

# Intonation des phrases interrogatives et affirmatives en langue vietnamienne

Vũ Minh Quang - Trần Đỗ Đạt - Eric Castelli

Centre de recherche international MICA  
IP Hanoi – CNRS/UMI-2954 – INP Grenoble  
1-Dai Co Viet - Hanoi, Vietnam

{Minh-Quang.Vu ; Do-Dat.Tran ; Eric.Castelli}@mica.edu.vn  
<http://www.mica.edu.vn>

## ABSTRACT

Interrogative and affirmative sentences in Vietnamese language, which have same tones and the same number of syllables, are recorded in order to analyse their intonation shape (F0 evolution) avoiding effect of lexical tones and of co-articulation. Comparisons permit us to characterise differences between question and non-question at sentence prosody level. Then, our work is completed by a perception study; its main goal is to check if sentence nature characteristics are included in the sentence prosody, allowing the auditor to classify questions and non-questions, despite the complex form of this prosody in tonal languages. Results show that information on sentence nature are present at the end of its last demi-syllable and that about 70% of sentences are good classified.

## 1. INTRODUCTION

Pour les langues occidentales non tonales (dont le français ou l'anglais sont des exemples) il a été validé que la prosodie de la phrase véhicule des informations extralinguistiques, comme les émotions ou l'état du locuteur ou bien comme la nature de la phrase (affirmative, interrogative ou exclamative) [1, 2]. Pour évaluer automatiquement le type des phrases à des fins de détection ou de classification, il est alors possible d'analyser directement le signal de parole, sans avoir besoin nécessairement du résultat lexical d'un moteur de reconnaissance automatique de la parole (RAP) mais en utilisant le contour prosodique de la phrase. Dans ce cas, l'essentiel des paramètres mesurés et analysés prennent en compte l'évolution de l'intonation pendant l'énoncé de la phrase : registre de F0, augmentation des valeurs de F0 en fin de phrase ou d'autres paramètres dérivés des valeurs de F0, par exemple [3, 4].

Cependant, dans le cas des langues tonales (comme le mandarin ou le vietnamien), le contour mélodique de l'intonation est complexe parce qu'il est composé de macro-variations correspondant à l'intonation de la phrase et de micro-variations correspondant aux tons lexicaux appliqués sur chacune des syllabes des mots mono (ou bi) syllabiques. C'est pourquoi, appliquer sur ces langues directement les méthodes d'analyse validées pour les langues non tonales semble ne pas permettre de

détecter ou de classer la nature des phrases avec suffisamment de fiabilité car les micro-variations tonales semblent brouiller l'information extra-lexicale de la phrase. Dans le cas de la langue vietnamienne d'ailleurs, pour différencier les phrases interrogatives des autres, l'emploi de mots spécifiques dit « classifieurs interrogatifs » (không, gì, ai, par exemple) est pratiquement systématique. La question alors posée peut être résumée ainsi : existe-t-il, pour le vietnamien, langue à tons dont la prosodie est complexe, des informations extralinguistiques caractérisant le type de phrases, véhiculées par la prosodie et utilisées pendant les actes de dialogue pour la classification de ces types de phrases ? La réponse, outre le fait qu'elle nous permettra d'approfondir nos connaissances de la langue, si elle est positive, nous permettra d'envisager la réalisation de classifieurs automatiques indépendants des moteurs de reconnaissance.

## 2. ANALYSE DU CONTOUR INTONATIF

### 2.1. Méthodologie et préparation du corpus

Jusqu'à ce jour, très peu d'études ont analysé la phonologie de la langue vietnamienne en profondeur. Nous pouvons citer quelques travaux récents portant sur les tons lexicaux [5, 6, 7] et sur la prosodie de la phrase [8, 9]. Après analyse de phrases de corpus « lus » et « spontanés », Lê T. X [8] et Nguyễn Thị T. H. & Boulakia [9] constatent qu'il existe une différence de hauteur de F0 entre les types de phrases. En évaluant leur niveau, [8 & 9] précisent que les assertives sont prononcées avec un registre bas alors que les questions et les injonctives le sont avec un registre haut. De plus, [9] fait le constat qu'au niveau de l'allure générale de l'intonation de la phrase, une pente descendante ne correspond pas toujours à une phrase déclarative. Au niveau de la durée : les énoncés interrogatifs ont un débit plus rapide que les énoncés assertifs et injonctifs, la différence de durée entre ces deux derniers n'étant pas significative [9]. Quant à l'intensité, elle est d'une manière générale plus forte dans la phrase interrogative, et les syllabes finales ont souvent un niveau intensité plus important que les autres syllabes de la phrase [9].

Partant de ces constats, nous souhaitons préciser les différences prosodiques entre les phrases interrogatives et

les phrases affirmatives. Pour cela, nous avons construit un corpus spécifique constitué de paires de phrases interrogatives et affirmatives qui présentent, pour chacune des phrases de la même paire, un même contexte tonal et le même nombre de syllabes. Le fait de choisir les mêmes tons nous permet d'éliminer l'influence des tons des syllabes sur l'intonation générale de la phrase, et donc de maîtriser ainsi les micro-variations de l'intonation. Pour éliminer aussi tous les phénomènes de co-articulation qui pourraient interférer avec notre analyse prosodique, nous avons, dans la mesure du possible, gardé aussi les mêmes mots, ou bien nous avons utilisé des mots à la prononciation peu différente. Toutes ces phrases ont été intégrées dans des dialogues significatifs, afin que leur prononciation soit la plus naturelle possible (nous avons enregistré la totalité des dialogues, puis extrait les phrases choisies pour l'analyse). Chaque dialogue est répété cinq fois par six locuteurs (3 hommes et 3 femmes) originaires du Nord du Vietnam. Les phrases sélectionnées sont présentées dans le tableau 1 suivant :

**Tableau 1 :** Les paires de phrases affirmatives (a) et interrogatives (i) du corpus.

1a	Hôm nay là ngày ba mươi <i>Aujourd'hui nous sommes le trente</i>
1i	Hôm nay là ngày bao nhiêu ? <i>Quel jour sommes-nous aujourd'hui ?</i>
2a	Tên anh ta là Trì <i>Il s'appelle Trì</i>
2i	Tên anh ta là gì ? <i>Il s'appelle comment ?</i>
3a	Anh ăn cơm không <i>Tu manges du riz seulement.</i>
3i	Anh ăn cơm không ? <i>Tu manges du riz ?</i>
4a	Em ăn bánh Ché <i>Je mange du gâteau Ché.</i>
4i	Em ăn bánh nhé ? <i>Tu vas manger du gâteau ?</i>

En vietnamien, pour la construction de phrases interrogatives, en plus de l'utilisation pratiquement systématique de mots « classificateurs interrogatifs », le locuteur peut ajouter en fin de phrase certains mots qui sont normalement facultatifs mais dont l'usage éventuel dépend fortement de l'habitude, de la façon de parler du locuteur, du contexte dans lequel se produit le dialogue, d'une manifestation de respect et/ou de politesse avec l'interlocuteur, etc. Des exemples sont présentés dans le tableau 2.

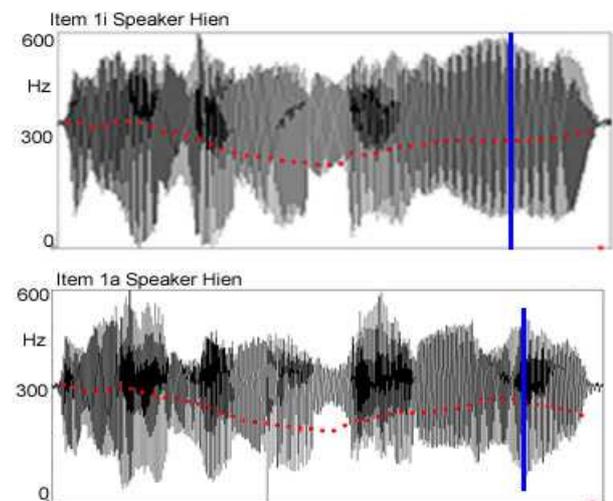
**Tableau 2 :** La même phrase interrogative avec des mots terminaux différents.

- Hôm nay là ngày bao nhiêu rồi ?
- Hôm nay là ngày bao nhiêu vậy ?
- Hôm nay là ngày bao nhiêu thế ?
- Hôm nay là ngày bao nhiêu hả ?
→ <i>Quel jour sommes-nous aujourd'hui ?</i>
- Hôm nay là ngày bao nhiêu hả anh ?
→ <i>Quel jour sommes-nous aujourd'hui, monsieur ?</i>
- Hôm nay là ngày bao nhiêu hả chị ?
→ <i>Quel jour sommes-nous aujourd'hui, madame ?</i>

Comme ces mots terminaux facultatifs peuvent porter n'importe lequel de six tons de la langue vietnamienne, alors la portion finale du contour intonatif de la phrase peut être fortement modifiée par le contour du ton du mot terminal. C'est pourquoi, pour chaque phrase interrogative sélectionnée, nous avons choisi d'incorporer au corpus un certain nombre de variations possédant des mots terminaux de différents tons afin d'étudier plusieurs formes de contours intonatifs possibles.

## 2.2. Résultats d'analyse

Pour chaque enregistrement nous analysons le contour de la fréquence fondamentale F0 avec le logiciel Praat (exemple présenté figure 1).



**Figure 1 :** Deux phrases à nombre de syllabes et tons identiques. En rouge le contour de F0. Figure du dessus : phrase interrogative, dessous : phrase affirmative.

En étudiant chaque paire de phrases présentées dans le tableau 1, nous remarquons que l'essentiel des différences d'intonation se situe à la fin de la phrase (zone située figure 1 après la barre verticale) : le contour de la dernière syllabe ou de la deuxième moitié de celle-ci semble être croissant pour les phrases interrogatives. Une étude statistique, présentée dans le tableau 3, confirme cette tendance : 80% des cas de phrases type I possède un contour de F0 croissant à la fin de phrase. Nous retrouvons là une tendance bien connue pour les langues non tonales comme le français. Cependant, pour un cas de phrase type affirmatif (A), le contour de la dernière moitié de la dernière syllabe est aussi croissant pour 29 enregistrements sur 30. Cette phrase « Em ăn bánh Ché » présente en fin de phrase deux mots au ton croissant qui influence le contour intonatif global de la phrase.

**Tableau 3 :** Nombre (et pourcentage) des contours de F0 de la dernière moitié de la dernière syllabe.

	Interrogative	Affirmative
Pente montante	96 (80%)	45 (37%)
Pente descendante	24 (20%)	75 (63%)

Nous nous intéressons aussi à la forme intonative des autres phrases de type interrogatives (I) du tableau 2. Bien que certains des mots terminaux portent des tons à contour descendant (ton 2 « descendant » ou ton 6 « grave »), les phrases porteuses de notre corpus présentent en grande majorité (minimum 83 %) un contour croissant à la fin de la phrase (tableau 4).

**Tableau 4 :** Nombre (et pourcentage) des contours de la dernière moitié de la dernière syllabe pour les phrases interrogatives avec des mots terminaux de différents tons (pas de mots terminaux avec les tons 1 et 3).

	Ton2	Ton6	Ton5	Ton4
Pente montante	45(90%)	80(88%)	75(83%)	28(93%)
Pente descendante	5(10%)	10(12%)	15(17%)	2(7%)

[8 et 9] ont suggéré que les phrases de type I sont prononcées avec un registre plus haut. Pour ce point, l'étude statistique de notre corpus montre qu'effectivement elles semblent montrer une valeur moyenne de F0 plus importante que celle des phrases de type A (tableau 5). Cependant, à part la locutrice Hien qui présente une différence assez importante (environ 40Hz), les différences pour les autres locuteurs sont faibles et plus petites que les valeurs des écarts-type correspondant. En contradiction avec les travaux de [8 et 9], l'effet de registre semble donc peu significatif dans notre corpus. Pour la durée, nous retrouvons la même tendance constatée dans [9] : la durée des phrases interrogatives est en moyenne plus petite de 10% que celle des affirmatives (12% dans [9]).

**Tableau 5 :** Fréquence fondamentale moyenne (et écart type) des phrases I et A pour les six locuteurs. (H = homme, F = femme).

Locuteur		Hien F	Quang H	Nam H	Huong F	Diep F	Khoa H
Type I	Hz	307 (16)	160 (15)	160 (8)	250 (18)	239 (11)	145 (16)
Type A	Hz	264 (10)	146 (14)	152 (10)	247 (13)	231 (23)	133 (11)

### 3. PERCEPTION DES PHRASES I ET A

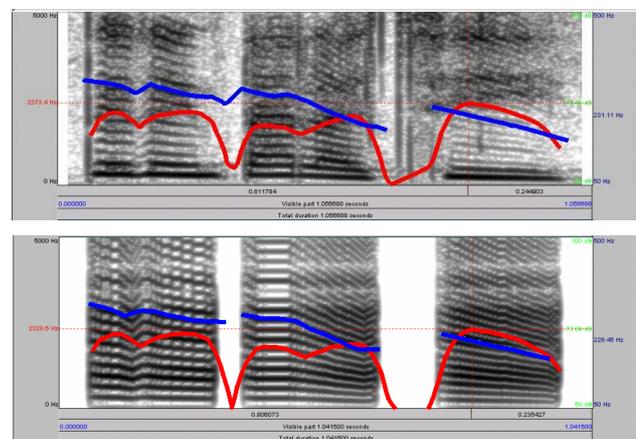
#### 3.1. Méthodologie et préparation du corpus

Nous souhaitons vérifier que les différences détectées dans notre analyse sont effectivement perçues comme un moyen pour l'auditeur de classer phrases interrogatives et phrases affirmatives, ou, en d'autres termes, que la prosodie de la phrase, malgré sa complexité due à la présence des tons, véhicule des informations permettant à l'auditeur de faire cette classification. Nous avons utilisé le même corpus décrit ci-dessus auquel nous avons rajouté d'autres phrases affirmatives pour obtenir 13 paires de question/non-question. Pour chaque phrase, après avoir extrait le contour prosodique, nous utilisons ce contour pour synthétiser une pseudo phrase dans laquelle toutes les syllabes sont remplacées par une

voyelle unique /a/. Etant donné que la signification sémantique des mots n'existe plus, nous éliminons ainsi la possibilité pour l'auditeur de reconnaître une question uniquement par la présence d'un mot «interrogatif ». Nous reproduisons le plus fidèlement possible, non seulement le contour de l'intonation, mais encore la durée des segments voisés/non voisés et le contour de l'intensité. Puis, nous demandons aux auditeurs de déterminer si la pseudo-phrase synthétisée entendue est interrogative ou affirmative.

#### 3.2. Synthèse des phrases pour le test de perception

L'extraction de ces informations prosodique est réalisée avec le logiciel Prat avec des fenêtres d'analyse de 20 ms pour F0 et 5ms pour l'intensité. Pour la synthèse, nous avons extrait deux périodes de signal de la voyelle /a/ de l'une des phrases prononcées par une locutrice de notre précédente étude.



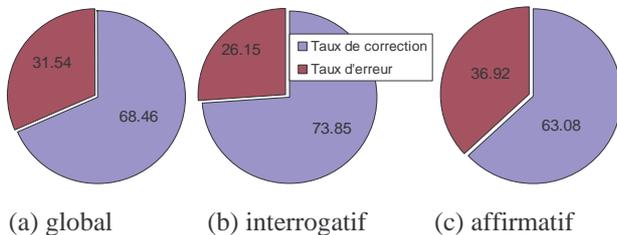
**Figure 2 :** Spectrogramme, contour de F0 (bleu) et contour d'énergie (rouge) du signal source (a) et du signal synthétisé correspondant (b). Phrase 2a du tableau 1.

L'algorithme TD-PSOLA est utilisé pour concaténer ces extraits de signal, tout en contrôlant le pitch (F0), l'énergie et la durée de chaque syllabe. Dans la phrase synthétisée, les zones correspondant à un signal non voisé (consonnes) sont remplacées par du silence. Nous obtenons ainsi un corpus composé de 13 pseudo-phrases synthétiques « interrogatives » et de 13 pseudo-phrases synthétiques « affirmatives », sans information sémantique. Six auditeurs (3 hommes et 3 femmes) participent à notre test de perception. Ils doivent choisir entre deux réponses « I » ou « A ». Chaque auditeur fait le test 5 fois, et pour chaque session, l'ordre des phrases qui lui sont proposées est aléatoire.

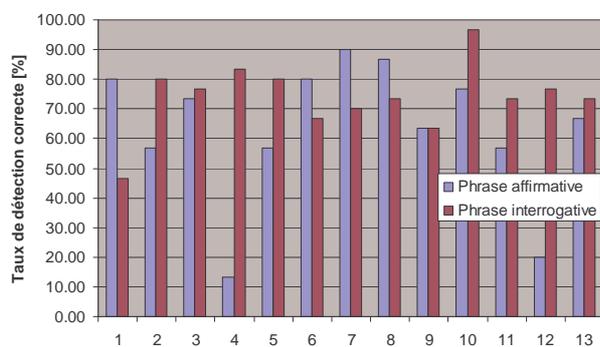
#### 3.3. Résultats de perception

Le résultat du test est présenté dans la figure 2. Le taux de bonne reconnaissance sur l'ensemble des I et A est d'environ 70% (figure 2.a). Les figures 2.b et 2.c présentent respectivement les taux de bonne classification des phrases interrogatives et des phrases affirmatives : nous pouvons remarquer que les phrases interrogatives semblent mieux reconnues (environ 74 % de bonnes réponses) que les phrases affirmatives (seulement 63%).

La figure 3 détaille les résultats pour les 13 paires de phrase I/A. Pour 10 de ces 13 paires, la phrase interrogative est bien reconnue avec un taux supérieur à 70%), pour la 10<sup>ème</sup> paire, ce taux atteint les 95%. Cependant pour les paires 4 et 12, la phrase affirmative est très mal reconnue (respectivement 12% et 20%).



**Figure 3** : Taux de détection correcte : (a) taux global (b) phrases interrogatives et (c) phrases affirmatives.



**Figure 4** : Taux de détection correcte des 13 paires I/A.

### 3.4. Discussion

En tentant de corrélérer ces résultats perceptifs avec ceux de notre analyse de la production des contours intonatifs, nous remarquons que l'auditeur semble juger une phrase comme étant interrogative, si elle présente une intonation croissante en fin de phrase, et juger la phrase comme étant affirmative dans le cas inverse. Cette hypothèse semble être valable pour expliquer le cas des paires 4 et 12 où le taux de reconnaissance des phrases type I est beaucoup plus élevé que celui des phrases type A. Pour ces deux paires, les phrases présentent toutes une dernière syllabe possédant le ton 5 montant, qui fait croître la partie finale du contour intonatif de la phrase, tant pour les interrogations que pour les affirmations. Le fait que le taux de bonne reconnaissance global des phrases A et I soit d'environ 70% (et que pour certaines d'entre elles, elles sont même reconnues à plus de 90%) montre que les paramètres prosodiques de la phrase vietnamienne transportent des informations extralinguistiques qui peuvent permettre à l'auditeur de discriminer le type de phrase. Comme pour les langues non tonales, ces informations sont essentiellement codées par le fait que l'intonation monte ou non en fin de phrase. Cependant, ces informations peuvent être brouillées par la modulation du contour prosodique par les tons lexicaux : des auditeurs peuvent mal classifier des affirmations si les phrases produites présentent une syllabe finale avec ton montant. Des questions peuvent être mal classifiées si leur syllabe finale porte un ton descendant. L'utilisation

de mots interrogatifs pour lever les ambiguïtés est donc nécessaire et logique.

## 4. CONCLUSION

Au niveau production, notre étude a permis de caractériser la prosodie des phrases simples de la langue vietnamienne (dialogue), en éliminant l'influence des tons : les différences entre questions et affirmations sont essentiellement une différence de pente de F0 (croissante ou décroissante) en fin de la phrase (deuxième moitié de la dernière syllabe), à laquelle s'ajoutent une modification du débit. Cependant, pour notre étude, le changement de registre semble plus faible que pour [8 et 9]. Au niveau perceptif, nous avons montré que, comme pour les langues non tonales, la prosodie de la phrase transporte des informations extralinguistiques sur la nature de la phrase, bien que celles-ci, à cause de la présence des tons lexicaux, ne soient pas toujours discriminatives.

## BIBLIOGRAPHIE

- [1] Rossi M. "L'intonation, le système du français : description et modélisation" Editions Ophrys, 1999, ISBN : 2-7080-0912-5
- [2] Hirst, D.J. & Di Cristo, A. (Eds.) "Intonation Systems. A Survey of 20 Languages" Cambridge: Cambridge University Press.
- [3] Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Jurafsky, D., Ries, K., Cocarro, N., Martin, R., Meteer, M. & Van Ess-Dykema, C. "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?" *Language and Speech* 41, pp. 439-487, 1998
- [4] Vu M.Q., Castelli E., Boucher A. & Besacier L. "Classification de parole en Question et Non-Question par arbre de décision" *SFC 05, 12<sup>èmes</sup> Rencontres de la Société Francophone de Classification* - Montréal, 2005
- [5] Nguyen Q.C., Pham Thi N.Y. & Castelli E. "Shape vector characterization of Vietnamese tones and application to automatic recognition" *ASRU 2001 Madonna di Campiglio*, cdrom
- [6] Pham Thi N. Y., Castelli E. & Nguyen Q.C. "Gabarits des tons vietnamiens" *JEP 2002 Nancy*, pp 23-26, juin 2002.
- [7] Michaud A. & Vu N.T. "Glottalised and non glottalised tones under emphasis: open quotient curves remain stable, F0 curve is modified" *Speech Prosody, Nara, Japan*. 745-748, 2004
- [8] Lê Thị X., "Etude contrastive de l'intonation expressive en français et en vietnamien". *Thèse en linguistique : Paris, Université Paris 7*, 1989
- [9] Nguyễn Thị T.H. & Boulakia, G. "Another look at vietnamese intonation" *14<sup>th</sup> International Congress of Phonetic Sciences, San Francisco, California*, pp. 2399-2402, 1999