

AVANT-PROPOS

Cet ouvrage est le fruit d'une réflexion et d'une expérience d'une bonne vingtaine d'années dans le domaine de l'analyse prosodique, acquise sur des corpus variés de lecture et de dialogue spontané. Orienté principalement vers l'étude de la prosodie de lecture, l'ouvrage élargira toutefois la perspective, lors du dernier chapitre, de la linguistique à la pragmatique, en proposant de nouvelles pistes de recherche dans le domaine de l'oral spontané, et de l'affectivité (émotion(s) et attitudes).

Dans cet ouvrage il sera question non seulement des caractéristiques acoustiques de la parole, mais aussi de syntaxe, de sémantique et de pragmatique. Avec l'analyse des stratégies des locuteurs, l'aspect psychologique sera également pris en compte. L'ouvrage est donc destiné en priorité aux linguistes, qu'ils soient spécialistes de la parole ou des textes, et aux psycholinguistes, mais aussi à tous les chercheurs des autres spécialités intéressés par les processus de la parole, notamment ceux qui travaillent en traitement automatique de la parole (compréhension, génération, synthèse ...). Ce livre s'adresse aux chercheurs confirmés et aux doctorants.

Sur un plan acoustique, la prosodie peut se définir en première approximation comme l'ensemble, au cours d'un énoncé, des modifications de la fréquence fondamentale (F0), de l'intensité (correspondant à la force de la voix), et de la durée. Sur le plan auditif, ces modifications contribuent à la perception discriminative de la mélodie et de l'intonation, de l'énergie et du rythme dans le discours. Intonation, mélodie et prosodie ne sont pas des équivalents : la mélodie dont le domaine est celui de la fréquence fondamentale, est l'un des paramètres de l'intonation (avec la durée et l'intensité), et l'intonation, est un processus linguistique associant une forme acoustique (F0, durée, énergie) à un sens (linguistique et/ou pragmatique). Selon Rossi (1999),

L'intonation [...], est un système linguistique destiné à organiser et hiérarchiser l'information que le locuteur entend communiquer à l'allocutaire ou aux allocutaires dans son message, et à linéariser la hiérarchie des structures syntaxiques (Rossi, 1999).

A un niveau général, la prosodie constitue l'élément essentiel de la parole car sans elle, la parole est fortement détériorée, ou même n'est plus audible. Dans ces conditions la prosodie constitue donc sur le versant perception comme sur le versant production, le support concret de la parole.

Mais comme les linguistes le croyaient dans les années soixante, la prosodie serait-elle uniquement le produit de notre affectivité ou de l'expression esthétique du langage, reflet d'une communauté culturelle ? Certains auteurs contemporains avancent au contraire l'idée que l'on ne peut pas penser sans avoir recours à une prosodie interne.

Médium entre le locuteur et l'interlocuteur, elle est aussi le support de l'interaction dans la lecture à voix haute comme dans le dialogue, et à ce titre, elle rentre nécessairement en relation avec la structure linguistique et la composante pragmatique.

Alors peut-on concevoir d'une part une parole liée au langage, d'autre part une prosodie si intimement liée à la fois à la parole et à la pensée, sans faire aussitôt l'hypothèse d'une prosodie liée à l'expression du sens ?

Mais par quel moyen pouvons-nous appréhender le sens ? En langue écrite, la signification est du seul ressort linguistique (syntaxe, sémantique), et l'objet de la compréhension est l'attribution d'un sens. Nous avons tous fait l'expérience à l'écrit

(mais aussi à l'oral), d'un mot mal identifié ou inconnu de notre répertoire personnel, qui suffit à faire manquer le sens de l'énoncé, voire du discours ...

La langue, il est vrai, est complexe : complexité de l'organisation syntaxique avec ses processus emboîtés et diversifiés, complexité des relations sémantiques, avec ses atomes sémiques et ses constellations de réseaux, complexité des processus pragmatiques, avec l'enchevêtrement des buts et motivations psychologiques en situation, et profusion des indices de tout niveau.

Les relations entre prosodie et linguistique sont moins étroites dans le cas de langues comme le français par exemple, que dans le cas des « langues accentuelles » comme l'anglais ou l'espagnol par exemple (dont l'accent fait partie de la structure du mot et permet d'opposer des classes morphologiques, voire d'identifier un nom d'un verbe), ou encore des « langues à tons » comme le chinois, chez lesquelles la hauteur, la configuration tonale du mot permet, à contexte phonétique équivalent, d'identifier son sens. Dans cet ouvrage, nous ferons référence uniquement à la prosodie du français.

L'hypothèse qui sous-tend cet ouvrage est que la prosodie apporte, dans le même temps où elle instancie le discours, une réduction de cette complexité : en effet en dotant certains mots de contours qui les différencient des autres, elle oriente l'attention de l'auditeur vers quelques uns d'entre eux, ce qui a pour effet de restreindre le champ de la signification et d'engager alors un processus d'interprétation. Opérant ainsi une sorte de filtrage à l'intersection d'un contenu linguistique et d'une intention sous-jacente, elle offre le meilleur accès au sens.

On peut alors se poser la question de savoir comment s'opère cette restriction de la signification menant à la saisie d'un sens. Selon nos hypothèses, ce phénomène peut être appréhendé sur trois plans, le plan linguistique, le plan psycholinguistique (et pragmatique) et le plan prosodique.

Sur le plan linguistique, il s'agit d'identifier les indices de l'émergence du sens contenus dans les diverses unités, de les trier, de les structurer. Sur le plan psycholinguistique, il s'agit d'identifier les processus qui commandent cette émergence. Quant au plan prosodique, il doit nous permettre de préciser non seulement quels sont les indices qui remplissent cette fonction de mise en valeur des unités lexicales, mais en relation avec le plan morphosyntaxique, de les localiser : l'indice est-il opératoire par exemple dans l'ensemble du mot, en syllabe finale, ou sur une portion de mot quelconque ?

Il ressort de tout ceci que notre objectif n'est pas uniquement linguistique, car le discours oral est avant tout une *appropriation subjective* des structures linguistiques. Pour le linguiste de l'oral, la saisie du sens s'opère par le moyen de la prosodie au carrefour du linguistique et du pragmatique. Autrement dit nous considérons la prosodie *comme l'indice révélateur de l'appropriation du sens* par le locuteur à destination d'un tiers et dans une situation donnée.

Nous mènerons donc de front deux types d'approche : une approche purement linguistique lorsque nous aborderons les structures syntaxiques et sémantiques, et une approche psycholinguistique lorsque nous aborderons les phénomènes qui n'appartiennent pas à ce type de structure. Pour faciliter la comparaison cependant, nous analyserons les processus en nous restreignant aux mots lexicaux. Ceci ne pose pas de problème puisque les mots lexicaux sont des entités à la fois sur le plan linguistique et sur le plan psycholinguistique.

La question se pose maintenant de savoir comment apparier le linguistique (au sens large) et le prosodique. Concernant la prosodie et tout particulièrement la mélodie, il suffit d'écouter quelqu'un parler pour remarquer que la notion de niveaux est

fondamentale : dans le discours oral, on perçoit sans difficulté une alternance de notes aiguës, moyennes ou basses.

Par ailleurs tout processus linguistique ancré sur une différence, une complexité, une hiérarchisation, une distance ... fait appel également à une notion de contraste de niveaux, même si elle est de nature plus intuitive. Il s'ensuit que dans l'évaluation subjective de ce qu'il importe au locuteur de transmettre en priorité à l'auditeur, on s'attend à ce que ce soient les processus linguistiques reposant sur un contraste de niveaux substantiel qui génèrent une mise en relief prosodique.

Et de fait ce contraste de niveaux, fondamental en langue, s'applique particulièrement bien au transfert de l'information : de la même façon qu'une structure syntaxique ou sémantique met en forme une opposition de niveaux dans une arborescence, de la même façon, en situation, le locuteur communique des notions plus ou moins centrales à son propos que d'autres, plus ou moins inattendues, plus ou moins complexes, ou encore plus ou moins informatives. Notre hypothèse globale est alors la suivante : *si dans le traitement cognitif de programmation de la parole, prosodie et signification sont liées, alors il doit exister une identité de leurs structures profondes¹, et cette identité est vérifiable numériquement.* Comme les composantes linguistique et prosodique possèdent des caractéristiques qui leur sont à chacune spécifiques, la relation entre les deux ne peut vraisemblablement être fondée que sur un modèle simple d'association et de correspondance. Il y a tout lieu de penser que le trait commun à ces deux composantes est celui du contraste quantitatif.

Se pose alors le problème de la granularité à accorder à la paramétrisation prosodique. Examinons la prosodie sur le plan des réalisations phonétiques. Outre les aspects formantiques, les unités phonétiques s'opposent les unes aux autres par le jeu des micro-variations en terme de hauteur, de durée et d'énergie. Cette diversité précisément prend effet sur des oppositions acoustiques relativement faibles dans l'absolu mais cependant hautement différenciées et significatives sur le plan phonétique. Il en résulte finalement une multiplication des niveaux contrastifs pertinents en phonétique et bien adaptés à la finesse de nos organes récepteurs.

Cependant reproduire cette multiplicité de niveaux dans le domaine sémantique aurait pour conséquence de définir sur le plan prosodique des oppositions quantitativement faibles, et finalement d'accéder de manière erronée à l'information phonétique. Au lieu donc de contribuer à la saisie du sens, cette "méthode" confondrait les informations des diverses composantes. Pour traiter de la relation entre sens et prosodie, et pour atteindre la couche de l'information qui lui correspond, il faut donc cibler les oppositions fondamentales qui se situent au-delà des différences phonétiques. En accord avec la recherche sur le plan international, quatre niveaux nous ont paru nécessaires et suffisants pour traiter du sens.

Pour mener à bien ce projet de mise en correspondance entre les signifiés linguistiques et la substance prosodique, il a fallu passer par certaines étapes qui constituent les racines de notre entreprise. Dans un premier temps, des a priori personnels m'avaient engagée d'une part à concevoir un modèle linguistique unique intégrant les aspects syntaxiques et sémantiques, et d'autre part à considérer que le cadre opératoire de la

¹ Les structures profondes dont il est question ici n'ont rien à voir avec l'analyse de la constituance de la grammaire générative. Il s'agit des structures linguistiques et des structures prosodiques fondamentales qui se constituent dans la phase de l'encodage de la parole lors des traitements cognitifs en vue de la production langagière.

relation linguistique / prosodie était la phrase. Les résultats très moyens m'ont fait revoir la méthodologie.

Plus précisément il m'a paru tout d'abord que le cadre de la phrase n'était pas suffisamment fin pour s'adapter à la mobilité du contenu linguistique et pragmatique du discours oral. Le groupe prosodique (syntactique ou pseudo-syntactique), défini a priori à partir du texte pour une base de comparaison stable entre les locuteurs, a donc été retenu comme cadre de l'étude. Ce groupe dans les réalisations mélodiques des locuteurs se comporte soit comme une unité à part entière soit comme un sous-ensemble d'une unité plus vaste. Ainsi tel modèle peut décrire les contours mélodiques aussi bien d'une petite portion de l'énoncé que de la totalité du discours. Dans la plupart des cas toutefois, plusieurs modèles se relaient dans la description des contours mélodiques lexicaux des énoncés. Ce processus qui est naturel puisque dans le texte les concepts véhiculés sont variables (information plus périphérique, plus inattendue, plus complexe etc.), se vérifie aussi statistiquement dans nos résultats. Au niveau de la méthode d'analyse, le groupe représente alors le lieu d'observation de la correspondance modèle / valeurs mélodiques et également l'unité de comptage de nos résultats.

Deuxième amélioration, nous avons pensé par ailleurs que les différents principes syntaxiques, sémantiques et pragmatiques devaient constituer nécessairement des modèles concurrents, et ce, précisément au sein d'un même domaine. C'est ainsi que nous avons défini six *modèles* linguistiques ou sémantico-pragmatiques. Les modèles linguistiques correspondent au principe de la hiérarchisation syntaxique (cf modèle en constituants), de la distance des constituants, de la hiérarchisation sémantique (deux modèles complémentaires reprennent l'organisation thème-rhème mais d'un point de vue non linéaire). Les modèles sémantico-pragmatiques portent sur l'évaluation de la complexité lexicale, et sur celle des informations attendues / inattendues.

Il est intéressant de remarquer que les propos de Ducrot (1984) sont confirmés par les résultats obtenus suite à ce changement de méthode. En effet si les modèles sont bien élaborés sur l'objet théorique qu'est la phrase pour en extraire des significations, ils doivent s'appliquer inversement sur un observable, à savoir les énoncés lus par nos locuteurs. En outre concevoir ces différents modèles comme concurrents, c'est tout simplement reconnaître avec lui que la signification n'est pas un donné, qu'elle est susceptible de plusieurs interprétations, et que ces interprétations ne sont pas autre chose que des hypothèses sur le sens. Chaque type d'interprétation ou chaque hypothèse, autrement dit chaque modèle, dont le choix (sans doute inconscient) résulte de l'exercice d'une compétence linguistique et d'une subjectivité, fonctionne comme un ensemble clos d'instructions auxquelles s'est conformé le locuteur pour attribuer un sens à la phrase à lire, et pour le communiquer à l'auditeur. Ce sont ces instructions-là que l'auditeur est amené à son tour à décoder et dont la prosodie, et en particulier la mélodie, porte trace.

Et de fait sur le plan des résultats, cette amélioration s'est traduite immédiatement : augmentation générale des taux de coïncidences, récurrence des phénomènes au fil des locuteurs, des phrases ou des consignes, très grande homogénéité générale des résultats et enfin évolution "intelligente" des stratégies moyennes en fonction du contenu des phrases et du type de consigne.

Un objectif-clef de l'étude est d'analyser comme nous l'avons vu, les stratégies de lecture. Le texte est un ensemble de signifiés qui forment un tout cohérent. Mais cet ensemble cohérent est laissé à l'interprétation personnelle qui peut faire émerger des sens différents, car tout locuteur en fonction de son histoire personnelle, de ses motivations, des contraintes qu'il perçoit, de son interlocuteur, va extraire un sens préférentiel qu'il

destine à autrui. Dans ces conditions, il n'est pas possible de prédire a priori les stratégies utilisées par le locuteur, ni l'ordre adopté.

Toutefois si la stratégie d'ensemble est imprédictible, il est certainement possible de prévoir les stratégies qui en fonction du contenu du discours et des conditions pragmatiques de réalisation, sont les plus probables. De la sorte si les principales sources d'interprétation possibles sont correctement recensées pour l'ensemble de l'énoncé, il est possible de suivre le cheminement du locuteur tout au long des stratégies successives qu'il emprunte lorsqu'il transmet son interprétation du texte à autrui. Dans notre étude, les stratégies successives ont pour support les différents modèles.

Il était donc important d'une part de prendre un nombre suffisamment élevé de locuteurs pour analyser ce qui est variable et ce qui ne l'est pas ou l'est moins, d'autre part de faire varier les conditions de production des énoncés par l'intermédiaire de différentes consignes de lecture. La situation expérimentale étant celle de lectures sous contraintes (de naturel puis d'intelligibilité), a pour effet d'orienter les motivations du locuteur dans sa lecture, et donc de faciliter une étude psycholinguistique de la prosodie.

Pour étudier la prosodie, deux attitudes sont possibles, soit privilégier une approche extensive, soit privilégier une approche fouillée et en profondeur. Comme on le comprend, notre choix repose sur la deuxième solution.

En effet pour étudier les stratégies sous ces types divers de variabilité, nous n'avions pas intérêt à choisir un texte long, car un texte court permet d'une part de ne pas introduire des paramètres supplémentaires de variabilité (par exemple extralinguistiques) qui pourraient biaiser l'analyse, et d'autre part de mieux maîtriser les signifiés. Autre avantage, ce choix d'un texte court nous permet d'évaluer plus objectivement la méthode d'analyse et finalement, au vu de la très grande cohérence des résultats, de la valider.

Ainsi si nous avons voulu privilégier au contraire l'aspect textuel, une expérimentation de même envergure aurait porté sur un texte de 1800 mots. Mais un tel choix avait plusieurs inconvénients : outre celui de ne pas accorder la priorité à l'aspect parole, il nous privait de l'aspect de la reproductibilité des processus, ensuite de la diversité des stratégies, et qui plus est de la vérification de notre méthode car c'est bien à l'occasion de la diversité des réalisations opérant sur un support textuel identique, que l'on peut efficacement tester les hypothèses que l'on émet sur la structuration linguistique et prosodique. Tout en permettant l'émergence d'une dimension linguistique ou pragmatique, cet espace limité du texte permettait au contraire d'étudier en profondeur le processus de "l'argumentation" prosodique mis en oeuvre.

Dans ces conditions, un texte court (50 unités dont 30 lexicales), extrêmement travaillé à tous les niveaux de l'architecture linguistique et notamment morphosyntaxique et sémantique (cf chapitre II), pouvait du fait qu'il recelait les processus linguistiques fondamentaux d'instanciation du sens, non seulement rivaliser avec un texte plus long, mais aussi nous permettre de tester à plus grande échelle les comportements élocutoires. De ce fait le texte a généré 36 énoncés qui ont été analysés grâce à 27 indices prosodiques.

Du fait de la petitesse du texte, la portée de l'analyse est sans aucun doute limitée. Admettons simplement que constituant une unité à part entière, représentatif d'un certain nombre de processus linguistiques fondamentaux sur les plans syntaxiques et sémantiques, ayant donné lieu à un grand nombre de lectures, donc d'appropriations subjectives diverses du contenu en vue du transfert du sens, ce texte court permet toutefois d'approcher de manière fine certains processus cognitifs d'encodage du sens dans la parole.

Le terme de *modèle* réclame également un commentaire car il est ambigu. Une des hypothèses majeures de notre étude repose sur le fait que les principes linguistiques ou sémantico-pragmatiques, énoncés précédemment, déterminent des contours mélodiques spécifiques au niveau local, puisque le type d'information à véhiculer s'articule sur les mots lexicaux que chaque principe prend en charge selon un point de vue particulier. Selon le cas, le modèle est descriptif et *prédictif* de la configuration mélodique de chacun des mots dans le groupe considéré, voire comme on l'a vu plus haut, d'un ensemble de groupes.

Toutefois si la notion de modèle se justifie totalement sur le plan de la description et de la prédiction des valeurs au niveau local, il n'est en rien, comme on l'a dit ci-dessus, un modèle psycholinguistique permettant d'expliquer une stratégie de lecture qui couvrirait l'ensemble du texte. Tout au plus peut-on dégager des tendances communes à plusieurs locuteurs, voire à une majorité de locuteurs, comme on le montre au chapitre XIV. S'il en était autrement, cela reviendrait à dire que tout texte n'induit qu'une seule interprétation, ce qui est évidemment impensable.

Dans ces conditions, un méta-modèle rendant compte de l'interprétation singulière d'un texte est une idée qui n'est pas recevable. Inversement disposer d'une palette de modèles pour rendre compte des réalisations mélodiques, semble se justifier pleinement dans la mesure où cela permet de prendre en charge la mobilité et la variabilité que les locuteurs mettent en œuvre pour adapter leurs discours au contenu. C'est ainsi dans la succession des modèles qui auront prédit² au mieux les valeurs mélodiques qu'une stratégie mélodique globale du locuteur aura des chances d'être approchée. Pour poursuivre la métaphore, la méthode utiliserait alors une sorte de *palette impressionniste* suffisamment sensible pour tenter de s'adapter à la mobilité des formes acoustiques et la fluidité des intentions sous-jacentes des locuteurs.

Dans cet ouvrage, nous n'aurons pas recours au processus bien connu de déclinaison. Celui-ci (Bolinger, 1964; Cohen and 'T Hart, 1967; Maeda, 1974 ; Pierrehumbert, 1979) se caractérise par un abaissement progressif des valeurs maximales de la phrase. Or nous ne traitons pas ici de la mélodie en relation avec la structure de phrase (donc d'intonation), mais avec le lexique. Comme ce processus de déclinaison n'est pas constant (il peut être remis en cause lorsque le locuteur transmet une information qui lui semble prioritaire), et que les 4 niveaux constituent un maillage suffisamment large pour contenir ces variations, nous avons considéré qu'il n'affecte pas fondamentalement le phénomène de mélodie lexicale.

En résumé, l'intérêt de cet ouvrage réside dans la présentation d'une nouvelle méthode d'analyse de la prosodie, et tout particulièrement de la mélodie. C'est, semble-t-il, la première fois que dans le courant des recherches internationales, la méthode utilisée élargit l'analyse à l'ensemble des unités lexicales, et conduit l'investigation dans plusieurs réseaux par nature imbriqués, comme sont les réseaux syntaxiques, sémantiques et pragmatiques. Dépassant la voie de loin la plus souvent explorée dans les travaux actuels en prosodie, celle du *donné / nouveau* opérée dans une perspective linéaire, la méthode s'ouvre également, pour mieux rendre compte des stratégies inconsciemment mises en place par les locuteurs, à plusieurs réseaux d'analyse concurrents dans le domaine du

² Ce taux de prédiction qui montre dans quelle mesure un modèle est en accord avec les réalisations des locuteurs, est décrit dans nos résultats par la notion de "prédictions satisfaites" (ou encore de taux de "coïncidence").

sens, combinant les approches hiérarchiques (holistiques) et linéaires (analytiques). Cette étude se présente donc comme une recherche de validation de la méthode, et d'autre part comme une tentative d'analyse des processus psycholinguistiques qui gouvernent la production de lecture.

Le plan de l'ouvrage est simple : après avoir présenté les bases de notre analyse sur le plan bibliographique (chapitre I) puis technique (chapitres II et IV), nous définirons les modèles qui dans notre perspective sous-tendent les stratégies successives des locuteurs (chapitre III). Ces modèles définissant des niveaux de hauteur mélodique pour chacun des mots lexicaux, nous étudierons ensuite en fonction des consignes, dans quelle mesure ces modèles s'appliquent à la description des valeurs mélodiques et grâce à quels indices de F0 (chapitres V, VI, VIII, IX, XI, XII). Nous étudierons ensuite les stratégies linguistiques et mélodiques des locuteurs dans les différentes consignes (chapitres VII, X, XIII). Puis nous traiterons des indices de l'énergie (chapitre XV) et des indices de la durée (chapitre XVI). Enfin nous récapitulerons (chapitre XVII) les résultats principaux sur le plan linguistique, et nous ouvrirons quelques pistes de recherche sur le plan pragmatique (chapitre XVIII), en élargissant la perspective à des corpus de parole spontanée. L'ouvrage se clôturera (chapitre XIX) sur l'élargissement actuel du domaine de la prosodie à l'affectivité, émotions et attitudes, tant du point de vue bibliographique que des recherches personnelles. Nous présenterons ainsi l'outil de calcul et de codage automatiques MELISM (Caelen-Haumont et Auran, 2004) intégré sous Praat, syllabant et annotant les proéminences mélodiques, souvent impliquées dans les processus subjectifs. MELISM rend possible une analyse fine et précise des modulations et de ce fait, nous présenterons les premières analyses qui en découlent. Chapitre XX, une annexe générale propose une présentation individuelle des diverses stratégies des locuteurs (modèles / indices), en fonction de chaque consigne.

