

Les indices mélodiques de la reconnaissance de la question en français

Monique Vion & Annie Colas*

Université de Provence CNRS UMR 6057 - Parole et Langage,
équipe Psycholinguistique
monique.vion@up.univ-mrs.fr

ABSTRACT

Linguistics studies of intonational devices of total questions in French showed that, in questions containing more than two stress groups, interrogative intonation is characterized by a sequence of lowered pitches or downstepped tones which precedes the final rise. The gating paradigm was used to determine whether subjects who are listening French NP utterance containing three stress groups can indicate whether the utterance is an affirmation or a question before the final rise is available.

Although the task was difficult, findings indicate that listeners are effectively able to recognize, up to a certain extent, with mid confidence ratings, the intonational device of a question while they are hearing the downstepped tones preceding the final rise.

1. INTRODUCTION

Di Cristo & Hirst [3] ont analysé les régularités prosodiques caractéristiques de différents types de questions en français (les questions qui appellent une réponse par « oui » ou par « non » « tu veux du fromage ? » ; les questions elliptiques « et du dessert ? » ; celles qui proposent un choix alternatif « tu veux du fromage ou du dessert ? » et celles qui comportent un détachement à gauche « du dessert, tu en veux aussi ? »). Leur analyse contraste deux types de contours finaux. Alors que les questions qui appellent une réponse par oui ou par non ainsi que le premier terme des questions alternatives (tu veux du fromage ...) et la partie rhématique des questions qui comportent un détachement à gauche (tu en veux aussi ?) présentent un contour final ascendant, les questions elliptiques ainsi que le deuxième terme des questions alternatives et la partie thématique des questions à détachement à gauche (du dessert...) présentent un contour final descendant. En outre quel que soit le type de la question, lorsque les unités intonatives comportent plus de deux unités tonales, la description d'ensemble des variations de hauteur, montre la présence d'abaissements successifs pour les unités tonales intermédiaires. Lorsque la question est formée d'une ou deux unités tonales, le patron est similaire à celui de l'affirmation à l'exception de la montée finale. Lorsque la question comporte plus de deux unités tonales le patron des unités tonales intermédiaires est caractérisé par une séquence d'abaissements successifs de la hauteur. Selon les auteurs, ce patron d'abaissements successifs de la hauteur pour les unités tonales intermédiaires, pour ne pas être seulement lié à un contour ascendant subséquent,

constituerait une propriété phonologique de la question en français et pas seulement une caractéristique liée à l'anticipation du mouvement ascendant.

Nombreux sont les cas dans les échanges verbaux ordinaires où l'intonation porte à elle seule l'indication de la modalité de phrase et de l'intention du locuteur (« Tu as bien dormi ? »). Le but de la présente expérience est de montrer que les abaissements successifs de la hauteur qui précèdent le contour final constituent à eux seuls des indices pertinents pour l'auditeur tout venant lorsqu'il reconnaît l'énoncé d'une question. Cette recherche des indices prosodiques permettant de détecter au plus tôt la modalité de tels énoncés constitue également, au plan des applications, une étape préalable à la mise au point de systèmes de compréhension de la parole.

2. METHODE

L'expérience utilise la méthode de dévoilement graduel du stimulus (*gating*) (Grosjean [4]). Une série de stimulus comportant des portions de plus en plus longues d'une phrase est constituée en partant à chaque fois du début du signal. Le premier stimulus est très bref et le dernier correspond à la totalité de la phrase. L'auditeur a pour tâche d'identifier successivement chaque stimulus et d'exprimer sa confiance dans la réponse donnée.

2.1. Matériel

Les phrases à partir desquelles les huit séries de stimulus test et les trois séries de stimulus d'entraînement ont été constituées sont des phrases sans verbe, composées de trois unités tonales (exemple : « le début du roman de Zola »). Chaque unité tonale comporte trois syllabes de type Consonne voisée/Voyelle.

Chaque phrase a été enregistrée deux fois par un locuteur féminin (l'un des auteurs) : une fois avec l'intonation de l'affirmation et une fois avec l'intonation de la question. Les figures 1 et 2 présentent chacune en haut, la courbe de variation de la fréquence fondamentale, et en bas le signal acoustique. Les enregistrements ont été vérifiés du point de vue de l'homogénéité de leur prononciation et de leur conformité aux patrons linguistiques du français ; à savoir pour l'affirmation : " *a basic pitch pattern (ie an Intonation Unit) ending on a final low pitch (such as a statement) contains a rising pitch movement (from low to high) at the end of each stress-group, except the last which is pronounced with a falling pitch movement from the mid (or lowered high)*

to low." Di Cristo [2], p 201; et pour la question ("an initial peak situated on the last syllable of the first stress-group followed by a declining pitch until the final rise." Di Cristo [2], p 203. Voir dans Vion & Colas [8] le détail des études préalables.

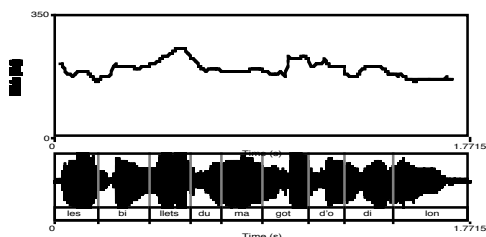


Figure 1: Affirmation. Analyse acoustique de la phrase « les billets du magot d'Odilon ».

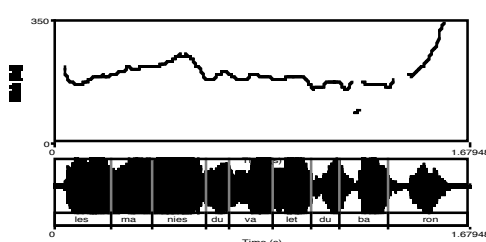


Figure 2 : Question. Analyse acoustique de la phrase « les manies du valet du baron ».

Pour chaque enregistrement, un ensemble de onze stimulus a été constitué. Le premier stimulus (s1) comporte les trois syllabes de la première unité tonale (UT1). Les cinq stimulus suivants comportent, outre les trois syllabes de s1 un fragment progressivement de plus en plus long des syllabes qui constituent la seconde unité tonale (UT2). Le pas d'accroissement est de 20 % : les stimulus s2, s3, s4, s5 et s6 comportent respectivement 20 %, 40 %, 60 %, 80 % et 100 % de la durée de l'UT2. Les cinq derniers stimulus comportent outre les six syllabes des deux premières unités tonales un fragment progressivement de plus en plus long des syllabes qui constituent la troisième unité tonale (UT3). Le pas d'accroissement est de 20 % : les stimulus s7, s8, s9, s10 et s11 comportent respectivement 20 %, 40 %, 60 %, 80 % et 100 % de la durée de l'UT3. Au dernier stimulus (s11) la totalité de la phrase est disponible. La confrontation par paires des stimulus A et Q constitués à chaque pas d'incrémentations montre que, quel que soit le contenu des phrases, les paires s5 et s6 (80% et 100% de l'UT2) ainsi que les paires s10 et s11 (80% et 100% de l'UT3) présentent à l'audition des différences mélodiques nettement perceptibles. Il est à noter que pour les séries Q, s9 offre la totalité du patron d'abaissements.

2.2. Procédure

La passation, pilotée par ordinateur, est individuelle. L'expérimentateur délivre la consigne page d'écran après page d'écran et la commente si besoin. Puis il invite l'auditeur à placer les écouteurs qui lui permettront d'entendre les stimulus. L'auditeur est ensuite confronté aux trois items d'entraînement, puis aux huit items de

l'expérience proprement dite. Pour chaque stimulus, le programme enregistre la réponse d'identification, la réponse de certitude ainsi que le temps mis pour donner chacune des réponses. L'expérience dure en moyenne trente minutes.

Les essais d'entraînement ainsi que ceux de l'expérience proprement dite ont la même structure. Un stimulus est d'abord diffusé. Immédiatement après la fin de celui-ci, une première page d'écran invite à donner au clavier une réponse d'identification en choix forcé (« A » ou « Q »). Simultanément les deux versions possibles de la phrase s'écrivent l'une en dessous de l'autre dans la partie haute de l'écran (l'affirmation est précédée de l'étiquette A et comporte un point final ; la question est précédée de l'étiquette Q et comporte un point d'interrogation). Une fois la réponse d'identification saisie, une seconde page d'écran s'affiche. Elle invite à exprimer (en appuyant sur l'une des 9 touches du pavé numérique) le degré de confiance en la réponse donnée (1 signifie : "pas sûr du tout" et 9 : "tout à fait sûr"). Une troisième page d'écran invite alors à passer au stimulus suivant en appuyant sur la barre d'espace. En fin d'item, une page d'écran indique la fin de ce dernier et invite à passer à l'item suivant en appuyant sur la barre d'espace.

Les programmes, écrits grâce au logiciel PsyScope (Cohen, Mac Whinney, Flatt et Provost [1]), gèrent l'ensemble du recueil. Les affirmations et les questions sont présentées dans un ordre tiré au hasard sur l'ensemble des huit items.

2.3. Facteurs et plan

Deux variables sont manipulées : l'intonation (Affirmation vs. Question) et la quantité d'information entrante (de s1 - première unité tonale - à s11 - totalité de la phrase). Les variables qui concernent : l'identité du locuteur (voix féminine) ; le nombre d'unités tonales des phrases (trois unités) et la taille des unités tonales (trois syllabes) sont maintenues constantes. À ces deux ensembles de variables, s'ajoutent des variables de contrôle : la liste d'items, qui contrebalance les éventuels effets liés au contenu des phrases (un même contenu est présenté dans une liste dans sa version A et dans l'autre liste dans sa version Q) ; l'ordre d'affichage des phrases à l'écran, qui contrebalance les éventuels effets d'ordre d'apparition (au-dessus vs. au-dessous) des phrases ; l'ordre des touches de réponse sur le clavier, qui contrebalance les éventuels effets d'ordre (Gauche - Droite) de disposition des touches de réponse.

Chaque participant est confronté aux deux types d'intonation. Il entend chaque contenu sous une seule des intonations possibles (soit 4A et 4Q). Il est confronté à un ordre d'affichage des phrases à l'écran lors de l'invitation au choix forcé. Il est confronté à une seule disposition des touches de réponse au clavier.

2.4. Participants

La passation de l'expérience a donc nécessité au total la constitution de huit groupes de quatre participants. Trente deux jeunes adultes ayant le français comme

langue première, étudiants de l'Université de Provence ont participé à l'expérience.

2.5. Attentes

Pour les phrases du matériel, qui comportent trois unités tonales, l'hypothèse est que l'auditeur peut commencer à différencier le patron de la question de celui de l'affirmation dès l'audition des abaissements de hauteur de l'unité tonale intermédiaire. Les attentes concernent donc l'évolution des réponses au fur et à mesure du dévoilement de l'information prosodique supposée discriminante. Si au cours du dévoilement des phrases, les réponses données sont liées au traitement de l'information prosodique pertinente, le nombre d'identifications exactes devrait augmenter au fur et à mesure que s'accroît la taille de la partie du signal porteuse des abaissements de hauteur.

3. RESULTATS

Pour étudier l'évolution des réponses au cours des essais, les huit séries de réponse données par chaque participant (formant chacune un patron de réponse à onze places) ont été examinées. Si, pour une série de stimulus, l'auditeur a décidé de répondre d'emblée « A » et de maintenir ce type de réponse jusqu'à preuve trouvée dans le stimulus que ce choix ne convient pas et si la série proposée est une série A, alors les réponses sont toutes pertinentes. Si confronté à une série Q l'auditeur n'a pas découvert d'indices contraires à son choix initial « A », le patron des réponses est une suite de réponses non pertinentes. Si l'auditeur a modifié à une ou à plusieurs reprises son choix au cours des essais, le patron des réponses est mixte. Il en va de même pour un auditeur qui aurait décidé de répondre d'emblée « Q » au premier stimulus de la série.

L'examen des 256 patrons (8 items x 32 participants) montre que, dans 38,3 % des cas, les participants ne modifient pas leur choix initial au cours de l'audition de la série de stimulus. Ils répondent systématiquement plutôt « A » (28,9 %) que « Q » (9,4 %). Cette observation est une indication de la difficulté des auditeurs à focaliser leur attention sur les caractéristiques prosodiques offertes par les stimulus. Ils donnent une réponse par défaut : a priori le stimulus provient plutôt d'une phrase A que d'une phrase Q.

Malgré cela, des patrons mixtes se rencontrent chez tous les participants (61,7 % des patrons). Ces patrons peuvent être le résultat ou bien d'un changement de la réponse au cours de la série de stimulus non fondée sur des indices perceptifs ou bien du prélèvement en un point de la série d'un indice en faveur de l'une des intonations possibles que les stimulus suivants confortent éventuellement. Quoi qu'il en soit, deux types de patrons mixtes peuvent être observés. Ou bien le patron se présente comme une suite qui alterne des réponses pertinentes et non-pertinentes (patron "irrégulier"). Ou bien le patron commence par des réponses non-pertinentes et se poursuit ainsi jusqu'à ce que pour un stimulus donné une réponse pertinente émerge et se stabilise ultérieurement (patron "régulier").

Les patrons irréguliers, bien que globalement minoritaires, représentent près de la moitié des patrons mixtes (48,8 %). Ils sont plutôt observés en réponse aux séries A. Alors que les patrons réguliers qui sont majoritaires (53,16 % des patrons), sont plutôt observés en réponse aux séries Q.

En outre, dix participants donnent des patrons réguliers en réponse à une série A alors que vingt-neuf en donnent en réponse à une série Q. La présence de patrons réguliers donnés en réponse aux séries Q par une large majorité d'auditeurs (90,6 % des participants), permet de rechercher quelle est la quantité d'information prosodique entrante qui les a conduits à reconnaître une question.

Sur les 73 patrons réguliers observés en réponse à une série Q, 31 sont tels, que la première réponse pertinente est donnée pour le dernier stimulus. Ce résultat souligne le caractère crucial pour l'auditeur de l'information prosodique portée par la toute fin du contour ascendant. La distribution des 42 autres patrons est bimodale. La présentation des stimulus s3 et s6 a été l'occasion de donner la première réponse pertinente de la série pour près de la moitié des patrons restants. Alors que la réponse au stimulus s6 (100 % de l'UT2) peut être expliqué par la présence de différences prosodiques perceptibles entre les stimulus A et Q en ce point, la réponse donnée au stimulus s3 ne peut pas être expliquée de cette manière. Elle semble plutôt relever d'un souci de ne pas répondre toujours pareil en l'absence d'indice valide.

L'analyse de la variance conduite avec le facteur *Quantité d'information entrante* et la fréquence moyenne de la réponse pertinente en chaque point de la série Q pour variable dépendante montre un effet significatif de ce facteur [$F(10,280) = 42,244$; $p < .00001$]. Une comparaison par paires au moyen du test Neuman-Keuls signale (au seuil de .01) un ensemble de différences significatives entre certaines paires (figure 3).

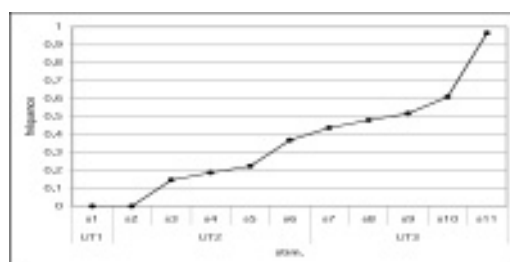


Figure 3 : Séries Q. Patrons réguliers : évolution de la fréquence moyenne des réponses pertinentes au cours des essais.

En premier lieu la comparaison confirme le caractère décisif de l'audition de la fin du contour ascendant. La fréquence de la réponse donnée pour le stimulus s11 est très élevée (.97) et elle diffère significativement des fréquences observées pour tous les autres stimulus. Dans ce stimulus, l'information mélodique prévaut (l'information segmentale est déjà toute disponible dans le stimulus précédent, s10). Il est à noter aussi que la fréquence de la réponse pertinente est seulement égale à

.61 pour le stimulus s10 et dépasse .50 seulement à partir du neuvième stimulus (.51), soit après que 60 % de la troisième unité tonale ait été entendue. En ce point, la totalité des abaissements qui précèdent la montée finale est disponible. En second lieu, la fréquence de la réponse observée pour le stimulus s6 qui représente 100 % de la deuxième unité tonale (.37), est supérieure et diffère significativement de celle observée pour les stimulus s4 et s3. Et la fréquence observée pour le stimulus s5 (.22), qui représente 80 % de l'UT2, diffère significativement de celle observée pour les stimulus s1 et s2. Dans la mesure où les paires de stimulus A - Q en s5 et s6 présentent des différences mélodiques nettement perceptibles, ces deux derniers résultats vont dans le sens de l'hypothèse d'un accroissement de la fréquence de la réponse pertinente lié à la détection de caractéristiques mélodiques propres à l'organisation prosodique de la deuxième unité tonale.

Du point de vue de l'expression de la confiance sur l'échelle en 9 points, l'analyse de la variance conduite selon le même plan que ci-dessus montre que le degré de certitude augmente au fur et à mesure des essais et la comparaison par paires au moyen du test Neuman-Keuls montre (au seuil de .01) l'ensemble de différences suivant (figure 4).

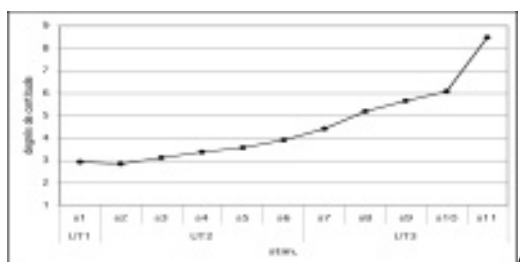


Figure 4 : Séries Q. Patrons réguliers : évolution du degré moyen de certitude au cours des essais.

Le degré de certitude exprimé à l'issue de la réponse au dernier stimulus, qui est très élevé (8,5), diffère significativement de celui donné pour tous les autres stimulus. De même, le degré de certitude exprimé à l'issue des stimulus s10 et s9 (respectivement 6,1 et 5,6) diffère significativement de celui donné à l'issue des stimulus s1 à s7. Il est à noter que le degré de certitude associé à la réponse d'identification donnée pour le stimulus s7 se situe au centre de l'échelle (4,4) et ne diffère pas significativement de celui associé aux réponses d'identification données pour les stimulus s5 et s6.

CONCLUSION

Le principal résultat souligné par la présente expérience est que la reconnaissance du marquage prosodique de la modalité de phrase dans le cadre du paradigme du dévoilement graduel est une tâche difficile pour les auditeurs tout-venant. Tous les auditeurs ont montré des persévérations dans l'emploi d'une réponse stratégique initiale pour certaines séries de stimulus. Interrogés à l'issue de la session expérimentale, nombre d'entre eux confirment le caractère stratégique de leurs réponses initiales.

Malgré la difficulté de la tâche, tous les auditeurs ont également montré qu'ils étaient attentifs aux informations dérivées de l'entrée sensorielle. En premier lieu, ainsi que l'ont fait des études pour d'autres langues que le Français comme le Suédois, le Hollandais, l'Anglais Américain et Australien (Hadding-Koch [5] ; Hadding-Koch & Studdert-Kennedy [6] ; Wales & Taylor [9] ; Swerts, Bouwhuis & Collier [7]), les présents résultats confirment le rôle décisif du contour final. En second lieu, l'analyse des patrons réguliers montre qu'une partie des réponses aux séries Q s'organise selon un patron compatible avec la découverte progressive de la prosodie du signal d'entrée. Les auditeurs qui reconnaissent une question à l'audition des abaissements mélodiques qui précèdent le contour final ne sont que moyennement sûrs des réponses qu'ils donnent. Ainsi, malgré le fait largement déclaré de modifier leur stratégie de réponse initiale à la perception d'une montée mélodique, il semble bien que les auditeurs se soient également fondés sur les abaissements successifs qui précèdent le contour final pour reconnaître l'énoncé d'une question.

BIBLIOGRAPHIE

- [1] J. Cohen, B. MacWhinney, M. Flatt. and J. Provost. An interactive graphical system for designing and controlling experiments in the psychological laboratory using MacIntosh computers. *Behavior Methods, Research, Instruments and Computer*, 25: 257-271, 1993.
- [2] A. Di Cristo. Intonation in French. In D. Hirst & A. D. Cristo (Eds.), *Intonation systems : A survey of twenty languages* (pp. 195-218).: Cambridge University Press, Cambridge;1998.
- [3] A. Di Cristo and D. J. Hirst. *Prosodic regularities in the surface structure of French questions*. Paper presented at the European Speech Communication Association Workshop on Prosody, Lund.1993.
- [4] F. Grosjean. Gating. *Language and Cognitive Processes*, 11: 597-604, 1996.
- [5] K. Hadding-Koch. *Acoustico-phonetic studies in the intonation of southern Swedish*. Lund: Gleerup, 1961.
- [6] K. Hadding-Koch and M. Studdert-Kennedy. An experimental study of some intonation contours. *Phonetica*, 11; 175-185, 1964.
- [7] M. Swerts, D. Bouwhuis and R. Collier. Melodic cues to the perceived « finality » of utterances. *Journal of Acoustical Society of America*, 96(4): 2064-2075,1994.
- [8] M. Vion et A. Colas. La reconnaissance du pattern prosodique de la question : questions de méthode. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, 21: 83-106, 2002.
- [9] R. Wales and S. Taylor. Intonation cues to questions and statements: how are they perceived? *Language and Speech*, 30: 199-211, 1987.