

## Les cycles de sonorité comme paradigme de reconnaissance automatique des emprunts lexicaux formels en arabe standard

Sonority cycles as a paradigm for automatic recognition of formal lexical borrowings in standard Arabic

دورات النغمة كنموذج للتعرف التلقائي على المقترضات المعجمية الصورية في اللغة العربية الفصحى

\* Hicham Ouardi

Langues, Littératures, Arts et Cultures (LLAC),  
Université Mohammed V de Rabat, Maroc  
University of Mohammed V de Rabat, Morocco  
hicham.ouardi@flsh.um5.ac.ma

d/dep: 24/02/2022

d/ acc : 02/04/2022

d/ pub : 02/06/2022

### Résumé :

La problématique de la reconnaissance des emprunts formels, qu'elle soit automatique ou assistée, se veut l'une des questions les plus épineuses de la linguistique appliquée. En effet, la systématisation du processus d'identification exige un traitement fin des structures morphophonologiques afin d'en extraire des variables susceptibles de définir plus nettement celles jugées exogènes. Dans cette perspective, le présent papier s'interroge sur la portée sonore des radicaux trilitères en arabe standard. Par le biais d'une approche quantitative puisant dans un corpus phono lexical comptant 6512 racines trilitères, nous nous placerons sur le plan de la perception et essayerons de dévoiler les corrélats structurant les différentes lexies endogènes et exogènes au système sonore de l'arabe. Notre objectif est de bâtir le terrain à une systématisation de la reconnaissance automatique des emprunts formels trilitères sur la base des substrats sonores des entités testées.

**Mots-clés :** Emprunt formel, identification, corpus, échelle de sonorité, perception.

### Abstract :

The problem of formal borrowing recognition, whether automatic or assisted, is one of the most challenging issues in applied linguistics. Indeed, the systematization of the identification process requires a fine processing of morphophonological structures in order to extract from them variables likely to define more clearly those considered exogenous. In this perspective, the present paper investigates the sound scope of trilateral radicals in Standard Arabic.

\* Hicham ouardi: hicham.ouardi@flsh.um5.ac.ma

Through a quantitative approach drawing on a phono-lexical corpus of 6512 trilateral roots, we will place ourselves on the level of perception and try to unveil the correlates structuring the different lexies endogenous and exogenous to the sound system of Arabic. Our objective is to build the ground for a systematization of the automatic recognition of trilateral formal borrowings based on the sound substrates of the tested entities.

**Keywords :** Formal borrowing, identification, corpus, sonority scale, perception.



## I. Introduction

La question d'identification des lexies exogènes à un système linguistique est l'une des problématiques les plus rigoureuses de la linguistique contemporaine. Certes, les travaux qui s'y sont investis en vue de mettre la main sur certains corrélats révélateurs de l'aspect allogène ne datent pas d'hier ; mais, la majorité des résultats obtenus ne dépassent l'aspect classificatoire et descriptif des données linguistiques sans aboutir de façon satisfaisante à une interdépendance formelle ou substantielle susceptible de servir de moyen à une reconnaissance automatique des lexies étrangères.

Dans le cadre de cet article, nous tenterons de dévoiler la corrélation entre la structure sonore des observables et leur origine potentielle. Pour ce faire, nous présenterons tout d'abord la méthodologie du travail ainsi que les outils engagés ; puis nous examinerons les données collectées tout en adoptant une démarche quantitative. Dans un dernier temps, nous discuterons les résultats obtenus.

## II. Contexte de la recherche

Le présent travail s'inscrit dans le cadre d'un programme de recherche visant la mise en place d'un prototype de reconnaissance automatique des lexies exogènes au système de l'arabe standard basée sur la simulation de la capacité humaine à percevoir les sons de la parole. En effet, le protocole d'identification se nourrit des informations idiosyncrasiques des entités testées pour se prononcer quant à leur origine. Notre hypothèse de base s'appuie sur le fait que la structure sonore des lexies revêt des informations acoustiques et articulatoires susceptibles d'orienter le locuteur natif vers le statut d'un item lexical. Les différents travaux menés jusqu'à présent ont pu démontrer la consistance de ce postulat<sup>†</sup>. D'autre part, sur le plan des recherches acoustiques, les modèles quantitatifs de perception catégorielle et

<sup>†</sup> CLEMENTS G. N. (2009), KAYE (1979) et Ouadi, (2021, 2018 & 2017)

continue de traits acoustiques de la parole dévoilent un champ d'études en ce sens. Les résultats obtenus par Dominic W. Massaro (1983), qui achemine la quête vers l'aspect continu du mode perceptif, présentent des arguments supplémentaires en faveur du rôle de l'information continue sur les traits acoustiques dans le traitement de la parole. Lesquels contrecoups nous mènent à remettre en cause la démarche de notre travail et réorienter notre quête vers la distribution des classes articulatoires structurant les racines collectées dans le corpus.

### III. Matériels et méthodologie

Nous nous intéresserons dans le cadre de cette section à la description du protocole expérimental qui constitue la trame de fond de la présente recherche. *Ipsa facto*, nous présenterons succinctement l'outillage numérique des données et la méthodologie de collecte et de classification des observables.

#### 1. Les outils de collecte et de traitement numériques

L'étape de la collecte constitue la trame de fond de l'ensemble du processus du traitement du phénomène et conditionne systématiquement la phase d'exploitation des données. De ce fait, la constitution d'un corpus commensurable et adéquat à notre problématique, notamment dans sa phase de collecte doit indéniablement répondre aux principes de représentativité, d'exhaustivité et encore plus d'adéquation.

Notre préoccupation principale est la collecte des observables que nous passerons en peigne fin. C'est pour cela que les références de notre corpus doivent être authentiques et vérifiables. Le choix de *Lissan Al-Arabe* d'*IBN-MANDOUR* et du traité grammatical *AL-MOUARAB* d'*AL-JAOUALIQUI* comme sources de collecte et de vérification de l'origine des entités testées se montre naturel. En réalité, le premier ouvrage encyclopédique englobe une panoplie de dictionnaires extensifs introduisant dans la majorité des cas l'origine de l'entité lexicale. Quant au second, il représente la première contribution voire la première thèse tentant d'établir une relation de corrélat entre les critères intrinsèques des lexies et leur origine possible.

En effet, la première étape a consisté en un brassage minutieux des ouvrages puis la collecte et la vérification des différentes ressources lexicales. Puis, vient la phase d'enregistrement des données. Les lexies collectées sont enregistrées dans des tables tout en leur affectant des clés automatiques en vue de faciliter la manipulation dans les étapes ultérieures. Cette tâche est rendue possible par le biais du gestionnaire *MySQL*. Une base de données a été créée en vue de structurer le travail. Celle-ci dévoile

un ensemble de tables reliées par des règles assurant des clés et des relations. Chaque table est dédiée à réunir les informations correspondant au maximum à un seul type d'objet. De cette façon, il sera facile de manipuler les données relatives à l'origine séparément des autres informations. D'autant plus, une table consacrée à la collecte des radicaux structurant chacune des lexies est créée. Celle-ci est initiée par l'index identifiant la lexie qui résulte de l'association des trois radicaux R1&R2&R3. Ces identifiants sont ensuite notés sur la liste principale de métadonnées ligotée au corpus et mise à jour au fil de l'acquisition des données. Elle fait état des trois racines consonantiques formant la lexie.

Étant donné le nombre des lexies qui remonte à 6512 entrées, l'automatisation se montre inévitable. Pour ce faire, une série de formules d'extraction automatique des radicaux selon leurs positions sont programmées. Cette solution se veut essentielle puisque la segmentation se fait automatiquement après ajout de l'entrée à la première table destinée à la cueillette des lexies. Encore plus, la table affichant la matrice des données phonétiques fournit tous les traits phonétiques étiquetant chacun des sons radicaux collectés. Chacune de ses entrées porte un index unique et spécifique facilitant la création de liens transparents avec les autres tables et objets ce qui favorisera, entre autres, les processus d'étiquetage, de filtrage, d'importation, d'exportation des données et aussi les traitements statistiques. La matrice mère permet de visualiser toutes les données dans un même endroit. Toutes les données sont regroupées sur la même table à l'aide des identifiants uniques affectés lors de l'enregistrement dans le gestionnaire.

Quant aux traitements statistiques, ils se feront à l'appui du logiciel *SPSS* 17.0 qui se présente en parfaite conformité avec nos besoins en termes de statistiques combinatoires. Cette alternative nous sera propice du fait que les analyses sont orientées vers l'investigation des corrélats reliant entre trois et cinq observables ce qui n'est pas le cas pour d'autres logiciels d'analyses statistiques. D'autant plus, l'exportation des résultats sous formes numérique ou graphique se fait sans difficultés liées à la transportabilité ou compatibilité.

## **2. Méthode de classification des données**

Une fois la construction du corpus achevée, nous visualisons les données sur la matrice mère en vue de procéder aux traitements nécessaires. En réalité, les entrées collectées sont classées dans une matrice comme suit : les lignes représentent les entrées lexicales trilitères arabes et empruntées et les

colonnes indiquent respectivement, les trois consonnes formant le radical, la consonne tout en indiquant son emplacement dans le continuum sonore de la lexie, sa transcription phonétique, sa classe articulatoire et son degré de sonorité ; puis, l'origine, et les paramètres linguistiques générés automatiquement et qui constituent la charpente de notre travail ; en l'occurrence, la structure sonore et la courbe sonore formée autour du noyau consonantique (Tableau 1).

Tableau 1. Illustration graphique du mode de classification des observables.

Lexie	racine	R1	API1	ClassePhon1	SON1	R2	API2	ClassePhon2	SON2	R3	API3	ClassePhon3	SON3	Structure sonore	Origine	StructureC1C2	StructureC2C3
أبأ	ʔbʔ	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	أ	ʔ	[Glottal]	3	3-3-3	Arabe	Monotone	Monotone
أبب	ʔbb	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ب	b	[Labial]	3	3-3-3	Arabe	Monotone	Monotone
أبت	ʔbt	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ت	t	[Coronal]	1	3-3-1	Arabe	Monotone	Descandante
أبخ	ʔbx	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	خ	x	[Dorsal]	4	3-3-4	Arabe	Monotone	Ascandante
أبد	ʔbd	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	د	d	[Coronal]	3	3-3-3	Arabe	Monotone	Monotone
أبر	ʔbr	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ر	r	[Coronal]	7	3-3-7	Arabe	Monotone	Ascandante
أبز	ʔbz	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ز	z	[Coronal]	4	3-3-4	Arabe	Monotone	Ascandante
أبس	ʔbs	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	س	s	[Coronal]	2	3-3-2	Arabe	Monotone	Descandante
أبش	ʔbʃ	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ش	ʃ	[Dorsal]	2	3-3-2	Arabe	Monotone	Descandante
أبص	ʔbʕ	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ص	ʕ	[Coronal]	2	3-3-2	Arabe	Monotone	Descandante
أبض	ʔbʔ	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ض	ʔ	[Coronal]	3	3-3-3	Arabe	Monotone	Monotone
أبط	ʔbtʕ	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ط	ʔ	[Coronal]	3	3-3-3	Arabe	Monotone	Monotone
أبغ	ʔby	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	غ	ɣ	[Dorsal]	4	3-3-4	Arabe	Monotone	Ascandante
أبق	ʔbq	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ق	q	[Dorsal]	3	3-3-3	Arabe	Monotone	Monotone
أبك	ʔbk	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ك	k	[Dorsal]	1	3-3-1	Arabe	Monotone	Descandante
أبل	ʔbl	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ل	l	[Coronal]	6	3-3-6	Arabe	Monotone	Ascandante
أبن	ʔbn	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ن	n	[Coronal]	5	3-3-5	Arabe	Monotone	Ascandante
أبي	ʔbzj	أ	ʔ	[Glottal]	3	ب	b	[Labial]	3	ي	j	[Dorsal]	8	3-3-8	Arabe	Monotone	Ascandante
أتا	ʔʔʔ	أ	ʔ	[Glottal]	3	ت	t	[Coronal]	1	أ	ʔ	[Glottal]	3	3-1-3	Arabe	Descandante	Ascandante
أتب	ʔtb	أ	ʔ	[Glottal]	3	ت	t	[Coronal]	1	ب	b	[Labial]	3	3-1-3	Arabe	Descandante	Ascandante
أتن	ʔtn	أ	ʔ	[Glottal]	3	ت	t	[Coronal]	1	ن	n	[Coronal]	4	3-1-4	Arabe	Descandante	Monotone

La distribution des étiquettes linguistiques sur des cellules séparées a le mérite de cibler les classes observables lors des traitements statistiques. De surcroît, cette représentation se montre fine, claire et détaillée ce qui nous permettra de percevoir systématiquement les dissimilitudes qui existent entre chacune des classes mises en observation (oppositions horizontales) et de mettre la main sur les divergences constitutionnelles discernées dans le cadre de la même classe (oppositions verticales). De telle manière, lors d'analyse numérique, il sera facile de focaliser sur une classe afin d'en extraire les divergences internes ou opter pour une étude multivariée cherchant à dégager les corrélats sonores en faisant appel à plusieurs classes instantanément.

#### IV. Nature des données collectées

La collecte des ressources phono-lexicales a donné lieu à un corpus constitué de 6512 racines trilitères sourdes et défectueuses (Tableau 1 et Figure 2). La distribution des observables en termes de leur appartenance ou non au système de l'arabe classique dévoile indéniablement la dominance

des entrées endogènes (92,2%). Toutefois, bien que le score rattaché aux emprunts reste très limité, ce qui paraît tout à fait naturel du fait que l'emprunt ne doit aucunement déstabiliser la consistance du système linguistique, il paraît significatif sur le plan quantitatif. 505 items empruntés ont gardé leur aspect allogène contre 6007 formes arabes.

Tableau 1. Illustration numérique de la fréquence de l'emprunt dans le corpus.

	Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Arabe	6007	92,2	92,2	92,2
Emprunt	505	7,8	7,8	100,0
Total	<b>6512</b>	<b>100,0</b>	<b>100,0</b>	

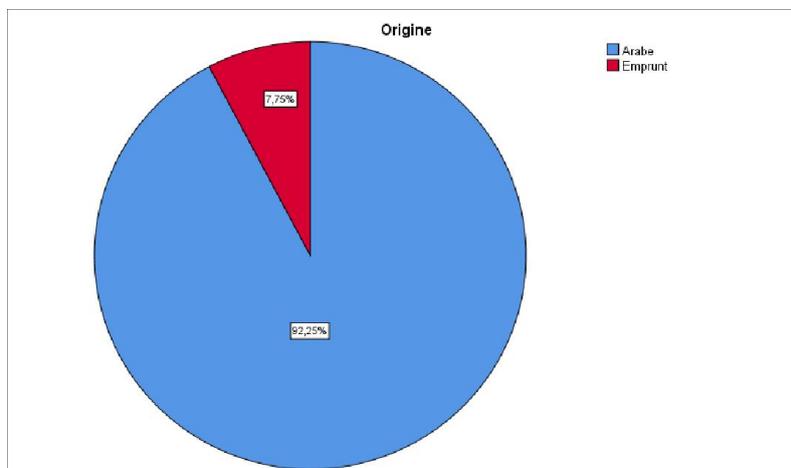


Figure 1. Représentation graphique des sous-classes du verbe trilitère sain en arabe classique.

Pour mieux comprendre les sous-bassements du phénomène de l'emprunt, nous essayerons de décortiquer chacune des catégories testées en maintenant un même principe de segmentation en vue de mettre la main sur les invariances susceptibles de discriminer l'une et l'autre classe. De cette manière, nous appliquerons une première expérience sur l'ensemble du

‡ La fréquence cumulée permet de déterminer le nombre d'observations qui se situent au-dessus d'une valeur particulière dans un ensemble de données. Elle est calculée sur la base des lignes supérieures du tableau de distribution de fréquences : Le nombre constaté dans la ligne en cours plus la somme des lignes supérieures. Voir plus loin ; notamment la section V dédiée à l'interprétation des résultats.

corpus et examinerons le développement de la courbe sonore dans les entités arabes et celles empruntées.

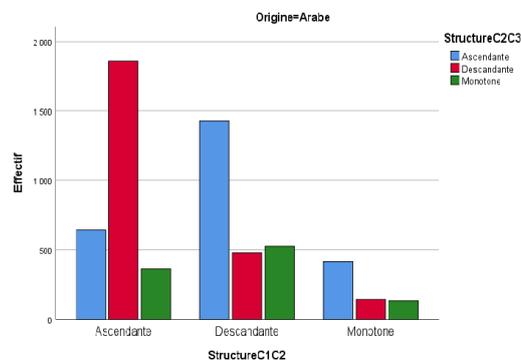
## V. Résultats

### 1. Distributions des structures sonores autour du noyau consonantique R2

Une lecture rapide des données collectées (Tableau 3 et figures 2 et 3) nous révèle la conformité du comportement sonore dans les deux classes lexicales examinées. La courbe la plus préférée dans les lexies arabes et empruntées dévoile un engouement à la distribution égalitaire. Une structure sonore ascendante est suivie dans la majorité des cas par une structure descendante (3180 cas). Le résultat le plus frappant est l'abondance des lexies initiées par une structure sonore descendante (1514 lexies) suivie, comme il a été signalé, d'une forme différente en l'occurrence la structure ascendante. Par ailleurs, la monotonie sonore se voit moins fréquente.

**Tableau 2.** Représentation numérique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Origine.

		StructureC2C3			Total
		Ascendante	Descendante	Monotone	
StructureC1C2	Ascendante	703	2081	396	3180
	Descendante	1514	499	561	2574
	Monotone	449	156	153	758
Total		2666	2736	1110	6512



**Figure 2.** Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Arabe

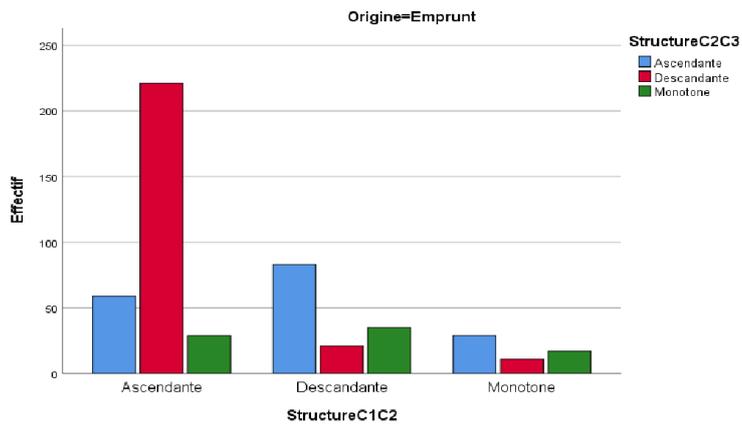


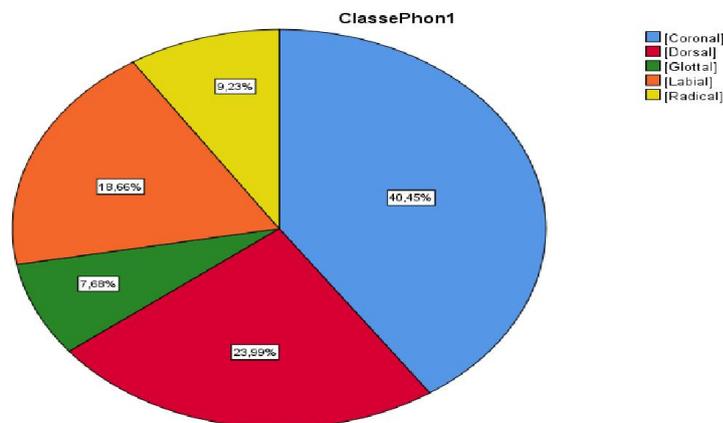
Figure 3. Représentation graphique du Tableau

croisé StructureC1C2 \* StructureC2C3 \* Emprunt

Ces distributions quoiqu'elles se montrent frappantes, elles ne puissent rendre compte du phénomène de la reconnaissance systématique de l'origine. L'opacité des données exige, du point de vue méthodologique, cibler une classe et tenter de spéculer les différentes possibilités de combinaison afin d'identifier les invariants susceptibles de dévoiler l'aspect allogène des lexies. Pour ce faire nous essayerons de restreindre le champ de notre investigation en nous focalisant sur la classe des radicaux manifestant un coup de glotte. Ce choix est motivé, en fait, par les données explicitées en infra.

**Tableau 3.** Représentation numérique des occurrences des classes articulatoires dans la première position radicale (R1)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	[Coronal]	2634	40,4	40,4	40,4
	[Dorsal]	1562	24,0	24,0	64,4
	[Glottal]	500	7,7	7,7	72,1
	[Labial]	1215	18,7	18,7	90,8
	[Radical]	601	9,2	9,2	100,0
	Total	6512	100,0	100,0	



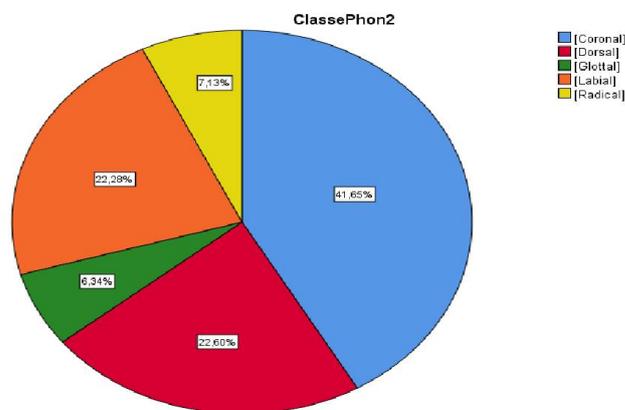
**Figure 4.** Représentation graphique des occurrences des classes articulatoires dans la première position radicale (R1)

Le premier site articuloire semble être dominé par les lexies manifestant une consonne coronale (40,4%) ou dorsale (24%). En effet, le cumul de la fréquence des deux traits articuloires restreint le reste des catégories à 35,6 %. Une dominance frappante qui nous pousse à nous demander sur les contraintes phonotactiques qui pourraient limiter l'occurrence des traits dans ce site d'attaque. D'autant plus, la restriction des classes [Labial], [Radical] et [glottal] paraît très significative du fait que le système linguistique de l'arabe engage des contraintes militent dans le sens d'exclure ou restreindre toutes les entités manifestant ses trois classes articuloires.

De ce fait, nous essayerons dans notre analyse statistique de spéculer l'occurrence des traits dans les sites médian et final pour pouvoir la consistance de cette remarque.

**Tableau 4.** Représentation numérique des occurrences des classes articuloires dans la seconde position radicale (R2)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	[Coronal]	2712	41,6	41,6	41,6
	[Dorsal]	1472	22,6	22,6	64,3
	[Glottal]	413	6,3	6,3	70,6
	[Labial]	1451	22,3	22,3	92,9
	[Radical]	464	7,1	7,1	100,0

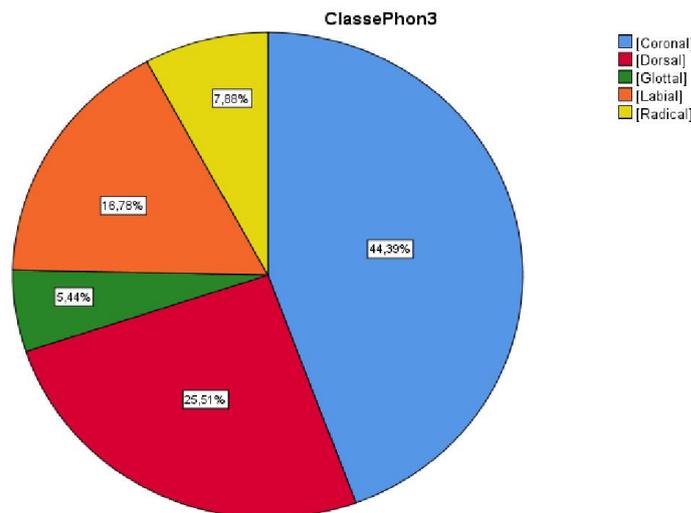


**Figure 5.** Représentation graphique des occurrences des classes articuloires dans la première position radicale (R2)

Une première lecture des résultats permet, sans contredit, de dévoiler la consistance de la remarque soulevée dans la section précédente. Le cumul des occurrences relatives aux deux classes dominantes [Coronal] et [Dorsal] aboutit à un résultat très frappant (64,3%). Malgré la solidité de l'ordre de l'occurrence des classes articulatoires [Coronal]>>[Dorsal]>>[Labial]>>[Radical]>>[Glottal], il faut signaler l'augment perçu au niveau de la classe labiale au dépend des glottales et radicales.

**Tableau 5.** Représentation numérique des occurrences des classes articulatoires dans la troisième position radicale (R3)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	[Coronal]	2891	44,4	44,4	44,4
	[Dorsal]	1661	25,5	25,5	69,9
	[Glottal]	354	5,4	5,4	75,3
	[Labial]	1093	16,8	16,8	92,1
	[Radical]	513	7,9	7,9	100,0
	Total	6512	100,0	100,0	



**Figure 6.** Représentation graphique des occurrences des classes articulatoires dans la troisième position radicale (R3)

Au niveau de la localité R3, les deux classes dominantes se voient gagner plus de fréquence par rapport au reste des classes articulatoires (69,9%). Aussi, les classes [Radical] et [Glottal] dévoilent une restriction explicite.

Une lecture synthétique des résultats divulgue incontestablement la dominance de la classe coronale dans le corpus au niveau des trois positions radicales R1, R2 et R3. L'occurrence des autres classes selon l'ordre dorsal, labial, radical et/ou glottal se montre naturelle et confirme prosaïquement des résultats affirmés par d'autres recherches [1], [2],[5].

Nous focaliserons dans le cadre des paragraphes suivants sur la classe articulatoire [glottal] pour la bonne raison que les autres classes manifestent une hétérogénéité frappante des sous-classes. L'analyse de la classe articulatoire [Dorsal] nécessite systématiquement l'étude des sous-classes qui la spécifie entre autres les catégories [Palatal], [Vélaire] et [Uvulaire]. De même, un traitement fin de la classe [Coronal] exige la déclinaison de l'étude des sous-classes [Alvéolaire], [Dental], [Interdental] etc. sans compter les caractéristiques d'emphase et de sifflement qui peuvent intervenir pour spécifier la classe les consonnes coronales. D'autant plus, afin de mieux cerner notre problématique, il a été jugé fructueux du point de vue méthodologique d'entamer le processus d'analyse en canalisant le filtre autour d'une classe présentant moins d'ingrédients.

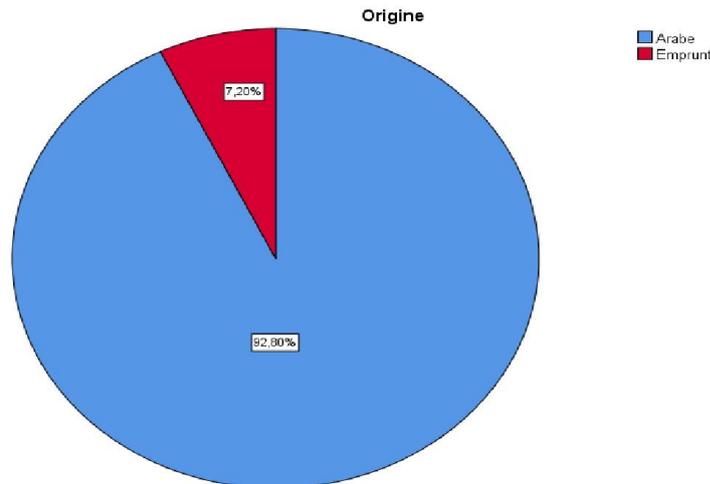
## 2. Étude du cas : [glottal]

Afin de mieux comprendre la corrélation liant l'émergence du coup de glotte dans la structure sonore d'une entité et son origine potentielle, nous tenterons dans le cadre de cette section de retracer l'occurrence du trait de position [Glottal] selon sa localité dans le schème radical. L'objectif est de quantifier le poids de cette distribution sur l'origine étymologique de la lexie. Pour ce faire, nous polariserons les filtres en direction des trois sites R1, R2 et R3 manifestant ce trait articulatoire.

### a. Occurrence d'un segment glottal dans la position tête

Tableau 6. Représentation numérique de l'abondance des lexies initiées par un coup de glotte (R1)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Arabe	464	92,8	92,8	92,8
	Emprunt	36	7,2	7,2	100,0
	Total	500	100,0	100,0	



**Figure 7.** Représentation graphique de l'abondance des lexies initiées par un coup de glotte (R1)  
Le résultat le plus frappant est l'analogie des scores. Les lexies empruntées comportant un coup de glotte dans la première position représentent 7,20% des observables ; ce qui est, en fait, très proche du pourcentage des lexies empruntées dans le corpus en général (cf. Tableau 2). Cette remarque mérite d'être examinée en faisant appel aux combinaisons sonores et l'origine de l'entité lexicale.

**Tableau 7.** Représentation numérique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Origine ;  
dont R1=[Glottal]

		StructureC2C3			Total	
		Ascendante	Descendante	Monotone		
StructureC1C2	Ascendante	Effectif	81	200	47	328
		Effectif théorique	131,9	141,0	55,1	328,0
		% dans StructureC1C2	24,7%	61,0%	14,3%	100,0%
		% dans StructureC2C3	40,3%	93,0%	56,0%	65,6%
		% du total	16,2%	40,0%	9,4%	65,6%
	Descendante	Effectif	77	3	18	98
		Effectif théorique	39,4	42,1	16,5	98,0
		% dans StructureC1C2	78,6%	3,1%	18,4%	100,0%
		% dans StructureC2C3	38,3%	1,4%	21,4%	19,6%
		% du total	15,4%	0,6%	3,6%	19,6%
	Monotone	Effectif	43	12	19	74
		Effectif théorique	29,7	31,8	12,4	74,0
% dans StructureC1C2		58,1%	16,2%	25,7%	100,0%	
% dans StructureC2C3		21,4%	5,6%	22,6%	14,8%	
% du total		8,6%	2,4%	3,8%	14,8%	
Total	Effectif	201	215	84	500	
	Effectif théorique	201,0	215,0	84,0	500,0	
	% dans StructureC1C2	40,2%	43,0%	16,8%	100,0%	
	% dans StructureC2C3	100,0%	100,0%	100,0%	100,0%	
	% du total	40,2%	43,0%	16,8%	100,0%	

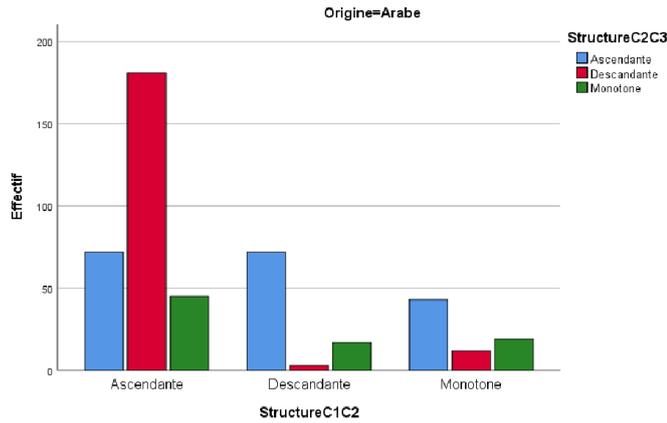


Figure 8. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Arabe ; dont R1=[Glottal]

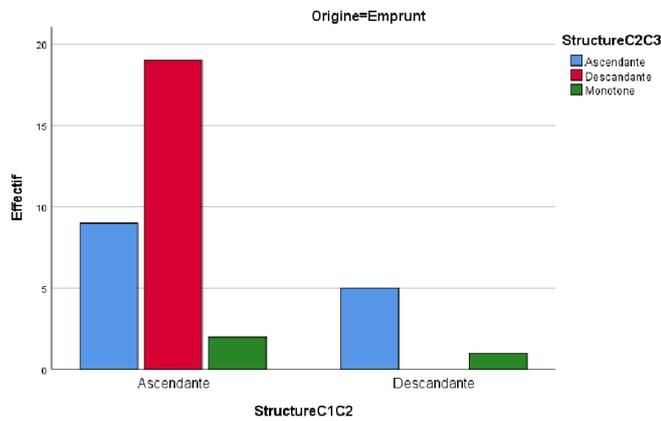


Figure 9. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Emprunt ; dont R1=[Glottal]

L'exploration multivariée des variables explicitées ci-dessus permet de discriminer les invariants sonores susceptibles de dévoiler l'aspect allogène. Outre la remarque signalée dans la section 5.1 et qui prône l'alternance sonore autour du noyau radical (Ascendant-Descendant ou Descendant-Ascendant), l'emprunt manifestant un coup de glotte dans la position initiale déjoue les structures monotones. De plus, les lexies qui commencent par une structure sonore précédant le noyau du radical se montrent fort probables par rapport aux autres candidats.

### b. Occurrence d'un segment glottal dans la position médiane (R2)

Tableau 8. Représentation numérique de l'abondance des lexies manifestant un coup de glotte dans la position médiane (R2)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Arabe	394	95,4	95,4	95,4
	Emprunt	19	4,6	4,6	100,0
	Total	413	100,0	100,0	

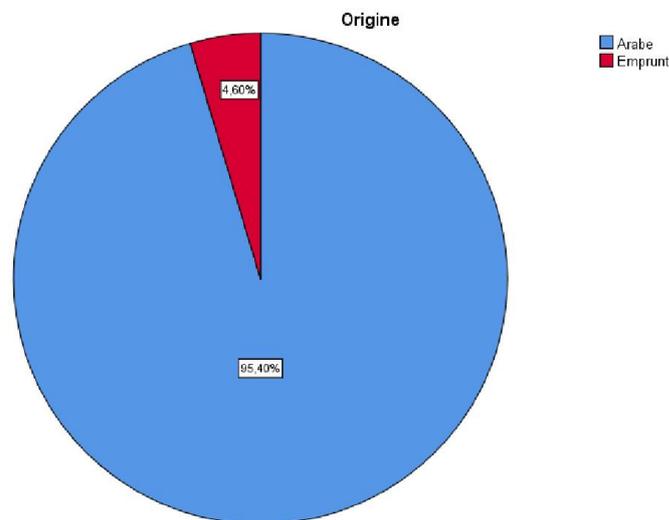


Figure 10. Représentation numérique de l'abondance des lexies manifestant un coup de glotte dans la position médiane (R2)

Le rendement des entités empruntées manifestant le trait articulatoire [glottal] au niveau médian se veut très faible. Seules 19 lexies empruntées contre 394. Cela s'explique par le fait que cette localité est assujettie aux règles phonotactiques plus strictes.

Tableau 9. Représentation numérique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Origine ; dont R2= [Glottal]

		StructureC2C3			Total
		Ascendante	Descendante	Monotone	
StructureC1C2	Ascendante	41	12	15	68
	Descendante	175	31	50	256
	Monotone	59	14	16	89
Total		275	57	81	413

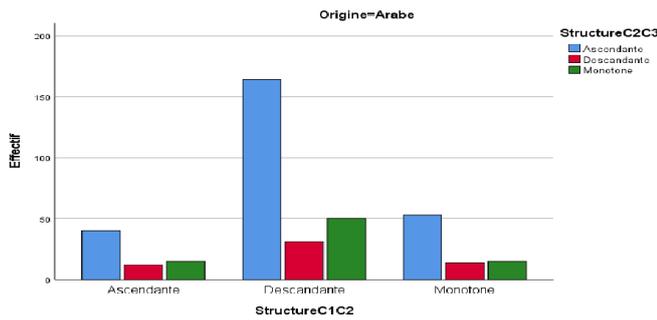


Figure 11. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Arabe ; dont R2= [Glottal]

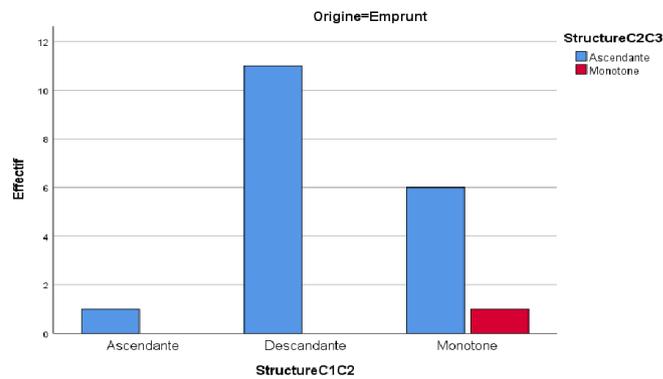


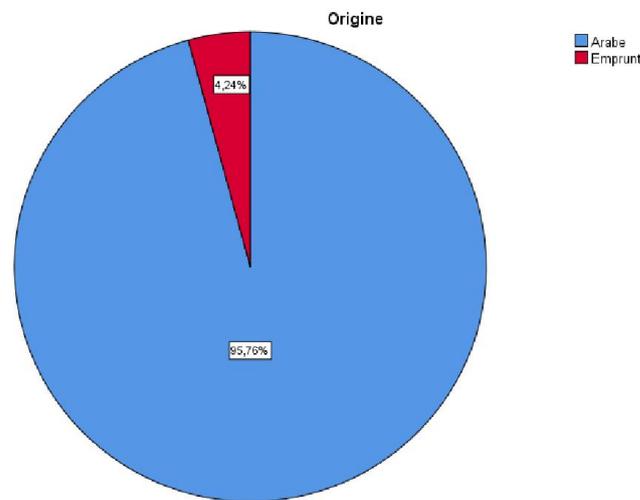
Figure 12. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Emprunt ; dont R2= [Glottal]

Les données ci-dessus dévoilent incontestablement la sujétion de ce site médian à des contraintes strictes prohibant l'occurrence des structures sonores monotones. Aucune des trois combinatoires ne révèle la structure descendante juste après le noyau consonantique. De même la monotonie sonore est prohibée après un pic autour du coup de glotte.

### c. Occurrence des classes phonétiques dans le schème radical

**Tableau 10.** Représentation numérique de l'abondance des lexies manifestant un coup de glotte dans la position queue(R3)

		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Arabe	339	95,8	95,8	95,8
	Emprunt	15	4,2	4,2	100,0
	Total	354	100,0	100,0	



**Figure 13.** Représentation graphique de l'abondance des lexies manifestant un coup de glotte dans la position queue(R3)

La remarque soulevée dans la section précédente se confirme de plus en plus. Le site R3 semble aussi être contraint. Parmi les 354 candidats, seuls 15 emprunts (4,24%) contenant une consonne radicale marquée par le trait [glottal] et occupant la position finale.

**Tableau 11.** Représentation numérique du Tableau croisé StructureC1C2 \* StructureC2C3 \*  
Origine ; dont R3= [Glottal]

		StructureC2C3			Total
		Ascendante	Descendante	Monotone	
StructureC1C2	Ascendante	6	151	23	180
	Descendante	63	39	33	135
	Monotone	12	14	13	39
Total		81	204	69	354

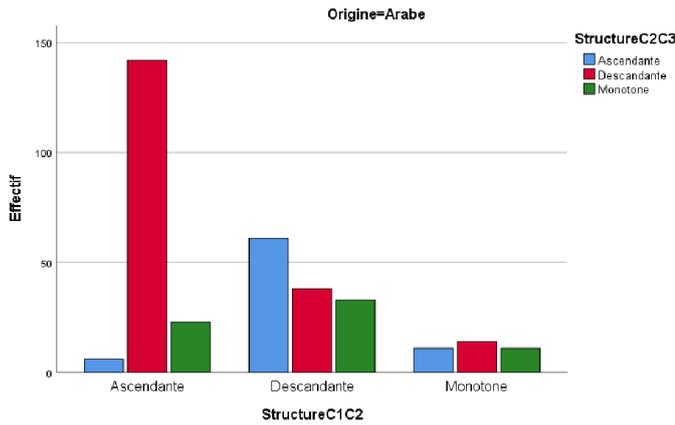


Figure 14. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Arabe ; dont R3= [Glottal]

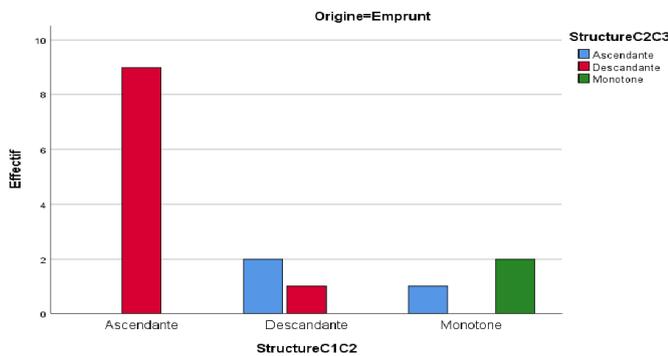


Figure 15. Représentation graphique du Tableau croisé StructureC1C2 \* StructureC2C3 \* Emprunt ; dont R3= [Glottal]

L'alternance sonore, elle aussi, se confirme plus nettement dans la liste des lexies dont l'origine est arabe. Par ailleurs, la classe des emprunts qui constitue la clé de voûte de notre réflexion se montre plus assujettie aux contraintes phonotactiques du système linguistique arabe. Cela s'explique par le nombre réduit d'observables. D'ailleurs, les combinatoires constatées dévoilent un refus inéluctable des structures sonores monotones. Le système phonologique de l'arabe montre un engouement considérable aux structures alternant deux rythmes sonores différents. Les alternances Ascendant-Descendant, Ascendant-Monotone et Descendant-Monotone se veulent des critères d'identification automatique sans conteste.

## VI. Conclusion

Le présent papier s'est dressé comme objectif la quantification des substrats sonores des lexies arabes trilitères dans une perspective d'automatisation de la reconnaissance. Le postulat de base qui a soutenu le travail est la capacité d'un locuteur natif à reconnaître systématiquement l'appartenance d'une entité lexicale au système phonologique de sa langue en se référant intuitivement à la perception des cycles sonores. Ainsi, nous avons tenté de spéculer les différentes localités structurant le schème consonantique trilitère en vue d'identifier les traits chauds susceptibles de corrélérer avec l'origine de la lexie.

Eu égard à l'hétérogénéité des classes et sous-classes articulatoires identifiant les différents schèmes, nous avons axé le travail sur la classe phonétique [glottal]. À l'issue de ce travail, nous avons pu, tout d'abord, quantifier le phénomène de l'emprunt en arabe standard. Puis, nous avons dégagé un ensemble de substrats sonores capables d'identifier l'origine des entités en prenant en considération la distribution des sonorités autour de la consonne médiane formant la racine trilitère. L'approche empirique adoptée présente en réalité des avantages et des inconvénients. D'une part, elle permet de concrétiser les données ce qui permet de mieux cerner le phénomène tout en prenant en considération les différentes variables susceptibles de contrecarrer l'analyse du phénomène de l'emprunt. D'autre part, cette analyse présuppose un brassage des différentes classes et sous-classes. En effet, ce protocole présente le risque d'alourdir la cadence du travail et de stigmatiser les résultats obtenus.

## Références

Blanchet, P. (2000). *Linguistique de terrain. Méthode et théorie. Une approche ethnosociolinguistique*. Rennes: Presses universitaires de Rennes.

- Brandão de Carvalho, J. (1997). Primitives et naturalité. (A. Colin, Éd.) *Langages*, 125, 14–34. Récupéré sur <http://www.jstor.org/stable/41683748>
- Bruce, B. (1989). The Crossing Constraint and 'Backwards Languages. *Natural Language & Linguistic Theory*, Vol. 7(No. 4), pp. 481-549.
- Cantineau, J. (1950). Racine et schème. Dans *Mélanges William MARÇAIS* (pp. 119-124). Paris: G.-P. Maisonneuve.
- Clements, G. N. (2003 ). Feature Economy in Sound Systems. *Phonology* , 20, 287-333.
- Clements, G. N. (2009). Chapter The role of features in speech sound inventories. Dans *Contemporary views on architecture and representations in phonological theory* (pp. 19-68 ). Cambridge (MA: The MIT Press.
- Dalbera, J.-P. (. ([En ligne], 1 | 2002, mis en ligne le 15 décembre 2003, consulté le 23 décembre 2021. ). Le corpus entre données, analyse et théorie. *Corpus*(DOI : <https://doi.org/10.4000/corpus.10>)).
- Dominic W. Massaro, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2(1), 15-35.
- Dugua, C., & Kanaan-Cailloil, L. (2021, février 12). Introduction. *Corpus*, [En ligne], 22 | 2021, mis en ligne le 12 février 2021, consulté le 2 novembre 2021. URL : <http://journals.openedition.org/corpus/5885> ;(DOI : <https://doi.org/10.4000/corpus.5885>). Consulté le novembre 2, 2021, sur DOI : <https://doi.org/10.4000/corpus.5885>
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language* , Vol. 26( 2)), 210-231.
- Jaubert, A. (2021). Corpus et champs disciplinaires. Le rôle du point de vue. *Corpus [En ligne]*, 1 | 2002, mis en ligne le 15 décembre 2003, consulté le 20 novembre 2021 DOI : <https://doi.org/10.4000/corpus>.
- Kaye, J. &. (1979). Loan words and abstract phonotactic constraints. *Canadian Journal of Linguistics/ Revue canadienne de linguistique*, 24(2), 71-93.
- Noël, N., Wauquier, Sophie, & Durand, J. (2005). La perception de la parole. Dans *Phonologie et phonétique* (pp. 425-447). Hermès Cognition et traitement de l'information. fhal00142953f.
- Ouardi, H. (2017). Identification-detection assistée des formes trilitères allogènes en arabe standard classique. *Colloque international de la linguistique appliquée*. Rabat.
- Ouardi, H. (2018). Pour une reconnaissance automatique des formes trilitères allogènes en arabe. *INTERNATIONAL CONFERENCE ON LANGUAGE AND CURRENT ISSUES - Issue 1 : "Language and Education"*, *Laboratoire Valeurs, Société et développement, Université Ibn Zohr, Agadir.*, 272-304.

- Ouardi, H. (2018). Pour une reconnaissance automatique des formes trilitères allogènes en arabe standard classique. *1*(Issue I : "Language and education"), 272-304.
- Ouardi, H. (2021). *Approche morphophonologique de l'emprunt en arabe classique: Implémentation d'un identificateur-détecteur des emprunts lexicaux trilitères dans un corpus textuel*. Editions universitaires européennes.
- Ouardi, H. (2021, ISBN : 978-9920-230-9 ). Le degré de complexité syllabique : une apostille apodictique. Dans *La lecture syllabique : Fondements théoriques*.