

Perception de la colère dans un corpus de français spontané par des apprenants portugais et tchèques

Sophie de Abreu, Catherine Mathon, Daniela Perekopska

EA333 “Atelier de Recherches sur la Parole”
Université Paris 7 – Denis Diderot
2 place Jussieu, 75251 Paris Cedex 05 France
{deabreu ; mathon ; perekopska}@linguist.jussieu.fr

ABSTRACT

The aim of this paper is to show how prosody can provide a sufficient amount of information which allows recognizing the emotion of anger in a French spontaneous corpus for foreign learners of French. We present the results of a perception test carried on two groups of foreign learners of French, Portuguese and Czech. They had to listen to French sentences and evaluate the presence, or lack thereof, and the degree of anger in these sentences. We chose to use a real spontaneous corpus in order to keep the intonation of emotions in French intact. The semantic content was neutralised to put aside the information given by the content.

1. INTRODUCTION

L'étude que nous présentons ici est à l'intersection de deux domaines de recherche : l'émotion et ses manifestations vocales d'une part et l'enseignement apprentissage du français langue étrangère d'autre part. La communication relie ces deux domaines. C'est une des pierres angulaires de la recherche en FLE [1]. Or la communication passe aussi par l'interprétation de l'émotion [2]. Néanmoins, il y a peu d'études concernant la perception et la production des émotions en enseignement des langues étrangères. Il s'agit pourtant d'une compétence qu'il faut qu'un apprenant d'une langue étrangère acquière afin de réagir de manière appropriée dans des situations de communication quotidienne où l'émotion est présente. Le but de cette communication est de montrer l'importance de la prosodie dans la détection des émotions dans la parole spontanée. Cette étude fait partie d'un projet plus large qui consiste à évaluer la compétence perceptuelle d'étudiants de FLE à identifier des émotions. Nous voulons souligner la nécessité d'inclure la prosodie dans l'enseignement des langues. Nous présentons ici l'élaboration d'un test de perception basé sur la parole spontanée en français, et qui montre toute l'importance de la prosodie dans la perception de l'émotion.

2. PROTOCOLE EXPÉRIMENTAL

Le but de notre étude est de montrer que la prosodie contient suffisamment d'informations pour renseigner

sur l'état émotionnel du locuteur. Nous avons émis l'hypothèse que la prosodie était un des paramètres de l'expressivité des émotions. Nous avons aussi voulu évaluer la capacité d'apprenants d'une langue étrangère à percevoir une émotion avec pour seule information la prosodie de la langue. Pour finir, nous avons émis l'hypothèse que les locuteurs de différents groupes de langue perçoivent différemment la colère.

2.1. Description du corpus d'origine

Nous avons choisi d'utiliser un corpus de parole spontanée pour construire notre test de perception. Nous voulions travailler sur de l'émotion réelle et non pas jouée. Une série de canulars radiophoniques nous a permis d'extraire un corpus intéressant. Un animateur radio appelle des professionnels ou des institutions. Il joue le rôle d'un client et demande un service qui ne correspond pas à ce que propose le professionnel. Il crée ainsi une situation d'incompréhension et de mésentente, qui entraîne une réaction de colère de la part de la victime. Nous avons collecté 24 canulars, ce qui correspond à 1 heure et 4 minutes de parole. Nous en avons éliminé 9, qui ne contenaient pas suffisamment de colère. Nous avons transcrit orthographiquement les 15 restants à l'aide de Transcriber 4.0. Nous avons ensuite annoté chaque tour de parole de chacun de ces dialogues selon trois catégories : Colère (C) associée à un chiffre de 1 à 5, marquant l'intensité de la colère perçue; Neutre (N) pour décrire un tour de parole dans lequel aucune expressivité particulière n'est perceptible; Autre Emotion (AE) qui décrit un tour de parole dans lequel le locuteur exprime une émotion autre que la colère. Cette dernière annotation est également associée à une échelle de 1 à 5. Cette annotation préliminaire, basée sur notre expérience de locuteurs natifs du français et de spécialistes en phonétique, nous a permis de faire une première sélection des tours de parole.

2.2. Sélection des phrases (pré-test)

A partir de cette première annotation, nous avons effectué un pré-test perceptif pour déterminer l'état émotionnel perçu dans chacun des tours de parole que nous avons sélectionné. Nous avons demandé à 5 sujets francophones natifs d'écouter 81 tours de parole. Ils

devaient juger si l'état émotionnel contenu dans les énoncés qu'ils entendaient était de la colère, une autre émotion ou neutre. En fonction des jugements obtenus à ce pré-test : nous avons choisi 13 énoncés clairement perçus comme étant de la colère à au moins 80% et 13 autres comme ne contenant pas de la colère.

3. TEST DE PERCEPTION

3.1. Stimuli

Pour montrer que la prosodie était suffisante pour percevoir la colère contenue dans nos énoncés, nous avons décidé d'isoler la prosodie du contenu segmental. Celui-ci est ainsi neutralisé, de manière à ce qu'il n'interfère pas dans la perception de l'émotion. Pour ce faire, nous avons décidé de masquer le contenu segmental avec du bruit blanc [3]. Ce n'est sans doute pas la méthode de masquage la plus couramment employée, mais elle proposait une alternative plus acceptable que les autres à nos yeux. L'une de ces méthodes consiste à re-synthétiser les énoncés. Nous étions toutefois très attachées à garder l'aspect spontané de notre corpus et nous ne voulions pas le modifier en le re-synthétisant. L'autre méthode consiste en un filtre passe-bas. C'est une bonne méthode de masquage. Mais elle coupe l'énergie dans les hautes fréquences. Or, l'énergie dans les hautes fréquences est un des paramètres acoustiques de la colère [4]. Et il nous paraissait important de conserver ce paramètre dans un premier temps. Nous avons donc préféré la méthode de masquage avec le bruit blanc. Pour chaque énoncé que nous avons choisi à partir du pré-test, nous avons créé un bruit blanc de même durée avec Soundforge 7.0, que nous avons mixé au son d'origine. Nous avons réglé l'intensité du bruit blanc de manière à ce qu'il cache l'information linguistique sans masquer pour autant la voix du locuteur. Ces stimuli ont été doublés, c'est-à-dire présentés deux fois aux sujets, afin de vérifier pour chaque sujet la cohérence des réponses. Nous avons également rajouté 5 stimuli d'entraînement, afin que les sujets puissent s'accoutumer à la sonorité particulière des stimuli, perçus comme des sons de « mauvaise qualité ».

3.2. Sujets

Nous avons demandé à 4 groupes de locuteurs d'effectuer le test de perception : 2 groupes de francophones natifs, et 2 groupes d'apprenants en Français Langue Etrangère.

Le groupe contrôle 1 (Français - sans bruit) est constitué de locuteurs francophones natifs (6 femmes et 4 hommes) ayant passé le test avec les sons originaux, sans addition de bruit blanc. Il s'agissait d'établir les réponses correctes au test. Le groupe contrôle 2 (Français - avec bruit) est constitué de 10 locuteurs francophones (6 femmes et 4 hommes) ayant passé le test avec des stimuli

masqués par le bruit.

Le premier groupe testé est composé de 7 femmes et 3 hommes locuteurs du portugais, étudiants à l'Université de Lisbonne, le second de 10 locuteurs tchèques (8 femmes et 2 hommes), étudiants à l'Université de Pilsen. Les deux groupes d'apprenants ont été choisis par les enseignants en fonction de leur niveau, B2 selon le Portfolio Européen des Langues [5]. Aucun n'était déjà venu en France.

3.3. Tâche

Nous avons demandé aux auditeurs d'effectuer une double tâche : une tâche de décision et une tâche d'évaluation.

La première tâche consistait à décider si le son entendu véhiculait de la colère ou non. Si oui, il s'agissait ensuite de déterminer le degré de colère sur une échelle de 1 (faible) à 5 (fort).

Les auditeurs étaient préalablement informés de la mauvaise qualité des sons entendus, et ce afin d'éviter une adaptation trop longue à ces stimuli.

3.4. Interface

Le test de perception a été présenté sous format électronique à l'aide du logiciel EasyPhp. Nous y avons attaché une importance particulière afin de contrôler au maximum l'impact de la forme prise par le test sur les réponses des auditeurs.

Le format électronique nous a notamment permis de contrôler toutes les étapes effectuées par les auditeurs : le temps passé sur chaque décision, les hésitations, sont mesurés, les retours en arrière (incontrôlables avec un test papier) sont impossibles, les auditeurs doivent donner une réponse pour chaque son avant de passer au suivant.

Par ailleurs, nous avons veillé à ce que le test soit pertinent du point de vue de la didactique des langues étrangères. Ainsi, nous avons présenté les consignes dans la langue maternelle (LM) des auditeurs de sorte que celles-ci soient bien comprises. Les stimuli proposés étant en français, nous avons choisi de ne pas utiliser la LM des auditeurs pendant la phase de test. Nous leur avons donc proposé de donner leurs réponses avec des icônes afin d'éviter toute surcharge cognitive due aux multiples changements de code. L'icône « écouter le son » 🎧 disparaissait après deux écoutes successives (ou après la décision), les icônes « Colère » 😡 et « pas colère » 😐 apparaissaient au moment d'effectuer le choix concernant l'émotion du stimulus. Pour finir, l'échelle de 1 à 5 était présentée en écriture numérique. Cette interface, relativement flexible, devrait nous permettre d'effectuer d'autres tests avec d'autres émotions et d'autres groupes de locuteurs.

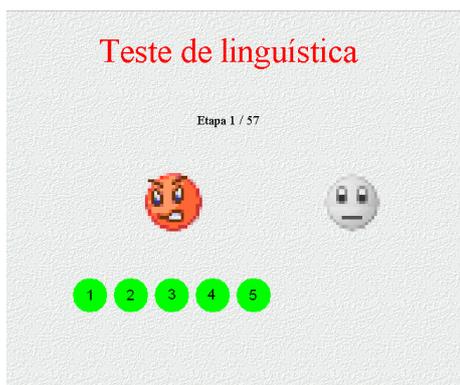


Figure 1 : Exemple d'une étape du test

4. RÉSULTATS

Nous allons présenter dans cette section les résultats obtenus jusqu'ici. Pour commencer, nous avons vérifié la cohérence entre les réponses données à la première et à la seconde écoute des stimuli en effectuant un test de corrélation de Spearman (ordonné par rang) avec le logiciel Statview 5.0. La table 1 montre que cette corrélation est significative, c'est-à-dire que les réponses des 4 groupes sont cohérentes pour les deux écoutes.

Table 1 : Résultats tests de corrélation de Spearman.

Groupe	ρ_s	p
Français – sans bruit	0.763	$p < 0.0001$
Français – avec bruit	0.848	$p < 0.0001$
Portugais FLE	0.833	$p < 0.0001$
Tchèques FLE	0.691	$p < 0.0001$

Nous avons ensuite comparé les réponses des auditeurs ayant passé le test bruité avec les réponses des francophones dans le cas des sons non bruités. Ce test nous a permis de montrer que les 3 groupes reconnaissent la colère malgré le bruit : les 13 phrases de colère ont été reconnues comme telles à 84 % par les Français et les Tchèques, et à 77% par les Portugais. On peut donc conclure que la prosodie seule permet de reconnaître cette émotion. Dans un second temps, nous avons examiné la catégorisation de la colère en degrés par les locuteurs.

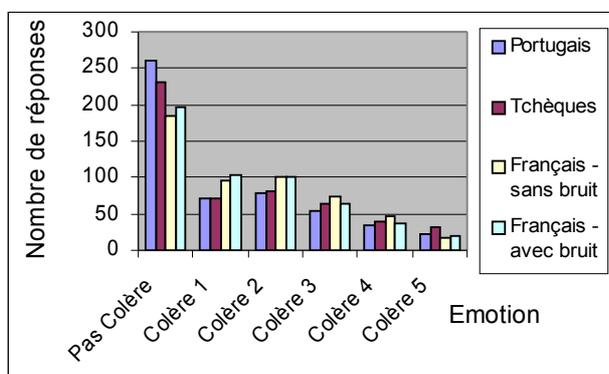


Figure 2 : Nombre et type de jugement par groupe d'auditeurs

Nous pouvons observer dans la figure 2 que les auditeurs ont répondu « Pas colère » dans une plus grande proportion que les autres réponses, ce qui est conforme à ce qui était attendu, la moitié des stimuli proposés ne véhiculant pas de colère. En revanche, on peut noter une différence de l'ordre de 10 % entre les réponses des francophones et celles des apprenants de FLE, différence répercutée dans les catégories « Colère 1 » et « Colère 2 ». Ceci semble indiquer une différence de perception du degré de colère. Tandis que les francophones natifs vont disperser leurs réponses de manière plus homogène entre les différentes catégories de colère faible - Colère 1 (C1) et Colère 2 (C2) - les auditeurs non natifs (ici Tchèques et Portugais) vont percevoir la faible colère comme n'étant pas du tout de la colère. De la même façon, les auditeurs tchèques ont émis un jugement de forte colère (C4 - C5) plus souvent que les Français et les Portugais. Un test statistique sur l'ensemble des résultats (test de Friedman) indique qu'il n'y a pas de différence suffisamment grande pour être significative entre les réponses globales des différents groupes. Cela confirme notre hypothèse première : la prosodie semble suffisante d'après ce test de perception, pour percevoir la colère. Cependant, des tests plus ponctuels (tests t) nous indiquent qu'il y a parfois une différence de perception entre les groupes. Ainsi, la différence entre les réponses des Tchèques et des francophones (test avec bruit) pour la Colère 1 est significative, $p=0.0207$. Ceci corrobore notre interprétation précédente, les Tchèques ont une perception plus approximative du degré de colère que les francophones natifs. La différence entre les Français (test sans bruit) et les Portugais pour la Colère 3 et la Colère 4 est également significative (respectivement $p=0.0338$ et $p=0,0301$). Ce résultat nous indique que les informations sémantiques ont probablement influencé les locuteurs francophones lors du test non bruité pour ces catégories. Des phrases qui ont été jugées comme Colère 3 et 4 dans les conditions de test non bruité ont été catégorisées comme Colère 1 ou 2 lorsque les auditeurs n'avaient pas accès au sens. On peut supposer que lorsque l'information émotionnelle est suffisamment marquée sémantiquement, la prosodie est quant à elle moins marquée car elle serait alors redondante.

5. ANALYSE PROSODIQUE

Afin de déterminer quels étaient les paramètres prosodiques qui pourraient servir d'indices pour la perception de la colère, nous avons procédé à quelques mesures globales, notamment sur la F0 et sur l'intensité. Nous ne présenterons ici que des résultats préliminaires. Une analyse plus affinée est en cours.

Nous avons mesuré pour chacun de nos énoncés la moyenne de la F0, les valeurs minimale et maximale, l'amplitude de la F0. Nous avons fait les moyennes de ces mesures par type de locuteur, homme ou femme, en opposant d'une part les énoncés reconnus comme étant de la colère et d'autre part ceux reconnus comme n'en

étant pas. Si l'on compare les productions des hommes pour les deux catégories, on s'aperçoit qu'au niveau de la F0, il y a une réelle différence. Les énoncés reconnus comme « colère » présentent une moyenne et une amplitude de F0 plus élevées que les énoncés perçus comme « pas colère ». Au niveau de la F0 moyenne, il y a une différence de 40 Hz (environ 4 demi-tons) et pour l'amplitude la différence s'élève à presque 60 Hz (soit trois demi-tons et un quart). En revanche, pour les femmes, il n'y a aucune différence au niveau de la F0 entre les énoncés « colère » et les énoncés « pas colère ». Nous pouvons cependant expliquer ce résultat par le fait que nous avons moins de locutrices que de locuteurs dans notre corpus. Il est donc plus difficile d'établir un résultat certain.

Ces mesures préliminaires globales nécessitent une recherche plus approfondie. Il serait notamment intéressant de voir les différences de registres entre chaque énoncé de « colère » et le registre moyen du locuteur qui prononce cet énoncé. Il nous paraît en effet difficile d'obtenir des résultats stables, en isolant complètement l'énoncé produit du contexte du dialogue. Cette analyse est en cours. Toutefois, nous avons déjà quelques exemples intéressants de courbes mélodiques qui correspondent bien aux patrons de la colère.



Figure 3 : Courbe mélodique d'un énoncé reconnu comme « colère » (Winpitch Pro)

La figure ci-dessus montre la courbe mélodique obtenue avec Winpitch Pro d'un énoncé, prononcé par un locuteur masculin, reconnu comme étant de la colère forte (C4-C5) à 84%. En observant cette courbe, on remarque un contour de F0 très abrupt, avec des montées et des descentes brutales [6]. L'amplitude de F0 sur certaines syllabes est de l'ordre de 100 Hz. La variation de F0 est donc très importante. Le registre de la voix du locuteur pour cet énoncé est très étendu : il passe de 88Hz à 318 Hz (22 demi-tons de différence) dans le même énoncé. Si d'un point de vue global les mesures prises pour l'instant ne sont pas très significatives, en s'intéressant à une analyse plus fine de chaque énoncé, de véritables différences devraient apparaître, notamment au niveau de la courbe de F0.

6. CONCLUSION ET PERSPECTIVES

Cette étude nous a permis de montrer que la prosodie jouait un rôle important dans la perception de la colère en français spontané. Elle a mis également en avant la capacité d'apprenants de FLE à percevoir l'émotion dans la langue étrangère.

Cette étude n'est que la première partie d'un projet plus global. Nous souhaiterions dans un second temps, travailler sur la reproduction des émotions par les apprenants de FLE, en essayant de voir les interférences de la langue maternelle dans leurs réalisations. D'autre part, nous souhaitons mener quelques études de perception secondaires : nous allons tester par exemple les différentes méthodes de masquage du son, pour déterminer quelle est la plus adaptée à notre étude. Un test est en cours avec un filtre passe-bas.

Ce test a été créé dans l'idée d'être adaptable à des apprenants de différentes langues maternelles, et à des émotions différentes, et nous aimerions dans un proche avenir pouvoir le développer.

BIBLIOGRAPHIE

- [1] J. Beale, Is communicative language teaching a thing of the past ?, *Babel*, Vol. 37, No. 1, Winter 2002, pp. 12-6.
- [2] E. Galazzi et E. Guimbretière, Intonation et attitudes: une question de perception, *Studi di Linguistica, Storia della lingua Filologia francesi*, Edizioni dell'Orso, Milan, 1994.
- [3] Miller and Licklider, The intelligibility of interrupted speech, *J. Acoust. Soc. Am.* 22: 167-173, 1950)
- [4] T. Bänziger, and K. R. Scherer, Relations entre caractéristiques vocales perçues et émotions attribuées. Actes des Journées Prosodie 2001, Grenoble, France, 119-124, 2003.
- [5] http://www.enpc.fr/fr/international/eleves_etrangers/portfolio.pdf
- [6] P. Léon, Précis de phonostylistique : Parole et expressivité, Nathan, Paris, France, 1993.