



HAL
open science

Une approche méthodologique de la reconnaissance de l'origine géographique de locuteurs arabophones

N. Riaibi, E. Lhote

► **To cite this version:**

N. Riaibi, E. Lhote. Une approche méthodologique de la reconnaissance de l'origine géographique de locuteurs arabophones. *Journal de Physique IV Proceedings*, 1994, 04 (C5), pp.C5-513-C5-516. 10.1051/jp4:19945109 . jpa-00252783

HAL Id: jpa-00252783

<https://hal.science/jpa-00252783>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une approche méthodologique de la reconnaissance de l'origine géographique de locuteurs arabophones

N. RJAIBI et E. LHOTE

*Laboratoire d'Analyse de la Parole, Centre de Linguistique Appliquée, 6 rue Gabriel Plançon,
25030 Besançon, France*

Abstract

The purpose of this study is to search a means for geographical characterization of Arabic speakers when speaking standard Arabic. Different approaches are used to attempt to define a methodology of the identification of geographic origin by means of dialectal marks in standard Arabic. 114 utterances produced by 18 arabic subjects have been studied by means of sonagraphic analysis and 16 utterances by long term average spectrum (LTAS). The sentences have been studied by interspectral correlation coefficient. Among several approaches, the LTAS appears as a reliable means of characterization.

1. Introduction

Quand une population se reconnaît dans une langue commune, la description de la langue parlée doit répondre à une double exigence : mettre en évidence les propriétés invariantes dans le groupe qui conditionnent la communication et déterminer les marques de variabilité qui caractérisent des sous-groupes.

Abou Haidar [1] a tenté de décrire le fonds commun de la langue arabe actuelle en commençant par le système vocalique, considérant que s'il existe un consensus réel concernant la perception linguistique de l'arabe standard, il devrait être possible de construire un test d'audiométrie vocale permettant d'évaluer la déficience auditive de tout public arabophone.

Le travail présenté ici vise à caractériser des aspects liés à la variabilité dialectale arabe, et plus spécialement ceux qui sont liés à l'origine géographique des locuteurs. Nous avons confronté dans un premier temps les observations dégagées dans des études à caractère historique à une analyse phonologique de type synchronique [2]. Nous proposons une double approche acoustique : la première, que nous appellerons analyse locale, est appliquée à la syllabe au moyen d'un CSL (Kay Elemetrics). La seconde, dite approche globale, exploite la notion de spectre moyen appliquée à des énoncés longs. C'est à partir de la corrélation entre les deux types de résultats que nous proposons d'envisager la caractérisation de l'origine géographique des arabophones.

2. Les variantes dialectales dégagées par les approches diachronique et synchronique

Une première approche diachronique traite de la variabilité des parlers des anciennes tribus arabes; elle est fondée sur les études de deux grammairiens des VIII^e et X^e siècles après J.-C (Sibawayhi et Ibn Jinni) [3]-[4]. Cette étude a permis de constater que certaines variantes consonantiques, détectées par l'écoute dans les parlers actuels avaient des racines dans les anciens parlers de la Presqu'île arabique; les variantes les plus systématiques concernent en particulier les fricatives interdentes (/θ/, /ð/, /ʕ/).

Une étude synchronique des systèmes phonologiques a porté sur 8 parlers arabes (Yémen, Iraq,

Liban, Syrie, Egypte, Soudan, Tunisie, Maroc) et a confirmé que chaque pays possède des variantes correspondant aux fricatives interdentes (tableau 1).

Tableau 1: Les variantes dialectales possibles des 3 fricatives interdentes arabes.

Phonème	/θ/					/ð/			/ʕ/		
Variantes	[θ]	[s]	[t]	[θ]	[ð]	[z]	[d]	[ð]	[ʕ]	[d]	[ʕ]

Pour vérifier la présence de ces variantes, des tests perceptifs sur des passages radiophoniques extraits de 6 radios arabes (Tunisie, Maroc, Egypte, Emirats Arabes Unis, Algérie et Syrie) ont été effectués auprès de 6 auditeurs arabophones non phonéticiens dont la tâche consistait à essayer de reconnaître l'origine géographique des speakers en indiquant dans la mesure du possible les critères ayant permis cette reconnaissance. Les résultats sont relativement satisfaisants puisque le taux de reconnaissance est de l'ordre de 90%.

3. Analyse acoustique locale

Une analyse acoustique segmentale au niveau de la syllabe a été effectuée afin de vérifier la présence des variantes dialectales. Le corpus était constitué de 114 énoncés en arabe standard répétés 3 fois dans une chambre insonorisée, la distance bouche-micro étant maintenue constante. Les informateurs étaient au nombre de 18 choisis dans 9 pays arabes. Les résultats de cette approche ont démontré que certains locuteurs maintenaient la prononciation normative neutralisant ainsi leurs différences dialectales.

Il est toutefois intéressant de comparer les sonagrammes obtenus à partir des réalisations du même mot prononcé par des locuteurs d'origines géographiques différentes. L'observation de la majorité des tracés a démontré l'existence des variantes annoncées précédemment. Les figures 1 et 2 illustrent des prononciations différentes du mot arabe /ʒuðɑ:m/¹: elles présentent la réalisation d'un seul locuteur de chacun des 2 pays : l'Algérie et l'Iraq. La fricative interdente sonore se manifeste habituellement par la présence de zones de résonance rappelant la structure formantique des voyelles [5].

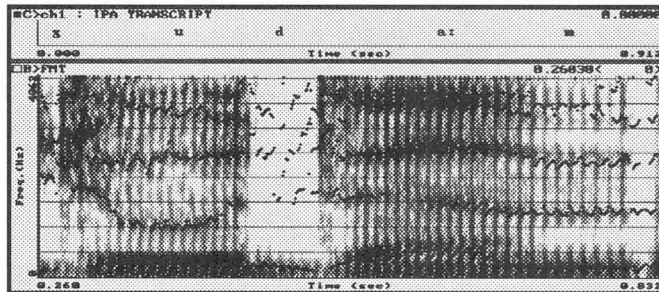


Figure 1: Le mot /juðɑ:m/ prononcé par un locuteur algérien.

La figure 1 dénote la présence de l'occlusive sonore [d], remplaçant la consonne [ð].

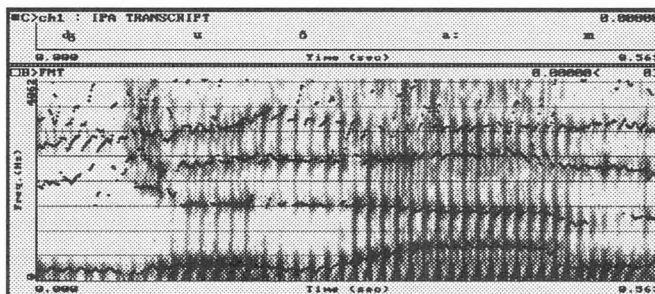


Figure 2: Le mot /juðɑ:m/ prononcé par un locuteur iraquien.

Sur la figure 2 on peut voir le maintien de [ð] par le locuteur iraquien.²

1. /ʒuðɑ:m/: maladie contagieuse.

2. Les fricatives interdentes sont maintenues dans le parler iraquien, elles n'ont pas de variantes dialectales.

Bien que ces variantes aient été détectées chez la majorité des locuteurs enregistrés, l'utilisation du niveau segmental pour la caractérisation géographique ne paraît pas suffisante étant donné que les locuteurs peuvent changer de comportement linguistique et effacer des traces dialectales en revenant à une prononciation normative.

4. Analyse acoustique globale

Nous avons soumis l'ensemble des énoncés à une analyse par spectre à long terme et nous avons calculé les spectres moyens [6]-[7] en appliquant un coefficient de pondération avant de procéder à l'analyse statistique et aux représentations graphiques.

Le traitement statistique était destiné à des comparaisons intra et inter-textes, intra et inter-locuteurs de même origine géographique et entre des locuteurs d'origines géographiques différentes. C'est le coefficient de corrélation interspectrale qui a été retenu pour la comparaison des spectres. Rappelons que le coefficient de corrélation r est un indice de similarité variant entre -1 et +1 et qui permet de préciser le degré de relation linéaire entre les séries statistiques (valeurs d'intensité des SMLT des 2 locuteurs) à comparer, il est obtenu par l'équation suivante :

$$r = \left(\sum_{i=1}^{257} (x_i - \bar{X}) (y_i - \bar{Y}) \right) / \sqrt{\sum_{i=1}^{257} x_i - \bar{X}^2 \sum_{i=1}^{257} y_i - \bar{Y}^2} \quad (\text{Eq.1})$$

où \bar{X} et \bar{Y} sont les moyennes arithmétiques des séries statistiques x_i et y_i associées au caractère intensité exprimé en dB.

Les comparaisons intra-locuteur, inter-textes (figure 3)¹ ont révélé une similarité dans le nombre et la position des pics. Ce résultat a permis de déterminer dans un premier temps, un spectre caractéristique du locuteur.

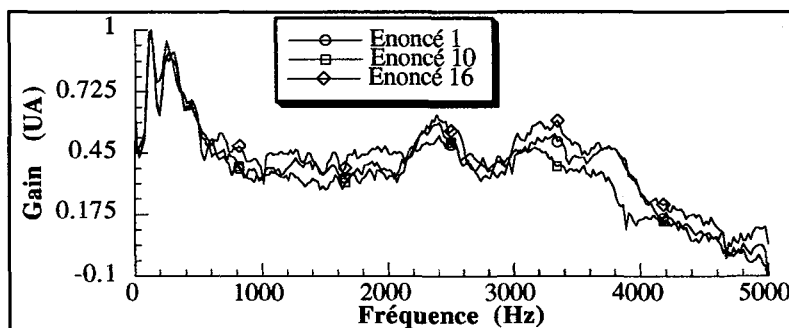


Figure 3: Comparaison inter-textes, intra-locuteur.

Ensuite, nous avons voulu vérifier si le SMLT pouvait être caractéristique du pays. Pour cela les SMLT moyens des locuteurs du même pays ont été comparés et le résultat a montré une similarité dans la position des pics et des zones stables (figure 4).

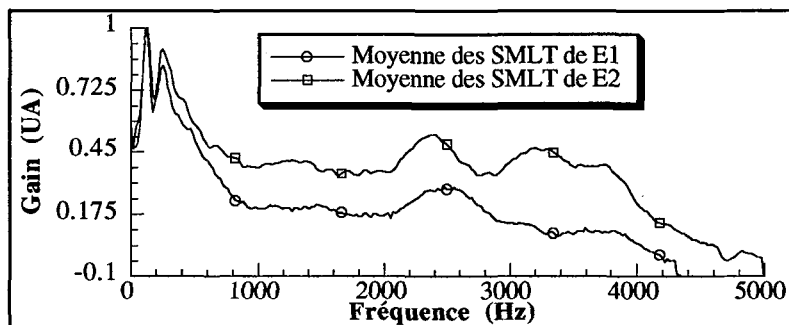


Figure 4: Comparaison inter-locuteurs, intra-pays.

1. Nous représentons ici l'exemple des locuteurs égyptiens, ils sont représentés par E suivi de 1 ou 2.

Les comparaisons inter-locuteurs intra-pays ont abouti à des coefficients de corrélation significatifs permettant de confirmer l'appartenance des locuteurs à la même origine géographique. (figure 5).

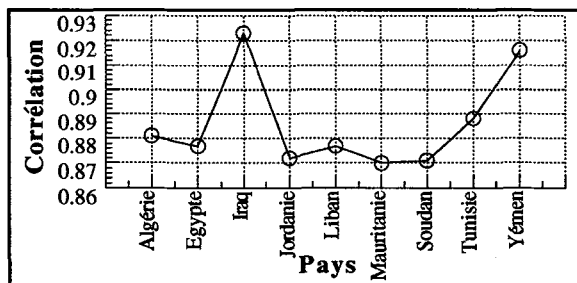


Figure 5: Coefficients de corrélation intra-pays.

4.1. Conclusion

Les différentes approches utilisées se complètent dans la mesure où elles ont fourni les premiers éléments d'une approche acoustique complexe. Si l'on recherche des caractéristiques géographiques il n'est pas bon de se contenter d'une approche locale. Ce travail fait apparaître la nécessité de faire appel à une approche bi-dimensionnelle (locale et globale) qui prenne en compte à la fois des caractères des sous-groupes (dialectes) et des individus.

Références Bibliographiques

- [1] ABOU HAIDAR L., 1991, Variabilité et invariance du système vocalique de l'arabe standard, Thèse de Doctorat, UFR des Sciences du Langage, de l'Homme et de la Société, Besançon, 390p.
- [2] RJAIBI N., 1993, Approches historique, phonologique et acoustique de la variabilité dialectale arabe, Thèse de Doctorat, UFR des Sciences du Langage, de l'Homme et de la Société, Besançon, Tome I : 419p, Tome II : 457p.
- [3] IBN JINNI, IVe siècle, *الخصائص* (Al *Ḥaṣā'ir*) 3 vol. ouvrage en langue arabe: Les particularités, Dār al Kutub al Maṣriyya, éd. Dār al Kitāb al 'arabi, Beyrouth, Liban. Le Caire, 1952.
- [4] SIBAWAYHI, VIIIe siècle, *الكتاب* (Le Livre). Réimpression de l'édition originale de Abdou s-Salam Mouhammad Haroun avec l'autorisation de l'éditeur, Imprimerie Al-Madani, Al-Mou'assassa As-Sa'oudiyya, Le Caire, distribué par Dar Suḥnoun Lilnachr wa-t-Tawzi', Tunis, 1990, Tomes I, II, III, IV et V.
- [5] AL-ANI S. H., 1970, Arabic Phonology: An Acoustical and Physiological Investigation. Indiana University, Série Practica, The Hague, Mouton, Paris. 104 p.
- [6] HARMEGNIES B. et LANDERCY A., 1986, Comparison of spectral similarity indices for speakers recognition, 12th International Congress of Acoustics, 2p.
- [7] HARMEGNIES B. et LANDERCY A., 1988, Intra-speaker Variability of the Long Term Speech Spectrum, Speech Communication 7, North-Holland, 81-86.