessairement sous la forme d'une énumération. Cependant, l'élément *les grands chefs religieux* est d'emblée signalé comme le premier élément d'une liste, à cause de son contour intonatif particulier : il

<sup>13</sup> Deux éléments peuvent entrer dans (« saturer ») un effet de liste ; par ailleurs une liste à deux termes se distingue d'une coordination (*acheter : du pain ; du lait. Acheter du pain et du lait*).

<sup>14</sup> Voir les *dispositifs* macrosyntaxiques (BLANCHE-BENVENISTE 1990).

est séparé de son verbe recteur par une pause (de 121ms) ; il porte un accent initial de niveau haut ; il est réalisé avec un dynamisme accru par rapport à ce qui précède (les cibles tonales sont plus hautes, les glissandos internes aux tons dynamiques plus marqués) ; il porte enfin un contour continuatif (ce contour, fréquent en français, permet simplement de signaler que l'unité se poursuit). On remarquera en outre que l'accent initial de niveau haut est placé sur l'article défini, élément clitique qui échappe normalement à l'accentuation : *LES grands chefs religieux*. Cet accent initial ouvre une frontière gauche qui est fermée par la frontière du groupe intonatif – délimitant ainsi un « arc accentuel » (DI CRISTO 1999).

Pour résumer, on peut dire que, en français, la présence d'un ou de plusieurs des éléments prosodiques suivants concourt à désigner un constituant comme le premier élément (potentiel) d'une liste :

- une attaque marquée détachant l'élément de ce qui précède (prise de souffle, pause) ;
- un accent initial de niveau haut (porté par la première syllabe de l'élément, fût-ce l'article ou la préposition qui l'introduit) ;
- un accent dynamique en position finale, éventuellement précédé d'une syllabe pénultième atone haute ;
- ce contour étant lui-même réalisé avec une *modulation* mélodique et rythmico-temporelle forçant ou exagérant certains traits ou écarts ; cela installe un point d'ancrage pour l'identification ultérieure des autres membres.

Ces caractéristiques prosodiques disposent d'un pouvoir projectif élevé. La prosodie et/ou la syntaxe s'allient donc pour organiser une suite de syntagmes en liste, en travaillant tantôt de manière complémentaire, tantôt de manière redondante. Pourtant, dans certains cas, on observe un **conflit** entre les instructions données par la syntaxe et l'empaquetage prosodique. Dans (2) l'élément *des dictionnaires* est présenté comme annonciateur d'une énumération, à cause du contour mélodique particulier porté par les deux dernières syllabes : pénultième atone de niveau haut et accent final de niveau haut abaissé (contour régulièrement désigné comme *call contour*, MERTENS 1987). On peut remarquer en outre que l'acte *qui # publie souvent des dictionnaires* fait l'objet d'un traitement rythmique particulier. Le pronom *qui* est isolé, car il est suivi d'une pause ; cette pause donne l'impression que l'attaque de ce segment se fait en deux temps, une première fois sur le *qui* et une seconde fois sur la syllabe initiale de *PUBlie*<sup>15</sup>. Enfin, cet acte est segmenté en deux groupes intonatifs de quatre syllabes chacun (*publie souvent* et *des dictionnaires*) ; cette scansion

---

<sup>15</sup> La première syllabe de *publie* est réalisé avec + 20Hz par rapport au *qui*, alors que ce *qui* a déjà fait l'objet d'une première réinitialisation mélodique.

installe déjà un effet de répétition (pattern rythmique) annonçant une énumération potentielle<sup>16</sup>.

```
(2) c'est l'homme qui s'occupe de cette collection qui a ce
      bb      /BB  \b.....b /BB  \b.....b
      qui   publie souvent → des dictionnaires
      ↑   b  # B /BB  bb HH      b.....h \HH
                                       dictionnaire d(e) la bêtise
      b.....b /BB      \b..h \HH
                                       dictionnaire des erreurs politiques
      b....b /BB      \b.....h \HH
                                       et cætera
      b....b B-B-
```

Sans l'information dispensée par la prosodie, on pourrait contester cette interprétation faisant de *des dictionnaires* le premier élément d'une liste. En effet, syntaxiquement parlant, il ne se situe pas au même plan que les éléments *dictionnaire de la bêtise*, *dictionnaire des erreurs politiques*, et *cætera*. Dans une transcription orthographique standard, cette différence de statut pourrait se marquer par une ponctuation du type : *qui publie souvent des dictionnaires (dictionnaire de la bêtise, dictionnaire des erreurs politique, et cætera)*. Pourtant, on est amené à reconnaître que la similarité répétée d'un contour particulier (b...h \HH) provoque un effet de cohésion et inclut rétrospectivement ce premier élément dans la liste, même si son statut syntaxique ne devrait pas l'y autoriser. Dans ce cas, on peut dire que les instructions prosodiques prennent le pas sur le format syntaxique.

Le fragment (3) présente une résolution similaire du conflit entre instructions prosodiques et format syntaxique :

```
(3) ils ont oublié à quel point      de Molière
      b.....b /BB  \b..b   /BB  @ H \hh BB
                                       à Baudelaire
      H \hh   BB
      Maupassant
      @ H \h   B\B
                                       à Montherlant
      bb H \hh  BB
                                       à Dutourd enfin on peut les citer tous presque
      b..h /HB  b-.b- b...../b   HB   B/B

      tout le monde est misogyne
      @ bb      /BB  b....\b BH
```

La préposition *de* projette un dispositif du type *de X à Y* qui, en français, est parfois complété par *en passant par Z* : la morphosyntaxe annonce la projection d'une série d'éléments (en l'occurrence des noms propres). Outre cette contrainte morphosyntaxique, on note la présence d'une organisation prosodique très marquée, aux trois niveaux :

<sup>16</sup> Nous reviendrons sur cet exemple pour décrire le traitement prosodique, informationnel et attitudinel des trois autres éléments de la liste : *dictionnaire de la bêtise*, *dictionnaire des erreurs politiques*, et *caetera*.

- du cliché mélodique (accent initial de niveau haut – séquence de syllabes atones de niveau haut rabaissé – accent final de niveau bas, éventuellement dynamique Haut-Bas) ;
- du déploiement rythmique : alternance de groupes rythmiques ternaires (Fort-faible-Fort) (regroupés pour les deux premiers), séparés par une prise de souffle (notée @) : @ F-f-F / F-f-F @ F-f-F @ (f)F-f-F / F-f-F ;
- de la modulation particulière de la voix (dynamique, cibles hautes) – mélange des deux précédents.

La scansion rythmique très marquée opère un regroupement en liste qui prend le pas sur le regroupement syntaxique binaire : la locutrice commence en utilisant ce dispositif *de... à...*, mais elle poursuit ensuite la liste en ajoutant simplement de nouveaux éléments prosodiquement intégrés, sans plus tenir compte des contraintes syntaxiques pesant pour une continuation du type *en passant par...*<sup>17</sup>. Ensuite, BG se permet parfois, pour des raisons d'eurythmie, d'omettre la préposition (*Maupassant* au lieu de *à Maupassant*), ou de la réaliser de façon très atténuée, dans *Montherlant*, afin de garder le rythme ternaire.

Plus que la syntaxe, c'est donc ici la prosodie qui aligne les éléments en liste, qui permet de court-circuiter le dispositif syntaxique, et d'amputer une préposition au profit d'une régularité rythmique. La réalisation très atténuée du *à* (*Montherlant*), outre qu'elle est due à une régularisation rythmique, fait apparaître que la « négociation » se règle non seulement au bénéfice de la prosodie (satisfaction des contraintes d'eurythmie), mais également avec des moyens propres à la prosodie. La syntaxe en effet est binaire : un élément est là ou non ; la prosodie, quant à elle, est graduelle et analogique : la préposition est là, mais « en tout petit ». Cette solution ménage la chèvre et le chou – avec les méthodes de la chèvre.

Mais la prosodie interagit également avec d'autres sphères de l'organisation du discours que le niveau syntaxique. Les éléments d'une énumération, on l'a vu, sont prosodiquement plus ou moins marqués, par des contours mélodiques, des patterns rythmiques, ou des modulations de la voix fonctionnant de manière plus ou moins autonome. Même si la notion de « neutralité » est problématique, on peut considérer que certaines listes sont réalisées de manière peu marquée : intonativement, cela va de pair avec une séquence de groupes intonatifs juxtaposés (portant un ton de niveau bas rehaussé ou haut en position d'accent final), souvent regroupés par le dernier d'entre eux, portant un ton plus haut que les précédents, ou alors infra-bas ; un ton

---

<sup>17</sup> Le regroupement par deux des éléments de la liste (*de... à ...*) n'est pas effacé complètement : (i) les allitérations et les rimes groupent les noms par deux (*Molière – Baudelaire* ; *Maupassant – Montherlant*) ; (ii) les respirations forment deux groupes de souffle binaires, différents d'ailleurs des précédents (\*[de Molière à Baudelaire] \* [Maupassant] \* [à Montherlant à Dutourd]). Mais ces « petits effets de phase » ne priment pas sur la mise en liste que réalise la prosodie.

« suffisamment démarqué » (plus haut ou plus bas ; plus long), pour opérer une clôture. Les exemples (4) et (5) illustrent cela :

- (4)      heu    ils ont    été    écrits    par de grands esprits si j'ose dire hein  
 @    BH    ↑ bb    /BB    \bb/BB    \bb HH    b....b    H    bb HB-    b-.....b-    H/H
- qu'il s'agisse      de philosophes  
 b.....b /BB      \b.....b /BB  
                           d'écrivains  
                           \b..b /BB  
                           d'historiens  
                           \b.....b HH  
                           de scientifiques              c'est à peine croyable  
                           b.....b HH                    ↑ H(H)    b    HH    b.b BB-
- (5)      donc le dénigrement d(e) la femme  
 @ bb h..b H bb BB              bb HH  
 n'est pas une question      de niveau social  
 b.....b HH                    b.....b HH  
   ni de religion  
   b.....b BB  
   ni de castE  
   b...b BB-

Ce regroupement intonatif « neutre<sup>18</sup> » des éléments de la liste n'empêche pas un travail au niveau rythmique : alignement temporel des attaques ; « compression » syllabique pour donner une impression de régularité. *De scientifiques* semble prononcé sur un tempo accéléré ; le *e* final de *castE* est prononcé, en partie pour soutenir la détente consonantique, mais aussi pour produire un GI de quatre syllabes et ainsi l'aligner rythmiquement sur les deux précédents ; la réalisation de la préposition *de* avec ou sans élision vocalique varie en fonction du nombre de syllabes du complément : *d(e) niveau social, ni d(e) religion, ni de castE*<sup>19</sup>.

Ces deux énumérations réalisées par RF interviennent dans des interventions initiatives de question : elles se présentent comme une reprise d'information (tirée de l'ouvrage de BG) ; et elles sont formulées sur un ton « affiliatif » : RF énumère pour faire valoir qu'elle connaît les membres, et non pour asserter (l'existence de) chacun d'eux. À côté de ça, les énumérations produites par BG dans ses interventions réactives de réponse paraissent fortement marquées (voir listes 1-2-3 décrites ci-dessus : accent initial, cliché mélodique, etc.). Cette « prosodie marquée » apporte, au contenu référentiel - désignatif des éléments de la liste, un supplément de sens qui se rapporte à deux dimensions : (i) la structure **référentielle/informationnelle** et (ii) l'engagement **attitudinel**.

<sup>18</sup> Dans (4), le dernier contour se distingue par sa hauteur : il regroupe les GI qui précèdent et annonce une continuation réalisée d'ailleurs par un *rush* vers la fin de l'intervention. Dans (5), le dernier GI regroupe les deux précédents et signale en même temps l'achèvement de l'intervention (ton conclusif B-B-).

<sup>19</sup> Voir AUER *et al.* (1999 : 121 i.a.) sur ces aspects.

Dans (1), les éléments de la liste sont gratifiés d'un accent initial de niveau haut sur le déterminant défini (a)<sup>20</sup>. Cet accent initial aurait pu être placé sur toute autre syllabe (non finale) de chaque groupe intonatif, avec des effets interprétatifs très différents (voir b, et dans une moindre mesure c<sup>21</sup>) :

- (a) LES grands chefs religieux, LES hommes de sciences, LES hommes de loi, LES hommes de lettres
- (b) les grands CHEFS religieux, les HOMMES de science, les HOMMES de loi, les HOMMES de lettres
- (c) les grands chefs RELigieux, les hommes de SCIENCes, les hommes de LOI, les hommes de LETTRes

Au niveau référentiel, l'accent initial tel que réalisé dans (a) précise la portée, la raison argumentative de cette énumération : par son phrasé mélodique et accentuel, la locutrice entend faire comprendre que c'est « *l'ensemble* des grands chefs religieux, *l'ensemble* des hommes de sciences, etc. », qui est misogyne, et que cette « tare » n'affecte pas seulement une partie de ceux-ci.

Simultanément à cette interprétation au niveau référentiel, la force avec laquelle sont réalisés les accents initiaux et leur position imprévue<sup>22</sup>, ainsi que le phrasé mélodique, exagéré dans l'amplitude de la variation mélodique et dans la distinctivité des cibles (chaque syllabe a une valeur tonale propre) montrent un haut degré d'**implication** de la locutrice, et manifestent son attitude vis-à-vis de ces catégories d'intellectuels qu'elle « place dans le même panier ». Elle indique que si elle en fait une énumération, c'est parce qu'elle a la même attitude (de dénigrement ?) vis-à-vis d'eux. Ceci concerne le second niveau – attitudinel – auquel les instructions prosodiques peuvent être rapportées. La description de ce processus expérientiel et interprétatif complexe fait l'objet du point suivant.

### 3.1.2. Attitudes manifestées dans les listes

Les contours mélodiques ou les schémas rythmiques particuliers qui caractérisent les listes constituent également une marque de l'engagement des locuteurs, d'un point de vue tant quantitatif que qualitatif (cfr §2.3). La prosodie très marquée de la liste (3) signale avant tout un haut degré d'engagement de la locutrice dans son discours, qui est simultanément et nécessairement perçu par son interlocutrice. En d'autres termes, la prosodie attire l'attention sur le simple fait qu'il se passe quelque chose.

<sup>20</sup> Ce qui, il faut le noter, « enfreint » les règles de l'intono-syntaxe, lesquelles prédisent que les clitiques ne portent préférentiellement pas d'accent.

<sup>21</sup> Pour *science*, *loi* et *lettres* l'accent initial se confond avec l'accent final – sauf à accorder un poids phonétique important au caractère sonore des *e* muets et à la particule *heu* après *loi*.

<sup>22</sup> Voire même agrammaticale.

(3) ils ont oublié à quel point de Molière à Baudelaire Maupassant  
 b.....b /BB \b..b /BB @ H \h BB H \hh BB b @ H \h B\B  
 à Montherlant, à Dutourd enfin on peut les citer tous presque  
 @ b H \h.h BB b..h /HB b-.b- b...../b HB B/B  
 tout le monde est misogyne  
 @ b...b /BB b....\b BH

L'**énergie** investie dans la prosodie de ce passage (proportion importante des syllabe accentuée par rapport aux syllabes atones ; effort pour reproduire le même pattern mélodique sur des « supports » différents ; rimes et allitérations ; etc.) est perçue par l'interlocuteur (l'analyste) comme signe (iconique) de l'investissement de la locutrice, de l'importance qu'elle accorde à ce sur quoi elle attire l'attention. La perception (expérientiation) de cet engagement de la locutrice ne s'intègre pas dans le cadre strict de l'interprétation conceptuelle du contenu de son discours, elle s'intègre comme un éprouvé, déterminé en quantité, dans l'expérience que l'on fait de l'engagement de BG : une certaine ardeur, une certaine vigueur.

En ce qui concerne non plus la quantité de cet engagement mais ses propriétés qualitatives et globalement sa valeur hédonique (positive ou négative), elles ne semblent pas pouvoir être récupérées à partir de « morphèmes intonatifs » ou du rythme utilisé ; à part quelques morphèmes attitudeaux de type « doute » ou « évidence »<sup>23</sup>, la signification conventionnellement attachée aux formes est limitée, et concerne plutôt les aspects structurels du discours (focalisation, continuation, fin).

D'un côté, on perçoit nettement l'attitude que BG manifeste en pointant chaque item de la liste – et qu'elle prend de ce fait à l'égard de leurs référents (dédain – mépris) ; dans l'interprétation-perception, il y a à la fois une certaine unité attitude-nelle, et une détermination suffisante et suffisamment univoque. Mais, d'un autre côté, ces déterminations qualitatives, cette perception attitude-nelle et sa direction hédonique, résultent plutôt d'une élaboration complexe, faite d'inférences tant conceptuelles (déclenchées par la présence de quelque chose de prosodiquement marqué) qu'émotionnelles – expérientielles (comme la perception de la variation d'intensité), élaboration mettant en relation (en phase ou déphasée) contenus discursifs, actuels et issus du déroulement antérieur de l'interaction.

Les déterminations qualitatives de l'attitude de BG sont « suffisantes » *dans le cours du déroulement du discours*, c'est-à-dire notamment en contraste plus ou moins immédiat avec une autre focalisation attitude-nelle différemment déterminée (avant, après) et donc en tenant compte également de la durée de sa manifestation, brève. Mais leur saisie explicite, réflexive, révèle différents glissements, des effets de fondu, de fusion, d'empilements... Rappelons l'exemple : *on m'a dit mais pas du tout faites votre enquête interrogez les jeunes ils ont oublié à quel point de Molière à Baudelaire (...) tout le monde est misogyne et comment nous avons fait pour nous débarrasser au fond de cette opinion péjorative et généralisée*. À y regarder de près,

<sup>23</sup> LACHERET-DUJOUR & BEAUGENDRE (1999) ; voir également *infra* § 4.

l'attitude est floue, au sens où tant la source ou l'origine de l'attitude que l'objet visé par elle, sont flous...

L'attitude adoptée, « dédain-mépris », pourrait aussi être qualifiée d'« indignation », voire, éventuellement, d'« étonnement ». Prise isolément, la perception de l'attitude peut se focaliser de façon fort différente sur tel ou tel élément, alors que, dans le phrasé même du discours, seule une possibilité émerge. Différentes ré-expérientiations de la séquence rappelée ci-dessus exhibent des rapports différents entre les plans discursifs engagés. La « netteté » de la perception attitudinelle *initiale et suffisante* semble ainsi elle-même un effet complexe, résultant des conditions *séquentielles* du traitement des données, autant que de la complexité « interne » des conditions elles-mêmes.

L'attitude présentée par BG sur *de Molière* est, virtuellement, attribuée à la voix citée (*on m'a dit ...*), dont BG théâtralise la véhémence sur *mais pas du tout*. Ici, tout en l'incarnant, BG ne montre pas qu'elle adhère au point de vue présenté par *mais pas du tout*. Par contraste, le dédain-mépris affiché sur la liste *de Molière...* semble pris en charge par BG elle-même. La réalisation effective de l'énoncé pose le problème de savoir à qui attribuer cette attitude, dans la « confluence » prosodique qui entremêle ces voix.

Enfin, la cible même visée par l'attitude exhibée est, elle aussi, relativement imprécise : (les jeunes<sup>i</sup> (ont oublié<sup>j</sup> (à quel point<sup>k</sup> ((de Molière<sup>11</sup> à Baudelaire...<sup>12</sup>) tout le monde<sup>m</sup>) est misogyne<sup>n</sup>)). Le dédain-mépris peut viser : les jeunes<sup>i</sup>, le fait qu'ils aient oublié<sup>j</sup>, qu'ils aient oublié à quel point<sup>k</sup>, ou encore la « brochette-des-auteurs-français-classiques »<sup>11-12-13</sup>... Mais d'une certaine façon, de la même façon que les sources fusionnaient, les cibles aussi sont « en enfilade » les unes des autres, comme un domino japonais : en toucher une entraîne la contagion – dispersion de l'attitude sur les autres<sup>24</sup>.

À la première occurrence de la liste, on ne connaît cependant pas encore le point commun, ce que les jeunes auraient oublié, et que, répétitivement, partageraient les auteurs cités, à savoir la misogynie. BG éprouve du dégoût pour la misogynie qui réunit ces membres dans l'ensemble qu'elle construit ; voilà pourquoi son énumération pourrait attribuer, à chaque élément de la liste, une expression de dégoût-rejet. Il est cependant impossible, après 200 ré-écoutes, d'assurer un souvenir de la première occurrence de ces 3 ou 4 secondes. On n'a là qu'un donné hypothétique. Seul un examen empirique, interrogeant des sujets non informés, qui auraient entendu l'interview jusqu'à ce passage inclus (mais à l'exclusion de la chute *tout le monde est misogyne*), répondant à des questions ciblées et judicieusement ordonnées, pourrait peut-être attester de la contamination de la liste par l'attitude de dégoût de BG.

---

<sup>24</sup> Les quilles du bowling seraient une autre métaphore, tout à fait acceptable ici.

Ce qui, dans l'expérience discursive, se donne comme homogène et unifié émerge vraisemblablement de la convergence complexe de flux attitudinels distincts, visant des cibles distinctes, mais rendus solidaires à la faveur de leur intégration au séquençement de la chaîne verbale.

### 3.1.3. Les listes, donnée expérientielle (vs interprétative-inférentielle)

On ne peut pas réduire la saisie des patterns intonatifs et rythmiques qui accompagnent (constituent) les listes à leurs impacts sur les aspects textuels (syntaxiques, informationnels) et attitudinels (quantitatifs et qualitatifs). En effet, les listes constituent également et avant tout une donnée brute de l'expérience discursive. Elles donnent l'occasion au locuteur de produire, de manière répétée et suivie, plusieurs occurrences d'un « même » contour – pattern exhibant une certaine cohérence. Cela peut déboucher sur un **cliché mélodique ou rythmique local**, mis au premier plan, et donc sur un partage (éphémère, événementiel) d'une même temporalité explicitement « scandée », « accordée » entre les deux interlocuteurs, par la liste. Ceci constitue au moins une première attestation brute des organisations prosodiques comme productrices d'expérience.

D'un point de vue hiérarchique-périodique, la liste instaure un double mouvement temporel, à la fois de progression (les contours projettent leur continuation, jusqu'au dernier élément de la liste, la coda), et d'empilement-accumulation (chaque élément de la liste, même s'il entre dans une série cohérente et plus importante, constitue un nouvel item et, au niveau hiérarchique, un acte à part entière qui active une information nouvellement pertinente (la question est discutée dans SIMON, GROBET & AUCLIN ici-même, §4). À cause de la temporalité particulière ouverte par une liste (impossibilité pour l'interlocuteur d'interrompre la liste, et donc de prendre la parole, « suspension » de la progression du discours), il est nécessaire d'avoir une indication de leur achèvement – rôle que joue la coda. Ce rôle cependant ne doit pas être réduit à celui d'une indication topographique : la coda accomplit, exécute, le passage de l'une à l'autre de ces deux temporalités<sup>25</sup>.

## 3.2. Les changements de tour de parole

Les phénomènes qui nous intéresseront ici sont d'ordre interactionnel : ils ont trait aux réglages et ajustements prosodiques mutuellement accessibles dans les interactions, qui témoignent du degré de synchronisation entre les interactants<sup>26</sup> et des conditions d'établissement de cette synchronisation (ou empathie). Leur appréhension est intimement associée au déclenchement de petites émotions interactionnelles

---

<sup>25</sup> AUCLIN 1996.

<sup>26</sup> Ou « inter-synchronisation » : COSNIER 1994 ; KERBRAT-ORECCHIONI 1990.

(COSNIER 1994), qui se signalent à la fois par le caractère ténu de leurs substrats voco-prosodiques, leur fugacité - fluidité dans le cours de l'échange et, néanmoins, par une certaine « obstination ».

Il faut considérer l'interaction de façon générale comme la mise en contact d'organismes dotés à la fois de caractéristiques partagées (comme une langue commune, notamment) et de caractéristiques plus ou moins singulières, un style, un timbre, un ton : un ethos, entendu comme assemblage de l'ensemble de ces caractéristiques (AUCHLIN 2000).

En vertu de ces prédispositions, les organismes en contact vont occuper des « places » et des « rôles » d'interaction l'un vis-à-vis de l'autre (VION 1992). Il est bien entendu important de décrire ces places et rôles : on est relativement loin ou près (axe horizontal), en haut ou en bas (axe vertical), amical ou hostile, exerçant ou subissant plus ou moins de pouvoir ; il faut également décrire ces places dans leurs effets identitaires (BURGER 1995, 1997), leurs déterminations mutuelles (complémentarité des places) et les désirs de reconnaissance qui en meuvent le jeu (FLAHAULT 1978). Ces effets de places identitaires sont un enjeu essentiel de toute interaction, d'un point de vue interne (*qui je suis pour toi qui tu es pour moi – maintenant*), mais aussi d'un point de vue externe, dans la biographie des personnes : dire que *on a été invité par Pivot* montre que l'image de soi s'alimente des échanges qu'on a eus.

Mais si les places dans l'interaction, les identités respectives, s'établissent et se modifient au fur et à mesure qu'elles se manifestent, on peut convenir que les mécanismes d'établissement régissent les effets de place, et sont en un sens « premiers » par rapport à eux. Dès son seuil toute interaction négocie : est-ce qu'on est prêt à collaborer (empathie d'action) ? est-ce qu'on se convient/ aime les mêmes choses (empathie d'affect) ? est-ce qu'on se comprend (empathie de pensée)<sup>27</sup> ? Et cette négociation se déroule par pas, par cycles de régulation – évaluation – enregistrement par le système de son état présent.

Les changements de tour de parole sont évidemment des endroits sensibles pour l'établissement du degré de synchronisation (et donc pour sa mesure ou son observation). Dans l'analyse des changements de tour de parole présentée ci-dessous, nous montrons une fois de plus que les différents sous-systèmes prosodiques – qu'on peut envisager ici comme synchronisation rythmique, accordage tonal et partage d'un certain niveau d'intensité – peuvent être dissociés dans les niveaux d'empathie ou de désynchronisation qu'ils pointent (indexent). Les locutrices peuvent simultanément faire preuve d'une empathie d'action (un passage de tour de parole fluidement mené est soutenu par une bonne synchronisation rythmique) et d'un désaccord au niveau de

---

<sup>27</sup> « L'empathie désigne le partage synchronique d'états psycho-corporels, c'est-à-dire le fait qu'à un même instant, les partenaires de l'interaction vivent et éprouvent un état semblable. Ainsi peut-il y avoir *empathie de pensée, empathie d'action, empathie d'affects* » (COSNIER 1994 : 86).

la pensée (qui serait inféré sur la base d'un désaccordage du registre tonal utilisé pour enchaîner sur l'interlocutrice précédente), comme dans l'exemple suivant, déjà discuté dans WIDAYANTI (2002) et GROBET & AUCHLIN (2001) :

- (6) RF (...) c'est à peine croyable  
BG mais si ça avait été des imbéciles ça n'avait aucune importance

À l'approche de la fin de son tour de parole, on entend RF se rapprocher, par la conjonction de la baisse d'intensité et de variation de F0 (rétrécissement vers le bas de l'espace tonal), à l'occasion pourtant d'une expression de haut degré (de jugement et d'implication : *à peine croyable*). Simultanément, et dans une attitude inverse, BG reste en arrière, ne partage pas le mouvement de rapprochement vocal initié par RF, adoptant un comportement partiellement désaffiliatif<sup>28</sup>. L'examen du rapport avec le contenu montre que ce que vise BG, c'est la haute valeur d'intérêt accordée par son interlocutrice à un fait qu'elle-même estime trivial. Il s'agit d'un désaccord partiel, portant spécifiquement sur l'attitude de RF et qu'on pourrait situer au niveau de l'empathie de pensée ou des affects.

Ce qui donne, peut-être, à cet enchaînement un air relativement inamical, c'est le caractère tranché de la réaction de BG. Ce caractère tranché s'explique par la conjonction du texte (le *mais* en tête de réplique indique un désaccord ; le raisonnement hypothétique a contrario, voire par l'absurde, recourt à un terme classifiant péjoratif : *imbéciles*) et de la rupture prosodique, par quoi BG accomplit *hic et nunc* un refus de suivre RF – elle ne la suit ni dans la baisse d'intensité ni dans la baisse de F0. On observe là un effet de mise en phase typique, mais complexe, dans la mesure où il fait intervenir données textuelles-sémantiques et données prosodiques, mais ces dernières essentiellement dans leurs *valeurs relatives locales*. Ce ne sont donc pas des données prosodiques intrinsèques, mais des données prosodiques relatives car, ici encore, la détermination de *ce à quoi elles se portent comme relatives* (quelles inférences elles permettent de déclencher – ou de ne pas déclencher) est une affaire de mise en phase, par la re-présentation d'un ensemble suffisant d'indices prosodiques dotés de caractéristiques identiques à celles de l'élément à associer<sup>29</sup>.

Une comparaison rapide avec d'autres changements de tour renforce une impression générale de relatif désaccordage affectif entre BG et RF, de non-mutualité empathique, de divergence quant au ton à adopter. Dans (7), le changement de tour

<sup>28</sup> Il ne s'inscrit pas dans les gabarits, tonal notamment, proposés par la fin du tour de RF ; voir la description dans GROBET & AUCHLIN (2001 : 180 sq.) ; WICHMANN (2000 : 142-143).

<sup>29</sup> Le fonctionnement des données prosodiques tel que nous le décrivons ici partage certaines des propriétés attribuées par AUER (1992 : 30-34) aux indices de contextualisation : leur signification est non référentielle (elle n'existe pas hors contexte) ; elle est contrastive (montrer une rupture par rapport à ce qui précède) avec parfois des éléments de signification inhérents pour certains indices, qu'ils soient naturels (non arbitraires) ou partiellement conventionnalisés.

laisse entendre une prise de souffle de RF sur la finale de *science*, et un départ de RF immédiatement après – il n’y a pas de pause :

- (7) BG alors que ça devrait être objectif, heu, la science @  
RF ➔ donc le dénigrement de la femme n’est pas une question de ...

Le tour de parole de RF, en accord avec son rôle d’intervieweuse, occupe une position initiative d’un échange (quelle que soit la description structurale de l’enchaînement qu’opère *donc*), et il est totalement dans les gabarits du tour précédent. Dans (8), l’enchaînement de RF, qui pose à nouveau une question initiative d’échange, est également rapide :

- (8) BG non de rien euh ni même de civilisation hé [petit rire]  
RF ➔ mais qu’est-ce qui vous a poussé(e)...

Ces deux transitions font globalement apparaître une « vivacité » de RF (qu’on peut mettre sur le compte de son professionnalisme) et une « rigidité » de BG (qui peut-être n’est pas sans lien avec son âge avancé). Au niveau de l’articulation des interventions en échanges, on s’attend théoriquement à ce que les frontières « structurales » entre les interventions d’un même échange (6) soient moins importantes que celles entre échanges adjacents, coordonnés (7-8).

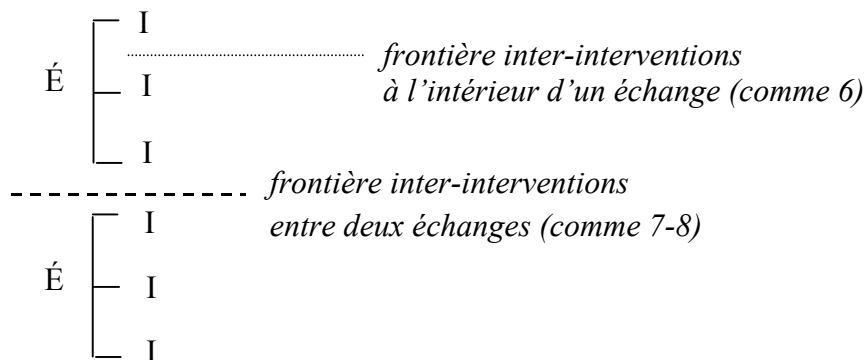


Figure 2. Échanges prototypiques coordonnés, composés chacun de trois interventions (par ex. question / réponse / évaluation)

Or, dans ce corpus, les interruptions de la continuité apparaissent au milieu des échanges, entre deux interventions interdépendantes, et non pas entre deux échanges coordonnés. Par conséquent lorsque, comme dans (7), la frontière de tour correspond à un décrochement ascendant dans la structure (*donc* enchaîne en subordonnant l’échange qui précède), ou qu’elle correspond à un décrochement descendant, ouverture d’un échange subordonné (voir la question de clarification *c’est qui on ?*), on ne sait pas quel aspect est pertinent – lequel serait souligné-renforcé par la promptitude de RF.

Enfin, même quand BG réagit de façon confirmative, elle n’en adopte pas pour autant une attitude affiliative, de plein accord. Par son attaque traînante sur *heu*, par la

montée sur *oui* articulée très en avant produisant un timbre « creaky »<sup>30</sup>, sa confirmation semble conserver une réticence, un relent « concessif » - même si celui-ci ne dispose pas véritablement de cible ou d'objet apparent :

- (9) RF est-ce un petit peu ça qui ressort de votre étude  
 BG (.) heu oui parce que je crois que tout ce qui conforte le pouvoir est utilisé

Globalement, il apparaît donc utile de distinguer entre une empathie (ou une antipathie) d'action (en l'occurrence les locutrice coopèrent assez efficacement pour agir dans le discours) et une empathie d'affects (on note ici un décalage dans les valeurs et intérêts investis dans les propos tenus – décalage « idéologique »). L'exemple (10) est emblématique de la gestion différenciée de ces niveaux, indexée par des indices prosodiques (et syntaxiques) potentiellement divergents :

- (10) RF donc le dénigrement de la femme n'est pas une question de niveau social  
 ni de religion ni de caste  
 BG non de rien euh ni même de civilisation hen [petit rire]

La (relativement) bonne coordination rythmique existant entre les locutrices permet à BG d'enchaîner sur le format de liste proposé par RF et d'intégrer sa réponse, formellement, en reprenant une projection syntaxique disponible dans le contexte<sup>31</sup>. On peut donc dire qu'une empathie d'action existe, qui permet une coordination conversationnelle efficace. Pour autant, l'accord manifesté au niveau rythmique n'est pas reconduit au niveau du registre, ni à celui de la « qualité » de voix (cfr les micro-effets de timbre des segments liés au mode phono-articulatoire évoqués supra, §1 et analysés par d'Alessandro et al., ici-même). Il y a chez RF dans le ton comme une sincérité, une envie de comprendre et d'expliquer, et chez BG un dénigrement de ceux qui dénigrent les femmes (surtout cette particule *hen !* à la fin de sa réplique, prononcée avec un grand et manifeste dédain). C'est cette capacité des données prosodiques (*des prosodies*) à indexer simultanément plusieurs foyers pertinents dans l'interaction en cours (plusieurs *contextes* pour reprendre le terme de Gumperz), à la fois au niveau du contenu verbal mais aussi au niveau de la relation interactionnelle, que nous désignons par la notion de « fractalité ».

<sup>30</sup> Dans les termes de la figure (1) ci-dessus, l'énergie manifestée sur l'axe ergotropique se disperse en bruit, plutôt qu'en signal ; par ce timbre, ce qui est montré c'est l'effort dissipé dans des mouvements antagonistes (« bras de fer ») ; une réaction pleinement confirmative conduirait à une dissipation de la même énergie en signal périodique.

<sup>31</sup> Il n'est par certain qu'il s'agisse à proprement parler d'une coénonciation (au sens de JEANNERET 1999), car l'articulation du segment co-construit par les locutrices est passible d'une double analyse, en intervention (monologique dialogale), ou en échange (dialogique dialogale). La suite de la liste produite par BG est à la fois une continuation de celle entamée par RF, mais son orientation un peu différente (voir le degré d'empathie) permet de l'interpréter comme une réponse négative à une question à orientation négative (soutenant cette interprétation, la présence de la particule *non* – fournissant une réponse, avant de procéder à l'achèvement de la liste).

#### 4. CONCLUSION : *FRACTALITÉS PROSODIQUES*, LE CAS D'UN ACCENT D'INSISTANCE

Nous avons émis (SIMON & AUCHLIN 2001) l'hypothèse du caractère « coopératif » des indicateurs prosodiques, dans leur fonctionnement avec les autres dimensions verbales. Il faut certainement ajouter que les différentes dimensions prosodiques et phono-articulatoires entretiennent entre elles également des relations de type coopératif, qui, elles aussi, peuvent aller du renforcement par redondance, à l'enrichissement par différenciation, mais aussi au blocage par contradiction<sup>32</sup>. Un exemple permettra d'illustrer cette idée :

(11) RF mais<sup>H</sup> qu'est-ce qui vous a poussé(e) à entreprendre....

L'accent initial Haut, qui focalise l'élément interrogatif de l'énoncé de RF<sup>33</sup>, présente des caractéristiques intéressantes :

- c'est une cible haute *relativement haute* (le *mais* est à 298Hz, *qu'est-ce* à 392Hz ; environ 4 demi-tons d'intervalle<sup>34</sup>), à laquelle est associée une valeur-fonction de « reset » des paramètres de déclinaison (voir GROBET & SIMON 2003), signalant par là l'ouverture d'une nouvelle macro-unité (GROBET, SIMON & AUCHLIN ici-même) ;
- plus une cible est haute, plus elle demande proportionnellement d'effort ; cependant, la réalisation de cette cible-là [kes] est perçue comme engageant relativement peu d'énergie ;
- dans l'occlusive et le timbre du segment [kes], la voix ne fait pas sentir l'effort constrictif qui accompagne les cibles hautes (par comparaison avec les segments environnants) ; ces caractéristiques de timbre « coopèrent » avec la valeur de F0, et la valeur d'intensité, pour indiquer les propriétés de cette cible-là : atteindre une cible haute sans effort est un symptôme d'enthousiasme, de détente, d'engagement sans défense ; l'atteindre sans

---

<sup>32</sup> Comme le sens commun le saisit très bien : une intonation de « fausse surprise », par exemple – l'intuition est claire, ce qui ne l'est pas ce sont les mécanismes par lesquels elle s'établit, comme nous essayons de le montrer plus haut § 3.1.3.

<sup>33</sup> Conformément aux observations de WUNDERLI (1992).

<sup>34</sup> La montée est perçue comme abrupte, bien supérieure aux 4 demi-tons détectés par l'analyse de F0 opérée par Praat ; on perçoit une ambivalence, comme s'il y avait un saut d'octave en plus. Ce doute est lié à un possible changement de mode de phonation de la part de RF sur ces deux cibles, comme si elle passait d'une voix de poitrine à une voix de tête (voir D'ALESSANDRO *et al.* ici-même), ce que peut exiger une montée de 4 demi-tons si elle se situe près du plafond du mode de phonation en cours ; de semblables variations, particulièrement difficiles à analyser, semblent illustrer un effet de renforcement fractal – ni le timbre, ni F0, pris isolément, ne confirment l'observation d'une montée importante ; mais pris ensemble différentes de leurs caractéristiques négligeables par ailleurs produisent un effet de montée.

accroissement de la valeur de dB est un symptôme de proximité corporelle dans l'interaction.

Dans la réalité du signal, ces différents ingrédients sont, évidemment, étroitement entremêlés, comme ils le sont avec la réalité linguistique ; ce que nous en « percevons » est une réalité complexe non décomposée, plus ou moins homogène, plus ou moins « identifiable ». Mais comme le fait remarquer WICHMANN (2000 : 144-148) si l'on cherche à décrire cette réalité comme « achevée » ou comme « figée », on aboutit à une profusion d'étiquettes distinguant, au gré des compositions entre dimensions, différents types d'attitudes, profusion qui fait écran à notre compréhension de ces phénomènes.

On peut opportunément faire un parallélisme avec la position développée il y a déjà quelques années par Scherer et son équipe sur les émotions<sup>35</sup> : ce qui faisait (ou *fait* ?) écran à une compréhension des processus émotionnels, c'est la catégorisation a priori de celles-ci en « chunks » discrets – on peut parler de vision « sémiotique » des émotions<sup>36</sup>. Dans cette vision, « une » émotion est un état d'un organisme, qui peut être clairement caractérisé et identifié, auquel correspond un ensemble de traits partagés par ses manifestations. En dépit de l'intérêt de travaux tels que ceux de LÉON (1993) par exemple, qui l'ont développée, une telle approche sémiotique n'est pas apte à rendre compte ni de l'intrication des aspects (composantes) de la réponse émotionnelle, ni de sa fluidité, ni – surtout – de la relation par quoi la parole manifeste (ou trahit) des états émotionnels et, par là et à cause de ça, les contrôle : les suscite, les évite, les partage.

La conception des émotions par processus composants (SCHERER) fournit un modèle pour concevoir les rapports entre les indicateurs prosodiques et l'émergence d'attributions d'attitudes. Élaboré dans une certaine isolation – idéalisation des données concrètes de l'interaction langagière, et des conditions d'émergence - dissipation des émotions, ce modèle pourrait grandement contribuer à une saisie du fonctionnement des indicateurs prosodiques et attitudinels dans les attributions d'attitudes, à condition que son domaine d'application soit élargi et que le mécanisme descriptif rende compte de cet élargissement :

- si des phénomènes constituants d'un niveau inférieur (timbre, durée, intensité) peuvent renforcer ou contredire des phénomènes émergents d'un niveau supérieur, on peut métaphoriquement dire que les caractéristiques des niveaux inférieurs présentent des associations fractales avec celles des niveaux supérieurs ;

---

<sup>35</sup> Voir entre autres SCHERER (1989).

<sup>36</sup> Soutenue par exemple dans les travaux phonostylistique de P. LÉON, cette conception des émotions comme entités discrètes, objet de taxinomie populaire, est toujours largement en vigueur, si l'on examine la façon dont LACHERET-DUJOUR & BEAUGENDRE (1999) rendent compte des fonctions « ecto-linguistiques » de la prosodie.

- si tel est le cas, une étude détaillée du fonctionnement d'un niveau qui omettrait l'étude des processus inférieurs le composant devrait être réputée inutile, tant les résultats établis à un niveau sont soumis aux interactions, convergences et divergences, entre processus composants.

Autrement dit, les ensembles de caractéristiques prosodiques interagissent les uns avec les autres à la fois « horizontalement », entre ensembles distincts mais de même niveau de fonctionnement, et co-présents (accentuation – durée - hauteur), et « verticalement », ou « fractalement », dans un ensemble de caractéristiques, entre phénomènes constituants et phénomènes émergents.

Antoine AUCHLIN  
& Anne Catherine SIMON\*  
Université de Genève  
Département de Linguistique  
CH-1211 Genève 4  
antoine.auchlin@lettres.unige.ch  
\*Chargée de recherche du FNRS  
Université de Louvain, VALIBEL  
Place Blaise Pascal, 1  
B-1348 Louvain-la-Neuve  
simon@rom.ucl.ac.be

## RÉFÉRENCES BIBLIOGRAPHIQUES

- AUBERGÉ, V. & L. LEMAÎTRE. 2000. « The Prosody of Smile », paper presented at ISCA Workshop on Speech and Emotion, Newcastle.
- AUCHLIN, A. 1996. « La chute de la narration et la participation au discours », in M. LAFOREST (éd.), *Autour de la narration*, Québec : Nuit Blanche éditeur, 47-71.
- AUCHLIN, A. 2000. « Ethos et expérience du discours : quelques remarques », in M. WAUTHION & A.C. SIMON (éds), *Politesse et idéologie. Rencontres de pragmatique et de rhétorique conversationnelles*, Louvain : Peeters, BCILL n° 107.
- AUER, P., E. COUPER-KUHLEN & A. DI LUZIO. 1990. « Isochrony and 'uncomfortable moments' in conversation », in M.A.K. HALLIDAY, J. GIBBONS *et al.* (eds), *Learning, keeping and using language*, vol. 2, Amsterdam-Philadelphia : John Benjamin, 269-281.

- AUER, P., E. COUPER-KUHLEN & F. MÜLLER. 1999. *Language in time. The rhythm and tempo of spoken interaction*, New York-Oxford : Oxford University Press (chap. 6 « Rhythm in turn construction. Scansions in Italian conversations »).
- BLANCHE-BENVENISTE, C. 1990. *Le français parlé. Études grammaticales*, Paris : Éditions du CNRS.
- BOLINGER, D. 1986. *Intonation and its parts, melody in spoken English*, Stanford, Cal. : Stanford University Press.
- BURGER, M. 1995. « L'identité négociée : rapports de place(s) dans un entretien télédiffusé », *Cahiers de linguistique française* 17, 9-33.
- BURGER, M. 1997. « Positions d'interaction : une approche modulaire », *Cahiers de linguistique française* 19, 11-46.
- COSNIER, J. 1994. *Psychologie des émotions et des sentiments*, Paris : Retz.
- DI CRISTO, A. 1999. « Vers une modélisation de l'accentuation en français (seconde partie) », *Journal of French Language Studies* 10/1, 27-45.
- ERICKSON, F. 1992. « They know all the lines », in P. AUER & A. DI LUZIO (eds), *The contextualization of language*, Amsterdam-Philadelphia : John Benjamin's, 365-397.
- FLAHAULT, F. 1978. *La parole intermédiaire*, Paris : Le Seuil.
- FÓNAGY, I. 1983. *La vive voix, essais de psycho-phonétique*, Paris : Payot.
- GROBET, A. & A. AUCHLIN. 2001. « À l'attaque ! Vers une typologie des différentes prises d'élan du discours », *Cahiers de linguistique française* 23, 165-187.
- GROBET, A. & A.C. SIMON. 2003. « Réinitialisations (*resetting*) et unités prosodiques maximales : une évidence ? », in V. AUBERGÉ, A. LACHERET & H. LOEVENBRUCK (éds), *Journées prosodie 2001*, Presses Universitaires de Grenoble, 79-84.
- JEANNERET, T. 1999. *La coénonciation en français. Approche discursive, conversationnelle et syntaxique*, Berne : Peter Lang.
- KERBRAT-ORECCHIONI, C. 1990-1994. *Les interactions verbales*. Paris : A. Colin.
- LACHERET-DUJOUR, A. & F. BEAUGENDRE. 1999. *La prosodie du français*, Paris : Éditions du CNRS.
- LÉON, P. 1993. *Précis de phonostylistique*, Paris : Nathan.
- MENAHÉM, R. (1989 [1983]). « La voix et la communication des affects », in B. RIMÉ & K. SCHERER (éds), *Les émotions*, Neuchâtel : Delachaux et Niestlé, 203-230.

- MERTENS, P. 1987. *L'intonation du français. De la description linguistique à la reconnaissance automatique*. Doctorale dissertatie : KULeuven.
- NÚÑEZ, R.E. 1997. « Eating soup with chopsticks: dogmas, difficulties and alternatives in the study of conscious experience », *Journal of Consciousness Studies* 4/2, 143-166.
- OGDEN, R. 2001. « We speak prosodies and we listen to them », Symposium on Prosody and Interaction, Uppsala, November 10-11, 2001.
- SCHERER, K. 1989. « Les émotions : fonctions et composantes », in B. RIMÉ & K. SCHERER (éd.), *Les émotions*, Neuchâtel : Delachaux et Niestlé, 97-134.
- SIMON, A.C. & A. AUCHLIN. 2001. « Multimodal, multifocal ? Les hors-phase de la prosodie », in CH. CAVÉ, I. GUAÏTELLA & S. SANTI (éd.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication*, Actes du colloque ORAGE 2001, Aix-en-Provence, 18-22 juin 2001, Paris : L'Harmattan, 629-633.
- SIMON, A.C., A. GROBET & A. AUCHLIN. 2003. « *Le discours, scène globale? Constituance prosodique et (ir)régularités structurelles* », dans ce volume.
- SPERBER, D. & D. WILSON. 1989. *La pertinence. Communication et cognition*, Paris : Minuit.
- VION, R. 1992. *La communication verbale. Analyse des interactions*, Paris : Hachette.
- WIDAYANTI D. V. 2002. « Étude comparative et expérientielle : les émotions des interactants dans deux entretiens radiophoniques ». *Mémoire de DEA*, Université de Genève.
- WICHMANN, A. 2000. *Intonation in Text and Discourse. Beginnings, middles and ends*, Harlow : Longman.
- WUNDERLI P. 1992. « Interrogation et 'accent d'insistance' », in PH. MARTIN (éd.), *Mélanges Léon*, Toronto : Mélodie, 569-583.