

Patterns prosodiques et intentions des locuteurs : le rôle crucial des variables temporelles dans la parole

C. GERARD et C. RIGAUT

Laboratoire de Psychologie Expérimentale, Université René Descartes, URA 316 du CNRS, 28 rue Serpente, 75006 Paris, France

Abstract: All spoken communications imply a particular intent in the transmission of the content of the message, e.g. simple information, question, approval, regret, admiration... The prosodic patterns which are likely to translate such various intents are frequently studied from an "intonative" standpoint (e.g. melodic contours), but less frequently from a temporal standpoint. The purpose of our study is to define better the time-related regulations of speech when the same statements are expressed by the same speakers but with different intents. The affirmative and interrogative forms of such statements by 12 speakers have been used as reference forms for comparison with expressions of the same statements conveying joy, regret, admiration. The average value of F_0 , the melodic contours and the statement durations are measured at the global level of a whole sentence, then at a more local level, word after word, thanks to a sound signal editor, and submitted to statistical analyses. The results provide support for the hypothesis of a gradual construction of specific patterns related to each intention as the message unfolds.

1. INTRODUCTION

La compréhension du langage ne relève pas seulement du décodage phonétique, de l'identification lexicale, des traitements syntaxiques et sémantiques. Une composante dite pragmatique s'y ajoute. La pragmatique forme un sous-domaine assez hétérogène du champ de la psycholinguistique et recouvre à la fois 1 - les connaissances générales partagées par les interlocuteurs, 2 - la dimension implicite du langage (présuppositions, implications), 3 - les règles gouvernant les échanges conversationnels, 4 - les paramètres non linguistiques de la situation d'énonciation, 5 - les intentions et émotions du locuteur. Grice (1) fut l'un des premiers à élaborer une théorie pragmatique supposée rendre compte de la façon dont la signification intentionnelle peut être dérivée de ce qui est explicitement dit, à partir du contexte général d'énonciation. L'idée centrale est que le langage est une activité coopérative qui, par conséquent, obéit aux mêmes lois que toute interaction sociale. Des règles ou "maximes conversationnelles" constituent ces lois ou normes partagées qui gouvernent l'organisation des messages et qui peuvent être éventuellement délibérément violées pour tromper (mensonge) ou pour créer des effets spécifiques (ironie). Dans cette logique, la signification d'un énoncé est dépendante des intentions du locuteur et ce sont ces intentions qui doivent être identifiées pour que le message soit compris. Divers chercheurs, (2, 3) et de façon plus générale, l'école de Genève, ont tenté de systématiser les méthodes d'analyse linguistique des conversations. Un travail analogue doit être envisagé sur le plan non plus du langage mais de la parole: en effet, les mouvements conversationnels témoignant des croyances ou des intentions se marquent par des formes prosodiques spécifiques. La construction de systèmes artificiels susceptibles de simuler différents aspects des performances de production, perception, compréhension (intelligence artificielle) contribue à faire ressentir l'importance de ces questions. "Pour que la communication orale homme-machine devienne une réalité, il est indispensable de progresser dans la connaissance des processus humains d'interprétation de la parole" (4). Pour la

reconnaissance automatique de la parole en effet, on sait que des différences acoustiques importantes existent entre les réalisations d'un même énoncé prononcé par des locuteurs différents ou entre diverses réalisations d'un même énoncé produit par un même locuteur mais à des moments différents. L'utilisation de connaissances lexicales, syntaxiques et sémantiques permet de corriger certaines erreurs ou imprécisions du décodage acoustico-phonétique. Mais pour "comprendre" le message véhiculé par un énoncé, il ne suffit pas d'être linguistiquement compétent, il faut en outre posséder un certain nombre d'informations pragmatiques. C'est également dans cette optique qu'ont été conçus les MGTS (message generation to speech), qui en synthèse annotent le texte à reproduire de marqueurs prosodiques "interprétables par le système de synthèse à partir du texte (TTS) en terme de Fo, pause, durée et intensité du signal associé." (5). Si les MGTS diffèrent des TTS du fait qu'ils sont capables de prendre en compte certaines variables pragmatiques, ils manquent encore de précision quant aux règles prosodiques de l'intentionnalité. Sur le plan de la théorie linguistique, on peut distinguer deux catégories de processus pragmatiques, primaires et secondaires (6, 7). Le premier type de processus renvoie essentiellement à l'instanciation contextuelle des variables, le second type est constitué de processus inférentiels appliqués après compréhension du message et portant sur l'information implicite véhiculée par ce message, en particulier sur les intentions des locuteurs. C'est ce second type de processus qui nous intéresse.

La production, la perception et la compréhension du langage sont envisagées dans le cadre de la psychologie cognitive comme des opérations portant sur des représentations mentales (8). Les formes prosodiques témoignant des intentions des locuteurs pourraient faire partie de ces représentations mentales et servir de formes de référence auxquelles sont comparées les réalisations sonores actuelles d'un énoncé donné. Si ces représentations mentales gouvernent réellement perception et production du discours, des invariants structuraux doivent pouvoir être mis en évidence malgré la variabilité intra- et inter-locuteurs. Notre étude cherche donc à cerner les régulations temporelles et mélodiques du discours lors de l'expression des mêmes énoncés par les mêmes locuteurs mais selon des intentions différentes, et à dégager ces invariants prosodiques. Les formes affirmative et interrogative de ces énoncés réalisés par 12 locuteurs nous servent de formes prosodiques de base auxquelles sont comparées les expressions des mêmes phrases mais manifestant trois autres intentions, la joie, le regret, l'admiration, qui sont ensuite comparées entre elles à leur tour. Les valeurs moyennes de Fo, les contours mélodiques, les marges de variation des valeurs de Fo et les durées d'énonciation sont mesurés au niveau global de la phrase entière et au niveau plus local, mot par mot, grâce à un éditeur du signal sonore, puis soumis à analyses statistiques. Le but de ces analyses est de montrer 1. une stabilité de divers indices au travers des changements de locuteurs aussi bien qu'une stabilité de ces mêmes indices lors des diverses répétitions de la tâche par les mêmes locuteurs, 2. une variation systématique par contre de ces indices d'une intention à l'autre, et surtout, 3. une construction progressive, au cours du déroulement temporel de l'énoncé, des formes prosodiques recherchées.

2. METHODE

Trois phrases de huit syllabes, toutes composées d'un sujet, d'un verbe et d'un complément d'objet ou d'un complément circonstanciel ont été choisies : (P1) "Ils ont découvert le secret"; (P2) "Ils repartent à Londres mardi"; (P3) "On emmène Michel en vacances". Les phrases sont précédées de "mots introducteurs" tirés du langage courant et facilitant l'expression de cinq intentions : *affirmation* introduite par "Tu sais"; *interrogation* introduite par "Quoi ?"; *joie* introduite par "Chouette"; *regret* introduit par "Zut"; *admiration* introduite par "Oh". Les 12 locuteurs (6 hommes et 6 femmes) sont des étudiants en psychologie ne présentant pas de trouble auditif ni articulo-phonatoire. L'enregistrement s'est fait en chambre sourde. Les locuteurs prenaient préalablement connaissance des mots introducteurs et des phrases, inscrits sur des feuilles différentes, pendant quelques minutes. Pendant l'enregistrement, l'expérimentateur présentait chacune des phrases l'une après l'autre et indiquait oralement l'intention dans laquelle la phrase (que le locuteur gardait sous ses yeux) devait être dite, en même temps qu'il présentait au-dessus de celle-ci le mot introducteur de cette intention. Les locuteurs ont énoncé ainsi chaque phrase dans les cinq intentions avant de passer à la phrase suivante. Après une courte pause, les locuteurs ont effectué exactement la même tâche une seconde, puis une troisième fois. L'ordre de présentation des 3 phrases P1 à P3 et l'ordre de présentation des intentions pour chaque phrase étaient contrebalancés sur l'ensemble des 12 sujets. Soient S les locuteurs, X leur sexe, P les 3 phrases, I les 5 intentions, R les trois répétitions, le plan de l'expérience s'écrit: $S6 < X2 > * P3 * I5 * R3$. Après analyse grâce à un éditeur du signal sonore (Unice, Vecsys), les indices suivants ont fait l'objet d'analyses de la variance : 1 - indices "globaux" calculés sur

l'ensemble de la phrase = durée totale, Fo moyenne, contour mélodique (indice de Cooper & Eady), valeur du pic de Fo, gamme de variations de Fo. 2 - indices "locaux" = durée et Fo moyenne de chaque mot de chaque phrase.

3. RESULTATS

Quel que soit l'indice mesuré, global ou local, la répétition de la tâche n'entraîne jamais de différence significative, et ne s'exprime jamais non plus sous forme d'interaction significative avec aucun des autres facteurs (groupe de sujets, type de phrase, type d'intention exprimée). Donc les indices globaux et locaux mesurés "résistent" à la variabilité *intra*-locuteurs lors de la répétition de l'énoncé à un moment ultérieur. Le deuxième résultat essentiel est que, toujours par rapport aux indices que nous mesurons, la variabilité *inter*-locuteurs est également faible. Ainsi par exemple, le facteur "groupe de sujets" (qui renvoie au sexe des sujets) se marque bien sûr par des différences des valeurs de Fo brutes (moyennes, ou pics) mais ne joue ni sur les gammes de Fo par phrase et intention, ni sur les contours, ni sur les durées. Aucune interaction avec les autres facteurs (phrases, intentions, répétitions) n'étant en outre observée, on peut conclure que cette variabilité interlocuteurs s'exerce sans doute plus au niveau segmental que suprasegmental. Le seul et unique facteur significatif (qui donc l'emporte sur les différences inter-individuelles) est de façon systématique le type d'intention exprimée, et ce facteur conduit à des différences significatives répétées et constantes quelque soit le contenu lexical et sémantique de la phrase.

Les analyses "locales" effectuées ensuite avaient pour but de "cibler" les marques de l'intentionnalité, en recherchant si l'un ou l'autre des mots pleins de la phrase (noms, verbes) pouvait à lui seul être porteur des modifications de durée ou de fréquence enregistrées. La durée intrinsèque d'un mot donné, ou sa hauteur tonale moyenne pourraient être seuls responsables des différences significatives globales spécifiques de telle ou telle intention. En ce cas, nous aurions été renvoyées aux marques traditionnelles de l'emphase ou insistance sur tel ou tel mot de contenu. Les résultats des analyses de durée et de fréquence réalisées mot par mot montrent qu'il n'en est rien: un élément donné de l'énoncé n'est que rarement (statistiquement significativement) représentatif d'une intention, à tout le moins de façon suffisamment marquée pour franchir ce que l'on connaît des seuils de détection des auditeurs, et peut même évoluer de façon non conforme au pattern global. C'est donc le cumul au long de l'énoncé de microphénomènes de débit et la sommation "en escalier" au cours du temps de microphénomènes de fréquence qui produit les résultats significatifs des analyses globales. Pour nous en assurer, nous avons en dernier lieu fait porter les analyses sur des segments successifs de plus en plus longs, cumulés à partir du début de la phrase. En procédant ainsi, on voit bien émerger progressivement des différences significatives (tant en fréquence qu'en durée) de plus en plus nombreuses et de plus en plus typiques des différentes intentions au fur et à mesure du déroulement de l'énoncé. Il semble donc bien qu'il existe des formes prosodiques globales correspondant à diverses intentions expressives, et susceptibles d'être décrites par des règles simples, à condition de prendre en considération le déroulement temporel de l'énoncé. Compte tenu du grand nombre de données chiffrées et du type d'effet recherché (le franchissement du seuil de significativité des tests statistiques), le lecteur voudra bien nous excuser de ne pas fournir, dans ces 4 pages, de données quantitatives. L'allure générale des principales formes prosodiques est résumée en conclusion.

4. CONCLUSION

Par rapport à la forme affirmative, la forme interrogative s'est révélée être de durée plus courte et de contour montant. Les interrogatives se sont distinguées de toute autre forme expressive à la fois par le contour et la valeur moyenne de Fo (ce qui est connu depuis longtemps) mais aussi par la durée. La durée d'énonciation des autres intentions, et la fréquence fondamentale moyenne des mots qui composaient la phrase, se sont organisées en formes typiques: la joie était plus longue à exprimer que le regret, l'admiration plus longue que la joie et plus courte que le regret, ce dernier se manifestant aussi par un ton moyen plus grave que la joie et l'admiration, et des marges de variations de Fo plus restreintes. Ces phénomènes ne sont pas locaux, mais se cumulent progressivement tout au long de la phrase, même si ça et là des différences significatives locales apparaissent. C'est ce cumul progressif au cours du temps qui nous a semblé le phénomène essentiel d'expression des intentions, phénomène cohérent par rapport à ce que l'on connaît des nécessités de continuité rythmique et mélodique du discours. Nos buts plus lointains seraient d'établir une loi de variation du débit associé aux montées (ou descentes) progressives du ton, permettant de décrire les formes prosodiques générales que le locuteur utilise pour signaler ses intentions à l'auditeur. Les intentions que nous avons étudiées ont été envisagées comme des

produits des conventions culturelles, et non comme des témoins d'une implication affective du locuteur dans l'énoncé. Il faut avouer qu'une répugnance manifeste existe chez les psychologues cognitivistes à traiter des émotions. Les raisons profondes de cette répugnance méritent cependant analyse et ceux qui la surmontent ont déjà fourni des données intéressantes (9, 10, 11, 12). Ce n'est pas en commençant par amputer les processus de communication de leur composante émotionnelle que l'on améliorera la connaissance de la prosodie des dialogues. Chacun sait qu'une prosodie "naturelle" contribue à améliorer l'efficacité et la résistance à la dégradation du signal parlé. C'est donc une affaire à suivre que l'intégration des composantes intentionnelles et émotionnelles de l'énoncé dans les processus prosodiques naturels.

Références:

- (1) Grice, H.P. Logic and conversation, in P. Cole and J.L. Morgan, Eds, Syntax and Semantics, vol. 3, Speech Acts. (1975) N.Y: Academic Press.
- (2) Jayez, J. Un aspect de l'analyse automatique de conversations: croyances, mouvements conversationnels, mouvements discursifs, *Cahiers de Linguistique Française*, **10**, (1989), 147-170.
- (3) Moeschler, J. L'analyse pragmatique des conversations, *Cahiers de Linguistique Française*, **12**, (1991), 7-30.
- (4) Pierrel J.P. & Carbonel N. La reconnaissance automatique de la parole, *Le Courrier du CNRS*, **79** (1993) p.11
- (5) COZANET, A. Message to speech, Note interne, *Centre National d'Etudes des Télécommunications*, (1991).
- (6) Récanati, F. Insinuation et sous-entendu, *Communications*, **30**, (1979) 95-106.
- (7) Récanati, F. La pragmatique linguistique, *Le courrier du CNRS*, **79** p.21.
- (8) Sorin, C. Perception de la parole continue, in M.C. Botte, G. Canevet, L. Demany, & C. Sorin, *Psychoacoustique et perception auditive*, série audition, INSERM/SFA/CNET, (1989) 123-139.
- (9) Scherer K.R. Vocal Affect Expression: a review and a model for future research, *Psychological Bulletin*, **99**, (1986) 143-165.
- (10) Scherer K. et Zei B. La voix comme indice affectif, *Revue médicale de la Suisse Romande*, **109**, (1989) 61-66.
- (11) Johnson, W.F., Emde, R.N., Scherer, K.R. and Klinnert, M.D. Recognition of emotion from vocal cues, *Arch Gen Psychiatry*, **43**, (1986).
- (12) Scherer, K.R.; Les émotions: fonctions et composantes, *Cahiers de Psychologie Cognitive*, **4**, 1, (1984) 9-39.