

LES MARQUEURS ACOUSTIQUES DE L'ÉNONCÉ EN FRANÇAIS QUÉBÉCOIS

C. Ouellon, C. Paradis et L. Duchesne

CIRAL, Université Laval

ABSTRACT

We intend to identify some acoustic parameters involved as sentence boundary markers in spontaneous speech. The function of the subsequent pause as a marker is emphasized as well as particular aspects of the energy curve and comparison of F_0 values calculated on the vowel nucleus before and after pauses which appear to be important in the perception of sentences, at least in the idiolects considered in this study.

1. INTRODUCTION

Les travaux de recherche que nous poursuivons dans le cadre du projet PROSO visent à mettre en évidence les facteurs d'ordre acoustique qui caractérisent les énoncés du discours oral spontané en français québécois. Nous faisons l'hypothèse que la variation des facteurs acoustiques favorise la variation des taux de perception de la frontière d'un énoncé.

L'analyse des marqueurs acoustiques nécessite, dans la perspective retenue, que le discours soit au préalable découpé en énoncés. Dans une première étape, nous avons donc procédé à la délimitation d'unités théoriques, en retenant une approche inspirée de Nespors et Vogel [3] selon lesquelles le discours s'organise en un ensemble fini d'unités phonologiques organisées hiérarchiquement. De ces unités, le

syntagme intonatif (SI) et l'énoncé (E) nous intéressent tout particulièrement. Le SI est formé par l'assemblage de groupes phonologiques; il est le domaine d'un contour d'intonation et ses limites coïncident avec les positions où peuvent intervenir des pauses. L'unité de niveau supérieur E est constituée d'un ou plusieurs SI et occupe habituellement la longueur de la chaîne dominée par le plus haut noeu de l'arbre syntaxique [3]. D'autres approches de définition du SI ont été considérées [6].

Comme nous voulions dégager les facteurs d'ordre acoustique susceptibles de favoriser la perception des frontières d'unités en discours spontané, nous avons en second lieu soumis le corpus au jugement d'un groupe d'auditeurs, ce qui devrait permettre de dégager des unités concrètes du discours oral spontané et de les mettre en rapport avec les unités abstraites issues de l'application du modèle prosodique de Nespors et Vogel [3].

2. MÉTHODE D'ANALYSE

2.1 Le corpus

Le corpus d'analyse contient dix extraits tirés de deux entrevues réalisées dans la région de Chicoutimi-Jonquière (Québec-Canada) à l'occasion d'une enquête sociolinguistique [5]. Les deux locuteurs (cinq extraits pour chacun) sont de sexe masculin et

appartiennent à la classe moyenne. Chacun des extraits a une durée moyenne de dix secondes et ne comporte aucune interruption de la part d'un tiers. Le corpus a ensuite été numérisé à partir du logiciel d'analyse CSL de Kay Elemetrics.

2.2 Le test d'audition

La bande sonore soumise au test d'audition présentait trois enregistrements de chaque extrait, les deux premiers ayant été filtrés (résidu LPC) de façon à rendre inintelligible le message, le troisième étant la sortie numérisée originale de l'extrait. Vingt-trois juges, des étudiants de niveau universitaire, ont reçu la consigne d'indiquer par marquage électronique le moment qu'ils jugeaient correspondre à un début d'énoncé. Le protocole était inspiré de celui proposé par Lehiste [2] et Kreiman [1]. Au résultat, le corpus a été découpé en 57 énoncés, avec des taux d'accord sur la perception de frontières variant de 4.3% à 91.3% selon l'énoncé. Une procédure de normalisation a été appliquée pour tenir compte des divergences de temps de réaction des divers juges.

3. LE RAPPORT ENTRE ÉNONCÉS PERCUS ET UNITÉS PHONOLOGIQUES

Malgré le nombre relativement restreint de juges (23) et la taille modeste du corpus (10 extraits de 10 secondes chacun), certaines tendances ressortent de l'analyse. En premier lieu, on constate que les débuts d'énoncés perçus (dorénavant EP) coïncident toujours avec des débuts de SI, mais pas nécessairement avec des E (unités phonologiques de niveau supérieur). En second lieu, on peut noter une certaine hiérarchie entre les SI, en lien avec leur statut par rapport à E. Ainsi, le taux moyen de

perception décroît lorsqu'on passe d'un EP qui correspond à un SI constituant à lui seul un E (51.7%), puis à un EP coïncidant avec un SI qui occupe le premier rang d'un groupe de SI formant un même E (43.1%) et enfin à un EP correspondant à un SI de rang n dans le groupe de SI relevant d'un même E (16.4%). Comme les écarts-types sont relativement élevés, il ne peut s'agir là que de tendances qui sont de nature à appuyer l'existence d'une interaction entre données perceptuelles et modèle phonologique.

4. L'ANALYSE DE FACTEURS ACOUSTIQUES

4.1 Procédure

Les extraits numérisés ont été analysés à l'aide du logiciel CSL. Après avoir délimité les unités phonétiques sur les tracés oscillographiques et spectrographiques et sur les listes numériques, nous avons analysé les paramètres suivants:

- la durée des pauses
- la courbe de F_0
- la courbe d'énergie.

Nous avons tenu compte de la variation des valeurs de F_0 et d'énergie à l'intérieur même d'un EP et entre deux EP. Ce choix s'explique par l'objectif que nous nous sommes fixé, ce qui impliquait qu'il faille porter une attention plus grande aux phénomènes qui surviennent aux frontières de cet énoncé.

4.2 La fréquence

On admet communément que les énoncés se caractérisent, en contexte énonciatif, par la décroissance de la valeur de F_0 en finale [8]. Nos données confirment cette observation et concordent avec celles que nous avons présentées au cours d'une recherche antérieure [4].

On calcule une chute moyenne de 2.1 tons entre le Fo maximal de l'énoncé et le Fo du dernier noyau vocalique. Toutefois, dans la perspective où nous nous plaçons, il ne semble pas que ce facteur suffise à marquer la frontière de l'EP; il ne paraît guère exister de lien entre l'importance de la décroissance des valeurs de Fo et le degré de perception de la frontière d'un EP ($r=.085$), dans les corpus filtrés et non filtré.

D'un autre point de vue, la mesure de la pente de Fo donne, comme attendu, une pente négative (-2 tons/sec.) sans qu'il y ait de relation significative entre la variation de pente et la variation du taux de perception ($r=.06$).

Enfin, l'analyse des relations entre frontières de deux énoncés successifs fait voir une hausse de fréquence moyenne de .05 ton, qui confirme l'hypothèse de la réinitialisation de Fo en début d'un EP dans des énoncés consécutifs. On ne note cependant pas de relation entre les variations de Fo, dans ce contexte, et le taux de perception ($r=.02$).

4.3 La courbe d'énergie

Nous avons déjà signalé [4] que la variation de la courbe d'énergie semblait jouer un rôle dans la perception des frontières d'EP, en ce sens que les frontières d'énoncés à haut taux de perception étaient marquées par une chute d'énergie plus forte; cette chute correspond à la chute d'énergie en dB entre la voyelle affichant la plus haute valeur de l'énoncé et la dernière voyelle. Il existe peu d'études sur la variation d'énergie en discours oral, a fortiori en discours oral spontané. Nous avons quand même voulu vérifier la validité des résultats présentés dans Ouellon 90 [4].

Il semble effectivement exister une relation entre les valeurs de chute

d'énergie dans un énoncé et le taux de perception de la frontière de cet énoncé. Pour chacune des deux versions du corpus (filtrée et non filtrée), le calcul de la régression linéaire permet d'obtenir des données comparables, avec $r=-.3061$ et $r=-.3829$ respectivement. La droite résultante fait voir une chute d'énergie qui va de -5.2 dB à -8.8 dB pour des taux de perception qui vont de 20% à 100%. De telles variations d'énergie paraissent significatives, une variation de 1 à 2 dB par rapport au signal naturel étant auditivement détectable [7].

Comme la finale d'énoncé semble marquée par une chute d'énergie, on peut faire l'hypothèse qu'il y aura augmentation d'énergie entre la finale d'un énoncé et le début de l'énoncé subséquent, ou réinitialisation de la courbe d'énergie. Nous avons donc examiné les valeurs d'énergie de part et d'autre de la frontière d'EP. Au résultat, nous pouvons observer, en premier lieu, un écart positif important entre les valeurs d'énergie mesurées sur la voyelle finale de l'EP puis sur la voyelle initiale de l'EP suivant; cette augmentation est de l'ordre de +2dB pour les frontières d'EP à 20% et de +8dB pour celles d'EP à 100%. Donc, à une forte chute d'énergie en finale correspond une forte augmentation d'énergie en initiale d'EP subséquent.

En second lieu, le calcul de la régression linéaire fait voir une relation ($r=.47$) entre le niveau d'augmentation de l'énergie entre deux EP et le taux de perception des frontières.

On peut donc estimer que la variation de la courbe d'énergie dans un EP et entre deux EP favorise la perception des limites d'EP, du moins dans le style de discours et dans les idiolectes du

corpus analysé.

4.4 La pause

L'importance de la fonction de marquage de la pause a été maintes fois signalée, dans Vaissière 1988 [8] entre autres. Dans Ouellon 1990 [4], nous posions l'hypothèse qu'il y avait également une relation entre durée de la pause et taux de perception de l'énoncé. Notre recherche valide cette hypothèse. En effet, la durée des pauses varie entre 122ms et 980ms dans nos exemples et il existe une forte corrélation ($r=.7264$) entre la durée des pauses et le taux de perception des frontières d'EP dans les corpus filtrés et non filtré. La pertinence de la fonction marquage de la pause est confirmée, mais il semble aussi que l'allongement de la durée des pauses favorise un meilleur taux de perception de la frontière d'EP.

5. CONCLUSION

Nous avons fait ressortir l'importance des facteurs que sont la chute de la courbe d'énergie dans un EP, l'augmentation des valeurs d'énergie au passage d'un EP à un autre, de même que la durée des pauses pour le repérage des frontières d'énoncés. Les facteurs liés aux variations de Fo ne paraissent pas, curieusement, jouer un rôle particulier dans la perspective que nous avons privilégiée d'analyser les marqueurs acoustiques de frontières d'EP en regard du taux de perception des énoncés.

Il serait sans doute intéressant de vérifier si nos observations peuvent s'appliquer à d'autres idiolectes français en contexte de discours oral spontané. D'un autre point de vue, nous pensons qu'il faudra solutionner le problème du découpage en unités du discours spontané et approfondir les

connaissances sur l'interaction entre perception et grammaire.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] KREIMAN, J. (1982), "Perception of sentences and paragraph boundaries in natural conversation", *Journal of phonetics*, 10.
- [2] LEHISTE, I ET W.WANG (1976), "Perception of sentence boundaries with and without semantic information", *Phonologica*, Dressler-Pfeiffer.
- [3] NESPOR, M. ET I.VOGEL (1986), "Prosodic Phonology", Dordrecht-Hollande, Foris publications.
- [4] OUELLON, C. (1990), "Les marqueurs acoustiques de l'énoncé en discours québécois spontané", *Actes des XVIIIèmes JEP*, Université de Montréal.
- [5] PARADIS, C. (1985), "An acoustic study of variation and change on the vowel system of Chicoutimi and Jonquière (Québec)", thèse de Ph.D inédite, University of Pennsylvania.
- [6] POIRÉ, F., J.M.SOSA, H.PERREAULT et H.J.CEDERGREN (1990), "Le syntagme intonatif en langage spontané; étude préliminaire", *Revue québécoise de linguistique*, 19-2, Université du Québec à Montréal.
- [7] SORIN, C. (1981), "Functions, role and treatment of intensity in speech", *Journal of phonetics*, 9.
- [8] VAISSIERE, J. (1988), "The use of prosodic parameters in automatic speech recognition", *Recent advances in speech understanding and dialog systems*, NATO ASE, F46, Springer Verlag, Berlin.