

Du phonostyle à l'éthos, les prosodies comme interfaces entre sens et corps

Pour repenser l'acte de langage à partir de l'activité de parole¹

Antoine Auchlin, Université de Genève

0. Introduction

Les ressources auxquelles puise l'éthos sont nombreuses et variées : l'éthos fait feu de tout bois, pour emprunter un mot de C. Kerbrat-Orecchioni, le Symposium l'a amplement démontré. Parmi ces ressources, dans les manifestations orales du discours, la prosodie joue sans doute un rôle décisif. Kerbrat-Orecchioni & De Chanay (2007 : 311) soulignent l'« importance du rôle que joue dans la construction de l'éthos le matériau voco-prosodique et mimo-gestuel ». De très simples manipulations de la hauteur de quelques cibles tonales², sur lesquelles se construit la mélodie et le *ton* du discours, peuvent modifier de fond en comble l'appréhension que l'on a de ce que le locuteur est en train de faire. Ainsi par exemple, le discours des candidats aux élections présidentielles françaises S. Royal et N. Sarkozy peut-il être transformé afin de donner à la première un ton convaincu et présent à ce qu'elle dit, en augmentant certains écarts mélodiques, et en ajoutant des « accents initiaux » par l'élévation, de quelques demi-tons, de la syllabe initiale de mots ou de syntagmes. Ou d'altérer en ton plaintif le ton assertif et déterminé du second, en réduisant globalement les écarts mélodiques importants en intervalles mineurs³.

On ne fait là que des modifications phonostylistiques, dont aucune ne change le contenu présenté (ce qui est *dit*) ni l'interprétation à laquelle on parvient ; elles ne sont pas linguistiquement ou interprétativement pertinentes. Mais elles interviennent dans la perception, l'appréhension que l'on construit, de la personne qui parle, de la relation qu'elle entretient à la circonstance de parole, à ce qu'elle dit, et à l'auditoire à qui elle le dit. Autrement dit on modifie l'éthos même des locuteurs.

Mon propos ici n'est pas tant de distinguer l'éthos du phonostyle (voir cependant §2.2.) que de discuter ce qu'implique, pour l'analyse du discours et pour la recherche en prosodie, la prise en considération de phénomènes complexes, évanescents, « subjectifs », qui plus est mal définis, comme l'éthos et le phonostyle.

¹ Ce texte est une version remaniée de la présentation faite au Symposium ; l'expérience directe des exemples audio-visuels n'est pas disponible dans le texte et a dû être remplacée par quelques détours.

² Manipulations « artisanales », sous Praat, de *F0 uniquement* ; ajouter des manipulations de *durée* accroît considérablement ces effets.

³ Passer de huit demi-tons ou plus à un ou deux.

0.1. *Prosodie et expérience discursive*

La thèse générale est que le donné prosodique du discours détermine et s'intègre dans l'*expérience discursive*, expérience temporalisée de l'occurrence de segments linguistiques en séquence. Cette thèse concerne autant la façon de considérer le donné phono-prosodique, que la définition même du « discours », entendu ici comme expérience sensible temporellement organisée autour du *maintenant* de son occurrence, selon l'état du déploiement des structures linguistiques en cours, (micro-)syntaxiques, et/ou textuelles-discursives⁴. Un tel cadre me semble requis pour traiter de phénomènes pragmatiques émergents, comme le sont, chacun à sa manière, l'éthos, et le phonostyle.

0.2. *Prosodie et « overt communication »*

Une part des manifestations prosodiques peut être couplée à la signification linguistique ou à l'interprétation (Wilson & Wharton 2005) ; mais cette part-là, aussi large et sémiotiquement variée que la conçoivent Wilson & Wharton, ne rend compte ni la complexité de l'élaboration voco-prosodique, ni de celle de ses rapports avec le déploiement linguistique.

De même, considérer les manifestations voco-prosodiques sous l'angle de leurs propriétés de signes et/ou signaux (Wilson & Wharton 2005), ou en tant que *porteurs d'information*, c'est, me semble-t-il, ignorer une évidence : dans le complexe des sous-systèmes prosodiques dont le jeu coopératif (Simon & Auchlin 2001) produit le perçu prosodique homogène du discours, certaines dimensions déterminent *directement* l'expérience du discours, sans médiatisation par quelque processus sémiotique ou inférentiel d'aucune sorte. C'est le cas du *débit* de parole, et c'est le cas de l'*énergie* produite/perçue. Dans ces deux cas, la communication implique et contient une co-expérienciation, synchronisation mutuelle de deux expériences.

Dire ce partage d'expérience *non-médiat* ne revient évidemment pas à dire que ces dimensions échappent à tout *contrôle*. Elles sont à la fois socio-culturellement déterminées, et régulées par adaptation tant aux besoins du parleur (auto-régulation, ou boucle de régulation courte), qu'à ceux, dans une mesure variable, de l'auditeur (hétéro-régulation, boucle longue) ; mais ce qui intervient, cependant, dans ces deux boucles de régulation, c'est l'expérience directe et immédiate du discours.

0.3. « *Overt communication* » et communication « globale »

De sorte que si, après Wilson & Wharton, on admet l'existence d'un continuum entre la significationⁿⁿ (au sens de Grice) et la monstration – ostension⁵ *fig.a.* ci-dessous), ce continuum doit être prolongé, en deçà de *showing*, et inclure, également, ce qui est *présenté* par une

⁴ V. Auchlin (1998 ; 2000a, i.a.).

⁵ « Relevance Theorists (...) have consistently argued that there is a continuum of cases between showing and meaning^{NN}, all of which may fall within the domain of pragmatics and contribute to a speaker's meaning (Sperber and Wilson, 1986/1995, chapter 1, section 10). » (Wilson & Wharton 2006 : 1564).

séquence orale. Ce qui est *présenté* n'est pas ce que Wilson & Wharton nomment des « natural signs », symptômes, informations qui ne sont pas intrinsèquement communicatives, « non intentionnellement rendues manifestes », qui font partie de la « communication ouverte », et donnent lieu, de la part de l'auditrice, à des inférences (Vs. un *décodage*, comme les signaux, ou les signes linguistiques). Les « signes naturels », dans le sens ci-dessus, sont, en tant que tels, des manifestations prosodiques *présentées* et interprétées comme *renvoyant à autre chose* (ainsi le tremblement de la voix *saisi comme* symptôme de peur). Mais cette intégration dans une fonction de symptôme ne retient qu'un sous-ensemble du donné prosodique présenté ; un débit relativement rapide, une intensité relativement forte, font partie de ce qui est présenté, mais ne donnent lieu à aucune inférence, et ne sont donc signes naturels de rien, sinon d'eux-mêmes.

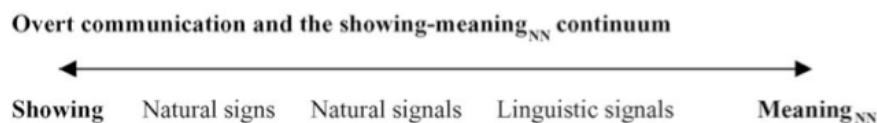


Fig. 2. Varieties of overt communication.

Fig. a. Le continuum montrer-signifier non-naturellement (Wilson & Wharton 2006 : 1564)

Or ces données font partie de ce que *présente* le signal acoustique, tout au moins de ce qu'on en reconstruit perceptivement. Le *grain de la voix*, son *timbre*, ne sont pas *a priori* objets d'inférences, sauf déviation par rapport à une attente (*Mary est enrrouée*) ; mais, comme le tempo et l'intensité, il sont *présentés*, donnés à percevoir, et cette perception détermine des pans décisifs de l'expérience du discours oral (*ça va vite ; c'est trop fort ; j'aime sa voix...*) voire, comme le fait apparaître *a contrario* la synthèse de la parole, ce qui fait le « naturel » de la parole, et qui alimente l'expérience spécifique d'une interaction avec un humain vs une machine. Il semble donc non seulement légitime, mais également nécessaire, de prendre en considération le donné prosodique de façon globale : les manifestations *médiatisées*, par codage, linguistique ou iconique, ou par inférences impliquant une connaissance encyclopédique (symptômes), font partie, mais ne sont qu'un sous-ensemble, des manifestations *présentées*, effectives, qui déterminent les différentes *facettes* de l'expérience du discours.

De plus, une partie tout au moins des manifestations *interprétables* comme symptômes ne réduisent pas leur activité communicative à celle de figurer comme symptôme éventuellement pertinent pour l'interprétation. Par exemple, un timbre « dégoûté » peut être produit par une articulation fortement antériorisée. Ce timbre est la conséquence mécanique de l'avancement et du réhaussement de la partie postérieure de la langue, et de la constriction du larynx, qui sont une conséquence réflexe du refus d'ingérer (même imaginativement) quelque chose de déplaisant. A côté du fonctionnement du résultat acoustique comme symptôme de dégoût, voire peut-être même dans le but de procéder à cette identification, l'auditeur effectue une copie

motrice, plus ou moins inhibée, de la posture et du mode phonatoire susceptibles d'avoir produit ce timbre, procédant à une sorte d' « analyse spectrale motrice », ou incarnée.⁶

Une étude empirique réaliste de la communication ne peut pas légitimement être restreinte à la communication ouverte, mais doit s'étendre à la communication « globale » (*fig. b*) :



Fig. b. Communication « globale », continuum présenter-montrer-signifierⁿⁿ

Si l'on envisage la *communication* comme un *processus de co-expérimentation*, le champ des données prosodiques prises en compte est considérablement plus vaste, et, du point de vue d'une appréhension d'ensemble du fonctionnement voco-prosodique dans l'usage du langage, ne souffre pas d'une réduction liminaire de ses données⁷. Envisager les manifestations prosodiques dans leurs « fonctions » expérientielles revient à en proposer une description *émique*, et non plus *étique*⁸, une description fonctionnelle, et non plus « naturaliste ». Une proposition éactive consiste à poser que la fonction globale des prosodies est d'énacter l'expérience du sens, i.e. accomplir l'intégration du sens comme un fait corporel et sensible (ce que la prosodie fait apparaître du genre d'activité verbale en cours, du type de contact engagé, de la nature de la relation, de l'appartenance du locuteur à un groupe social, etc.).

1. Perception, illusion, réalités de premier / deuxième ordre : intégration expérientielle

Il ne faut pas comprendre la « perception » comme un processus d' « extraction de traits » de l'environnement (sens commun), mais plutôt, dans le sens de l'énaction (Varela & al. 1993 ; Bottineau 2007) comme une élaboration interne à l'organisme, partiellement déterminée par les propriétés de l'environnement, par les propriétés de notre dotation biologique⁹, et par l'environnement culturel, l'histoire des *couplages structurels* de l'organisme avec l'environnement et avec ses pairs. Indépendamment de ses aspects éactifs sur lesquels je reviens plus bas, cette élaboration perceptive *lie* les unes avec les autres différentes propriétés de l'environnement (*binding*, au sens de Bache 2005), de façon telle que ces propriétés ne sont pas dissociables

⁶ Expression empruntée à Auchlin & Simon 2008 ; voir Skipper & al. 2006, et n.14 ci-dessous. Des modifications de timbre vocalique « artificiellement » produites par abaissement / élévation des deux premiers formants entraînent chez l'auditeur un très manifeste effet d'imitation posturale, comme le (dé)montraient les exemples audio présentés.

⁷ C'est dans cette direction que devrait être examiné le fonctionnement des « codes biologiques » (Gussenhoven 2002), et non seulement, comme le suggère Hirschberg (2002), en les considérant dans le cadre de la communication ostensive (« overt communication »).

⁸ Voir Di Cristo & al (2004) et l'idée de 'grammaire écologique'. La phonologie intonative (Ladd 1996) est une description émique des éléments intonatifs pertinents pour l'interprétation – et non pour l'expérimentation.

⁹ Par exemple nous pourrions, comme les chauve-souris, percevoir des sons de 110kHz ; que serait le « silence » de la place déserte, la nuit, quand les chauve-souris sortent en bandes piaillantes pour s'alimenter...

perceptivement¹⁰. L'intégration perceptive, *binding*, est considéré par Bache comme le niveau fondamental des processus de *blending* conceptuel (Fauconnier & Turner 2002 i.a.), selon une gradation de complexité cognitive. Sommairement, on peut dire que l'énonciation décrit la dynamique des processus perceptifs - cognitifs engagés dans l'activité de parole, et que l'intégration cognitive (*blending theory*) saisit leurs aspects structurels - fût-ce dans la dynamique des blends, qui est de comporter des éléments « évanescents », constructions *temporaires* ; et dans la réversibilité (partielle) des assemblages, *dis-integration*, avancée par Bache (2005).

1.1. « Nos sens nous trompent »

Une illusion visuelle (*fig.c*) permet d'illustrer simplement mon propos :

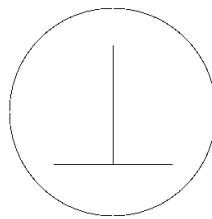


Fig. c. Divergence entre accès perceptif et accès métrique

Cette figure appelle plusieurs constats :

- a. nous *percevons* (voyons) la ligne verticale comme sensiblement plus longue que la ligne horizontale, alors que
- b. les « mesures objectives » (métriques) indiquent que ces lignes sont de même longueur ;
- c : a. et b. sont contradictoires et pourtant simultanément vrais ;
- d. l'évidence *perceptive* résiste à l'évidence *métrique* : sauf à déformer la figure, on a beau savoir que les droites sont de même longueur, on ne parvient pas à les (perce)voir comme ça ;
- e. enfin, tout aussi remarquable, notre perception ne rend pas compte des éléments impliqués dans l'illusion : la position de la figure dans le cercle, ou celle des droites l'une par rapport à l'autre : perçues, mais pas dans le rôle qu'elles jouent dans la perception des droites.

Watzlawick (1980 : 48) nomme *réalité de premier ordre* celle que l'on mesure, celle où les deux droites ont la même longueur ; et *réalité de deuxième ordre* la figuration interne que nous en élaborons; l'illusion est un cas paradigmatique de *diffraction* entre les deux. Cela permet de comprendre facilement en quoi, et pourquoi, les deux accès sont également nécessaires, complémentaires. La prosodie de la parole comme nous la percevons est évidemment une réalité de deuxième ordre ; son étude outillée, de plus en plus accessible et à certains égards facile (*Praat*, Boersma & Weeninck, en ligne), donne un accès de plus en plus clair à la réalité de

¹⁰ Comme les différents formants d'un timbre vocalique ne sont pas *séparables* les uns des autres par l'oreille humaine, même si leur composition est jusqu'à un certain point imitable (mais voir certaines difficultés dans l'acquisition de langues secondes).

premier ordre, celle du signal acoustique comme séquence alternant sons harmoniques (« voyelles ») et non harmoniques (« consonnes sourdes ») et leurs différentes variations¹¹.

Ce qui nous intéresse, comme dans le cas de l'illusion, est à la jonction des deux ordres de réalité ; une grande partie des questions de prosodie outillée concerne cette articulation entre réalités de premier et de deuxième ordre. Des résultats, il faut attendre qu'ils nous informent sur notre perception (et ses biais) de la prosodie, telle qu'intégrée – fusionnée – dans le traitement (interprétation-expérienciation) du discours, autant que sur la substance acoustique « de premier ordre » qui l'alimente. L'étude du phonostyle de France-Info (§2.2.) en est un cas de figure, où se complètent et s'éclairent mutuellement description perceptive intuitive et description « outillée », automatisée¹².

1.2. *Binding multi-modal*¹³ et illusion auditive

L'illusion auditive connue comme « effet McGurck-McDonald » (Mc Gurck & Mc Donald 1976)¹⁴ est un cas intéressant où l'intégration perceptive résulte de la fusion de deux sources (ou modalités) perceptives distinctes, visuelle et auditive. L'expérience est très simple, un visage filmé « prononce » les syllabes *ga-ga* trois fois de suite ; mais la bande son présente les syllabes [ba-ba]. La grande majorité des auditeurs entendent [da-da], qui n'est ni dans le signal visuel ni dans le signal acoustique. On peut schématiser cela ainsi (*fig.d*) :

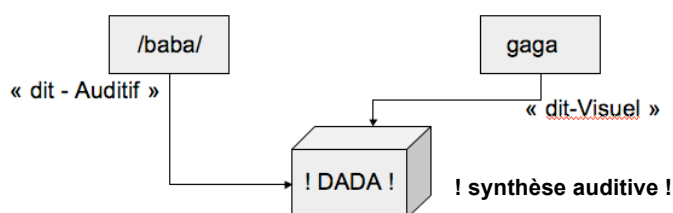


Fig. d. « L'effet Mc Gurck », représentation schématique

Certes, il s'agit là d'un cas de figure particulier ; il n'en est pas moins exemplaire de la condition de la perception auditive de la prosodie, comme réalité de second ordre. Avant d'illustrer la façon dont les différents « sous-systèmes » prosodiques interagissent les uns avec les autres et avec la

¹¹ Même si, comme le sait tout utilisateur de logiciel d'analyse du signal, Praat par exemple, l'accès métrique à la réalité de premier ordre exige des choix : il faut déterminer ce que l'on veut « voir », et sous quelles conditions (v.§ 2.3.2).

¹² Une autre étude, qui ne sera pas abordée ici, concerne le traitement des proéminences dans la parole, et met vis-à-vis plusieurs systèmes de détection automatique des proéminences et l'annotation humaine (Avanzi & al. 2008).

¹³ « Cross-modal binding » chez Bache.

¹⁴ Documentation accessible en ligne à

http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large. Skipper & al. (2006) proposent une analyse détaillée des flux nerveux associés aux deux modalités, visuelle et auditive, dans le traitement de l'effet Mc Gurck. Leurs observations font état d'une pré-éminence de la modalité visuelle (via l'activation de « forward models » et de systèmes de « neurones miroirs ») dans le déclenchement de processus d'imitation motrice chez l'auditeur (voir l'échoïstation corporelle, Cosnier 2003).

chaîne verbale, j'aimerais présenter ce qui me semble être une dimension structurelle de l'expérientiation discursive.

1.3. *Intégration expérientielle et seuil qualitatif, la « spatialisation » de l'expérience discursive*

La symphonie que j'écoute avec mes écouteurs en stéréo est pratiquement la même dans chaque écouteur¹⁵ ; si je passe l'enregistrement en mono, la symphonie est encore la même, mais la perception *spatiale* a disparu. Notre perception auditive traduit les micro-différences entre les canaux droit et gauche en une perception spatiale, et l'expérience auditive franchit un seuil qualitatif, le son étant perçu dans un espace, et non plus sur un plan (mono)¹⁶.

De façon à peine métaphorique, un tel seuil qualitatif me semble en jeu dans l'expérience ordinaire du discours. Nous ne traitons pas la parole comme un flux unifié mais comme un faisceau de flux distincts ; leur co-occurrence, leur déploiement « parallèle », se traduit en l'élaboration d'un espace, et en son *parcours*, par une projection corporelle, qui le dote d'une *texture* et d'une *consistance*.

On suppose plus précisément (après Simon & Auchlin 2001) que le donné prosodique résulte du jeu coopératif de différents sous-systèmes prosodiques plus ou moins autonomes¹⁷, qui entretiennent des rapports de redondance, de contrariété, ou de complémentarité, entre eux, et dans leurs articulations avec les différentes dimensions de l'activité de parole en cours (Auchlin, Filliettaz, Grobet, Simon 2004), telles que les décompose l'approche modulaire de Roulet & al. (2001). Les sous-systèmes prosodiques entretiennent des *effets de phase* sélectivement avec telle ou telle dimension (« module » ou « forme d'organisation », dans la terminologie de Roulet & al.) de l'interaction en cours. Ces effets de phase entre facettes prosodiques et dimensions du discours reposent sur la synchronisation de deux ou plus systèmes de gestion de formes¹⁸ (linguistique-textuel ; prosodiques : ventilatoire, phonatoire, articulatoire, et les sous systèmes formels résultants, rythmique, tonal-intonationel, etc.), et c'est, selon nous, le mécanisme de leurs (dés-)ajustements successifs qui est à l'origine des effets de texture et consistance (voire de *moiré*), à grain fin, et de mise perspective spatiale.

Gardons encore à l'esprit que, via les prosodies, de nombreuses informations relatives au parleur, à son état motivationnel, vis-à-vis du fait de parler (à X) ou du contenu à évoquer (devant

¹⁵ Pas tout à fait identiques, mais les différences entre elles ne sont pas perceptibles comme différences.

¹⁶ Il en va de même de la vision binoculaire : voir Bateson (1984 : 75sq.).

¹⁷ Suivant Firth. Voir aussi Ladd & al (1985).

¹⁸ Parler (ci-dessus) de « co-occurrence » est évidemment une simplification des « chronologies » propres aux différentes dimensions en jeu : dès le niveau, plus ou moins évident pour les prosodiciens, de la syllabe, son éventuelle attaque consonantique et la durée des segments, la montée dynamique, l'amorçage phonatoire du noyau, sont autant de micro-événements « co-occurents » avec « la syllabe ». Il n'en reste pas moins que, moyennant différentes délimitations, partiellement données par la parole, dans les différents flux, tout effet observable procède d'une convergence temporelle entre deux (ou plus) sous-systèmes, prosodiques et linguistiques.

X), sont rendues accessibles, sensibles, bien *avant* la complétion ou l'intégration d'unités linguistiques, pour *chacune* des unités linguistiques successives¹⁹. En cela, les prosodies invitent à re-considérer le sens linguistique à l'aune de l'expérience temporelle qu'on en a : aussi bien en amont de sa formation, que dans la durée de son élaboration. Sous cet angle, le déploiement des efforts voco-prosodiques est, essentiellement, consacré à faire émerger une expérience de sens partagée, bien d'avantage qu'à « envelopper » (borner, hiérarchiser) du sens déjà configuré.²⁰

Ces supposées propriétés de l'expérience discursive (granularité, texture, consistance, volume et perspective) sont évidemment difficiles à « prouver » ou à attester empiriquement, sans recourir au donné intérieur de l'expérimenteur. De plus, même à accepter ce donné d'expérience, ces propriétés y relèvent de *l'évident*, du *naturel*. À ce titre elles sont en quelque sorte *transparentes*. On peut, au moins, voir une preuve *a contrario* dans l'état de l'art en synthèse de la parole. On l'a dit, son problème majeur consiste à donner « du naturel » à la sortie vocale synthétique. Mais la complexité, technique et théorique, du problème, entraîne l'industrie et les ingénieurs vers un compromis qui consiste à reproduire, comme « synthèse », les segments les plus longs que possible de bases de données de parole authentique d'autant plus vastes ; autrement dit, à utiliser « du naturel » pour fabriquer du naturel,²¹ montrant le relatif échec des techniques de synthèse à reproduire cette chose pourtant *si évidente*.

2. Éthos, phonostyle : énonciation, blends expérientiels, co-expérientiation

2.1. Prosodie et énonciation du sens, deux exemples

Pour illustrer la perspective énonciative sur l'intégration discursive de la prosodie, j'aimerais prendre un premier exemple : la lecture à haute voix par trois locutrices d'un passage du *Petit Prince* (Saint-Exupéry)²², qui produisent, par leur intonation, deux réalisations différentes d'un énoncé du texte.

L'énoncé est : « On ne connaît que les choses que l'on apprivoise, dit le Renard. » L'une des réalisations (deux lectures) pose une cible haute montante, continuative, sur la dernière syllabe de *apprivoise* (qui est aussi la fin du discours rapporté du Renard) et une cible basse, terminale, sur celle de *Renard*. L'autre réalisation pose une cible terminale sur *apprivoise*, la fin du

¹⁹ Et que la reconstruction perceptive, par l'auditeur, de l'état motivationnel du locuteur exerce un véritable effet de « boot-strapping » sur sa reconstruction des intentions, communicative et informative, du locuteur, pour reprendre les termes de Sperber & Wilson, Wilson & Wharton, en raison de ce décalage temporel.

²⁰ Ou « *meaning meaning* », opposé à « *meant meaning* ».

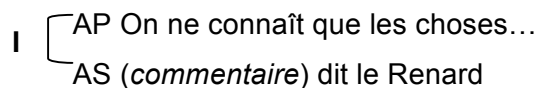
²¹ Le rôle joué par la technologie de la parole est ainsi bien différent de celui que lui voient jouer Keller 2004, Zellner-Keller 2004, dans une véritable dialectique de vérification-validation d'hypothèses linguistiques et pragmatiques (voir aussi Lacheret-Dujour 2004).

²² Ressource multilingue accessible en ligne (lectures du *Petit Prince* en plus de 100 langues) : <http://www.unige.ch/lettres/linguistique/prince/index.php> (Thanks J.-P. Goldman).

segment rapporté, et place le segment *dit le Renard* en appendice²³, réalisation atone, qui fait du groupe « dit le Renard » une information d'arrière-plan.

La première lecture « joue » la structuration du texte *écrit* (le regroupement prosodique maximal correspond à la fin de la « phrase » syntaxique); ce que *dit le Renard* n'est pas *dit* par les locutrices, mais seulement « *phrasé* ». La prosodie est ainsi en phase avec le texte écrit et le point de vue du narrateur. La structuration prosodique de la seconde lecture « joue » la voix du personnage Renard, et est en phase avec son point de vue. L'intonation en appendice du segment narratif *dit le Renard* met la voix du narrateur à l'arrière-plan ; la prosodie est globalement en phase avec « le sens » du discours, qui est de montrer le point de vue du Renard, et non de pointer qu'il s'agit d'une narration.

Du point de vue de leur structure textuelle hiérarchique²⁴, la première lecture accomplit un seul acte textuel, qui peut être paraphrasé par « le Renard dit qu'on ne connaît que les choses que l'on apprivoise ». La seconde lecture quant à elle accomplit *deux* actes textuels, qui s'articulent de la façon suivante :



et elle peut être paraphrasée par : *on ne connaît que les choses que l'on apprivoise.*²⁵

Ces deux lectures, par le biais de l'intonation (et du regroupement intonatif), énoncent deux versions différentes du texte écrit ; la première « joue » la partition du texte écrit, et énonce par là le point de vue du narrateur-scripteur ; la seconde met au premier plan le discours, rapporté, du Renard, et relègue à l'arrière-plan le discours du narrateur ; cette lecture rend disponible, et énonce par là, une tout autre expérience du sens de l'énoncé.

L'exemple suivant est tiré d'une émission de radio « phone-in » nocturne, et montre la question que l'animateur adresse à l'appelant, qui vient de se présenter comme entouré de personnes proches touchées récemment par la maladie : « et vous-même, heu-heu, personnellement, avez-vous été atteint d'une manière ou d'une autre par le destin aussi ? » La représentation utilisée est celle du « Prosogramme »²⁶ :

²³ « Appendice prosodique » selon Mertens (1992).

²⁴ Au sens de Roulet & al. (2001).

²⁵ Selon le principe de la supprimabilité relative des constituants subordonnés de Roulet & al.

²⁶ Voir Mertens (2004). Le Prosogramme fournit une représentation stylisée de l'intonation dans ses parties détectées audibles selon un modèle de perception. Quoique pertinente pour certaines tâches de description, la méthode du « filtre perceptif », ne laissant passer, pour la visualisation – et la re-synthèse – que les segments voisés et les modifications dans le temps (différences de hauteur) *suffisamment audibles*, ne tient pas compte des interférences avec les *autres* dimensions du traitement, prosodique, lexical, sémantique, syntaxique ou pragmatique (v. Auchlin & Simon 2008), qui interviennent également dans la perception (§1.1).

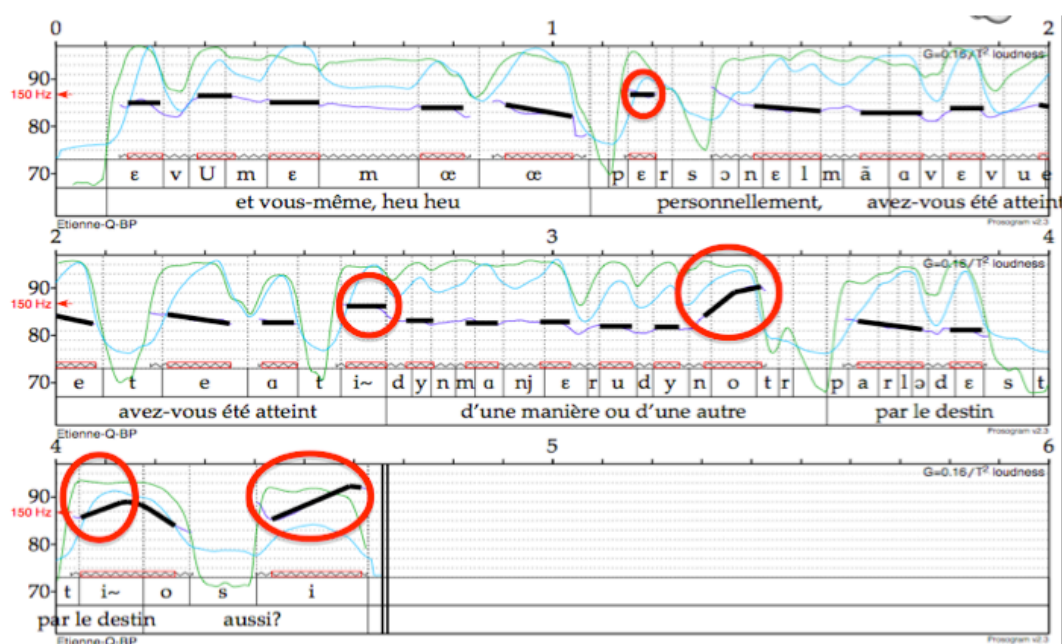


Fig. e. Prosogramme : énoncé d'une « question délicate et attentionnée »

Le locuteur utilise à trois reprises un mouvement mélodique montant /b... BH/, pattern typique de la demande d'information (« Et vous même, heu heu, personnellement, avez-vous été atteint d'une manière ou d'une **autre** par le destin **aussi** ? ») qui segmentent l'énoncé en trois Groupes Intonatifs (GI – Mertens 1992 ; Simon 2004), dont chacun effectue un mouvement de demande d'information, mais dont seul le dernier donne accès au contenu interrogé²⁷. La ré-itération du geste vocal, en elle-même, agit comme une forme d'atténuateur (adoucisseur) illocutoire (Kerbrat-Orecchioni). De plus, les deux cibles hautes (*autre* ; *si*) sont précédées de cibles (*atteint*, *destin*) qui activent un effet autotélique²⁸ de retour et de balancement-bercement, par leur assonance.

La perception de cette question comme délicate²⁹ est également construite à partir de l'élévation progressive des cibles tonales en fin de groupe (les trois dernières cibles entourées), par la diminution de la taille des groupes (en nombre de syllabes), ainsi que par le *découplage* des aspects tonals et dynamique, *l'élévation progressive des fins de groupe s'accompagnant d'une baisse progressive de l'intensité*, jusqu'à la fin du GI maximal, fin de l'énoncé, de l'intervention, et du tour de parole de l'animateur³⁰ (fig. f.) :

²⁷ Il y a quatre GI majeurs, mais le premier (*atteint*) joue moins clairement le contour 'b...BH' interrogatif.

²⁸ Roulet & al. (2001), après Jakobson.

²⁹ Il s'agit naturellement de ma perception, et non de celle de l'interlocuteur effectif. La perception de l'interlocuteur ne sera accessible qu'indirectement dans ce que renverra son tour de parole suivant ; quelle qu'elle soit, elle sera évaluée sur la base de sa convergence-divergence avec ma propre perception, comme effet d'empathie - *dyspathie*. (v. Auchlin 2000a).

³⁰ Autant de niveaux de « co-occurrence ».

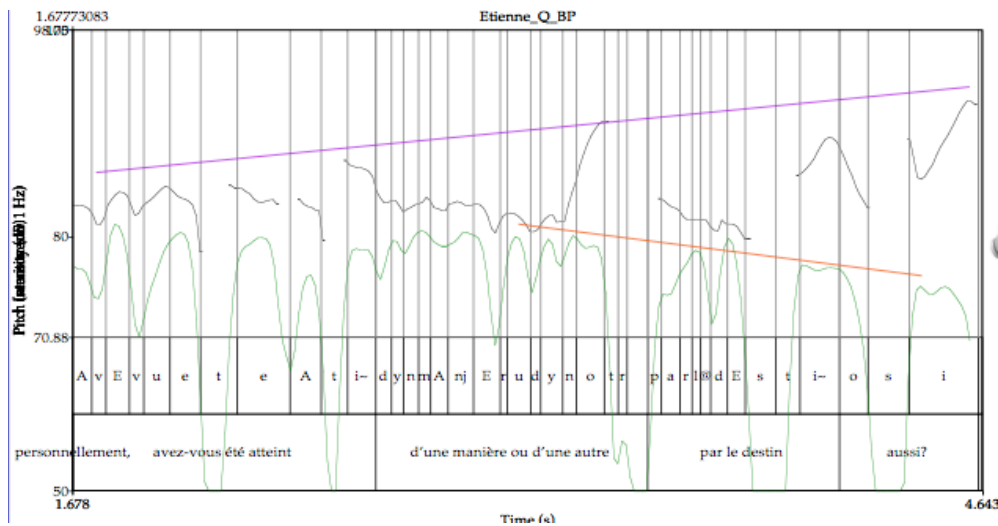


Fig. f. : découplage hauteur-intensité dans la succession des GI
'avez-vous été atteint d'une manière ou d'une autre par le destin aussi'

Ces différentes élaborations perceptives accompagnent la construction progressive des constituants linguistiques successifs ; elles s'intègrent au « contenu », dans l'expérience discursive, sous la forme de « blends expérimentiels »,³¹ où leur caractère perceptif assure un véritable *ancrage matériel* au sens de Hutchins (2005) aux processus d'intégration.

Ces exemples, de façon plus générale, montrent comment l'activité (de parole) et l'action (langagière) se configurent mutuellement, et qu'une description de l'une implique une description de l'autre³².

2.2. Phonostyle et éthos

Le phonostyle et l'éthos ont bien des points communs : réalités de deuxième ordre l'un et l'autre, phénomènes attestés dans l'*expérience* discursive, ce sont par ailleurs des constructions relatives à l'observateur-expérienceur³³, à plusieurs niveaux ou degrés d'assimilation-différenciation (langue / dialecte / sociolecte / réseau social / ...): la *relativité* à l'*observateur* est *relative* à des *collectivités* qui sont *elles-mêmes délinéées (produites) par les pratiques langagières et les variétés linguistiques*.

Une part de la variation conceptuelle affectant l'éthos (Auchlin 2000b : 86sq) touche également le phonostyle : l'un comme l'autre peuvent être plus ou moins singuliers, idiosyncrasiques, ou collectifs, partagés, « génériques » ; l'un comme l'autre peuvent être plus ou moins sémiotisés – conventionnalisés. L'éthos lui-même est sous-déterminé quant à sa dimension axiologique ou morale³⁴, notamment dans ses applications hors du champ de l'argumentation (Maingueneau

³¹ Le concept de « blend » est repris des « blends conceptuels » de Fauconier & Turner (2001), et adapté au traitement de processus articulant ressources conceptuelles et ressources sensori-motrices (Auchlin 2001)

³² Résumé peut-être caricatural du programme énonciatif en prosodie linguistique.

³³ « Observer-relative vs. intrinsic features of the world » Searle; Núñez.

³⁴ Laquelle ne fait pas partie intégrante de la définition du phonostyle.

1999). De sorte que de nombreuses descriptions phonostylistique décrivent en même temps des éthé, et la description d'éthos (en discours oral) repose, en grande partie, sur une description phonostylistique (Kerbrat & De Chanay 2007).

En suivant et généralisant P. Léon (1993), on peut définir un phonostyle comme la manifestation acoustiquement perçue d'un processus de différenciation – singularisation, dans une communauté de parole, d'un individu (par certaines de ses caractéristiques), d'une (sous-) communauté de parole, ou d'une circonstance. Un phonostyle peut être « hérité » (accent régional, par exemple) ou forgé, élaboré, et plus ou moins adopté – comme le phonostyle de France Info examiné ci-dessous. Le phonostyle peut être pensé à partir d'une catégorisation préalable (sociale ; situationnelle ; de genre, etc.), les différents échantillons examinés seront, alors, plus ou moins typiques, représentatifs du prototype. Il peut aussi être pensé à partir de l'émergence d'assemblages de propriétés singuliers, assemblages qui par leur récurrence peuvent être typifiés, accéder à une forme de réalité de second ordre, servir de cadre pour la perception de nouvelles occurrences.

Pour Fónagy (1987), le style vocal est une *nécessité* (toute parole sur-détermine la réalité linguistique qu'elle achemine), et un phénomène « parasite », dépendant et « greffé » sur le message linguistique. Pour autant, la sur-détermination (phonétique) du donné linguistique peut être (rendue) plus ou moins sensible - manifeste, plus ou moins constante, et le choix de ses propriétés est, plus ou moins, appuyé sur, ou régulé par, la perception du destinataire, effective ou supposée.

Cette dernière régulation, qui lie l'éthos à une forme d'intentionnalité, n'est pas nécessaire pour penser l'émergence du phonostyle. Cela me semble les distinguer fondamentalement : un phonostyle, bien que relatif à l'observateur (interlocuteur) n'est pas *constitutivement* dépendant de l'observateur-interlocuteur – l'éthos, lui, l'est³⁵ : c'est l'interlocuteur qui attribue, valide ou invalide l'éthos ; l'éthos est constitutivement dialogal, le phonostyle ne l'est pas.

2.3. Phonostyle: énonciation d'identité, l'exemple de France Info

Le phonostyle *supposé* de la chaîne de radio publique France Info (FI) a fait l'objet d'une description récente dans deux types de travaux. Une étude discursive, avec un repérage auditif de la prosodie³⁶, basée sur l'observation de phénomènes discursifs (articulation de la structure prosodique et de la structure textuelle, etc.), pose quelques hypothèses sur ce phonostyle ; ces hypothèses ont été testées par une étude prosodique outillée, automatisée³⁷ : le « phonostylographe » et « ProsoReport » (Goldman & al 2007, 2008).

³⁵ Développé dans Auchlin (2001).

³⁶ Burger & Auchlin (2007).

³⁷ Prémises et premières mesures automatiques dans Goldman & Auchlin 2006.

2.3.1. La première étude fait l'hypothèse que FI présente un phonostyle singulier, sur la base de l'observation informelle de propriétés récurrentes, et en propose une première description. Ce phonostyle, particulièrement bien représenté dans les chroniques, séquences courtes, sur des sujets d'intérêt général, voyage, science, etc, est qualifié de « pan-francophone didactique », en raison de ses propriétés saillantes : il est (intuitivement) perçu comme sur-articulé, sur-segmenté, jouant sur un registre tonal ample et parcouru avec agitation mélodique ; il affiche qu'il s'adresse à des francophones d'horizons variés, et, notamment par la récurrence d'accents initiaux « didactiques » (Callamand 1983, Léon 1993) montre un souci de « faire comprendre », souvent partagé par les médias.

Ces propriétés, ensemble hétérogène de traits phono-prosodiques³⁸, font l'objet d'un assemblage (ou blend), dont l'effet est l'émergence d'une identité phonostylistique singulière.

Certaines de ces propriétés phonostylistiques sont communes au genre « parole radiophonique »³⁹ (typiquement : les accents initiaux); certaines propres au genre « chronique », partagées par d'autres genres de la famille d'activités « divertissement » (typiquement : la « sur-mélodisation », enjouée), empruntant certains traits du genre « information » dans la culture radiophonique francophone française (par exemple régularité du tempo) ; cet enchevêtrement partiel d'activités (ou dimensions d'activités) est, finalement, *teinté* d'une forme de « signature » plus ou moins exclusive à FI (*fig. g.*) :

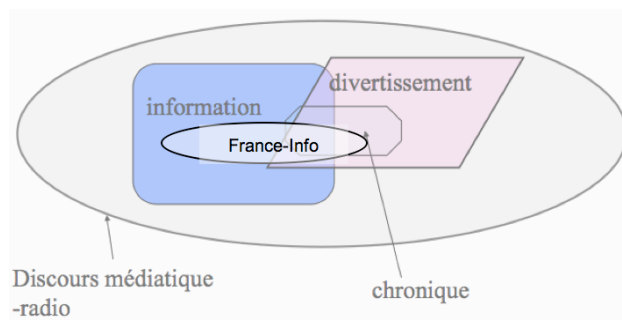


Fig. g. : La chronique radiophonique de FI – inclusion de genres et « signature »

En bref : par ses traits, la prosodie des chroniques *énacte* l'identité « France-Info » : les locuteurs se font être FI, dans le même temps qu'ils modèlent acoustiquement et font exister « FI ».

2.3.2. Pour valider ces observations, intuitives et informelles, un outil d'analyse semi-automatique a été élaboré, le « phonostylographe » et « ProsoReport » (Goldman & al 2007 ; 2008). Cet outil

³⁸ « Grappe » ou « cluster » de « phonostylèmes » (Léon 1993) (le suffixe *-ème* évoque une unité distinctive et oppositive, et donc un « système » de styles, comme le système des phonèmes, ce qui n'est vraisemblablement pas le bon modèle pour examiner la question, compte tenu de la productivité générique ou stylistique, le fait qu'un genre, un style, nouveau, peut parfaitement apparaître sans bouleverser les identités des autres genres ou styles existants. Cette idée de système n'est pas impliquée par le terme de grappe ou de trait phonoprosodique).

³⁹ Ici encore, une idéalisation vraisemblablement trompeuse : il y a un prototype de phonostyle radiophonique, plus ou moins présenté par les différents genres radiophoniques, plus ou moins distribué dans les autres médias oraux ou audio-visuels.

visé à faire apparaître et quantifier les caractéristiques prosodiques principales d'un corpus de parole, son profil phonostylistique. Pour être *parlantes*, les mesures doivent être comparatives, ou relatives ; en l'absence de norme de référence stable et légitime, le corpus étudié comportait trois chroniques de FI de deux minutes chacune (deux femmes et un homme ; total 1900 syllabes), et leurs lectures à haute voix par une locutrice unique, en contraste. Si le but était d'obtenir un portrait phonostylistique qui contraste globalement les trois échantillons de FI et la lecture à voix haute, les mesures des propriétés de chaque séquence peuvent être comparées à celles des autres, ce qui permet de quantifier également la variation intra-style inter-locuteurs (les trois séquences de FI), ou intra-locuteur (les trois lectures).

Je renvoie aux publications mentionnées pour une présentation détaillée des deux outils (*Phonostylographe*, et *ProsoReport*) et des résultats chiffrés. Je résume ici brièvement les observations faites (2.3.2.1), et présente les très grandes lignes des outils (2.3.2.2).

2.3.2.1. Nous supposons dans cette étude⁴⁰ que le phonostyle de FI se caractérise, par rapport à la lecture « neutre », (i) par une plus grande agitation mélodique, (ii) une sur-articulation, et (iii) une quantité plus élevée d'accents initiaux⁴¹. La prise de mesures automatiques, nécessaire pour traiter statistiquement des corpus importants et valider des observations, exige de déterminer précisément quoi, et comment, mesurer ce que l'on veut connaître, autrement dit comment passer de la réalité de 2^{ème} ordre où se forment impressions et hypothèses à la réalité de 1^{er} ordre qui la détermine.

(i) L'hypothèse d'une plus grande *agitation mélodique* a été confirmée sur quatre mesures tonales différentes :

- la *déviat*ion *standard* par rapport à la moyenne de F0, qui est de 5.40 demi-tons (abrégié *st*) pour la lecture à voix haute, et de 7.21 *st* pour FI ;

- l'*étendue du registre* de F0⁴², de 8.57 *st* pour la lecture, et de 12.90 pour FI ;

- le « *chemin mélodique* », parcours tonal cumulé, addition des mouvements intra-syllabiques, et des mouvements inter-syllabiques, en demi-tons par seconde : la lecture parcourt 16.9 demi-tons/seconde, et FI parcourt 21.6 *st*/sec ;

- enfin on compte, sur l'ensemble des syllabes, moins de *tons dynamiques* (cibles comportant un mouvement tonal perceptible) en lecture (13.30%) qu'en radio (17.10%), et animés de mouvements moins importants (19.4*st*/sec vs. 25.7 *st*/sec).

(ii) L'hypothèse d'une *sur-articulation* avait initialement été corrélée à une plus grande (intense) énergie consonantique pour FI ; les mesures ont permis d'invalider cette hypothèse : l'intensité moyenne, sur les consonnes est la même (FI et lecture), proportionnellement à celle des parties

⁴⁰ Après Goldman & Auchlin (2006).

⁴¹ Optionnels en français.

⁴² Espace tonal utilisé. La méthode utilisée consiste à ne considérer ni les 5% les plus bas, ni les plus hauts, où se logent des erreurs dans la détection de F0 ('pitch').

stylisées, une fois pondérées les différences entre les enregistrements⁴³. En revanche, l'impression de sur-articulation peut être expliquée par une plus grande *durée* des consonnes par rapport aux noyaux vocaliques dans les syllabes : pour FI, le noyau vocalique représente 40.77% de la durée de la syllabe, et la partie consonantique 59.23%, alors que pour la lecture le noyau représente 45.17%, et la partie consonantique 54.83%. Enfin, le calcul du *taux d'articulation* (proportion du temps d'articulation par rapport au temps de parole avec les pauses) montre une occupation temporelle du canal plus active pour FI que pour la lecture, le taux d'articulation de FI étant de 87.5%, et 81.8% pour la lecture. Cette différence ne concerne *ni* les *durées syllabiques* moyennes (qui sont les mêmes, 0.18 sec, avec sensiblement la même déviation standard de 0.07 pour la lecture et 0.08 pour FI) *ni* le *débit*, en syllabes par seconde, le débit moyen est sensiblement le même (4.57 syll/sec pour la lecture, 4.93 pour FI).

(iii). L'hypothèse d'une plus grande quantité d'accents initiaux est confirmée par deux mesures différentes, qui montrent une proportion plus élevée de syllabes proéminentes sur l'ensemble des syllabes (37% pour FI, 34.4% pour la lecture), ainsi qu'une plus grande proportion des syllabes proéminentes en position initiale (31.4% d'accent initiaux pour FI, 21.4% pour la lecture).

2.3.2.2. Le *ProsoReport* (*phonostylographe*) est un assemblage de quatre outils, plus ou moins automatiques, développés indépendamment les uns des autres (dans Praat). Ils sont utilisables distinctement, au gré des besoins de l'utilisateur, des tâches à accomplir et des résultats à obtenir, selon un degré de complexité (formelle) croissant.

(i) Une première étape indispensable à tout traitement de corpus audio de parole (à partir d'un enregistrement audio et de sa transcription orthographique) est l'alignement de la piste sonore et de la séquence de phonèmes tirée de la transcription phonétique du texte. Cette étape est prise en charge par *EasyAlign* (Goldman 2007), qui transforme un enregistrement et son texte orthographique en un « textgrid » de Praat phonétiquement aligné. Cet outil automatique ne demande qu'une vérification de l'alignement a posteriori (et quelques éventuelles corrections) ; il permet un gain de temps considérable dans le traitement des données de corpus.

(ii). La stylisation de F0, qui permet d'éviter certaines erreurs dans la détection du pitch (F0) : le *Prosogramme*⁴⁴ fournit un tracé de F0 segmenté représenté par des segments de droite ; ce tracé peut correspondre aux *pics de sonie*, si le *Prosogramme* est utilisé sans alignement phonétique préalable (TextGrid de Praat), ou aux noyaux limités aux voyelles du TextGrid. ProsoReport utilise une version légèrement modifiée du Prosogramme⁴⁵, qui autorise le noyau stylisé à s'étendre à toute la partie voisée de la syllabe (en cas de consonnes sonores), et ne le limite pas uniquement aux frontières de la voyelle.

(iii) Le repérage des *syllabes proéminentes*, de la « densité accentuelle » d'un enregistrement,

⁴³ Au demeurant, les mesures d'intensité demeurent très difficiles à contrôler, et sont du coup peu fiables.

⁴⁴ Mertens (2004).

⁴⁵ Avanzi & al. (2007).

est effectué à l'aide d'un troisième outil, nommé *ProsoProm* (Avanzi & al. 2007). Pour distinguer les syllabes proéminentes des autres, cet outil pondère différents paramètres prosodiques de la syllabe, avec un empan local, par rapport aux deux syllabes précédentes et la syllabe suivante : sa hauteur relative, sa durée relative, la présence ou non d'un mouvement intra-syllabique, et d'une pause avant-après. Les performances de l'outil (son taux de convergence avec les annotations de corpus faites par des juges), sont bonnes, comparables à celles des juges. Une partie des « erreurs », syllabes détectées proéminentes par la machine mais pas par les juges, ou non proéminentes alors qu'elle le sont pour les juges, est dûe (Auchlin & Simon 2008) à l'interaction, dans la perception humaine, avec les *autres* niveaux d'organisation, linguistique (morpho-syntaxe), discursif (type d'unité en cours), interactionnel (organisation du tour de parole), ou simplement prosodique (groupe rythmique ; groupe intonatif ; groupe de souffle ; ...). Ce n'est pas là une limite des approches strictement acoustiques de la prosodie (comme *ProsoProm*), bien au contraire : cette limite est celle du « trade off » de la perception humaine dans l'intégration linguistique, et l'examen des divergences nous informe à son sujet.

(iv). Enfin, par exemple pour identifier les *accents initiaux*, les proéminences doivent être repérées par rapport à la chaîne morphosyntaxique. Une tire supplémentaire est ajoutée au *TextGrid*, *Lex*, indiquant si l'unité est un mot plein ou un clitique⁴⁶, et permet de discriminer les proéminences, initiales ou finales.

Le *ProsoReport* est actuellement en cours d'utilisation pour tenter de discriminer plusieurs phonostyles, sur un corpus de 70 minutes (Avanzi & al. à paraître). Si la description qu'il produit permet un contraste fin entre deux (ou plus) styles, ou corpus, une limite qu'on peut lui voir est de ne pas donner accès à semblable contraste à l'intérieur d'un corpus unique, repérant des *moments* dans une séquence, différenciés les uns des autres par des ensembles de variations, dans telle et telle dimensions prosodiques identifiées. Ceci sera la tâche du *ProsoReport dynamique*, en cours de développement.

3. Conclusion

Les manifestations « de haut niveau de complexité » que sont l'éthos, et le phonostyle, demandent de considérer le discours dans toute son épaisseur expérientielle, au delà du hiatus entre les niveaux sensori-moteurs et les niveaux symboliques de l'expérience.

Pour cela, on plaide en faveur d'une dialectique entre modèles locaux, partiels, outillés, objectifs, et approches qualitatives, phénoménologiques, expérientielles. Cela demande, au passage, de dé-cloisonner quelque peu les paradigmes « disciplinaires » (phonétique-prosodie-intonologie ; analyse du discours-pragmatique), érigés, et souvent fermement maintenus, pour le bien de la science.

⁴⁶ Le terme de clitique, sa définition, et son extension, sont problématiques ; il n'y a pas aujourd'hui de définition stable et non-circulaire, ni d'inventaire exhaustif, des clitics en français.

Bibliographie

- Arbib, M. A. (2006) (ed.), *Action to language via the Mirror Neuron System*, Cambridge, Cambridge University Press.
- Auchlin A. (1998), « Les dimensions de l'analyse pragmatique du discours dans une approche expérientielle et systémique de la compétence discursive », in Verschueren J. (éd.) (1998), *Pragmatics in 1998: Selected papers from the 6th International Pragmatics Conference*, Anvers, IPrA, 1-22.
- Auchlin A. (2000a), "Grain fin et rendu émotionnel subtil dans l'observation des interactions : sur le caractère "trans-épistémique" des attributions d'émotions", in Plantin Ch., M. Doury & V. Traverso (2000) (éds), 195-204.
- Auchlin A. (2000b), "Sur l'éthos. Quelques remarques", in Wauthion M. & A. C. Simon (2000) (éds), *Politesse et idéologie. Rencontres de pragmatique et de rhétorique conversationnelles*, Louvain, Peeters « BCILL », 75-92. (trad. port. Auchlin 2001, « Éthos e experiência do discurso : algumas observações », in Mari, H., I. L. Machado & R. De Mello (2001) (eds), *Análise do discurso : fundamentos e práticas*, Núcleo de Análise do discurso, FALE/UFMG, Belo Horizonte, 201-225.
- Auchlin A. (2003) « Compétence discursive et co-occurrence d'affects: "blends expérientiels" ou (con)fusion d'émotions? », in Colletta J.-M. & A. Tcherkassof (éds) (2003), *Les émotions. Cognition, langage et développement*, Hayen, Mardaga, 137-152.
- Auchlin A. & A. C. Simon (2004), « Gabarits prosodiques, empathie et attitudes », CILL 30, 181-206.
- Auchlin A. & A. C. Simon (2008), « Prosody and experiential blending », EMUS atelier-satellite de Speech Prosody'08.
- Auchlin A., L. Filliettaz, A. Grobet & A. C. Simon (2004), « (En)action, expérientiation du discours et prosodie », Cahiers de linguistique française 26, 217-249.
- Auer P., E. Couper-Kuhlen, F. Muller (1999), *Language in time: the rhythm and tempo of spoken interaction*, Cambridge, Cambridge University Press.
- Avanzi, M., Goldman, J.-P., Simon, A.-C. & Auchlin, A. (à paraître), « Les phonostyles: une description prosodique des styles de parole en français », colloque Les voix du français, Oxford, septembre 2008.
- Avanzi, M., Goldman, J.-P., Lacheret-Dujour, A., Simon, A.-C. & Auchlin, A. (2008) : « Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé ». Cahiers of French Language Studies, 13/2, 2-30.
- Bache C. 2005, « Constraining conceptual integration theory : Levels of blending and disintegration », Journal of Pragmatics 37, 1615-1635.
- Bateson G. (1984), *La nature et la pensée*, Paris, Seuil.
- Boersma P. & D. Weeninck (en ligne), *Praat - logiciel d'analyse du signal vocal*, www.praat.org.
- Bottineau D. (2007 - sous presse), « Language and enaction », Stewart, J., Gapenne, O. & Di Paolo, E. (eds), *Enaction: towards a new paradigm for cognitive science*, MIT.
- Burger M. & A. Auchlin (2007) « Quand le parler radio dérange : remarques sur le phono-style de France Info », in Broth M., Forsgren M., Norén C. & Sullet-Nylander, F. (éds), *Le Français parlé des médias. Actes du colloque de Stockholm 8-12 juin 2005*, Stockholm, Acta Universitatis Stockholmiensis, 97-111.
- Cosnier, J. (2003), « Les deux voies de communication de l'émotion (en situation d'interaction de face à face) », in Colletta J.-M. & A. Tcherkassof (éds), *Les émotions. Cognition, langage, et développement*, Hayen, Mardaga, 59-67.
- Di Cristo A., C. Auran, R. Bertrand, C. Chanet, C. Portes & A. Régnier (2004), « Outils prosodiques et analyse du discours », CILL, 30, 1-3, 27-84.
- Fauconnier G. & M. Turner (2002), *The way we think. Conceptual blending and the mind's hidden complexities*, New York, Basic Books.
- Fónagy I. (1983), *La vive voix*, Paris, Payot.

- Fónagy I. (1995), « Musiques dans la parole », in Pozzi R. (éd.), *Tendenze e metodi nella ricerca musicologica*, Florence, L.S. Olschki, 129-142.
- Frijda N. (1988), « Les théories de l'émotion : un bilan », in Rimé B. & K. Scherer (1989) (éds), *Les émotions*, Neuchâtel-Paris, Delachaux et Niestlé, 21-72.
- Frijda N. (2003), « Passions : l'émotion comme motivation » in Colletta J.-M. & A. Tcherkassof (éds), *Les émotions. Cognition, langage et développement*, Hayen, Mardaga, 15-32.
- Goldman J.-Ph., A. Auchlin, A.-C. Simon, M. Avanzi (2008) : « ProsoReport: an automatic tool for prosodic description. Application to a radio style », *Speech Prosody*, mai 2008, Campinas, Brésil.
- Goldman J.-Ph., (2008), « Measuring and synthesising expressivity. Some tools to analyse and simulate phonostyle », EMUS atelier-satellite de Speech Prosody'08.
- Goldman J.-Ph., A.-C. Simon, M. Avanzi, & A. Auchlin (2007) : « *Phonostylographe*, un outil de description des phonostyles prosodiques. Chroniques radiophoniques et style lu », CLF28, 219-237. (<http://clf.unige.ch>)
- Goldman, J.-P., (2007) : « EasyAligner: a semi-automatic phonetic alignment tool under Praat. » Accessible en ligne à <http://latcui.unige.ch/phonetique>. (vérifié août 2008).
- Goldman J.-Ph., A. Auchlin (2006) "Le phonostyle France-Info et ses ingrédients prosodico-phonétiques. Approche croisée homme-automate", Colloque PFC2006, *Approches phonologiques et prosodiques de la variation sociolinguistique: le cas du français*. Louvain-la-neuve, juillet 2006
- Gussenhoven C. (2002), « Intonation and interpretation : phonetics and phonology », *Speech Prosody 2002*.
- Hirschberg J. (2002), « The pragmatics of intonational meaning », *Speech Prosody 2002*.
- Hutchins E. (2005), « Material anchors for conceptual blends », *Journal of Pragmatics* 37,1555–1577.
- Keller E. (2004), « La vérification d'hypothèses linguistiques au moyen de la synthèse de la parole », CILL30/1-3, 243-253.
- Kerbrat-Orecchioni C. (2001), *Les actes de langage dans le discours. Théorie et fonctionnement*, Paris. Nathan.
- Kerbrat-Orecchioni, C. & De Chanay H. (2007), « 100 minutes pour convaincre. L'éthos en action de Nicolas Sarkozy », in Broth M., Forsgren M., Norén C. & Sullet-Nylander, F. (éds), *Le Français parlé des médias. Actes du colloque de Stockholm 8-12 juin 2005*, Stockholm, Acta Universitatis Stockholmiensis, 309-329.
- Lacheret-Dujour A. (2007), « Prosodie-discours : une interface à multiples facettes », *Cahiers de linguistique française* 28, 7-40.
- Lacheret-Dujour A. (2004), « Structure communicative et géométrie intonative : que nous dit la synthèse de la parole ? », *Cahiers de l'institut de linguistique de Louvain*, 30, 1-3, 115-139.
- Ladd, R., K. Silverman, F. Tolkmitt, G. Bergmann & K. Scherer, (1985), « Evidence for the independent function of intonation contour type, voice quality and F0 range in signaling speaker affect », *Journal of the Acoustical Society of America* 78, 435-444.
- Ladd, D. R. (1996) *Intonational Phonology*, Cambridge, Cambridge University Press.
- Léon P. (1993), *Précis de phonostylistique. Parole et expressivité*, Paris, Nathan.
- Maingueneau D. (1999), « Éthos, scénographie, incorporation », in Amossy R. (1999) (éd.), *Images de soi dans le discours*, Paris-Lausanne, Delachaux et Niestlé, 75-100.
- Mc Gurck H. & J. Mc Donald (1976), « Hearing lips and seeing voices », *Nature* 264, 746-748.
- Mertens, P. (1987), *L'intonation du français. De la description linguistique à la reconnaissance automatique*, mimeo, thèse de doctorat, Université catholique de Louvain.
- Mertens P. (2004), « Le *prosogramme* : une transcription semi-automatique de la prosodie », *Cahiers de l'Institut de Linguistique de Louvain* 30, 1-3, 7-25.
- Núñez R. (1997), " Eating soup with chopsticks: dogmas, difficulties and alternatives in the study of conscious experience ", *Journal of consciousness studies* 4 (2), 143-166.

- Núñez R. (1999), "Could the future taste purple? Reclaiming mind, body, and cognition", *Journal of consciousness studies* 6, 41-60.
- Obin N., X. Rodet & A. Lacheret-Dujour (à paraître), « French prominence : a probabilistic framework ».
- Plantin C., M. Doury & V. Traverso (2000) (éds), *Les émotions dans les interactions*, Lyon, P.U.L.
- Rohrer T. (2005), « Embodiment and Experientialism », in Geeraerts D. & H. Cuyckens (eds), *The Handbook of Cognitive Linguistics*, Oxford, Oxford University Press.
- Roulet E., L. Filliettaz, A. Grobet, M. Burger (2001), *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne, Lang.
- Scherer, K. R. (1989), «Les émotions : fonctions et composantes», in Rimé, B. & K. Scherer (1989) (éds), *Les émotions*, Neuchâtel, Delachaux & Niestlé, 97-134.
- Scherer K., T. Bänziger & D. Grandjean (2003), « L'étude de l'expression vocale des émotions: mise en évidence de la dynamique des processus affectifs », in Colletta J.-M. & A. Tcherkassof (éds) (2003), *Les émotions. Cognition, langage et développement*, Hayen, Mardaga, 39-58.
- Simon A. C. (2004), *La structuration prosodique du discours en français. Une approche multidimensionnelle et expérientielle*, Berne, Lang.
- Simon A. C. & A. Auchlin (2001), « Les hors-phase de la prosodie », in Cavé C., Guaïtella I. & Santi S. (éds), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication*, Paris, L'Harmattan, 629-633.
- Skipper J. I., H. C. Nusbaum & S. L. Small (2006), « Lending a helping hand to hearing : another motor theory of speech perception », in Arbib M. (ed.) (2006), 250-285.
- Suciu I., I. Kanellos & Th. Moudenc (2007), « Expressivité et synthèse vocale. Isotopies expressives, cohérence discursive et structures prosodiques », *Cahiers de linguistique française* 28, 199-206. (<http://clf.unige.ch>)
- Varela F., Thompson E. & Rosch E. ([1991] 1993), *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, Paris, Seuil.
- Zellner (-Keller) B. (2004), « Réactions sur 'La vérification d'hypothèses linguistiques au moyen de la synthèse de la parole' », *CILL30/1-3*, 255-258.