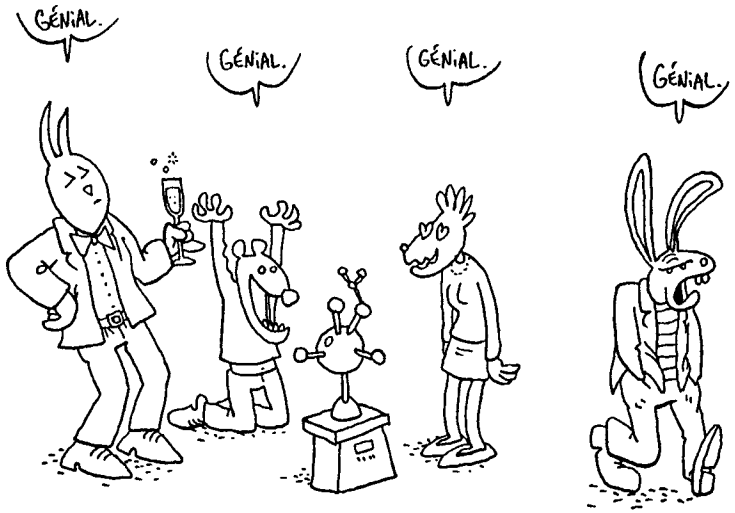

V

comme
la voix, les gestes, le corps

par Jacques Cosnier

La voix est le support sonore de la parole oralisée – par opposition à la parole écrite ou à la parole gestuelle – et la parole est l'utilisation concrète de la langue. Or, c'est une des grandes originalités des langues parlées (comme la langue française) – par rapport aux systèmes de communication animaux – que d'utiliser un ensemble limité (mais propre à chaque langue) de petites unités sonores non signifiantes (phonèmes) pour construire un nombre théoriquement illimité d'unités signifiantes (morphèmes ou, pour simplifier, « mots »). C'est la « double articulation » des langues orales. Cette double articulation est illustrée dans les textes rédigés en écriture alphabétique : succession de mots composés de lettres qui sont des indicateurs de prononciation. Les unités phonématiques (« sons de la langue ») sont ainsi des « idealtypes » qui, dans la pratique parolière, sont mis en voix sous forme de « sons de la parole ». Ces derniers sont reconnus grâce à un certain nombre de traits dictinctifs fondamentaux mais présentant de nombreuses variétés phono-acoustiques : les sons de la langue relèvent de la phonémique (ou phonologie) ; les sons de la parole relèvent de la phonétique et concernent autant les audiophonologistes, les psychophysiologistes, voire les

psychologues, que les linguistes. Il est à remarquer que chaque langue possède son propre système phonologique. La langue française, par exemple, est riche en voyelles (une douzaine) ; d'autre part, c'est une langue à articulation antérieure, sonore et douce, avec une prépondérance de syllabes ouvertes – ce qui la rend plus musicale que les langues anglo-saxones ou germaniques.

Les paroles sont donc doublement différenciées : elles utilisent les sons propres à une communauté linguistique, et ces sons se trouvent eux-mêmes modalisés par l'instrument phonogène propre à chaque parleur.

Au-delà de ce niveau d'organisation « segmentale » (ou articulatoire) de la prononciation, s'ajoutent les phénomènes « suprasegmentaux » de la prosodie (mélodie, accents, rythme) : à l'articulation se superpose ou se combine l'intonation.

Ces phénomènes jouent dans les conversations habituelles des rôles variés. Prenons, par exemple, le cas de « *tu parles* » : ce syntagme est formé de six phonèmes et possède une définition lexicale précise. Or, sur le plan de la parole, il peut correspondre à plusieurs énoncés selon l'intonation signifiant :

– l'affirmation après une longue attente (tu parles ! : enfin ! tu parles),

– l'interrogation (tu parles ? : est-ce que tu dis quelque chose ?),

– l'ironie (tu parles ! : ce que tu dis, quelle blague !),

– l'impatience (tu parles : vas-tu te décider à parler ?)

– l'étonnement (tu parles ? ! : toi, le muet, tu parles !).

La ponctuation essaie de rendre pleinement compte de ces variations, ce qui n'est pas toujours possible. En situation, l'intonation, les mimiques faciales, les gestes et évidemment le contexte – autrement dit des éléments voco-gestuels autant que verbaux – permettront aisément l'interprétation. On pourrait citer facilement d'autres exemples montrant que la voix peut en effet

indiquer, en plus de ses fonctions proprement linguistiques, des caractéristiques personnelles (par le timbre qui est aussi varié que les empreintes digitales et qui permet au téléphone d'identifier le correspondant dès les trois premiers phonèmes), des marques catégorielles (sexe, âge, appartenance régionale ou ethnique), mais aussi des dispositions affectives (sentiments, émotions, humeurs), enfin des modes relationnels liés aux situations sociales (voix « de composition »). Des observations diverses ont montré que l'audition d'une parole enregistrée suffit aux auditeurs pour définir un portrait physique, caractériel et social du propriétaire de cette voix. Or, si ces attributions ne sont pas toujours conformes à la réalité, elles présentent en revanche une grande convergence consensuelle, ce qui signifie que la voix est porteuse d'indices typifiés pertinents pour un groupe social donné.

Il est assez simple d'expliquer les raisons des variétés et variations vocales par les processus physiques et anatomo-physiologiques qui sont à la source de la production vocale. Celle-ci est réalisée par un instrument à vent constitué par une soufflerie (l'air pulmonaire), un système vibratoire (les cordes vocales), des caisses de résonance (pharynx, cavités buccales et naso-pharyngées). On comprend aisément qu'étant donné les variations morphologiques infinies des êtres humains, les timbres soient aussi variés, et qu'il en soit de même pour les « manières articulatoires », vu les variétés psycho-sociologiques et culturelles. Enfin, on comprend aussi qu'à ces aspects articulatoires (« segmentaux ») s'ajoutent les aspects « suprasegmentaux », plus fonctionnels, de l'expression prosodique. En effet, les états affectifs exprimés par la voix correspondent à des configurations corporelles définies (répartition du tonus, mimiques faciales, gestes, postures, rythme respiratoire...) qui vont conditionner des « mimiques vocales ». De nombreux auteurs ont, à juste titre, souligné que la voix était

le produit d'un geste. Comme nous l'avons vu, la mise en jeu de l'instrument à vent phonogène est une activité corporelle dynamique selon la ventilation, le rythme d'expiration, les contractions laryngées, le rétrécissement ou l'élargissement des cavités de résonance, les mouvements du palais et de la langue, les modifications de l'orifice buccal : la voix sera modulée de plusieurs façons et une même phrase sera produite avec des qualités vocales variables.

Ainsi des « mimiques vocales » sont associées aux « mimiques faciales », elles-mêmes liées aux patterns spécifiques des différents états affectifs. Il est fréquent d'identifier les mimiques faciales d'un locuteur au téléphone (« mimiques audibles ») d'après la perception de ses mimiques vocales. On reconnait la voix du sourire, de la contrariété, de la colère, de l'impatience et l'on « voit », à travers les perceptions acoustiques, l'activité faciale du locuteur. L'expérimentation confirme ces faits et il est aisé aujourd'hui de constater l'activité mimo-gestuelle variée des utilisateurs de téléphones mobiles. Leurs activités motrices sont liées à la production des modulations verbales (ainsi qu'au travail énonciatif, comme nous le verrons plus loin). Le dynamisme vocal est synchrone du dynamisme gestuel, et les dispositions corporelles conditionnent les types de production vocale : l'expérimentation montre qu'une voix « joyeuse » ne peut être produite par un corps exprimant la dépression, et qu'une voix « triste » ne peut être produite par un corps exprimant l'allégresse.

Se pose alors la question de la « motivation » des expressions vocales : sont-elles « naturelles » et donc « universelles » ou sont-elles « culturelles » ? Autrement dit, y a-t-il des « voix darwiniennes » comme certains affirment qu'il y a des « mimiques faciales darwiniennes » ? Aujourd'hui, grâce à de nombreuses observations, on sait répondre à cette dernière question en ce qui concerne les mimiques faciales : oui, il existe des

types expressifs émotionnels naturels (on ne voit pas pourquoi d'ailleurs l'animal humain différerait à cet égard des autres espèces animales). Toutefois, ces expressions naturelles subissent des modalisations culturelles qui encouragent certaines expressions et en répriment d'autres. On peut donc bien s'attendre à ce qu'il en soit de même pour les expressions vocales qui leur sont intimement liées. De même, alors que la langue française est parlée par l'ensemble de la communauté francophone, il n'est pas étonnant de relever des différences liées aux modes gestuels et expressifs régionaux.

De plus, l'espèce humaine possède une aptitude qui n'existe qu'à l'état rudimentaire chez les autres animaux : la capacité d'exprimer intentionnellement un état affectif. Ainsi y a-t-il des états affectifs dits spontanés, « vécus » ou « involontaires », et d'autres « fabriqués » ou « affichés ». Autrement dit, on peut aisément feindre ou mimer la tristesse ou la satisfaction, à tel point que les expressions affectives peuvent se conventionnaliser et que chaque culture possède un répertoire d'« emblèmes » qui ont acquis un statut « quasilinguistique ». Ces mimiques sont intégrées le plus souvent dans des configurations gestuelles plus ou moins iconiques : par exemple, le « ras-le-bol » français (hexagonal). Ces configurations ont des équivalents expressifs verbaux (ici « ras-le-bol ») qui peuvent accompagner le geste d'une voix adéquate. On peut donc parler aussi d'emblèmes vocaux, c'est-à-dire de configurations prosodiques devenues conventionnelles pour exprimer par exemple l'ironie, la lassitude, etc. Ici encore les modalisations culturelles sont importantes, les emblèmes « quasilinguistiques » ne sont pas liés organiquement à la langue mais au milieu culturel, à son histoire et à son écologie. Il n'y a pas de répertoire « quasi-linguistique » type de la langue française, même si on

peut évidemment en décrire un noyau commun substantiel.

Pour résumer les différents aspects de l'expression vocale tels qu'ils nous apparaissent en ce point de l'exposé, nous énumérerons les points suivants :

1. la voix produite par un instrument anatomo-physiologique phonogène sert à véhiculer la parole en transformant en phénomène vibratoire les phonèmes constitutifs de la chaine verbale : la mise en mots s'accompagne d'une mise en voix.

2. La voix présente des caractéristiques individuelles comme le timbre, ou catégorielles, indicatrices du sexe, de l'âge, de l'ethnie, de l'appartenance sociale, etc.

3. La voix surajoute à la chaine verbale des modulations dont certaines sont conventionnelles et font partie d'un système paraverbal prosodique (marques syntaxo-pragmatiques du questionnement, de l'ordre, de l'assertion, de l'emphase, de la clôture des propositions), ou émotif (emblèmes vocaux quasilinguistiques) ; certaines refléteront les troubles émotionnels spontanés ou les altérations de l'état physique (fatigue, maladies neurologiques...), d'autres enfin s'ajusteront au rôle social impliqué par le site et le type de relation.

4. Dans la mesure où la voix est le résultat d'une activité motrice, elle est intimement associée avec d'autres activités corporelles et, en particulier, aux mimiques faciales et posturo-gestuelles.

Les éléments précédents étaient généralement centrés sur l'objet « voix », son appareil de production et son contexte expressif. Nous devons maintenant aborder son aspect « impressif » : comment la voix est-elle reçue et interprétée ?

Une remarque préalable sera d'ordre psychophysiologique : n'est perçu d'un message acoustique que ce qui est permis par l'instrument récepteur auditif humain, ce dernier étant conçu de manière à capter les vibrations de 16 Hz à 16000 Hz. Au-delà, on entre dans la zone des

ultrasons. D'autre part, la mélodie produite par l'émission vocale est axée sur la fréquence de base dite « fondamentale » qui donne l'impression de hauteur d'un son en rapport avec cette fréquence (125 Hz en moyenne pour les hommes et 250 Hz pour les femmes). Cette fréquence fondamentale s'accompagne d'harmoniques qui enrichissent la production en la rendant plus ou moins mélodieuse, froide ou chaleureuse. Enfin, outre ses aspects fréquentiels, le message sonore doit être d'une certaine intensité. L'optimum conversationnel est évidemment variable selon le bruit de fond et la culture des interlocuteurs.

Une deuxième remarque concerne la réceptivité particulière de l'espèce humaine aux sons mélodiques et rythmés : le nourrisson sait très précocement distinguer les voix humaines entre elles et parmi les autres phénomènes acoustiques environnementaux. Il est en outre capable de distinguer, très rapidement, la voix de sa mère. Au départ (lors des trois premiers mois), il est également sensible à tout l'éventail des phonèmes possibles ; puis, au cours des mois suivants, il sélectionne les caractéristiques vocales de son milieu linguistique qu'il est, de plus, capable de reproduire dans son babil (dès le huitième mois, on peut distinguer le babil d'un bébé chinois du babil d'un bébé français).

Cette prédisposition constitutionnelle de l'enfant humain se complète par un autre phénomène caractéristique : la réceptivité aux effets dynamogènes des sons mélodiques et rythmés. Pris dans un bain musical, le corps humain a spontanément tendance à se synchroniser musculairement au rythme auditif (ce qui ne s'observe pas chez les autres espèces animales, sauf chez certains insectes et oiseaux assez éloignés de notre espèce).

Ces remarques nous incitent à évoquer la théorie motrice de la perception auditive de la parole. Selon cette théorie, la compréhension d'un message verbal oralisé passe par la reconnaissance des éléments constitutifs

de ce message, ce qui parait logique. Mais cette reconnaissance se fait à partir d'une répétition (plus ou moins explicite) du message : on comprendrait d'autant mieux que l'on serait capable d'émettre ce qu'on entend. Cela expliquerait entre autres le fait qu'un adulte, lors d'un apprentissage tardif d'une langue seconde, a une compétence (de compréhension) supérieure à la lecture qu'à l'écoute.

Si cette théorie peut être controversée – car il est probable que la perception de la parole ne se fait pas toujours et uniquement par ce phénomène d'« échoïsation » interne –, elle présente cependant un grand intérêt. D'une part, pour la compréhension des phénomènes d'acquisition précoce que nous avons signalés plus haut (il faut d'ailleurs ajouter que les adultes facilitent spontanément ce processus par le « motherese », le « parler-bébé », langage simplifié et ajusté aux capacités du jeune enfant, et qu'ils manifestent eux-mêmes des échoïsations nombreuses aux émissions enfantines). D'autre part, l'intérêt de cette théorie réside dans le rapprochement que l'on peut faire avec la compréhension de l'accompagnement mimo-gestuel de la parole.

Il est en effet bien connu que la parole orale s'accompagne de gestes : on ne peut parler sans bouger non seulement les organes phonatoires mais aussi le reste du corps, et en particulier les mains, les bras, le buste, la tête et la face. Cette activité mimo-gestuelle fait depuis une trentaine d'années l'objet de plusieurs études qui ont amené les auteurs à s'accorder sur les catégories fonctionnelles résumées dans le tableau suivant.

Tableau des catégories gestuelles

EMBLÈMES : gestes (et/ou vocalisations) quasi-linguistiques de forme et d'utilisation conventionnelle qui peuvent être utilisés avec ou sans parole.

CO-VERBAUX :

Phonogènes : liés à l'activité motrice productrice de la parole.

Illustratifs : liés au contenu propositionnel du discours.

On distingue :

– *Les déictiques* : désignant le référent présent ou symbolique.

– *Les iconiques* : représentant les objets concrets.

– *Les métaphoriques ou idéographiques* : représentant les objets abstraits.

Bâtons ou battements : mouvements en deux temps de la tête ou des mains, marqueurs pragmatiques.

Expressifs : principalement les mimiques faciales qui connotent le contenu propositionnel ou qui situent métacommunicativement la position de l'orateur.

CO-ORDINATEURS : assurent le copilotage de l'interaction (maintenance et passage de tours).

Phatiques : activité du parleur destinée à vérifier ou à entretenir le contact principalement par le regard et l'intonation, parfois par le contact physique.

Régulateurs : activité du récepteur en réponse aux précédents (« back channel ») : hochements de tête, sourires et courtes voco-verbalisations, en sont des exemples fréquents.

Ce tableau mérite quelques commentaires.

Comme on le voit, les catégories fonctionnelles des gestes communicatifs sont nombreuses et en général, très liées aux fonctions de la chaine parolière. Mais on peut se demander à quoi et à qui elles sont utiles, voire nécessaires. Au-delà de leur fonctions, quelle est donc leur importance pour l'intelligibilité des messages verbaux ? Cette importance semble très variable.

Il y a d'abord les gestes qui sont nécessaires à l'interprétation du message dans son aspect propositionnel : c'est le cas des gestes de pointage (déictiques) désignant le référent du discours (« cet homme », « cette table », « là »...) ; c'est aussi le cas des gestes emblématiques (quasilinguistiques) qui peuvent remplacer un syntagme verbal.

On trouve ensuite les gestes qui sont utiles pour la bonne coordination de l'interaction (le « co-pilotage ») : gestes de maintenance du canal et gestes assurant les tours de paroles (tels : hochements de tête, sourires, courtes émissions voco-verbales...).

Mais deux autres catégories sont plus problématiques : les expressifs et les illustratifs.

Les illustratifs qu'ils soient « iconiques » ou « métaphoriques » ne paraissent en effet la plupart du temps ni nécessaires ni utiles à l'intelligibilité du contenu propositionnel : on s'en passe au téléphone et les trois quarts des émissions télévisées ne montrent les orateurs qu'en gros plans d'où sont exclus les bras et les mains.

Quant aux expressifs, ils sont transcatégoriels : mimiques conventionnelles intégrées aux quasilinguistiques ; mimiques emblématiques associées au discours verbal pour en connoter le contenu ou en donner une évaluation métacommunicative parallèle ; mimiques expressives spontanées enfin, affichant le vécu affectif du parleur.

Mais ces deux dernières catégories que nous avons qualifiées de « fonctionnellement problématiques » ont cependant leur utilité : d'abord pour l'émetteur lui-même, ensuite pour la relation avec le receveur.

a) Utilité pour l'émetteur, car ces gestes sont constitutifs de la « pensée imagée ». Ils constituent également une puissante aide, souvent indispensable, à la mise en mots. Le corps et ses coordonnées spatio-temporelles, égocentriques et projectives, mettent en scène, par le geste, le contexte du discours et situe les objets présents, représentés ou virtuels dans l'espace scénique. On peut dire que cette gestualité illustrative est une « gestualité énonciative », c'est-à-dire une gestualité qui participe à la création de l'énoncé et à son embrayage contextuel.

Ajoutons que la catégorie des « bâtons » qui scandent certains passages du discours a un rôle plus syntaxique.

b) Utilité pour le receveur, parce que l'on peut appliquer la théorie motrice de la perception auditive à la perception de la gestuelle. Ici aussi le receveur échoïserait (plus ou moins ouvertement) la posturo-mimo-gestualité de l'émetteur, et par un processus de cinesthésie cette échoïsation induirait en lui des représentations et des affects synchrones à ceux de l'émetteur. On peut parler de processus d'« empathie inférentielle ». Cette théorie dite de l'« analyseur corporel » des productions communicatives concernerait donc à la fois une grande partie des phénomènes vocaux et des phénomènes gestuels – très associés, comme nous l'avons vu. Nous avons souligné la liaison étroite qui existe entre les mimiques expressives faciales et les mimiques expressives vocales (il est d'ailleurs remarquable que les imitateurs de voix reproduisent aussi, parfois à leur insu, les mimiques du personnage en question).

Ainsi, dans les situations d'interlocution de face-à-face agiraient deux systèmes étroitement intriqués : un système d'échange de signaux soit arbitraires (le système verbal), soit simplement conventionnels (emblèmes vocaux et gestuels, déictiques gestuels, mimiques métacommunicatives), traités sur le mode encodage/décodage cognitivo-inférentiel et un système de partage empathico-inférentiel utilisant de façon privilégiée les manifestations non verbales de la pensée imagée qui s'exprime par la gestualité

vocale et corporelle. Le premier système – digitalisé (ou digitalisable), rationnel, organisé, hautement programmé – rapproche l'esprit humain de « l'esprit computationnel ». Le second système – spontané et à base de motivations analogiques – rapproche l'esprit humain de « l'esprit animal », et c'est sans doute le plus souvent par lui que sont induits les affects positifs et négatifs.

Plusieurs arguments viennent corroborer cette hypothèse.

1. Le bon fonctionnement d'une interaction de face-à-face s'accompagne fréquemment d'un effet de convergence par symétrie ou complémentarité : rapprochement des locuteurs, échanges thématiques prolongés avec partage des tours de parole et, pour ce qui nous occupe particulièrement ici, fréquentes échoïsations mimo-gestuelles, convergence des qualités vocales avec ajustement des rythmes, des intensités et des fréquences fondamentales, le tout accompagné d'une subtile synchronisation interactionnelle.

Corollaire de l'observation précédente, la mauvaise entente relationnelle s'accompagne souvent de divergence et d'asynchronie…

2. L'aptitude empathique de l'espèce humaine la prédispose aux sources de plaisir que fournissent l'activité synchronisée intra- et intercorporelle. Ainsi en est-il du déclenchement des activités rythmiques et des mélopées scandées par des productions sonores corporelles (claquements de mains) ou vocales (refrains et scansions oralisées). Le chant choral, les effets dynamogéniques de la musique, les danses diverses fournissent donc de multiples sources de plaisir synergique : l'entrainement corporel au rythme, l'anticipation des événements mélodiques, l'échoïsation réciproque motrice et vocale.

3. À un niveau psychologique plus profond, le laisser-aller à l'entrainement rythmique musical, vocal et corporel supprime temporairement les défenses tout en donnant une sensation de maitrise de la cinesthésie, tandis que l'appartenance syntone corporelle à la collectivité permet une régression psychocorporelle sécurisante.

Ces actions d'induction psychophysiologique trouvent leurs exemples dans les deux cas extrêmes des effets d'endormissement des « berceuses » sur les nourrissons, et des effets de transe des « raves » contemporaines.

4. Valeur cathartique et régulatrice : l'activité phonogène permet aussi la décharge émotionnelle cathartique – ainsi les hurlements, les clameurs de foule, et de façon plus personnalisée les « logorrhées » post-traumatiques – cependant qu'en sens contraire la voix peut aussi servir de régulateur tensionnel. Dans ce cas, le rôle de la mise en mots est aussi important que ceux de la mise en voix et de la présence d'un auditeur.

Enfin, il faut encore rappeler que malgré leurs bases « naturelles » tous les phénomènes décrits ci-dessus ont des marques culturelles. Nous avons vu que la voix sert de support à la parole et qu'à ce titre, elle est nécessairement tributaire du système phonologique de la langue – ce qui donne des aspects spécifiques au style voco-acoustique et aux impressions que peuvent avoir les pratiquants d'un autre système. Les différences proviennent a) des phonèmes, b) de l'accent, c) de l'intensité sonore habituelle, d) des habitudes rythmiques et du débit plus ou moins rapide, e) des rituels conversationnels (respect plus ou moins rigoureux des tours de parole). On retrouve des différences analogues en ce qui concerne la mimo-gestualité plus ou moins retenue ou libérée, le tout étant compliqué par les règles culturelles très variables de l'usage du regard et des distances interpersonnelles.

Ces spécificités culturelles font que la langue française – bien que fondamentalement commune en tant que code linguistique aux différentes régions francophones – présente des variations assez notables quant à son accompagnement voco-gestuel. Sources particulières du charme empathique que l'on éprouve à reconnaitre le semblable rassurant, chez un autre qui nous surprend cependant par l'étrangeté de ses différences.

BIBLIOGRAPHIE

CALBRIS, G., PORCHER, L., 1989, *Geste et communication*, Paris, Hatier-Crédif.
CORNUT, G., 1983, *La Voix*, Paris, PUF.
COSNIER, J., BOSSARD, A. (éd.), 1984, *La Communication non verbale*, Neuchâtel, Delachaux et Niestlé.
COSNIER, J., KERBRAT-ORECCHIONI, C. (éd.), 1987, *Décrire la conversation*, Lyon, Presses universitaires de Lyon.
COSNIER, J., VAYSSE, J., 1997, « Sémiotique des gestes communicatifs », in *Nouveaux actes sémiotiques*, 7-28, 52.
FONAGY, I., 1983, *La Vive Voix, Essais de psychophonétique*, Paris, Payot.
FONTANEY, L., 1987, « L'intonation et la régulation de l'interaction » *in* COSNIER et KERBRAT-ORECCHIONI (éd.), *Décrire la conversation*, Lyon, Presses universitaires de Lyon.
GROSJEAN, M., 1991, *Les Musiques de l'interaction, Contribution à une recherche sur les fonctions de la voix dans l'interaction*, Thèse de doctorat, Université de Lyon 2.
KERBRAT-ORECCHIONI, C., 1992, *Les Interactions verbales*, 3 tomes, Paris, Armand Colin.
LANDERCY, A., RENARD, R., 1977, *Éléments de phonétique*, Bruxelles, Didier.
LÉON, P.,1993, *Précis de phonostylistique, parole et expressivité*, Paris, Nathan.
LÉON, M., LÉON, P., 1997, *La Prononciation du français*, Paris, Nathan, « Université ».
MOREL, M.-A., DANON-BOILEAU, L., 1998, *Grammaire de l'intonation, l'exemple du français*, Paris, Ophrys.
RITTAUD-HUTINET, C., 1987, « Les signes vocaux de la communauté énonciative » *in* COSNIER et KERBRAT-ORECCHIONI (éd.), *Décrire la conversation*, Lyon, Presses universitaires de Lyon.
ROSSI, M., DI CRISTO, A. et *alii*., 1981, *L'Intonation, de l'acoustique à la sémantique*, Paris, Klincksieck.
SCHERER, K., EKMAN, P., 1982, *Handbook of Methods in Nonverbal Behavior Research*, Paris, Cambridge University-Maison des sciences de l'Homme.
WIOLAND, F., 1991, *Prononcer les mots français*, Paris, Hachette.