

Expression et Reconnaissance de Onze Attitudes Assertives et Interrogatives en Persan Standard

Olivier Piot, Mehdi Lyaghat

Université Paris III, CNRS UPRESA 7018

opiot@msh-paris.fr

Abstract

This is a study of the way emotional and cognitive attitudes may be expressed by prosody in standard Persian. Six assertive and five interrogative attitudes were produced on target words included in a short sentence. These target words were chosen on the basis of a cross-variation of both place of accent and word length. The great variability of contours thus obtained were quite highly recognized by native subjects. This study thus shows how some currently used attitudes may be expressed in Persian so as to be robustly recognized, and also sheds some light on the controversial issue of the phonetics of accent in this language.

1. Introduction

Langue indo-européenne asiatique, le Persan, langue officielle en Iran mais aussi en Afghanistan et au Tadjikistan, est également très pratiqué au Pakistan. Il possède six voyelles et deux diphtongues, vingt-deux consonnes et une semi-voyelle. Il n'y a pas d'opposition de quantité vocalique, même si les voyelles fermées /i, u/ et la voyelle ouverte postérieure sont d'anciennes voyelles longues (et en gardent encore la trace selon [3]). Les consonnes incluent à peu près celles du français sur un plan strictement phonologique, avec en plus deux affriquées, et quatre pharyngales ou laryngales (dont le "coup de glotte", ou occlusive laryngée).

L'ordre des mots place en principe le groupe verbal en dernière position dans la phrase (par exemple, "Nahid be madreseh raft", trad. mot à mot: "Nahid à l'école est allé"), mais dans la langue parlée on trouve souvent le groupe verbal en première position, et l'ordre des mots apparaît relativement libre. La position de l'accent a une fonction contrastive, comme par exemple sur la paire suivante: *ferestadan* "envoyer" versus *ferestadan* "ils ont envoyé". Les études des corrélats acoustiques de l'accent en Persan standard sont quelque peu contradictoires: selon [2], la prééminence d'une syllabe est due à F_0 et à l'intensité, mais la durée ne joue aucun rôle; selon [5] au contraire, une voyelle en position accentuée est presque deux fois plus longue qu'en position inaccentuée; [1] montre sur des mots isolés que les variations de F_0 jouent un rôle prépondérant, l'intensité jouant un rôle secondaire, et la durée ne jouant aucun rôle.

Selon [3], les paramètres F_0 et intensité sont prédominants, et il ajoute qu'un accent situé sur la syllabe finale d'un mot est réalisé assez faible, mais qu'il est fortement marqué lorsqu'une ou plusieurs syllabes inaccentuées suivent dans le mot. La syllabe accentuable est le plus souvent la dernière syllabe du mot, mais les cas d'accent non final sont fréquents. Il existe quelques cas d'accent lexical non final, mais la plupart des accents non finaux sont dus à la présence dans le mot de

suffixes inaccentués, l'accent portant alors généralement sur la dernière syllabe de la racine lexicale.

2. Présentation de l'expérience

2.1. Première phase: l'enregistrement

2.1.1. Corpus

Il est constitué d'énoncés débutant avec un mot cible, qui est toujours un groupe nominal sujet, inséré dans une phrase "cadre", toujours la même, de la façon suivante: ["X" (mot cible)] *mireh Canada* ("X va au Canada"). Le mot cible possède de une à quatre syllabes, et la syllabe accentuable est située sur la dernière, la pénultième, ou l'antépénultième syllabe du mot. Les mots cible sont les suivants (le noyau vocalique de la syllabe accentuée est souligné):

/rad/	nom propre
/nahid/	nom propre
/duste/	"son ami"
/nilufar/	nom propre
/pesare/	"son fils"
/kargaremun/	"notre ouvrier"

Les enregistrements étaient effectués dans la chambre sourde de l'institut de phonétique, sur mini-disk (au format 44kHz, 16 bits), puis sous-échantillonnés à 22 kHz sur le logiciel SoundForge. Les sujets se tenaient assis. La distance tête-micro étaient réglée à environ 40 centimètres, afin de permettre ensuite des mesures d'intensité, sans toutefois atteindre à leur confort pendant cette phase d'enregistrement.

2.1.2. Méthode

Pour obtenir des réalisations naturelles, chaque énoncé était situé dans une situation où le sujet jouait un rôle face à l'allocutaire expérimentateur. La consigne était de produire l'attitude uniquement par la prosodie. Les attitudes retenues pour cette expérience l'ont été afin de couvrir un spectre aussi varié que possible, sans pour autant rechercher l'exhaustivité. Elles sont réparties entre attitudes assertives (impliquant un "apport d'information") et interrogatives (impliquant une "demande d'information"), et sont ainsi notées et définies pour les locuteurs:

Assertions

Assertion neutre (AN): "le locuteur apporte une information sans prendre de position quelconque face à son propos".

Assertion restreinte (AR): "le locuteur apporte une information, en réalisant une mise en relief simple sur le mot cible de l'énoncé".

Assertion de confirmation (AC): "le locuteur confirme ce que croit l'allocutaire".

Avertissement (AV): "le locuteur met en garde l'allocutaire".

Lassitude (LAS): "le locuteur répond pour la cinquième fois à la question que lui pose l'allocutaire".

Evidence (EVI): "le locuteur trouve ce qu'il dit évident".

Questions

Question simple (QS): "le locuteur pose une question pour la première fois".

Question restreinte (QSR): "le locuteur pose une question pour la première fois, en mettant en relief le mot cible".

Demande de confirmation (DC): "le locuteur reprend l'affirmation de l'allocutaire et lui en demande confirmation".

Doute-incrédulité (DI): "le locuteur reprend l'affirmation de l'allocutaire en la mettant en doute".

Question surprise (SUR): "le locuteur exprime sa stupéfaction devant l'affirmation de l'allocutaire".

Chaque attitude a été enregistrée une fois sur chacune des six phrases du corpus, soit au total $11 \times 6 = 66$ phrases par locuteur. Les enregistrements terminés, ils ont été soumis à un groupe de trois locuteurs natifs afin d'évaluer la qualité et le naturel des réalisations. Cette procédure a conduit à ne retenir les productions que d'une des trois locutrices, pour leur utilisation dans le test de perception subséquent.

2.1.3. Locuteurs

Les trois locuteurs étaient des femmes originaires de Téhéran, et ayant le Persan standard comme langue maternelle.

2.2. Seconde phase: le test de reconnaissance des attitudes

2.2.1. Stimuli

Un premier groupe de sujets (**groupe 1**) s'est vu proposé comme stimuli les phrases complètes, y compris l'appendice "mireh Canada", tandis qu'un second groupe de sujets (**groupe 2**) n'était confronté qu'aux seuls mots cible. Pour chacun des deux groupes l'expérience se divisait en deux parties, la première (**partie (a)**) concernait les attitudes assertives, la seconde (**partie (b)**) les attitudes interrogatives. Pour la partie

(a) il y avait 6 phrases \times 6 attitudes = 36 stimuli, et pour la partie (b) 6 phrases \times 5 attitudes = 30 stimuli.

2.2.2. Méthode

Les stimuli apparaissaient pour chacune des parties (a) et (b) avec un ordre aléatoire, toujours le même. Chaque stimulus apparaissait trois fois aux sujets, avec une seconde d'intervalle entre deux apparitions successives, et ne pouvaient pas être réécoutés. Chaque groupe de trois stimuli identiques était séparé des autres groupes par un intervalle de deux secondes. Les sujets étaient soumis à un choix forcé entre les six attitudes assertives AN, AR, AC, AV, LAS et EVI pour la partie (a), et entre les cinq attitudes interrogatives QS, QSR, DC, DI, SUR pour la partie (b). Les deux parties (a) et (b) s'enchaînaient sans pause entre elles, et l'expérience durait environ 6 minutes par sujet.

2.2.3. Sujets

Le groupe 1 (phrases complètes) est constitué de 3 sujets. Il s'est avéré que les résultats au test n'étaient pas différents pour les deux groupes, et on s'est donc reporté sur l'évaluation du groupe 2 (mots cible isolés), composé de 11 autres sujets. Tous sont originaires de Téhéran, et ont le Persan standard comme langue maternelle.

3. Analyse des taux de reconnaissance

Pour le groupe 1 les attitudes sont correctement reconnues dans 89.8% des cas pour la partie (a) (attitudes assertives), et dans 91% des cas pour la partie (b) (attitudes interrogatives). Pour le groupe 2 (auquel seuls les mots cible étaient présentés), ces scores sont respectivement de 91.9% et de 91.2%. La donnée de l'appendice "mireh Canada" n'apparaît donc pas apporter d'amélioration des scores, et nous nous limitons par la suite à l'analyse des résultats du groupe 2.

Attitude	AN	AR	AC	AV	LAS	EVI
AN	64	2				
AR		58	3			5
AC		6	59		1	
AV				66		
LAS			3		63	
EVI		7				59

Tableau 1: Matrice de confusion pour la reconnaissance des attitudes assertives (66 réponses par attitude).

Attitude	QS	QSR	DC	DI	SUR
OS	65	1			
QSR	2	61		3	
DC		1	57	2	6
DI		2	3	61	
SUR			9		57

Tableau 2: Matrice de confusion pour la reconnaissance des attitudes interrogatives (66 réponses par attitude).

Les matrices de confusion sont représentées sur les tableaux 1 et 2. Pour les assertions, la confusion se situe essentiellement entre AR et AC (9 cas), et entre AR et EVI (12 cas). Pour les

questions, elle se situe surtout entre DC et SUR (15 cas), et dans une moindre mesure entre QSR et DI (5 cas), ainsi qu'entre DC et DI (5 cas). Ajoutons que le nombre d'erreur, pour chaque mot cible, est compris entre 6 et 13 (sur 121). Il va sans dire que cette répartition des réponses est très significativement différente d'une équi-répartition ($p < 10^{-200}$ pour (a) comme pour (b)).

4. Analyse prosodique des réalisations assertives

Pour faciliter l'analyse, nous nommerons **x** les syllabes précédant la syllabe accentuée (s'il y en a), **X** la syllabe accentuée et **x'** les syllabes suivant X dans le mot cible (s'il y en a). L'analyse globale montre, sur le mot cible /kargaremun/ (choisi pour des raisons liées à une neutralisation maximale des influences d'ordre phonotactique), que l'intensité est majoritairement plus faible sur la syllabe accentuée que sur la syllabe précédente inaccentuée, en moyenne de 3 dB: cela correspond, rappelons-le, à une division par deux de l'énergie acoustique, et représente donc un amoindrissement notable. Précisons enfin qu'il nous est apparu impossible d'extraire des régularités temporelles faisant intervenir, les unes par rapport aux autres, l'ensemble des syllabes des mots cible. Par contre, la considération de la durée de la seule syllabe accentuée de ce mot s'est avérée déterminante dans l'extraction de régularités, en constituant un indice robuste vis-à-vis des changements de mot cible.

L'assertion "neutre" AN se caractérise par un contour plat sur **x**, montant sur **X**, et qui reste plat sur **x'**, au niveau atteint en fin de **X**. Il est intéressant de noter qu'ensuite le contour continue à suivre ce plateau sur "mireh", avant de descendre progressivement sur "Canada".

L'assertion restreinte AR se caractérise comme AN par un contour plat sur **x**, montant sur **X**, mais la chute tonale s'amorce sur **X**, et se poursuit sur **x'**, pour atteindre très vite le niveau bas du registre. La hauteur du pic tonal est très proche de celle atteinte pour AN. La durée de **X** reste relativement

courte, mais est légèrement plus longue (de 15% environ) en moyenne que celle de AN, et son intensité est supérieure à celle de AN d'environ 0.4 dB (en moyenne).

L'assertion de confirmation AC suit exactement le même contour que AR, mais le niveau tonal atteint (sur **X**) est ici plus haut (de 1 à 5 demi-tons), et le pic tonal apparaît ici plus tard dans **X**. La durée de **X** est importante, plus longue que celle de AN d'environ 45% (en moyenne), et son intensité est supérieure à celle de AN d'environ 2 dB (en moyenne).

L'avertissement AV suit un contour également très semblable à celui de AR, mais le niveau tonal atteint (sur **X**) est encore plus haut que pour AC (de 6 à 9 demi-tons), et le pic tonal apparaît lui aussi encore plus tard dans **X**; il peut même apparaître sur la syllabe **x'** suivant immédiatement **X**, comme dans /pesare/ ou /kargaremun/. La durée de **X** est longue, plus longue que celle de AN d'environ 130% (en moyenne), et son intensité est supérieure à celle de AN d'environ 2 dB (en moyenne).

La lassitude LAS est réalisée par la locutrice avec un contour plat sur **x**, légèrement montant ou plat sur **X** (au même niveau, ou très légèrement supérieur à celui de **x**), puis au même niveau ou rehaussé sur le début de **x'** (sur une portion pouvant atteindre une syllabe entière), avant d'optionnellement redescendre sur la suite de **x'**. Le niveau tonal atteint est celui de AN (ou AR). La durée de **X** est moyenne-longue, plus longue que celle de AN d'environ 80% (en moyenne), et son intensité est supérieure à celle de AN d'environ 3.3 dB (en moyenne).

L'évidence EVI commence par un contour plat sur **x**, très légèrement montant ou plat (au même niveau ou légèrement supérieur à celui de **x**) sur **X**, suivi d'une chute tonale qui ne commence qu'au début de **x'** (mais dans aucun cas sur **X**). Le niveau tonal atteint est proche de celui de AN. La durée de **X** est moyenne, plus longue que celle de AN d'environ 30% (en moyenne), et son intensité est supérieure à celle de AN d'environ 5 dB (en moyenne).

Attitude	AN	AR	AC	AV	LAS	EVI
Contour						
Durée X	*	**	****	*****	*****	***
Int. X	*	*	**	**	**	****

Tableau 3: Représentation schématique des tendances observées pour les énoncés assertifs. On a retenu cinq niveaux pour la durée comme pour l'intensité vocalique de **X**, du plus petit au plus grand: *, **, ***, ****, ***** (le niveau ***** pour l'intensité n'est atteint que dans la partie (b), cf. tableau 4).

5. Analyse prosodique des réalisations interrogatives

Toutes les tendances sont résumées sur le tableau 4. Pour la représentation des contours, on a utilisé exactement la même

portée à quatre niveaux que celle utilisée pour la partie (a), ainsi que la même méthode de catégorisation de l'intensité et des allongements syllabiques.

La question simple QS se caractérise par un contour plat sur **x**, montant sur **X** (et pouvant même continuer à monter sur la

première syllabe de x'), puis qui reste plat sur x', au niveau atteint en fin de montée tonale. Il est intéressant de noter qu'ensuite, sur "mireh Canada", le contour redescend doucement jusqu'au niveau de x, pour reproduire le même contour montant (avec en moyenne les mêmes hauteurs initiales et finales) sur la dernière syllabe de "Canada".

La question restreinte QSR se caractérise comme QS par un contour plat sur x et montant sur X, mais pouvant continuer à monter jusqu'à la hauteur initiale de la portion suivante, x', sur laquelle le contour est descendant: il descend jusqu'au niveau de départ, qui était celui de x. Le niveau tonal atteint (sur X) est ici nettement plus haut que pour QS (de 4 demi-tons en moyenne). La durée de X reste courte, mais est légèrement plus longue (en moyenne de 15% environ) que celle de QS, et son intensité est supérieure à celle de AN d'environ 1.3 dB (en moyenne).

La demande de confirmation DC se caractérise par le même contour que QSR, avec un niveau tonal atteint (sur X) encore

plus haut que pour QSR (de 8 demi-tons en moyenne). La durée de X est moyenne, plus longue que celle de QS (en moyenne de 40% environ), et son intensité est supérieure à celle de AN d'environ 6.5 dB (en moyenne).

Le doute-incrédulité DI se caractérise par le même contour que QSR, avec un niveau tonal atteint (sur X) plus haut que pour QSR (de 6 demi-tons en moyenne). La durée de X est longue, plus longue que celle de QS (en moyenne de 75% environ), et son intensité est supérieure à celle de AN d'environ 3.9 dB (en moyenne).

La question surprise SUR se caractérise par le même contour que QSR, avec le niveau tonal atteint (sur X) le plus haut des cinq (de 9 demi-tons en moyenne par rapport à QSR). La durée de X est longue, plus longue que celle de QS (en moyenne de 95% environ), et son intensité est supérieure à celle de AN d'environ 6.5 dB (en moyenne).

Attitude	OS	QSR	DI	DC	SUR
Contour					
Durée X	*	**	*****	****	*****
Int. X	*	**	***	*****	*****

Tableau 4: Représentation des tendances observées pour les énoncés interrogatifs. On a retenu les mêmes principes - et valeurs - qu'en (a) pour la représentation des contours, de la durée, et de l'intensité vocalique.

6. Discussion

Tout d'abord on peut remarquer que les taux de reconnaissance extrêmement élevés ne peuvent pas s'expliquer sur la base d'une simple identification des contours mélodiques utilisés: les contours sont, surtout pour la partie (b), très semblables les uns aux autres. Cependant, les confusions relevées semblent clairement concerner les paires de stimuli les plus proches perceptivement au niveau de la hauteur moyenne et du contour. Ainsi par exemple, LAS et EVI ont une hauteur moyenne proche, mais leurs contours sont très différents sur x' (leur grande différence de durée assure vraisemblablement leur discrimination en l'absence de x'), et ne sont pas confondus. De la même façon pour la partie (b), où la forme du contour est la même pour les cinq attitudes sauf QS, mais où les hauteurs moyennes forment une suite régulière, ce sont les paires constituées des contours de hauteurs moyennes les plus proches qui sont confondus: QSR/DI (5 cas), DI/DC (5 cas), et DC/SUR (15 cas). Le nombre de confusions est cependant décroissant avec la différence entre les stimuli au niveau de la durée et de l'intensité (différence représentée par exemple par le nombre d'* qui les séparent dans les tableaux). Quelles hypothèses peut-on formuler pour tenter d'expliquer les taux élevés de reconnaissance? Répondre à ces questions dépasse le cadre de cette étude introductive, et nous renvoyons pour cela, ainsi que pour une proposition d'interprétation de l'ensemble des contours prosodiques adoptés, à [4].

7. Conclusion

Retenir les productions d'une locutrice particulièrement "expressive", et ayant bien assimilé la tâche demandée, nous a permis d'obtenir des expressions reconnues de façon très robuste par les sujets natifs, bien que les situations voire les intitulés des attitudes retenues soient parfois très proches. Sur la question des paramètres prosodiques de l'accentuation, notre étude met en évidence que la montée mélodique par rapport à la syllabe précédente est le seul paramètre commun à ces onze attitudes, ce qui apporte une confirmation à [3] et [1], c'est-à-dire à un rôle prépondérant de ce paramètre dans la réalisation de l'accent en Persan.

8. References

- [1] Amirfiruz-kuhi, S., 1997. *Etude de la mélodie et de la durée des énoncés assertifs et interrogatifs en persan*, mémoire de DEA, Université Paris III.
- [2] Khanlari, N.P., 1961. *La grammaire de la langue persane*, Tousse, Téhéran.
- [3] Lazard, G., 1957. *Grammaire du persan contemporain*, Klincksieck, Paris.
- [4] Piot, O., 2002. *Vers une théorie unifiée de la prosodie du français et de l'anglais: des émotions à la phonologie*, Université Paris III.
- [5] Sepenta, S., 1970. Thèse de doctorat, Université de Téhéran.