



**HAL**  
open science

## Focus et tons lexicaux: de la variété des paramètres prosodiques

Alexis Michaud

► **To cite this version:**

Alexis Michaud. Focus et tons lexicaux: de la variété des paramètres prosodiques. A. Lacheret-Dujour, J. François. Fonction et moyens d'expression de la focalisation à travers les langues, Mémoires de la Société de Linguistique de Paris, nouvelle série, tome XIII, Peeters, pp. 13-47, 2003, Mémoires de la Société de Linguistique de Paris, nouvelle série, tome XIII. hal-00130159

**HAL Id: hal-00130159**

**<https://hal.science/hal-00130159>**

Submitted on 9 Feb 2007

**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.

## Focus et tons lexicaux : de la variété des paramètres prosodiques

Alexis Michaud, Laboratoire de Phonétique et Phonologie (UMR7018) CNRS / Sorbonne Nouvelle

Paru dans *Fonction et moyens d'expression de la focalisation à travers les langues*, A. Lacheret-Dujour et J. François, Mémoires de la Société de Linguistique de Paris, nouvelle série, tome XIII, Louvain, Peeters, pp. 13-47.

*Dans cette version (« postprint »), quelques corrections mineures ont été apportées, telles que la notation des tons en tons ponctuels ( ´, ˘, ˙) au lieu de 1, 1, 1.*

### Résumé

Dans les langues à tons d'Asie se manifeste une grande latitude de variation dans la réalisation prosodique des syllabes, ce qui n'empêche pas l'identification des tons lexicaux. L'étude de syllabes sous focus en naxi, langue tibéto-birmane, montre la diversité des paramètres prosodiques par lesquels se réalise la superposition des tons et de l'intonation. L'emploi de l'électroglottographie, mesure non invasive de la surface d'accolement des cordes vocales, permet deux mesures complémentaires des paramètres couramment relevés que sont la fréquence fondamentale, la durée et l'intensité. La première est celle du *quotient ouvert*, qui aide notamment à caractériser l'*accent d'insistance*, phénomène présent en naxi (et très certainement universel), distinct de la *focalisation*. La seconde mesure est celle de l'*amplitude des pics de fermeture* mesurée sur la dérivée du signal électroglottographique ; cette mesure baptisée DECPA, étudiée pour la première fois, contribue à éclairer les questions de l'*intensité* et de l'*effort vocal*.

### Abstract

In Asian languages, lexical tones may undergo extensive variation carrying prosodic indications. Studying syllables under focus in Naxi, a Tibeto-Burman language, shows that several parameters are involved. Electroglottography, a noninvasive means of measuring vocal fold contact surface, allows for the calculation of two parameters in addition to fundamental frequency, duration and intensity : one is the open quotient, which helps single out cases of *emphatic stress*, a (very likely universal) phenomenon distinct from prosodic *focalization*, and the amplitude of the closure peaks on the derivative of the electroglottographic signal. The latter measurement, DECPA, is attempted here for the first time ; it helps shed light on issues of *intensity* and *vocal effort*.

## 1. Introduction et cadre de recherche

### 1.1 Prosodie et physiologie

#### 1.1.1. La mesure des phénomènes glottiques

La présente étude repose sur une définition large de la prosodie : celle proposée par J. Vaissière, qui comprend *l'ensemble des phénomènes non prédictibles à partir de la connaissance des phonèmes*. La division de ces phénomènes en trois composantes fait ressortir leur variété : composante sous-glottique (variation de pression d'air sous-glottique, liée à l'activité des muscles respiratoires), glottique (action des muscles du larynx), et supra-glottique : l'articulation, dont les variations véhiculent des indications prosodiques<sup>1</sup>. L'observation de ces trois composantes

---

<sup>1</sup> La définition de la prosodie peut être étendue aux mouvements de l'ensemble du visage et du corps. Dans le cas de la langue naxi, sur laquelle porte l'enquête, les enregistrements vidéo montrent des mouvements de la tête et des bras pour appuyer un mot important, ainsi qu'une tendance à la montée

s'appuie sur diverses techniques exploratoires. Au plan supra-glottique, les modifications de l'articulation des voyelles et des consonnes peuvent être observées acoustiquement, sur spectrogramme, mais aussi par des mesures des mouvements des articulateurs : mâchoire, langue, voile du palais. Les mouvements de la langue et leur relation avec l'accent sont présentés par Mooshammer et Fuchs 2002. Le rôle prosodique du degré d'ouverture de la mâchoire est montré par Erickson, Fujimura et Menezes 2002:495-497 ; ces auteurs vont jusqu'à conclure que les modifications du degré d'ouverture de la mâchoire au cours d'un énoncé indiquent la « structure métrique » de l'énoncé, tandis que les variations de fréquence fondamentale leur paraissent laissés à la libre appréciation du locuteur. Au plan sous-glottique, il est possible, par des techniques invasives, de mesurer la pression d'air sous-glottique, et l'activité des muscles respiratoires. L'étude consacrée par Ladefoged, Draper et Whitteridge 1958 à la composante sous-glottique illustre la confiance accordée aux nouvelles techniques d'observation, en l'occurrence l'électromyographie, technique (fortement invasive) de mesure de l'activité musculaire. Après avoir exposé leurs résultats, les auteurs évoquent quelques hypothèses complémentaires, pour ajouter aussitôt : « Mais les spéculations sur ce sujet sont vaines, puisqu'une preuve objective peut être établie par n'importe quelle équipe de physiologistes et de phonéticiens qui prend la peine d'insérer une électrode dans le muscle concerné. » Les conclusions des études menées sur les muscles du larynx sont moins triomphales. La source sonore (qui est au centre de la présente étude) est complexe, tant au plan linguistique qu'au plan physiologique. Elle entre en jeu dans les diverses composantes de la prosodie : expression des émotions (voir en particulier le chapitre 2 de Fónagy 1983, qui décrit cinq états de la glotte correspondant à cinq émotions), réalisation des frontières syntaxiques, des phénomènes accentuels, et de la mise en valeur pragmatique. Les paramètres glottiques sont la fréquence fondamentale (paramètre central) et d'autres paramètres, plus difficiles à définir, qui constituent la « qualité vocale ». On trouve chez d'Alessandro la formule de « prosodie au sens étroit », qui recouvre mélodie, durée et intensité, l'ambition actuelle des chercheurs étant d'élargir la modélisation (en particulier à la qualité vocale : voir d'Alessandro, Boula de Mareüil et Prudon 2003).

### 1.1.2. Un paramètre glottique : la pente spectrale. La question de l'intensité.

Sluijter, Van Heuven et Pacilly 1997 soulignent une erreur de méthode dans les expériences qui avaient conduit, à partir des années 1950, à prêter à l'intensité du son un rôle linguistique très secondaire. Dans les expériences de manipulation de l'intensité menées par Fry 1955, 1958; Mol et Uhlenbeck 1956; Issatchenko et Schädlich 1966, l'intensité était modifiée par un changement de gain global, qui affectait également toutes les fréquences ; or dans la parole naturelle, l'augmentation de l'intensité vocale résulte de phénomènes plus complexes. Il est nécessaire de

---

de la tête sur les syllabes au ton haut et sa descente sur tons bas ; gestes sans mystère pour l'observateur. Il ne paraît pas exister de code gestuel qui serait associé à tel ou tel registre de parole. Dans l'ensemble, les Naxi ne semblent pas cultiver le geste comme appui du discours.

tenir compte de cette spécificité, celle de la présence ou absence d'*effort vocal*. Au plan perceptif, l'oreille sait reconnaître cet effort : un stimulus produit avec un effort vocal sera jugé « plus intense » qu'un stimulus prononcé sans cet effort, même si l'intensité globale est en fait identique dans les deux cas : voir Glave et Rietveld 1975 et références dans Vaissière à paraître. « L'oreille opère une sommation des composantes spectrales différente de celle qui régit la mesure de l'intensité globale » (Rossi 1971:143). Etudiant le néerlandais et l'anglais américain, Sluijter, Van Heuven et Pacilly montrent le rôle de la répartition spectrale de l'intensité (ou : pente spectrale), découverte qui relance la question des paramètres prosodiques. Au sujet du rôle linguistique de la pente spectrale, ces auteurs apportent une réponse très tranchée : beaucoup d'intensité dans les hautes fréquences (« high-frequency emphasis ») serait un corrélat de l'accent lexical (en anglais : « stress »), dont un autre corrélat important est la durée, tandis que les mouvements de fréquence fondamentale (ci-après  $F_0$ ) seraient des corrélats de l'accent intonatif (en anglais : « accent »). Dans la dichotomie établie par Sluijter et Van Heuven, ce dernier terme recouvre les phénomènes pragmatiques de mise en valeur d'une syllabe, et regroupe « intonation de focalisation » et « accent d'insistance » ; il est parfois remplacé par le terme de « cible de focus » ou d'« accent de phrase ».

Au plan linguistique, cette dichotomie est surprenante : l'expérience des recherches en prosodie semble au contraire montrer qu'il n'existe pas de correspondance terme à terme entre faits linguistiques et paramètres acoustiques. Les paramètres intonatifs sont liés, tant au plan de la production qu'au plan de la perception. Il ne manque pas d'observations qui contredisent la proposition de Sluijter, Van Heuven et Pacilly : la fréquence fondamentale est citée en premier par J. Vaissière parmi les (nombreux) corrélats de l'accent lexical ; parallèlement, la fréquence fondamentale n'est pas le seul corrélat de l'accent intonatif (« focus ») : pour le suédois, Gobl 1988 montre que le focus est associé à un surcroît d'énergie dans les hautes fréquences ; Fant et Kruckenberg 1995:623 décrivent, également en suédois, des différences spectrales sous l'accent lexical, mais précisent que ces différences sont plus marquées sur les mots qui reçoivent un accent intonatif<sup>1</sup>. De même, Heldner (Heldner, Strangert et Deschamps 1999; Heldner 2001a, b) utilise la mesure de pente spectrale pour détecter ce qu'il nomme « focus », « accent focal », « accent » ou « proéminence », phénomène d'ordre pragmatique correspondant au terme anglais d'« *accent* » chez Sluijter, Van Heuven et al. 1997. En anglais, Campbell 1995 observe une augmentation de l'énergie dans les hautes fréquences lorsqu'il y a mise en valeur.

Ces divers résultats confirment l'importance de la pente spectrale comme paramètre intonatif, démontrant qu'elle n'a pas pour seule fonction de signaler l'accent lexical. Il nous paraît particulièrement intéressant de tenter de prendre en compte ce paramètre dans la description du naxi, langue sans accent lexical, qui connaît des

---

<sup>1</sup> Ces auteurs préfèrent au terme de « focus » ou « accent intonatif » celui d'« accent emphatique ». Ce choix terminologique n'est pas retenu ici, parce qu'il risque de masquer la distinction (qui nous paraît indispensable, voir 2.2.3.2) entre focalisation (accent intonatif) d'une part, accent d'insistance de l'autre.

phénomènes locaux de mise en valeur pragmatique malgré la contrainte lexicale sur la fréquence fondamentale (chaque syllabe possédant un ton propre).

### 1.1.3. Importance de mesures « à la source ».

L'amplification des hautes fréquences est liée à la forme de l'onde de débit glottique. Une introduction à cette question se trouve dans Titze 1994:117-120. Il n'existe actuellement pas de méthode efficace pour la mesure de l'onde de débit glottique et de la pente spectrale. La méthode expérimentale choisie ici est l'électroglottographie, qui, à défaut du débit d'air passant par la glotte, mesure de façon très précise la surface d'accolement des cordes vocales. Un exposé préliminaire au sujet de cet appareil est nécessaire, « car c'est seulement en comprenant les techniques employées dans la recherche que le lecteur sera à même de juger par lui-même des propositions qu'on prétend lui prouver » (Ohala 1978:6). L'exposé porte principalement sur deux paramètres calculés à l'aide de l'électroglottographe, dont le premier (le quotient ouvert,  $O_q$ ) est relativement bien connu, tandis que le second (baptisé DECPA) est étudié ici pour la première fois.

## 1.2. *L'électroglottographie : intérêt et limites.*

### 1.2.1 Présentation de l'appareil

L'électroglottographe est un appareil d'utilisation aisée, qui se laisse assez facilement oublier au cours d'une séance d'enregistrement. Un courant de haute fréquence et de très faible intensité circule entre deux électrodes placées de part et d'autre du larynx. Plus la surface de contact entre les cordes vocales est grande, plus la résistance au courant est faible. Le signal est enregistré simultanément avec l'enregistrement audio. On peut le visualiser pendant l'enregistrement avec un oscilloscope, et par la suite avec les mêmes outils que le signal audio. L'électroglottographe (aussi appelé laryngographe) étant un appareil sans danger, les électrodes peuvent rester appliquées sur le cou tout au long d'une séance d'enregistrement. La réalisation d'un enregistrement électroglottographique de qualité satisfaisante est relativement facile, sans commune mesure avec l'électromyographie, technique fortement invasive et qui ne donne souvent que des résultats partiels (à ce sujet, voir Ohala 1978:15). Le filtrage inverse n'a pas été employé, car il repose sur une théorie sujette à débat.

De nombreuses études établissent que le signal électroglottographique (ci-après EGG) varie en fonction de la surface d'accolement des cordes vocales. Parmi les travaux récents, Gilbert, Potter et Hoodin 1984; Titze 1990; Baken 1992 proposent divers paramètres que l'on peut calculer à partir de l'EGG. Des études sont en cours pour en extraire de nouveaux paramètres (ainsi, l'équipe de Bell Labs entreprend de reconstituer l'onde de débit glottique à l'aide de l'EGG ; voir article de Kochanski et Shih soumis pour publication en 2003). Les paramètres relevés sur l'EGG demeurent très variés, et des spécialistes mettent en garde contre les interprétations erronées

de ce signal, qui ne renseigne qu'indirectement sur le flux d'air qui passe à la glotte : voir Colton et Conture 1990; Orlikoff 1998. Presque cinquante ans après la mise au point du premier électroglottographe, un travail très récent (Chen, Robb et Gilbert 2002) se limite encore à la mesure « manuelle » d'un seul paramètre, dont il est difficile de savoir dans quelle mesure il est significatif : la distance entre le point le plus haut et le plus bas au cours d'une période du signal EGG.

Il a pourtant été montré que l'on peut calculer certains paramètres glottiques à partir de l'EGG de façon précise et automatisable, par l'utilisation du signal *dérivé* de l'EGG (ci-après noté DEGG) : voir Henrich 2001, dont les principales conclusions sont résumées dans l'article de Henrich, d'Alessandro, Castellengo et Doval soumis pour publication, « De l'utilisation de la dérivée des signaux électro-glottographiques ». Le signal DEGG présente généralement deux pics par période : un pic positif, un pic négatif. Des expériences combinant électroglottographie et visualisation des cordes vocales, menées à l'Université d'Erlangen-Nuremberg, ont permis de montrer que l'un des pics de la dérivée du signal électroglottographique correspond à une accélération de la fermeture de la glotte (« pic de fermeture »), tandis que l'autre correspond à une accélération de l'ouverture de la glotte (« pic d'ouverture »). Des programmes de mesure de ces deux types de pics ont été mis au point par Vu Ngoc Tuân pour le propos de la présente étude.

### 1.2.2 Le quotient ouvert et son interprétation

Le premier des paramètres étudiés est le quotient ouvert ( $O_q$ ), égal à la durée de la phase ouverte (entre un pic d'ouverture et le pic de fermeture suivant) divisée par la longueur totale de la période (définie comme la durée séparant deux pics de fermeture successifs). Gardant en mémoire qu'« il n'y a pas de lien direct et univoque entre tel ou tel paramètre, le système musculaire de la phonation et l'effet produit » (Henrich 2001:103), il est néanmoins possible d'interpréter  $O_q$  comme indice de l'abduction des cordes vocales (Rothenberg et Mahshie 1988) : le quotient ouvert « renseigne sur la tension de la voix : pressée ou relâchée » (Vu Ngoc, d'Alessandro et Rosset 2002). Toutes choses égales par ailleurs, une montée de  $O_q$  signale un relâchement musculaire, une baisse de  $O_q$  un « serrage » de la voix.

Ce paramètre n'est pas indépendant de  $F_0$  et de l'intensité : en voix de poitrine (« mécanisme I », le seul qui importe pour ce corpus),  $O_q$  tend à monter avec  $F_0$ , et à baisser quand l'intensité vocale augmente (Henrich 2001). La plage de variation observée dans nos données va de 35% (voix très « pressée ») à 70%, voix très « détendue » (ce qui correspond aux chiffres cités par Titze 1988:236).

### 1.2.3. DECPA : l'amplitude des pics de fermeture sur le signal DEGG

Il n'existe pas encore de méthode éprouvée pour relier les signaux électroglottographiques à la pente spectrale. Une chose est sûre : une plus grande rapidité de fermeture des cordes vocales a pour effet un son plus riche en harmoniques (c'est-à-dire une pente spectrale moins forte) ; voir notamment Colton 1984. Cette rapidité peut s'exprimer par le quotient de vitesse, rapport entre la durée

de la phase d'ouverture et la durée de la phase de fermeture. Diverses méthodes d'approximation de ce paramètre directement à partir de l'EKG ont été proposées, mais la détection des points nécessaires à ce calcul, point d'ouverture maximale et point de fermeture maximale, repose sur le choix empirique d'un seuil ; or la forme des signaux EKG varie beaucoup, de sorte que la comparaison (même à l'intérieur du même enregistrement) est hypothéquée d'emblée.

En revanche, il est possible de mesurer de façon très précise l'amplitude du pic de fermeture sur le signal DEKG, qui donne la pente maximum atteinte par le signal EKG lors de la fermeture de la glotte. Cette indication sur la rapidité de fermeture des cordes vocales a toutes chances d'être liée à l'intensité du son, et en particulier à sa richesse en harmoniques. La validation définitive de cette hypothèse devra passer par une mise en regard de l'électroglottographie et d'images de la glotte (voir les recherches en cours de Nathalie Henrich). Pour l'heure, cette nouvelle mesure se trouve donc encore au stade des premières expérimentations ; dans les analyses qui suivent, cette mesure sert d'abord à compléter la mesure d'intensité acoustique globale, dont elle pallie certains inconvénients. On la baptisera (en anglais) DECPA, pour *Derivative-Electroglottographic Closure Peak Amplitude*<sup>1</sup>.

#### 1.2.4. La question des doubles pics

Dans certains cas, l'ouverture ou la fermeture sont moins bien marqués sur le signal DEKG ; ces phénomènes sont décrits par Henrich 2001:113-118. La détermination des instants d'ouverture et de fermeture à partir du signal DEKG pose alors problème. Le programme que nous avons utilisé, créé par Vu Ngoc Tuân, ne détermine le quotient ouvert que dans les cas où a été décelé un pic d'ouverture unique entre deux pics de fermeture ; dans les autres cas, la mesure doit être faite « manuellement », période après période, en repérant visuellement les instants de fermeture et d'ouverture sur le signal EKG. Dans le cas d'un double pic de fermeture, les valeurs de fréquence fondamentale, de quotient ouvert et de DECPA sont corrigées en ne retenant que le pic le plus élevé.

Le second volet introductif est consacré à une présentation du cadre proprement linguistique de la question étudiée, celle de la *pragmatique* (« structure de l'information »).

---

<sup>1</sup> La seule mention d'une mesure de l'amplitude des pics sur le signal DEKG dans les travaux publiés semble être celle de Dromey, Stathopoulos et Sapienza 1992:48 ; cependant, ces auteurs considèrent, non la valeur des pics en elle-même, mais le rapport *amplitude du pic de fermeture/amplitude du pic d'ouverture*. Ils nomment cette valeur  $SQ_{DEKG}$ , c'est-à-dire *quotient de rapidité calculé à partir du signal dérivé de l'EKG*. Il paraîtrait préférable de ne pas employer le terme de « quotient de rapidité » pour cette mesure (remarque due à Nathalie Henrich) ; en outre, l'amplitude du pic d'ouverture n'ayant pas d'incidence connue sur l'onde sonore créée, il paraît souhaitable d'étudier l'amplitude des pics de fermeture (DECPA) sans tenir compte de celle des pics d'ouverture.

### 1.3 La structure de l'information en naxi

#### 1.3.1. Moyens grammaticaux de structuration de l'information

Au seuil de l'étude des phénomènes de modification intonative des tons, il paraît utile de présenter sommairement les moyens grammaticaux de structuration de l'information en naxi. Le naxi est une langue tibéto-birmane parlée dans la province chinoise du Yunnan. Cette langue ne connaît pas de clivées, à la différence, par exemple, du chinois, étudié par Catherine Paris 1979, mais présente divers moyens syntaxiques de mise en valeur.

##### 1.3.1.1 Mise en valeur et mise en retrait

Le premier moyen est négatif : ne pas expliciter certains syntagmes, ce qui revient à mettre en valeur ceux qui demeurent. L'ellipse du sujet en naxi, décrite par He et Jiang 1985:100-101, illustre le paramètre « pro-drop » de la grammaire générative, étiquette glossocentrique que Launey 1994 propose de renverser en paramètre « NP-Add » au vu du fait qu'« il est probable qu'au bout du compte le type *pro-drop* apparaîtra largement dominant »<sup>1</sup>. Le naxi est, sur ce point, similaire au tamang, autre langue tibéto-birmane, qui connaît une ellipse du « référent actif » (ou « topic »), phénomène décrit dans Mazaudon 2003, « Du discours à la grammaire en tamang : topic, focus, intensificateurs et subordination », § 3.1.1. Voici un exemple simple en naxi :

(1) nù è tsū bē nù? « Qu'est-ce que tu es en train de faire ? »

*2<sup>e</sup> pers. sg.-pronom interrogatif « quoi »-faire-particule aspect/mode*

(2) hā t<sup>h</sup>v nù. « [Je] fais la cuisine. »

*nourriture-cuisiner-particule aspect/mode*

Dans les phénomènes de saillance linguistique, la mise en avant de tel ou tel élément d'un énoncé est solidaire de la mise en retrait d'autres éléments. Cela se vérifie dans le domaine de l'intonation : dans de nombreuses langues, le « post-focus » (ensemble des syllabes qui suivent le « focus ») est affaibli ; il présente des courbes (de fréquence fondamentale, d'intensité...) relativement égales, ce qui crée un contraste net avec le focus. Ce phénomène, rapporté par de nombreux auteurs dans l'ouvrage collectif dirigé par Hirst et Di Cristo 1998, se traduit également au plan grammatical. Par exemple :

(3) Tu es venu par le train de 6 heures ?

(4) Non, je suis venu par celui de 7 heures./ (4') Non, par celui de 7 heures.

En français, (4') paraît plus plausible que (4) dans ce contexte. Lors de l'enregistrement de notre corpus naxi, l'informateur a prononcé la forme complète (5) ; devant ensuite répéter avec pour consigne « tu l'as déjà dit plusieurs fois ; tu es agacé, impatient », l'informateur a répété deux fois (5) puis est spontanément passé à (5') :

(5) m̄ wà, η̄ ʂə̄ dzù ḡ ʂ<sup>h</sup>ū kjá ndzù pú ts<sup>h</sup>ù m̄.

---

<sup>1</sup> Launey 1994:42-47, 93.

non- 1<sup>e</sup> sg.-sept-heure-gén.-déictique-classificateur-être assis-gérondif [qui disparaît dans l'énoncé 5']-arriver-particule affirmative « Non, je suis venu par le train de sept heures. »

(5') m̄x wà, ŋ̀ s̄ə̄ dzù ḡx ts̄h̄w̄ kjá ndzũ ts̄h̄w̄ m̄x. (même signification)

Dans (5'), la particule /pú/ n'apparaît plus : il n'en reste plus qu'une trace tonale, dans le passage au ton montant du ton bas qui précède. Une description de ce phénomène morpho-phonologique est proposée dans Michaud 2003.

### 1.3.1.2 Ordre des mots

Un autre paramètre abondamment employé dans la structuration de l'information est l'ordre des mots.

(6) C'est Jean-Pierre qui est sorti avec Marie ?

(7) ū lí ū lí kó sū sū lá? (équivalent naxi de l'énoncé 6)

nom propre 1-nom propre 2-particule objet-connaître (redoublé)-particule interrogative

C'est nécessairement le premier des deux termes qui est mis en valeur. La prosodie à elle seule ne permet pas de changer le sens de l'énoncé : pour changer de focus, il faut changer l'ordre des mots.

### 1.3.1.3 Les particules

En naxi, plusieurs particules jouent un rôle de premier plan dans la structuration de l'information. La brève description qui en est proposée ici reprend les catégories utilisées par Mazaudon sous presse. La particule /sé/ marque la topicalisation, et le conditionnel (dualité qui rappelle les observations de Haiman 1978 : « Les conditionnels sont des topics »). Un rôle similaire est joué par /t<sup>h</sup>ũ/, particule homophone avec le pronom de 3<sup>e</sup> personne, et avec le déictique « lointain », dont elle provient sans doute (voir He et Jiang 1985:74-75, 88). Une particule /d̄d̄/, d'après les quelques exemples de notre corpus, sert à marquer le *topic contrastif*. Soit l'échange suivant, dans lequel la réplique est en décalage avec la question : « Tu as vu Jean-Pierre ? - J'ai vu son blouson dans l'entrée. » En naxi, la réponse a été traduite par :

(8) t<sup>h</sup>ũ ḡx b̄ā là d̄d̄ k<sup>h</sup>ũ k<sup>h</sup>ú t<sup>h</sup>è há j̄x.

3<sup>e</sup> pers. sg.-GEN.-vêtement-particule de topic contrastif-entrée-GERONDIF-se trouver-particule finale

que l'on peut paraphraser par : « son vêtement, lui, se trouve dans l'entrée. » L'emploi de cette particule n'est pas exclusif d'un marquage intonatif, au contraire : le mot « vêtement », /b̄ā là/, porte dans cet exemple l'« intonème de focalisation » dont la réalisation phonétique est détaillée au chapitre 2.

Enfin, il existe des particules qui marquent l'agent : /nũ/ et le patient : /tól~/kól/. L'ajout des particules n'est indispensable que lorsqu'il y a inversion sujet-objet (d'ordinaire, c'est l'agent qui vient en premier). Ces particules étant facultatives dans la plupart des cas, leur emploi, qui explicite le statut actanciel, constitue une forme de mise en valeur de l'argument concerné.

L'inventaire des particules est néanmoins plus limité qu'en tamang ; en particulier, il n'existe pas de particule qui marque spécifiquement le focus.

### 1.3.2. La prosodie des langues à tons d'Asie

Les langues à tons d'Asie sont du plus grand intérêt pour qui s'intéresse à la prosodie. Le fait qu'une syllabe porte un ton lexical n'exclut nullement une grande latitude de variation dans sa réalisation phonétique, cette variation étant porteuse de nombre d'indices. En chinois mandarin, la déformabilité tonale est décrite de façon très fine par Chao Yuen-ren (Chao 1932 ; Chao 1933), qui se fonde sur son intuition d'*Ohrenfonetiker* et de locuteur. Chao 1929 dégage vingt-sept « morphèmes intonatifs » qui reposeraient sur la variation de la hauteur et de la durée, et treize autres qui reposeraient sur la variation de l'intensité et de la qualité vocale. Pour rendre sensible à ses auditeurs la superposition des tons lexicaux, de l'intonation syntaxique et de l'intonation expressive (pragmatique et émotions), Chao Yuen-ren avait achevé l'une de ses conférences (Chao 1935) par la lecture d'un même texte de trois façons : syllabe après syllabe ; avec la seule intonation syntaxique ; avec intonation expressive. Kenneth Pike 1948 fournit également une brève analyse de ce qu'il baptise « changements mélodiques non phonémiques », phénomènes locaux d'intonation pragmatique. Le naxi et d'autres langues à tons d'Asie partagent cette caractéristique du chinois. La profonde influence de l'intonation sur les tons amène beaucoup d'auteurs à employer le mot d'« accent » dans la description de ces langues, qui pourtant ne possèdent pas d'accent lexical distinctif<sup>1</sup> : pour le chinois mandarin, Pike 1948:29 parle d'« accent d'insistance » ; en thai et vietnamien, René Gsell note trois catégories d'« accents » commandés par la syntaxe, un « accent emphatique » et un phénomène de « focus » (Gsell 1979a). Ces phénomènes dépassent la « coarticulation tonale » par laquelle les tons en contact s'influencent phonétiquement, phénomène étudié notamment par Xu 1997 pour le chinois, Han et Kim 1974 pour le vietnamien. Les descriptions du volet syntaxique de la prosodie sont maintenant assez nombreuses, cette recherche étant particulièrement encouragée dans le cadre des projets de synthèse de la parole (pour le chinois : Shih 2000; Shih et Kochanski 2000; Shih, Kochanski, Fosler-Lussier, Chan et Yuan 2001). La présente étude se concentre sur le volet pragmatique de l'intonation, étudiant la question du *focus* et de l'*insistance* en naxi : les divers degrés de mise en valeur prosodique d'une syllabe.

## 2. Mesures

Deux locuteurs ont participé aux enregistrements audio et électroglottographiques, réalisés au laboratoire de phonétique de l'Institut d'étude des minorités. Les mesures

---

<sup>1</sup> Le mandarin possède des syllabes au « ton neutre », c'est-à-dire sans ton ; une des possibilités pour la modélisation phonologique est de dire que toutes les syllabes portent un accent sauf les syllabes au « ton neutre ». On retrouve également des syllabes sans ton propre, ou « désaccentuées », dans d'autres langues de la région. En birman, par exemple, leur spécificité ressort du statut « extra-métrique » qui leur est accordé dans la versification (D. Bernot, communication personnelle).

concernent principalement un seul individu, la comparaison des résultats obtenus pour divers locuteurs se heurtant aussitôt aux considérables variations entre individus (voir Strik et Boves 1992, Vaissière à paraître ; sur la variabilité inter-sujets des enregistrements électroglottographiques, voir Chen, Robb et al. 2002). L'informateur, HX, a 24 ans. Sa prononciation hésite parfois entre le parler de son village de naissance et celui de la ville de Lijiang, où il a vécu dès l'adolescence<sup>1</sup>. Il possède une formation en chant. Sa voix se situe toujours en « registre de poitrine » (« mécanisme I » de Roubeau, Chevie-Muller et Arabia-Guidet 1987, 1991, Henrich 2001:104), sans changement de mécanisme laryngé en fonction de la plage de fréquence fondamentale.

Une présentation des tons isolés précèdera l'étude des tons en énoncé.

## *2.1. Tons isolés et syntagmes isolés*

### 2.1.1. Les quatre tons du naxi : fréquence fondamentale et quotient ouvert

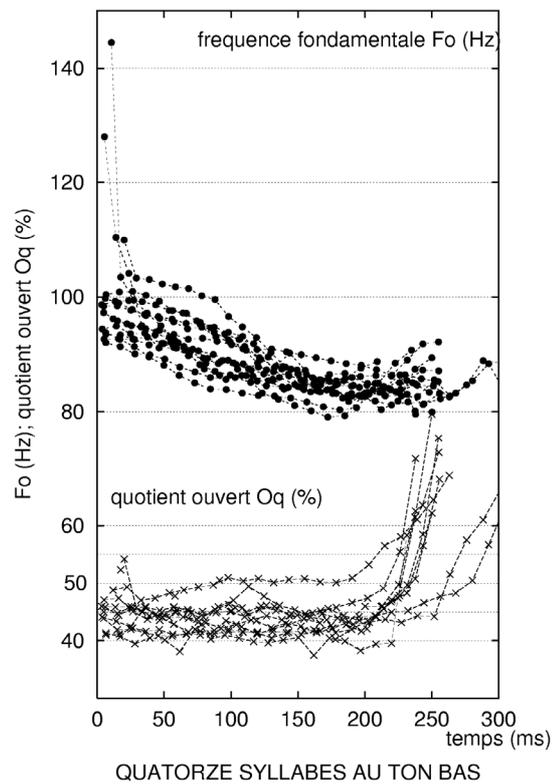
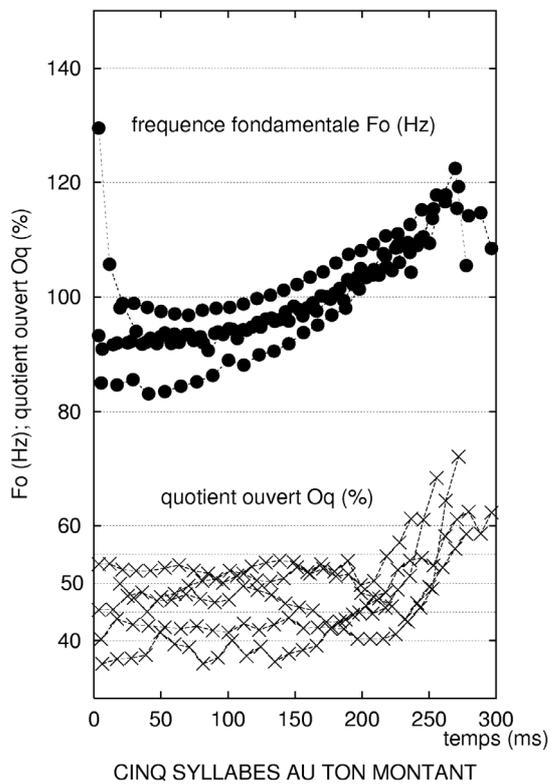
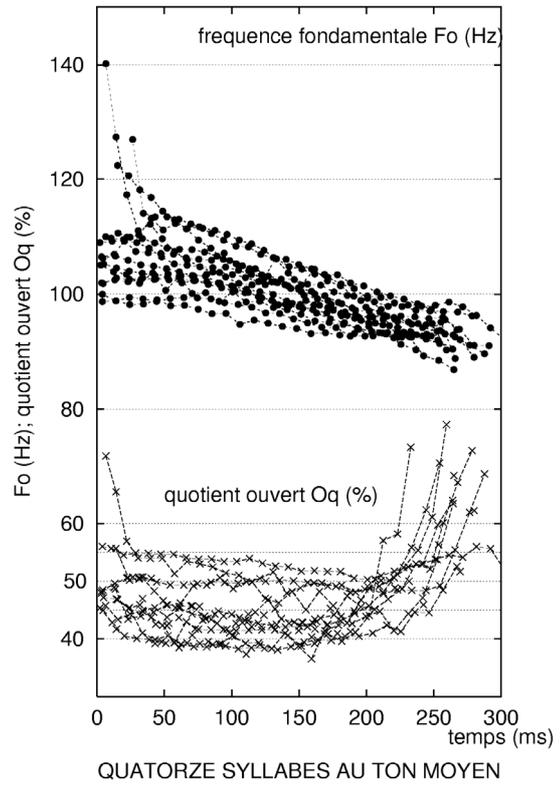
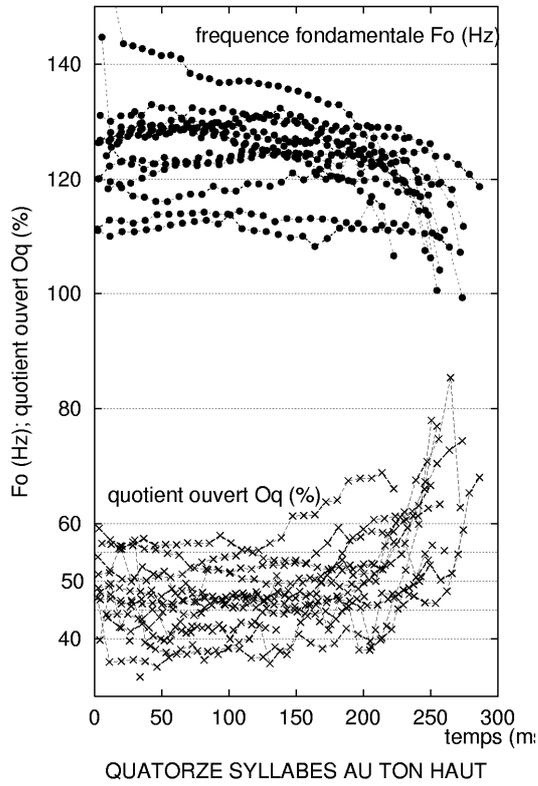
La structure syllabique de base du naxi est très simple : (C1)(C2)V, C2 ne pouvant être qu'une semi-consonne /w/ ou /j/. La langue possède quatre tons lexicaux : trois tons ponctuels Haut, Moyen et Bas, et un ton montant<sup>2</sup>. Un « quadruplet minimal » illustrera les quatre tons du naxi : lá « frapper » ; lā « tigre » (homophone : « aussi, également ») ; là « main » ; lǎ « pimenté ; piquant ».

Un graphique pour chacun des quatre tons présente les courbes superposées de plusieurs syllabes. Toutes les valeurs sont calculées sur la rime, sans tenir compte de l'initiale ; c'est-à-dire en fait sur la voyelle, le naxi n'ayant pas de syllabes fermées. La justification de ce choix vient du fait que la fréquence fondamentale (sur les consonnes voisées) ne joue pas de rôle prosodique, comme le remarque Pike 1948:30 : « Je n'ai jamais observé de mélodie significative (ni pour l'intonation, ni pour les tons lexicaux) sur une consonne initiale voisée (fricative ou occlusive). » Le ton montant, plus rare en syllabe isolée, n'est représenté que par cinq syllabes. Les courbes suivantes ont été calculées à partir de syllabes isolées en phrase-cadre, procédure usuelle pour stabiliser la réalisation du ton. La phrase-cadre est : « Je dis :... », /ŋɿ ... ʂɿ/. Le verbe étant final, ces deux syllabes suffisent à encadrer la syllabe sur laquelle porte l'analyse. Les paramètres représentés sont le quotient ouvert (plage de variation : de 30 à 70%) et la fréquence fondamentale (plage de variation : de 75 à 150Hz).

---

<sup>1</sup> Ainsi, HX prononce [à tsū] (« quoi ») et non [è tsū], [ā nè] (« qui ») plutôt que [ē nè] ; sa prononciation la plus fréquente de la particule [nù] (énoncé 1) est [nè], prononciation la plus courante dans la ville de Lijiang.

<sup>2</sup> Ce ton montant résulte de processus morpho-phonologiques et possède également un statut lexical, en particulier sur des mots d'emprunt ; voir à ce sujet Michaud 2003.



De ces graphiques, retenons trois observations.

- Les trois tons définissent trois bandes de fréquence fondamentale qui ne se chevauchent pas : aucun ton montant n'apparaît dans la même plage de fréquences que le ton haut ou le ton bas. Cette division en trois parties apparaît très tôt dans la voyelle, et se poursuit jusqu'à la fin de la syllabe.

- La dispersion des valeurs est plus forte au ton haut. Pour la fréquence fondamentale  $F_0$ , cette dispersion reste sensible même si l'on traduit les valeurs en écarts sur la gamme musicale, qui correspond mieux à la perception de la voix humaine. Quant au quotient ouvert, sa plage de variation est globalement la même pour les quatre tons (sur ce plan, le naxi contraste avec la spectaculaire variation des six tons du vietnamien ; voir Vu Ngoc, d'Alessandro et al. 2002), et l'on observe pour les quatre tons une montée finale de quotient ouvert, correspondant au relâchement de la tension des cordes vocales lors de l'interruption progressive du voisement ; mais les valeurs sont, là aussi, plus dispersées au ton haut. Cette asymétrie tient à la constitution de l'appareil vocal humain ; le fait qu'un ton haut possède une latitude de réalisation plus grande qu'un ton moyen ou bas fait peser une contrainte sur les possibilités de mise en valeur prosodique des syllabes.

- Le ton haut est stable. Les tons moyen et bas présentent, au plan phonétique, une descente (également sensible à l'oreille). Les mesures des fréquences fondamentales maximum, minimum et moyenne (mesures vérifiées à l'aide du logiciel PRAAT) confirment cette impression visuelle qui se dégage des tracés : aux ton bas et moyen, la valeur qui présente l'écart-type le plus faible est la valeur minimum (atteinte en fin de syllabe). Au ton haut, c'est la moyenne de  $F_0$  sur l'ensemble de la voyelle qui présente l'écart-type le plus faible. Le fait que des /tons statiques/ se réalisent phonétiquement comme des [tons descendants] n'est pas surprenant. Le même phénomène ressort de l'analyse du yoruba par Hombert 1976 ; en vietnamien, les tons *ngang* (Haut) et *huyền* (Bas) sont classiquement décrits comme formant une paire de tons ponctuels, mais le ton *huyền* (Bas) est nettement descendant, tandis que le ton *ngang* (Haut) ne présente pas de descente équivalente (voir Gsell 1980, et étude quantifiée dans Vu Ngoc, d'Alessandro et al. 2002).

Le ton montant présente un glissando montant dont l'ampleur dépasse celle de la descente du ton bas et du ton moyen (l'écart-type étant équivalent), surtout si l'on tient compte de l'asymétrie entre descente et montée de fréquence fondamentale, la première étant plus rapide et physiologiquement plus aisée (Sundberg 1979 ; Ohala et Ewan 1973). Ces mesures fournissent donc un argument phonétique pour décrire ce ton montant comme un *ton modulé*, l'opposant ainsi aux trois autres tons, *tons ponctuels*<sup>1</sup>.

---

<sup>1</sup> La tradition chinoise depuis Fu 1940, 1944 consiste à noter le ton le plus bas comme bas-descendant (31 ou 21 dans la notation de Chao Yuen-ren). Mais au plan phonétique le ton bas n'est pas sensiblement plus descendant que le ton moyen ; par ailleurs, des considérations de phonologie encouragent fortement à une description en trois tons ponctuels.

### 2.1.2. Valeur des pics de fermeture sur le signal DEGG (DECPA)

- Le paramètre DECPA présente une très grande amplitude de variation d'une syllabe à l'autre : son coefficient de variation est de l'ordre de 0,4, valeur qui contraste avec le coefficient de variation de la fréquence fondamentale, environ dix fois moins élevé. En outre, ce paramètre peut varier fortement au cours d'une même syllabe, même dans le « phonostyle » très contrôlé qu'est la production de syllabes en phrase-cadre.
- Des régularités ressortent néanmoins du calcul statistique. Les valeurs au ton haut sont supérieures aux valeurs observées au ton moyen, elles-mêmes inférieures aux valeurs au ton bas. La tendance générale est donc à une corrélation entre  $F_0$  et DECPA. Cette relation est de même nature que celle qui lie la pression sous-glottique et la valeur de fréquence fondamentale (décrite par Ohala 1978, Titze 1994) : ces paramètres ne peuvent être déduits l'un de l'autre, mais il est possible de proposer une quantification de leur relation de dépendance, toutes choses égales par ailleurs. La première quantification de cette relation proposée ici se fonde sur des mesures sur 62 syllabes, en cadre et hors cadre. Pour l'ensemble de la plage de  $F_0$ , on peut proposer l'approximation suivante : DECPA change de 1,7% lorsque  $F_0$  change de 1%.

### 2.1.3. Deux exemples de mise en retrait prosodique : phrase-cadre ; syntagme nominal.

#### 2.1.3.1. Comportement de la phrase-cadre

Une brève analyse du comportement de la phrase-cadre (« Je dis :... », /ŋʔ ... ʂʔ/ ; voir 2.1.1) donnera une idée de la souplesse tolérée par les tons lexicaux, et de son exploitation pragmatique. La saillance pragmatique de la syllabe « encadrée » s'accompagne d'une mise en retrait des syllabes « encadrantes ». Entre le cadre, qui ne varie pas, et la syllabe encadrée, la hiérarchie au plan informatif est claire. Phonétiquement, le verbe final est mis en retrait ; le phénomène est particulièrement sensible lorsque la « syllabe encadrée » porte, comme le verbe /ʂʔ/ final, un ton lexical haut (on peut s'en convaincre à l'écoute du fichier sonore « syllabes »). Dans cette configuration, le  $F_0$  est de 5% moins élevé en moyenne sur /ʂʔ/ que sur la syllabe au ton haut. Le quotient ouvert est légèrement plus élevé (de 7%), ce qui signale une voix légèrement moins « tendue/pressée ». Les mesures de DECPA sont plus basses de 10% environ.

Ce contraste entre deux syllabes qui se succèdent (et sont toutes deux au ton haut, de sorte que l'on ne peut parler de *downstep* ; voir Rialland 1997) montre que les phénomènes d'affaiblissement de la parole à mesure de l'énoncé (baisse de la fréquence fondamentale, de l'énergie, de l'ampleur des gestes articulatoires, etc.) ne peuvent se comprendre s'ils sont abordés uniquement sur une base temporelle. René Gsell souligne que le rétrécissement de l'espace tonal à mesure que la phrase se déroule s'observe également dans des phrases de courte durée (Gsell 1979a:62). Le phénomène est très profondément influencé par la composante pragmatique de la

prosodie, ce qui explique qu'il n'existe pas d'outil qui permette de calculer automatiquement un  $F_0$  (ou autre paramètre) corrigé de l'effet de la déclinaison.

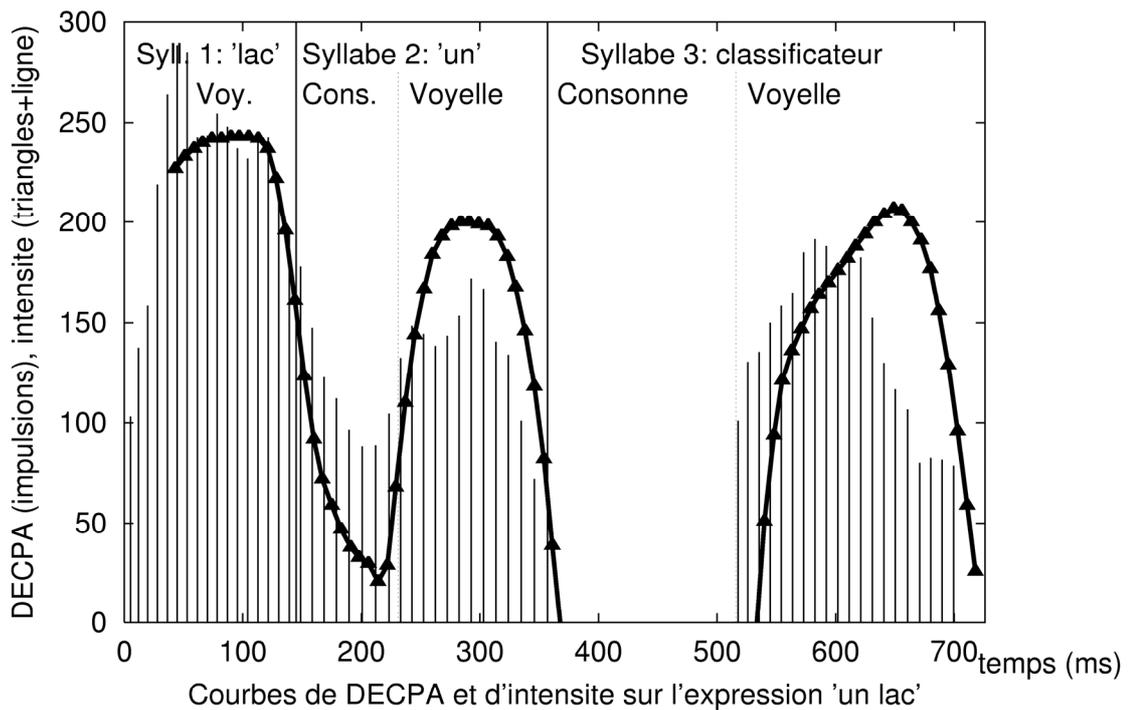
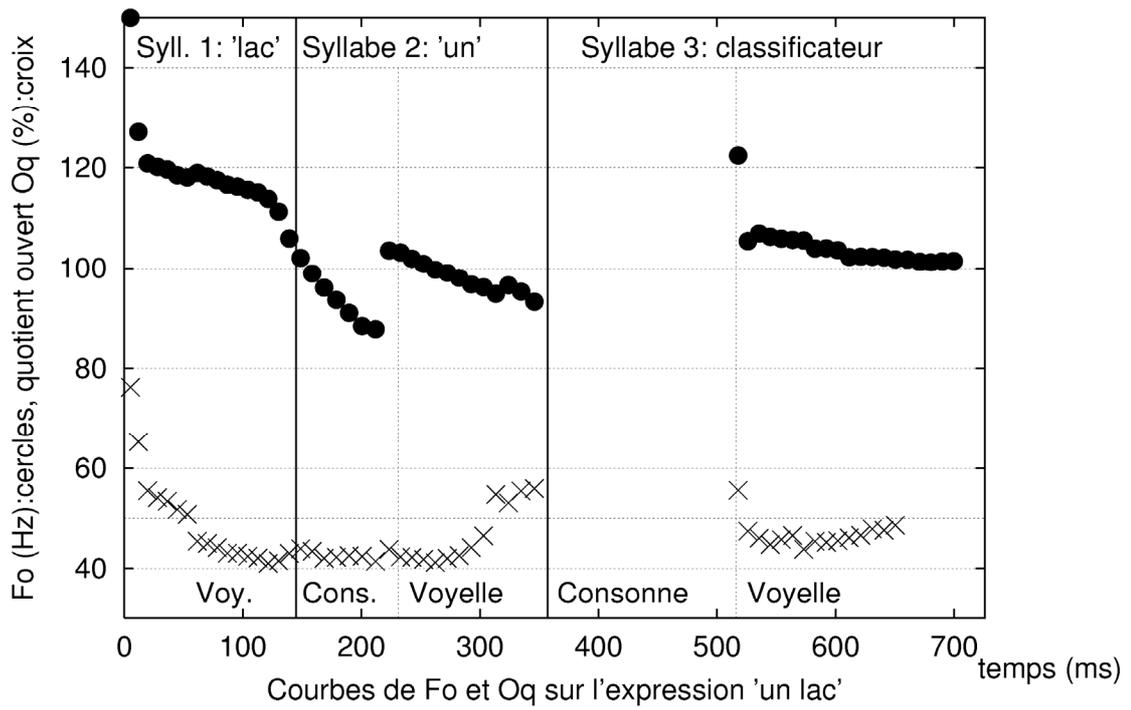
### 2.1.3.2. Syntagme nominal dont le classificateur est homophone du nom

Un autre prologue à l'étude des tons en énoncé est fourni par les structures dans lesquelles une même syllabe figure deux fois avec un statut différent. Ainsi, « un lac » se dit en naxi :

(9) hú dū hú

*lac/mer-un-classificateur des lacs et mers*

Le classificateur est issu du nom, dont il demeure homophone. La structure (9) permet une comparaison directe des paramètres intonatifs. Elle a été prononcée deux fois hors énoncé (9a, 9b) ; les deux répétitions présentent la même configuration. Voici les courbes pour (9a) :



Cet exemple illustre les variations prosodiques que peut connaître une même syllabe tonalisée en fonction de son importance et de sa place dans l'énoncé. Le classificateur, mot *grammatical*, est plus « faible » que le nom. La fréquence fondamentale est sensiblement plus basse sur le classificateur. Le quotient ouvert est plus élevé sur le classificateur, signalant, comme pour le verbe de la phrase-cadre, une phonation plus détendue. Les courbes de DECPA et d'intensité manifestent la même tendance. L'allongement final rend difficile une comparaison chiffrée ; en établissant une moyenne sur 100 ms, on peut proposer le chiffre d'une différence de 2,7dB (intensité globale calculée par le logiciel PRAAT, de P. Boersma et D. Weenink), et une différence de 40% du paramètre DECPA, entre nom et classificateur. La relation de dépendance entre  $F_0$  et DECPA proposée dans la section précédente prédit une différence de DECPA deux fois plus faible. Sur cette base, la valeur de DECPA effectivement mesurée sur le classificateur permet d'argumenter qu'il s'agit d'une syllabe *prosodiquement faible*, confirmant et précisant l'indication fournie par la mesure de  $F_0$ .

C'est dans le même esprit que vont être abordés les énoncés complets du corpus, l'étude veillant à ne pas se limiter à tel ou tel paramètre, et visant à établir l'importance relative de chacun.

## 2.2. Le corpus d'étude du focus

### 2.2.1. Présentation du corpus.

Ce corpus a pour point de départ le corpus réalisé par Claire Beyssade, Elizabeth Delais-Roussarie, Jenny Doetjes, Jean-Marie Marandin et Annie Rialland dans le cadre de leur travail sur la structure de l'information en français (voir notamment Delais-Roussarie, Rialland, Doetjes et Marandin 2002) . Ce corpus est composé de paires questions/réponses, permettant de fixer le focus dans la réponse, le terme de « focus » étant employé pour désigner *ce sur quoi porte la question*, et *ce qu'apporte la réponse*. La présente étude suit cette orientation terminologique ; précisons que les désignations de « focus », d'« accent d'insistance » et de « réalisation soignée » n'ont pas valeur explicative : la démarche consiste à établir des distinctions de fait, pour lesquelles sont proposées des désignations. Lorsqu'un même phénomène, observé indépendamment par plusieurs auteurs, reçoit divers noms (par exemple : « accent d'insistance », « accent emphatique » chez Coustenoble/Armstrong et Ohala, « accent contrastif et de niveau supérieur » chez Fant), on retient, dans un souci de clarté, un terme pour désigner de façon univoque le phénomène décrit, indiquant les équivalences de vocabulaire d'un auteur à l'autre, pour permettre au lecteur de catégoriser le phénomène en fonction du cadre de définitions qui lui paraît le plus éclairant.

Les cent trente énoncés du corpus de départ ont pour la plupart subi des modifications pour le propos de la présente enquête. Il leur a été donné un contexte, pour stabiliser leur interprétation. La langue de l'enquête étant le chinois, nous avons d'abord préparé, avec l'aide d'un locuteur natif du chinois, une traduction

respectueuse des habitudes du mandarin familial contemporain (« langue commune » de la Chine). La recherche d'équivalents convaincants de ces phrases en naxi a été menée avec l'informateur la veille de l'enregistrement.

### 2.2.2 Le ton bas

L'examen des trois tons ponctuels du naxi a montré l'asymétrie entre le ton haut, libre de se mouvoir dans une large plage de fréquence fondamentale, et les tracés moins dispersés du ton moyen et du ton bas, qui ne possèdent pas la même latitude de variation. Cette situation a une incidence sur les possibilités de mise en valeur prosodique d'une syllabe : une tendance universelle veut qu'un  $F_0$  bas signale une *faible importance* prosodique, la *mise en valeur* étant généralement associée à valeur élevée de  $F_0$  (voir Vaissière 1983, 1995). L'exemple du naxi révèle plusieurs stratégies pour résoudre le conflit entre ton lexical bas et mise en valeur prosodique.

#### 2.2.2.1. Première stratégie de mise en valeur

Une stratégie possible consiste, dans les cas où le contexte limite les risques de confusion, à passer outre au ton lexical, l'intonation l'emportant sur le ton ; la fréquence fondamentale se trouve alors dans la plage du ton moyen. En effet, la présence de tons lexicaux ne signifie pas que chacun doit être pleinement réalisé ; les tons sont en cela semblables aux phonèmes, qui (dans toutes les langues) peuvent, selon le contexte prosodique, être renforcés ou affaiblis, partiellement assimilés au contexte, ou ne laisser qu'une trace difficile à déceler. Ainsi de l'exemple sonore (10), dû à l'informatrice HQ :

(10)  $\eta\check{\zeta}$  sè bē lā̄ lē?

*1<sup>e</sup> pers. sg.-pronom interrogatif* « comment »- *devoir-particule interrogative*

La question signifie « Que dois-je faire ? ». Dans ce contexte, l'intonation met en valeur le pronom de première personne « je » ; le sens est : « Et moi, qu'est-ce que je vais faire ? ». La fréquence fondamentale sur cet item se trouve dans la plage du ton moyen de l'informatrice.

Ce phénomène demeure phonétique : lors de la transcription de cet énoncé, les informateurs sont unanimes pour dicter la syllabe  $/\eta\check{\zeta}/$ , « je », avec un ton bas. Mais il explique un phénomène phonologisé du naxi : la règle morpho-phonologique de remontée des pronoms personnels de 1<sup>e</sup> et 2<sup>e</sup> personne ( $/\eta\check{\zeta}/$ ,  $/n\grave{\zeta}/$ , tous deux au ton bas) lorsqu'ils sont suivis de la particule (optionnelle) qui le marque comme agent : la particule  $n\bar{u}$ . Ainsi :  $\eta\check{\zeta} + n\bar{u} > \eta\check{\zeta} n\bar{u}$ ,  $n\grave{\zeta} + n\bar{u} > n\bar{u} n\bar{u}$ . Le naxi présente d'autres phénomènes de sandhi tonal (terme est employé dans son étendue la plus large, suivant l'approche de Chen 2000) par lesquels certaines séquences *chiffre + classificateur* sont modifiées si le ton du chiffre est plus bas que celui du classificateur. L'origine phonétique de ces phénomènes paraît transparente : la faiblesse des mots grammaticaux s'accommode mal d'une mélodie plus élevée qu'un mot plus important pragmatiquement. Cette tendance générale, dont la présence phonétique est aisée à déceler dans la réalisation des tons naxi en énoncé, a ainsi déposé une trace dans la morpho-phonologie de la langue.

### 2.2.2.2. Deuxième stratégie de mise en valeur

Une autre solution pour résoudre le conflit entre ton lexical bas et mise en valeur prosodique s'observe dans des récits animés : le dévoisement d'une syllabe au ton bas qui porte un degré élevé d'insistance, dévoisement accompagné d'un serrage pharyngien. Ces exemples se rencontrent avec régularité aussi bien dans la voix féminine de HQ que dans la voix d'un informateur masculin, HXX. Le lecteur est invité à se reporter à l'illustration sonore de l'exemple suivant, extrait d'un récit<sup>1</sup> :

(11) k<sup>h</sup>ù kỳ kò à dū mē à k<sup>h</sup>ú...

à l'intérieur-poulet-un-*classificateur des animaux-particule d'aspect*-mettre

« Dedans, tu mettras un **poulet** » A ce passage du récit, le mot « poulet » est mis en relief ; l'enregistrement électroglottographique montre un dévoisement presque complet du /a/, le voisement ne s'installant que tout à la fin de la voyelle. Pendant la partie dévoisée, la source sonore n'est pas la vibration des cordes vocales mais une friction pharyngale très intense, qui signale efficacement la saillance prosodique de cette syllabe, là où un segment voisé aurait difficilement pu porter à la fois la même charge prosodique et le ton lexical, nécessaire, dans ce contexte, à la bonne identification du mot. Le phénomène n'est pas présent dans le corpus de l'informateur HX ; il n'est mentionné ici que pour montrer la variété des stratégies possibles pour la mise en valeur du ton bas<sup>2</sup>.

### 2.2.2.3 Troisième stratégie de mise en valeur

Dans la majorité des cas de mise en valeur locale d'une syllabe au ton bas, la charge intonative ne va pas jusqu'à dévoiser la voyelle, ni rendre le ton méconnaissable. L'idée de réaliser les mesures rapportées ici est motivée par l'impression auditive, pour un certain nombre d'énoncés naxi, d'un *effort supplémentaire qui n'est pas signalé uniquement (ni même principalement) par un changement de F<sub>0</sub>*. C'est dans l'intensité acoustique globale, et dans la mesure de DECPA, que l'on espère trouver trace de cet effort perçu. Soit l'exemple suivant :

(12) nu k<sup>h</sup>à dỳ t<sup>h</sup>v bū lá ?

*2<sup>e</sup> pers. sg.-Pékin-sortir-aller-particule interrogative*

« Tu vas jusqu'à Pékin ? »

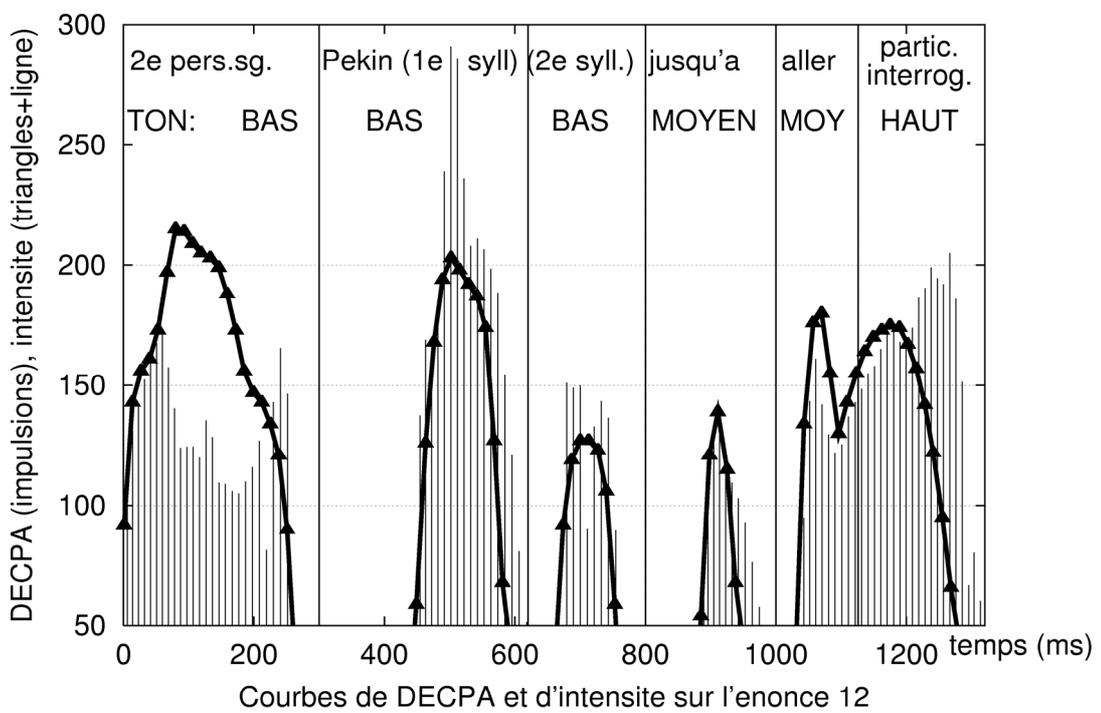
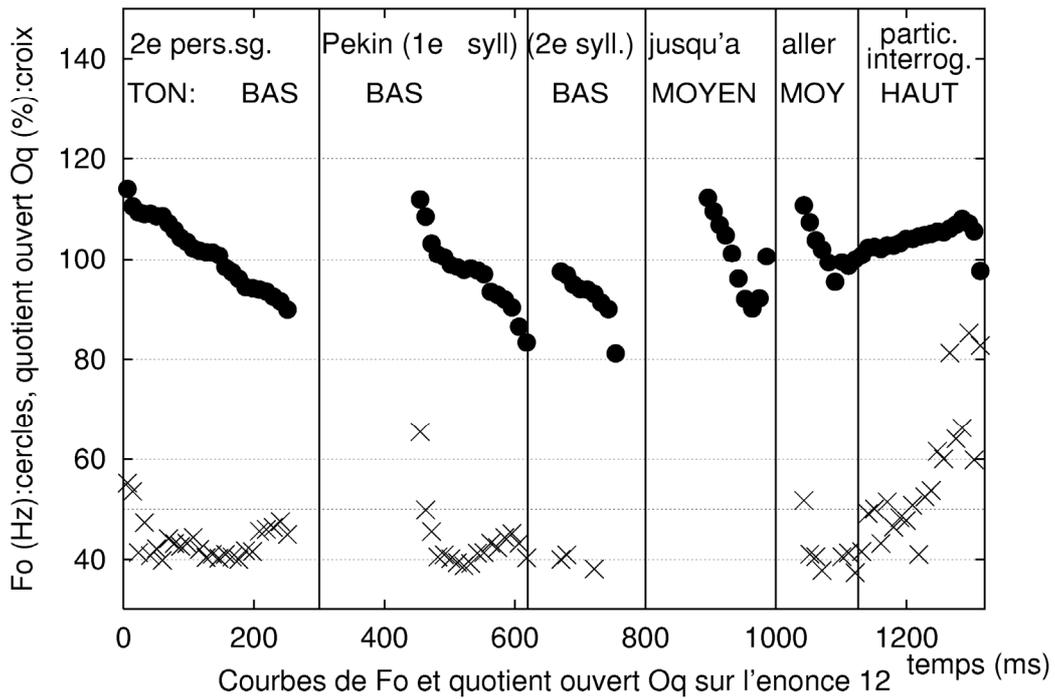
Avec mise en valeur intonative (« focus ») sur « Pékin » (/k<sup>h</sup>à dỳ/), la phrase signifie : « C'est bien à Pékin que tu vas ? » Les tracés obtenus sont les suivants<sup>3</sup> :

---

<sup>1</sup> A terme, les récits naxi (enregistrement et annotation) seront mis à disposition par l'intermédiaire du programme Archivage du LACITO (<http://lacito.vjf.cnrs.fr/archivage/index.html.fr>).

<sup>2</sup> Pour aller au bout de la description phonétique, et poursuivre par là le travail entamé par René Gsell 1979b dans son étude sur la réalisation des tons en l'absence de voisement (« Etude des tons chuchotés du thai standard »), des images du pharynx seraient nécessaires.

<sup>3</sup> C'est ici la première énonciation de la phrase qui est étudiée, plutôt que sa répétition. Dans la répétition, tâche méta-linguistique, l'énonciateur prend soin de bien répéter tous les mots, plutôt que de répéter l'acte communicatif. Les phénomènes de structure de l'information (intonation pragmatique) sont nettement moins saillants, et l'effort articulatoire est réparti plus également entre syllabes : on constate un moindre contraste entre syllabes « faibles » (subissant des réductions articulatoires) et syllabes « fortes » (articulées fortement, longues, etc.).



(plage d'intensité : 20dB ; de 50dB à 70dB)

Ni le quotient ouvert ni le  $F_0$  ne présentent de saillance particulière sur le mot « Pékin », /k<sup>h</sup>à dỳ/. En termes de hauteur, le  $F_0$  sur « Pékin » (/k<sup>h</sup>à dỳ/) est plus bas que sur le pronom ; en termes de glissando, « Pékin » est moins saillant que le verbe qui suit, /t<sup>h</sup>ṽ/. « Pékin » ne présente pas non plus de saillance en termes de durée. Au plan de l'intensité acoustique (mesurée sur le signal audio), la syllabe la plus intense de l'énoncé est la négation, en début d'énoncé. En revanche, la mesure de DECPA révèle une saillance sur la première syllabe de « Pékin », /k<sup>h</sup>à dỳ/.

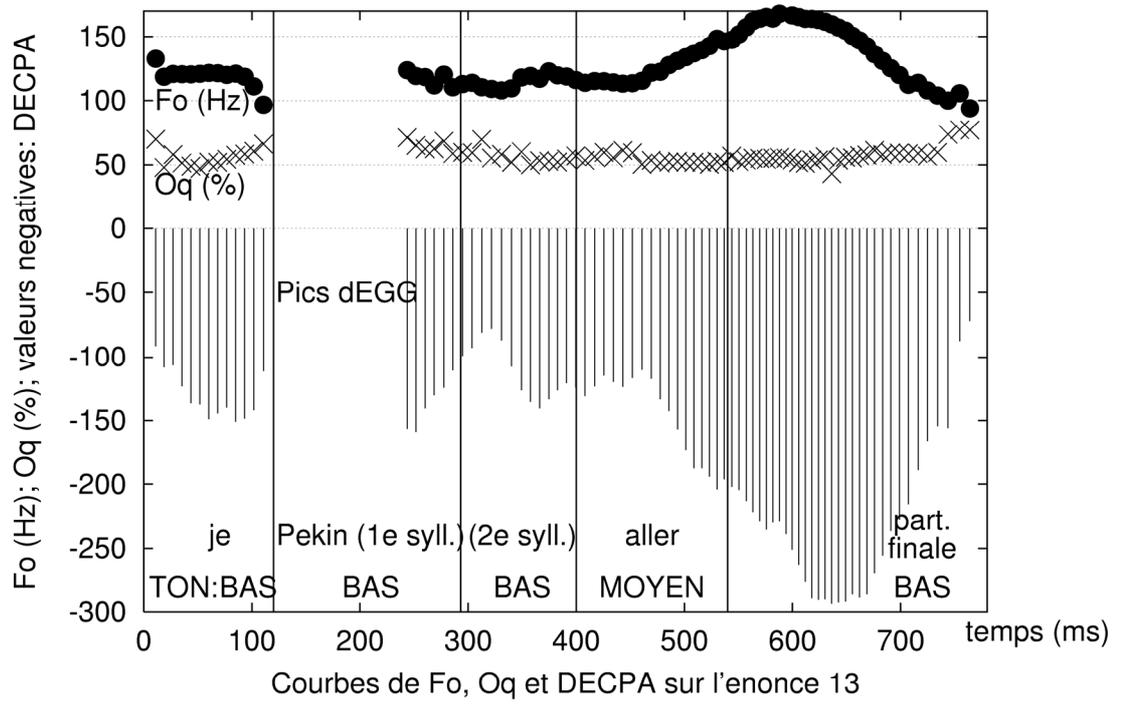
Il convient de signaler la marge de liberté qui existe dans la réalisation de semblable mise en valeur : le phénomène local sur le mot-focus, tel qu'il est illustré par (12), peut être remplacé par un phénomène global de montée de registre de  $F_0$ , la plus forte saillance prosodique se reportant sur la fin de l'énoncé. Telle est la configuration observée dans l'énoncé (13), dit « de façon impatiente »<sup>1</sup>. (Pour représenter trois paramètres sur un même graphique sans qu'ils se confondent, DECPA a été représenté en valeurs négatives.)

(13) ηṽ k<sup>h</sup>à dỳ bŭ mỳ.

*1<sup>e</sup> pers. sg.-Pékin-aller-particule finale d'affirmation* « C'est à Pékin que je vais. »

---

<sup>1</sup> Au sujet de la procédure d'élicitation des énoncés « emphatiques », voir la présentation du corpus en introduction de II. 2.



Les paramètres calculés ici montrent (en accord avec l'impression auditive) qu'il n'y a pas de phénomène local de mise en valeur du mot « Pékin », /k<sup>h</sup>à dỳ/, le quotient ouvert indiquant une phonation relativement détendue. Le ton lexical bas de la particule finale est méconnaissable : sa courbe de fréquence fondamentale domine celle de tout l'énoncé, et sa valeur de DECPA est plus saillante encore : le surcroît de DECPA sur la dernière syllabe est le double de celui prédit par la relation de dépendance proposée en II. 1. B ; cette augmentation ne peut donc être considérée comme la simple conséquence d'une augmentation de la fréquence fondamentale. L'intensité acoustique est également élevée (le tracé étant très similaire à celui de DECPA, il n'a pas paru nécessaire d'établir un graphique séparé).

### 2.2.3. Le ton moyen

L'étude du ton moyen sera l'occasion de poursuivre les observations sur DECPA et F<sub>0</sub>, et d'observer en détail le jeu d'un autre paramètre, le quotient ouvert.

Prenons pour point de départ de l'étude du ton moyen sous focus un échange qui, très classiquement, vise à mettre en valeur un des éléments de la réponse :

(14) nù hù mū k<sup>h</sup>v̄ t̚s<sup>h</sup>wá d̚zù g̃ hō t̚s<sup>h</sup>̃ ndzù pú ts<sup>h</sup>ù lá?

*2<sup>e</sup> pers. sg.-soir-six-heure-GEN.-train-être assis-"gérondif"-arriver-particule finale interrogative*

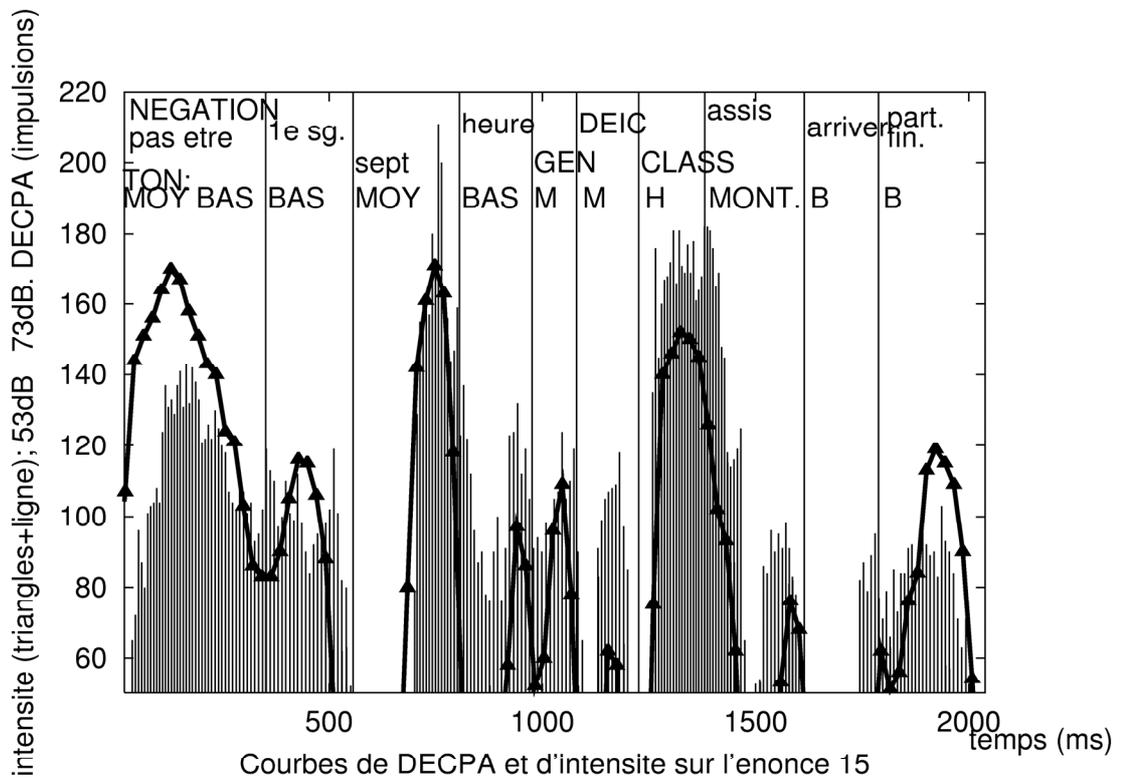
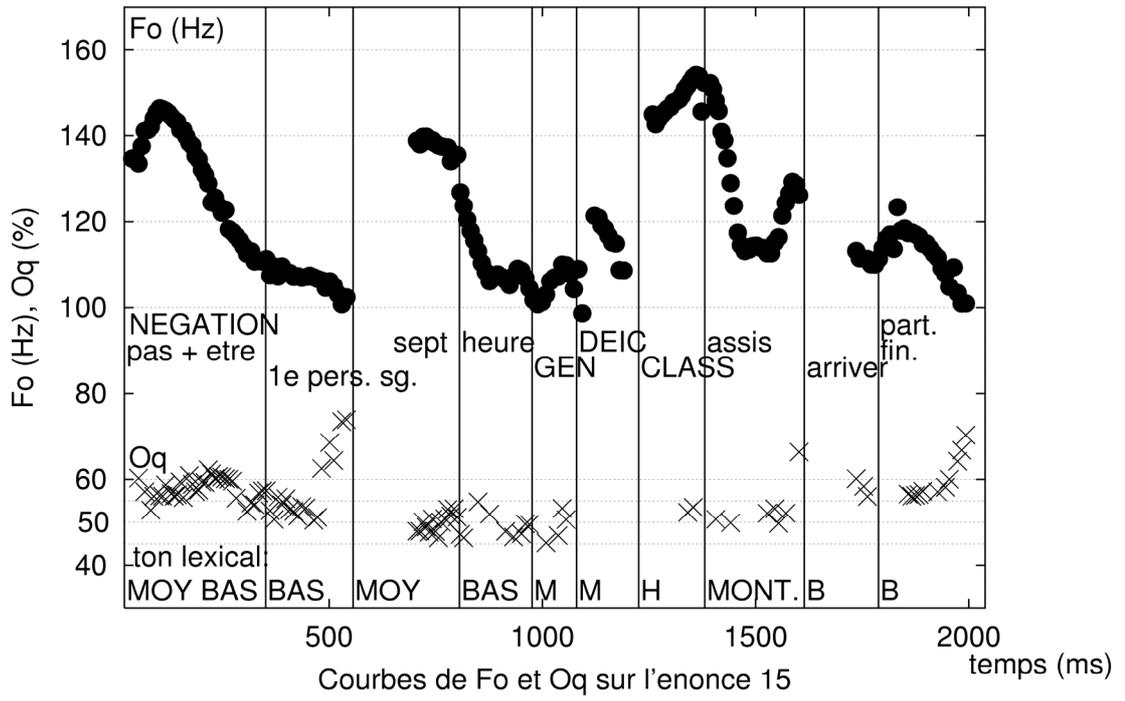
Tu es venu par le train de 6h du soir ?

(15) (=5') m̃ wà, η̃ s̃ə̃ d̚zù g̃ t̚s<sup>h</sup>ũ kjá ndzũ ts<sup>h</sup>ù m̃.

*nég.+verbe copule-1<sup>e</sup> pers. sg.-sept-heure-GEN.-déict.-classificateur-être assis-"gérondif" en ellipse-arriver-particule finale affirmative*

Non, c'est par celui de 7 heures que je suis arrivé.

#### 2.2.3.1. Mesures de F<sub>0</sub> et DECPA



Enoncé (15e)

Sur le premier tracé, on remarque que la fréquence fondamentale de la syllabe « sept »,  $\text{ʃə}^-$ , est inférieure à celle du classificateur  $\text{kj}^0$ . La valeur de DECPA, toutes choses égales par ailleurs, devrait également être inférieure. Or le second graphique révèle une brève saillance de DECPA sur la syllabe-focus « sept » ; on retrouve donc, comme dans l'exemple au ton bas (12), un phénomène local de mise en valeur prosodique, qui n'est pas décelable sur la seule courbe de fréquence fondamentale, ni sur la courbe d'intensité globale, sur laquelle le focus ne se détache pas plus que les deux premières syllabes de l'énoncé.

### 2.2.3.2. Mesures de quotient ouvert. Caractérisation de l'« accent d'insistance ».

Le quotient ouvert,  $O_q$ , fournit une indication sur le caractère tendu/pressé de la voix. La mesure de  $O_q$  était d'abord guidée par l'espoir d'utiliser ce paramètre comme indice de mise en valeur prosodique, escomptant une baisse de  $O_q$  sur les syllabes mises en valeur prosodiquement<sup>1</sup>. C'est en effet la tendance (discrète) qui se manifestait sur l'exemple (12) : à la mise en valeur de la première syllabe du mot « Pékin » correspondait un léger « ventre » de quotient ouvert, indiquant une voix plus tendue/pressée. Mais sur l'ensemble du corpus, cette recherche n'a pas abouti au résultat escompté. Apparaît au contraire un contraste de quotient ouvert entre deux ensembles d'énoncés, « insistants » et « non insistants » : la phrase (15) est en effet dite deux fois (15a, 15b), puis reprise trois fois avec pour consigne « cela fait plusieurs fois que l'interlocuteur te demande, il oublie à chaque fois; tu répètes de façon très insistante » (15c, d, e). Cette consigne visait à ajouter un paramètre à l'étude : l'« insistance », opposé à « réalisation non insistance ». Comme il a été constaté pour l'accent d'insistance en français, l'insistance est un phénomène scalaire, qui connaît un large degré de liberté individuelle ; néanmoins, dans les énoncés insistants, les syllabes focalisées présentent des valeurs de quotient ouvert élevées (tendance vérifiée par les autres exemples du corpus). Ainsi, les résultats observés sur la syllabe  $/\text{ʃə}^-/$  (« sept ») sont : quotient ouvert de 47% sur les énoncés (15a-b), et de 56% (donc plus *relâché*) sur les énoncés insistants (15c-d-e). La valeur sur (15a-b) est dans la moyenne des valeurs observées en syllabe isolée ; la valeur sur (15c-d-e) est nettement supérieure à cette moyenne. Retenons deux résultats :

- le quotient ouvert (à lui seul) ne peut être interprété directement comme indice de focalisation/mise en valeur prosodique
- le quotient ouvert est plus élevé sous insistance (différence de l'ordre de 12%).

Ce dernier résultat va à l'encontre des observations faites par Nathalie Henrich sur la voix chantée, qui montrent, chez un chanteur qui file une note, une baisse du quotient ouvert proportionnelle à l'augmentation de l'intensité : voir la figure 8.7.B3 de Henrich 2001. Pour comprendre ce phénomène particulier, propre à la voix parlée, il est éclairant de mettre en parallèle les mesures apportées par

---

<sup>1</sup> Voir par exemple Epstein 2002:xviii : « Both prominent words and phrase-initial words displayed a « tensor » voice quality than their non-prominent and phrase-final counterparts. »

l'électroglottographie avec les données qui ont été publiées, pour de nombreuses langues, au sujet de l'*accent d'insistance*.

L'« accent d'insistance » paraît être un *universel de l'intonation, qu'il importe de distinguer des phénomènes de focalisation*<sup>1</sup>. Les descriptions de ce phénomène se font écho d'une langue à l'autre. En français, les conclusions de l'enquête physiologique et perceptive de Carton, Hirst, Marchal et Séguinot 1976 sont résumées ainsi par Carton 1974 :

« Il n'y a pas de réalisation constante au niveau acoustique. C'est ce qui nous a amené à chercher s'il n'y a pas une réalisation physiologique constante. (...) On remarque que les accents d'insistance qui ont été repérés par un très fort pourcentage d'auditeurs sont ceux où apparaît une *brutale montée* de pression d'air sous-glottique qui commence sur une *consonne*. (...) »

Un dispositif expérimental recourant lui aussi aux techniques (invasives) de mesure de pression d'air sous-glottique a été mis en place par Benguerel 1973, qui conclut, reprenant la terminologie « accent emphatique/ accent non emphatique » de Coustenoble et Armstrong 1937 : « ...dans le cas de l'accent emphatique, il y a toujours une augmentation marquée de la pression sous-glottique pendant (ou juste avant) la syllabe accentuée. » Ce prototype physiologique de l'accent d'insistance ressort également de l'étude de John Ohala 1978:21 : son étude pléthysmographique (mesure du volume pulmonique) montre, de façon invariable, une baisse du volume pulmonique (indiquant un surcroît d'activité expiratoire) sur les syllabes accentuées de façon emphatique, et jamais sur les syllabes normalement accentuées. Les travaux récents de Fant, Hertegard et Kruckenberg 1996, dans lesquels on trouve la formule d'« activité pulmonique prosodiquement significative », distinguent également deux types de comportement : « une pression pulmonique (et partant sous-glottique) accrue s'observe dans les cas d'accent contrastif et de degré supérieur, mais n'est sans doute pas une composante nécessaire de l'accentuation focale. » Les dénominations varient ; il paraît néanmoins possible de retenir la convergence de ces observations approfondies : l'accent d'insistance apparaît comme une catégorie universelle<sup>2</sup>, définissable en termes physiologiques. Le wolof (voir contribution d'A. Rialland et S. Robert) n'y fait sans doute pas exception, même

---

<sup>1</sup> La description traditionnelle de l'accent d'insistance comme phénomène distinct, qui « se surajoute au mot comme un élément autonome » et « ne peut être considérée comme un fait accentuel » (Paul Garde 1968), a été un temps oubliée, du fait de la prééminence accordée à la fréquence fondamentale, paramètre aisé à mesurer, et à manipuler en synthèse de la parole. L'accent d'insistance est pourtant une composante essentielle de l'intonation, comme le rappellent aujourd'hui les chercheurs de l'université de Kiel : « ...it is necessary to complement the accent-pitch link by a force accent category (...) It is related to greater physiological and articulatory effort (...). » (Kohler 2003:2)

<sup>2</sup> Il ne paraît donc pas crédible d'affirmer avec Fónagy (1983:22) : « L'allongement consonantique et l'accent d'insistance, si courants dans le français moderne, étaient probablement inconnus avant la deuxième moitié du XX<sup>e</sup> siècle ». La fréquence de ces phénomènes peut certes varier ; leur emploi systématique dans le débit « professionnel » paraît en effet récent. En revanche, le mécanisme de l'accent d'insistance est sans aucun doute présent, avec un poids phono-stylistique inversement proportionnel à sa fréquence, dans toutes les langues.

si les auteurs ont choisi une notation de ce phénomène sous forme de « tons », suivant en cela le courant de la « phonologie intonative » (l'expression appartient à Robert Ladd). Le wolof, dans lequel il n'y a pas de marquage intonatif du focus, et qui présente néanmoins des phénomènes d'insistance, illustre clairement la distinction entre focus et insistance. Dans les énoncés insistants du corpus naxi, la fréquence fondamentale moyenne de l'énoncé est de supérieure de plus d'un ton musical à celle de la série non insistante ; sur la syllabe-focus, la fréquence fondamentale est supérieure de plus de deux tons dans la série insistante ; mais si l'on s'arrête à ce paramètre, on s'interdit de comprendre la spécificité du phénomène d'accent d'insistance. L'augmentation d'intensité globale de 9 dB en moyenne sur la syllabe concernée en fournit un indice. Définir le phénomène physiologiquement, par le surcroît d'action des muscles respiratoires, permet de le distinguer nettement parmi les phénomènes prosodiques (sans exclure l'existence de réalisations variées, en vertu des relations d'équivalence et de compensation entre l'ampleur du mouvement de fréquence fondamentale et l'effort des muscles expiratoires ; au sujet des équivalences fonctionnelles, voir Gussenhoven 2002; Vaissière 2004). Les données électroglottographiques sur la langue naxi, sans avoir l'ampleur des enquêtes qui viennent d'être citées, encouragent à ajouter à la caractérisation de l'accent d'insistance l'absence de *serrage* des cordes vocales, la voix n'étant pas « pressée/tendue » sur la syllabe qui porte l'accent d'insistance. Soulignons la correspondance avec la valeur relevée par Titze 1988:236 :

« ... pour des pressions élevées, une légère abduction des cordes vocales donne une puissance optimum en dB. Pour  $P_s=24\text{cm H}_2\text{O}$ , par exemple, un quotient ouvert de 56% paraît optimal. »

#### 2.2.4. Le ton haut

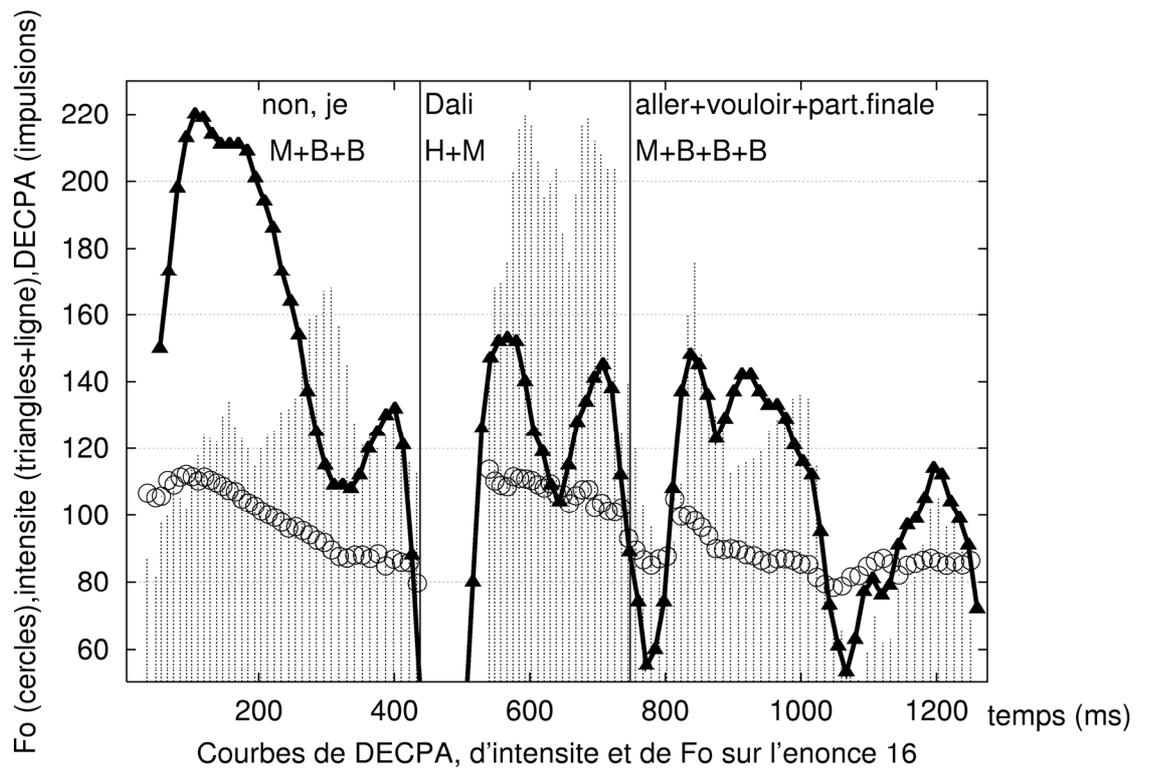
L'étude du ton bas et du ton moyen proposée ci-dessus doit une large part de son intérêt au conflit entre ton lexical et intonation pragmatique ; au ton haut, ce conflit n'existe plus. Les variations que connaît le ton haut en naxi rappellent beaucoup les phénomènes de langues sans tons lexicaux : en particulier, comme en allemand ou en anglais, la *position du pic de  $F_0$  à l'intérieur de la voyelle* a une grande importance pragmatique. En naxi, il semble possible de distinguer trois types de réalisations similaires à celles mises en lumière par Kohler 1991 pour l'allemand : haut-plain (réalisation neutre) ; avec descente rapide dès le début de la voyelle (valeur pragmatique/sémantique : exprimer l'évidence, l'acquis) ; avec « pic tardif » situé au-delà du milieu de la voyelle (valeur pragmatique/sémantique : arrêter l'attention sur l'item concerné ; exprimer un « contraste »). Cette question, qui nécessite une vérification par des tests de perception, fera l'objet d'une étude ultérieure.

On se bornera ici à commenter brièvement un seul exemple, qui fournit des arguments en faveur de l'utilisation de DECPA.

(16) m̄x wà, ηȳ tá lī b̄ū nà v̄ m̄ȳ.

NEG.-1<sup>e</sup> pers. sg.-Dali (nom de lieu)-aller-vouloir-PARTICULE FINALE

« Non, c'est à Dali que je veux aller. » (Réponse à la question : « Vous allez à Kunming ? »)



*Énoncé (16) : tracés de  $F_0$  (en Hz), d'intensité et de DECPA.*

*Les valeurs d'intensité, en dB, ont été ajustées à la fenêtre : de 46.25dB à 68.75dB, soit une fenêtre de 22.5dB.*

Outre la réalisation pleine de l'occlusion /t/ et sa durée, la mise en valeur locale des syllabes /tā li' / se reflète très nettement dans le tracé de DECPA, qui montre une saillance que F<sub>0</sub> ni l ne permettraient d'établir : l'intensité globale est nettement plus faible sur le focus que sur le début de l'énoncé /m̄ wà/ ; le F<sub>0</sub> n'est pas sensiblement plus élevé sur le focus que sur la négation /m̄/ (ton lexical moyen), première syllabe de l'énoncé.

### 3. Synthèse

#### 3.1. Les paramètres

Il serait illusoire de vouloir présenter ici l'intégralité des paramètres qui participent aux modifications intonatives de la syllabe sous focus. Les mouvements de la mâchoire et le timbre de la voyelle (position et largeur de bande des formants) ne peuvent être étudiés en détail ici. Par souci de synthèse, on s'attachera néanmoins à prendre en compte, outre les paramètres qui précèdent (F<sub>0</sub>, O<sub>q</sub> et DECPA), certains phénomènes de *durée*. Pour ce faire, une comparaison a été menée entre la syllabe /ʃə̄-/ « sept » telle qu'elle est réalisée dans (15a-e) et telle qu'elle apparaît dans deux autres contextes :

1. la syllabe /ʃə̄-/ « sept » dans des listes de nombres associés à un classificateur (un classificateur au ton bas : « montagne », un au ton moyen : « animal », un au ton haut : « arbre » ; les listes vont de 1 à 25)
2. la syllabe homophone /ʃə̄-/ « chose », dans un énoncé où elle n'est pas mise en valeur sémantiquement :

(17) è tsū ḡ ʃə̄ t<sup>h</sup>v̄ j̄?

*pronom interrogatif* « quel »-*gén.*-chose-se passer-particule finale « Qu'est-ce qui se passe ? »

Il apparaît que le rapport *durée de la consonne/durée de la voyelle* est nettement plus élevé pour les items focalisés (15a-e) (où il est de l'ordre de 1,5) que dans les deux autres contextes (valeur : environ 0,8)<sup>1</sup>. Ce rapport est plus élevé encore dans les énoncés « insistants/impatiens » (15c-e) que dans (15a-b). L'importance des durées respectives de l'initiale et de la rime d'une syllabe est connue notamment par les travaux de Vaissière 1977, 1986; Astesano 1999. Pragmatique et expression de l'émotion usent du même paramètre : « à l'intensité de l'émotion correspond l'intensité de l'activité musculaire, ainsi que le prolongement des voyelles et des consonnes (...) : celui des voyelles pour les émotions tendres, celui des consonnes pour les émotions agressives. » (Fónagy 1983:18) L'importance du rapport consonne/voyelle se vérifie sur l'ensemble du corpus naxi (mais la comparaison directe ne pouvant se faire qu'entre syllabes identiques au plan phonémique, on se limite ici à indiquer les chiffres pour la syllabe /ʃə̄/).

---

<sup>1</sup> Dans ce calcul, la durée de la consonne est définie comme l'intervalle non voisé entre la fin de la voyelle précédente et le début du voisement de la voyelle /ə̄/. Il n'est pas tenu compte de l'allongement de la rime de la syllabe précédente, qui contribue à la mise en valeur de la syllabe /ʃə̄/.

### 3.2. Les configurations

La récapitulation présentée ci-dessous ne revient pas à établir des catégories étanches ni une terminologie définitive : elle propose une dénomination pour des phénomènes observés régulièrement dans le corpus.

Deux ensembles se dégagent, les possibilités de *réalisation prosodiquement forte* d'une syllabe, s'opposant aux réalisations *prosodiquement faibles*. Ces dernières *ne possèdent pas de valeurs caractéristiques de  $F_0$ ,  $O_q$ , DECPA, ni de rapport consonne/voyelle* : elles présentent une très grande variabilité. La présente étude porte principalement sur les réalisations prosodiquement fortes ; les quelques mesures réalisées sur les réalisations faibles, telles que l'énoncé (17), plaident pour une description du phénomène dans les termes que proposent Shih, Kochanski et al. 2001 pour le chinois mandarin : dans leur système de synthèse du mandarin, chaque syllabe se voit affecter un coefficient de force ; plus une syllabe est faible, plus elle est soumise à l'influence des syllabes environnantes : un ton bas « faible » après ton haut pourra ainsi présenter une fréquence fondamentale élevée. C'est pourquoi, dans le tableau ci-dessous, seules des tendances générales sont indiquées pour la « position faible ».

En revanche, s'agissant des syllabes prosodiquement *fortes*, la description en termes de réalisation plus ou moins complète d'un « gabarit tonal » que proposent Shih, Kochanski et al. 2001 ne suffit plus. En effet, ces auteurs n'envisagent qu'une forme de mise en valeur de la syllabe : une réalisation canonique de son ton lexical (réalisation fidèle du « gabarit tonal », idéal et unique, défini pour chacun des quatre tons lexicaux). Le modèle phonologique de Chen 2000, toujours pour le chinois mandarin, érige cette conception en modèle phonologique, ne concevant les phénomènes pragmatiques (« emphasis and contrast ») qu'en termes de « déplacements d'accent » sur la grille métrique qui ont pour effet la réalisation affaiblie de certains des tons.

Ce modèle, utile dans l'état actuel des systèmes de synthèse de la parole, ne correspond absolument pas à la variété des réalisations observées en parole naturelle. Les limites imposées par les outils existants de manipulation du signal pour la synthèse de la parole ne doivent pas amener les linguistes à délaisser, dans leurs modèles, la variété des catégories existantes.

Dans le tableau ci-dessous figurent trois catégories de réalisation prosodiquement forte d'une syllabe : la « réalisation soignée », qui est celle observée en syllabe isolée ou dans des listes de chiffres, et correspond à la « réalisation forte » de Shih, Kochanski et al. 2001 ; et deux types de réalisation sous focus : focus simple (illustré par 15a-b), et accent d'insistance (illustré par 15c-d-e). Le rapprochement de ces deux derniers types au sein d'un même groupement se justifie dans la mesure où la syllabe qui porte l'accent d'insistance se trouve également, de ce fait, signalée comme focus. Au plan des réalisations, en revanche, la différence de quotient ouvert suffit à les distinguer : leurs corrélats physiologiques typiques sont distincts, voir 2.2.3.2. Quant à la « réalisation soignée », elle se distingue des autres par son

rapport consonne/voyelle nettement plus bas. Ainsi, aucun des trois types ne peut être réduit à une variante « affaiblie » de tel autre.

Type de réalisation \ Paramètre		F <sub>0</sub>	O <sub>q</sub>	DECPA	rapport C/V
Position faible		(plutôt bas)	(plutôt élevé, > 50)	(plutôt bas)	(plutôt bas)
<i>fortement influencé par le contexte</i>					
Position forte	« Réalisation soignée »	canonique (=valeur en syllabe isolée)	bas (40-50)	élevé	< 1 (≈0.8)
	Focus	Focus simple	canonique ou assez élevé	élevé	> 1 (≈1.3)
		Accent d'insistance	élevé	élevé (> 55)	élevé

*Proposition de catégorisation des phénomènes d'intonation pragmatique, d'après les réalisations de la syllabe /ʃə̃/*

Cette caractérisation porte exclusivement sur les phénomènes qui s'observent localement, sur la syllabe-focus. Dans les polysyllabes, c'est principalement la première syllabe qui est concernée par cette mise en valeur (suivant en cela une tendance universelle, voir Vaissière à paraître). Les variations prosodiques des autres parties de l'énoncé jouent bien sûr un rôle : la pause (ou allongement de la syllabe qui précède), phénomène largement représenté dans le corpus ; l'aplanissage intonatif du post-focus (*ibid.*), très visible sur trois syllabes environ après le focus, qui concerne également l'articulation des consonnes et des voyelles. Relevons quelques exemples : l'initiale /tʃ<sup>h</sup>/ du déictique perd son occlusion dans (15a-b), et se trouve voisée dans (15d). La syllabe /pú/ subit un voisement discret de l'initiale en (15c), un voisement plus ample en (15d), phénomènes qui préludent à la disparition totale de la syllabe en (15e), par un phénomène morpho-phonologique décrit en détail en 1.3.1.1. Ainsi, la réalisation des segments participe également à la prosodie, dont l'étude se confond par là avec le programme d'une « phonétique de la phrase » (formule reprise à Klaus Kohler).

### 3.3. Conclusion générale

La démarche de la présente étude a consisté, à partir de l'observation d'une même syllabe dans divers contextes, à relever les variations de quatre paramètres : fréquence fondamentale, quotient ouvert, amplitude des pics de fermeture du signal DEGG, et rapport de longueur entre initiale et rime. Ces paramètres ont aidé à cerner plusieurs types de réalisations prosodiques. L'exemple du naxi autorise à conclure qu'il n'y a pas de lien entre la présence ou l'absence de tons lexicaux et l'abondance des phénomènes intonatifs locaux. Le naxi appartient au groupe des langues à « intonème de focus »<sup>1</sup>, groupe qui paraît largement majoritaire parmi les langues du monde. Le wolof (Rialland et Robert, ce volume) est une langue sans tons et sans

<sup>1</sup> Le terme d'« intonème » est repris à Rossi 1999 ; on pourrait également parler d'« intonation de focus ».

intonème de focus ; le naxi est une langue à tons qui présente pourtant une grande variété de phénomènes d'intonation pragmatique.

Du détail des observations, il ressort que l'« intonème de focus » se reflète dans l'*ensemble des paramètres* qui ont pu être mesurés, y compris les modifications segmentales. Une *fréquence fondamentale* supérieure à la valeur attendue pour le ton lexical correspondant est un corrélat régulier de cette mise en valeur : la marge de variation est relativement importante, phénomène qui appelle une description *superpositionniste* du niveau lexical et du niveau prosodique. L'emploi d'une *intensité* accrue est régulier ; cette dimension très complexe ne peut être décrite uniquement par le paramètre d'intensité acoustique globale, et le paramètre DECPA paraît plus directement corrélé au phénomène linguistique qu'il s'agissait de soumettre à étude expérimentale.

### *3.4. Perspectives : les méthodes de mesure*

Les valeurs élevées de DECPA régulièrement relevées sur les items définis linguistiquement comme « focus » encouragent à poursuivre l'emploi de cette mesure, et à affiner son interprétation. Une meilleure intelligence des mécanismes de vibration permettra le choix d'une solution technique concernant la mesure des pics dédoublés (coefficientage, ou élimination du pic le plus bas).

Au plan de la description linguistique, les exemples détaillés qui ont été présentés ici invitent une étude des *relations de compensation* qu'entretiennent spectre, mélodie et durée dans la perception de la parole. Les variables pertinentes pour la description linguistique seraient plutôt des oppositions telles que *tension/détente* ou *douceur/force* (d'Alessandro, Boula de Mareuil et al. 2003:111) ; leur analyse instrumentale ne doit faire l'économie d'aucun paramètre pour pouvoir s'engager sur le chemin de la synthèse.

## *Remerciements*

L'idée de mesurer l'amplitude des pics de fermeture du signal d'EKG et de la rapporter à la saillance prosodique vient de Nathalie Henrich (Laboratoire d'Acoustique Musicale : LAM), qui a apporté son aide à cette étude jusqu'à la phase de rédaction. C'est également Nathalie Henrich qui a proposé le nom « DECPA ». Mes vifs remerciements également à Vu Ngoc Tuan (Laboratoire d'Ingénierie Mécanique pour les Sciences de l'Ingénieur : LIMS), qui a très généreusement donné de son temps pour les discussions scientifiques, et pour la mise au point des indispensables programmes en C.

Le LACITO a prêté du matériel d'enregistrement professionnel pour l'enquête menée en Chine en 2002, ainsi que du matériel informatique. Les enregistrements ont été re-numérisés dans le cadre du Programme Archivage du LACITO. Au plan scientifique, ce travail de recherche sur le naxi doit beaucoup à Martine Mazaudon et Boyd Michailovsky.

L'équipe de rattachement de cette recherche est l'UMR 7018, « Laboratoire de Phonétique et Phonologie » / ILPGA, Université Paris 3. Des remerciements particuliers sont dus à Jacqueline Vaissière et Annie Riolland ; à Cédric Gendrot ; à Lise Crevier-Buchman, du service du Pr. Brasnu à l'hôpital Georges Pompidou. Merci aux Pr. Sun Hongkai et Huang Xing pour l'invitation à mener des recherches en Chine ; au Pr. Kong Jiangping, responsable du laboratoire de phonétique de l'Institut des Minorités de l'Académie des Sciences Sociales de Chine, pour avoir autorisé l'utilisation du matériel (technicien : M. Chen Jiayou). Merci à He Jiren, Guo Dalie et He Xueguang à Kunming et Lijiang ; à Thomas Pinson, du SIL ; à He Jiezheng, chercheur à Lijiang ; au Centre de recherches sur la culture dongba et au Musée de la culture dongba de Lijiang. Les restrictions d'usage s'appliquent. Une version de ce travail a été présentée en mars 2003 à l'Institut de phonétique (IPdS) de l'université Christian-Albrecht de Kiel (Allemagne), à l'invitation du Pr. Klaus Kohler.

Une aide financière a été apportée par la Fondation de France, et par l'Université Paris 3 (Projet innovant « Travail de première main sur langues rares »).

## Références

- Astesano C., 1999, *Rythme et discours: invariance et sources de variabilité des phénomènes accentuels du français*, thèse, Université de Provence.
- Baken, 1992, Electroglottography, *Journal of Voice*, 6(2), pp. 98-110.
- Benguerele P., 1973, Corrélat physiologiques de l'accent en français, *Phonetica*, 27(1), pp. 21-35.
- Campbell W. N., 1995, Loudness, spectral tilt, and perceived prominence in dialogues, International Congress of Phonetic Sciences, Stockholm.
- Carton F., 1974, *Introduction à la phonétique du français*, Paris, Bordas.
- Carton F., Hirst D., Marchal A. et Séguinot A., 1976, *L'accent d'insistance*, Studia Phonetica n°12, Montréal, Didier.
- Chen M. Y., 2000, *Tone Sandhi : Patterns Across Chinese Dialects*, Cambridge Studies in Linguistics 92, Cambridge, Cambridge University Press.
- Chen Y., Robb M. P. et Gilbert H. R., 2002, Electroglottographic evaluation of gender and vowel effects during modal and vocal fry phonation, *Journal of Speech, Language and Hearing research*, 45, pp. 821-829.
- Colton R. H., 1984, Glottal waveform variations associated with different vocal intensity levels, Symposium on Care of the Professional Voice, New York, Voice Foundation.
- Colton R. H. et Conture E. G., 1990, Problems and Pitfalls of Electroglottography, *Journal of Voice*, 4(1), pp. 10-24.
- Coustenoble H. et Armstrong L., 1937, *Studies in French intonation*, Cambridge, Heffer.
- d'Alessandro C., Boula de Mareüil P. et Prudon R., 2003, Synthèse par sélection: prosodie, dialogue et qualité vocale, *Cahiers de l'Institut de Linguistique de Louvain*(28), pp. 87-117.
- Delais-Roussarie E., Rialland A., Doetjes J. et Marandin J.-M., 2002, The Prosody of Post-focus sequences in French, *Speech Prosody*, Aix-en-Provence.
- Dromey C., Stathopoulos E. T. et Sapienza C. M., 1992, Glottal airflow and electroglottographic measures of vocal function at multiple intensities, *Journal of Voice*, 6(1), pp. 44-54.
- Epstein M., 2002, *Voice Quality and Prosody in English*, Ph.D. dissertation, Linguistics, Los Angeles, University of California.
- Erickson D., Fujimura O. et Menezes C., 2002, Contrastive Emphasis: Comparison of Pitch Accents with Syllable Magnitudes, *Speech Prosody*, Aix-en-Provence (France).
- Fant G., Hertegard S. et Kruckenberg A., 1996, Focal accent and subglottal pressure, *TMH-QPSR*, 2, pp. 29-32.
- Fant G. et Kruckenberg A., 1995, The voice source in prosody, International Congress of Phonetic Sciences, Stockholm.
- Fónagy I., 1983, *La vive voix: essais de psycho-phonétique*, "Langages et Sociétés", Paris, Payot.
- Fry D. B., 1955, Duration and intensity as physical correlates of linguistic stress, *Journal of the Acoustical Society of America*, 27, pp. 765-768.
- Fry D. B., 1958, Experiments in the Perception of Stress, *Language and Speech*, 1, pp. 126-152.
- Fu M., 1940, *A Study of the Moso Language (Wei-Hsi Dialect), Part II: Grammar (in Chinese)*, University of Western China.

- Fu M., 1944, *A Study of the Moso Hieroglyphic Manuscript "The Genesis and History of Mankind"*, from *Likiang*, Wuchang, Wuchang University.
- Garde P., 1968, *L'Accent*, Paris, Presses Universitaires de France.
- Gilbert H. R., Potter C. R. et Hoodin R., 1984, Laryngograph as a measure of vocal fold contact area, *Journal of Speech and Hearing Research*, 27, pp. 173-178.
- Glave et Rietveld, 1975, Is the effort dependence of speech loudness explicable on the basis of acoustic cues?, *Journal of the Acoustical Society of America*, 58, pp. 875-879.
- Gobl C., 1988, Voice source dynamics in connected speech, *STL-QPSR*, 1, pp. 123-159.
- Gsell R., 1979a, *La prosodie du thai standard: tons et accent*, Institut de Phonétique, Paris, Université Paris 3.
- Gsell R., 1979b, *Sur la réalisation chuchotée des tons du thai standard*, research report, Institut d'Etudes Linguistiques et Phonétiques, Paris, Université de la Sorbonne Nouvelle.
- Gsell R., 1980, Remarques sur la structure de l'espace tonal en vietnamien du sud (parler de Saïgon), *Cahier d'études vietnamiennes, Département de Langues et Civilisations de l'Asie Orientale de l'Université Paris 7*, 4.
- Gussenhoven C., 2002, *Intonation and Interpretation: Phonetics and Phonology*, Speech Prosody, Aix en Provence.
- Haiman J., 1978, Conditionals are topics, *Language and Speech*, 54, pp. 564-589.
- Han M. S. et Kim K.-O., 1974, Phonetic variation of Vietnamese tones in disyllabic utterances, *Journal of Phonetics*, 2, pp. 223-232.
- He J. et Jiang Z., 1985, *Naxiyu Jianzhi (A Presentation of the Naxi Language)*, Beijing, Renmin Chubanshe.
- Heldner M., 2001a, *Spectral emphasis as a perceptual cue to prominence*, TMH-QPSR, pp. 51-57.
- Heldner M., 2001b, *Spectral Emphasis as an Additional Source of Information in Accent Detection*, Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, Red Bank, New Jersey.
- Heldner M., Strangert E. et Deschamps T., 1999, A focus detector using overall intensity and high frequency emphasis, International Conference on Speech and Language Processing, San Francisco.
- Henrich N., 2001, *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception*, thèse, Paris, Université Paris 6.
- Henrich N., d'Alessandro C., Castellengo M. et Doval B., soumis pour publication, On the use of the derivative of the electroglottographic signal, *Journal of the Acoustical Society of America*.
- Hirst D. et Di Cristo A., 1998, *Intonation Systems: A Survey of Twenty Languages*, Cambridge, Cambridge University Press.
- Hombert J.-M., 1976, Consonant types, vowel quality and tone, in *Tone : a Linguistic Survey*, V. A. Fromkin, New York, Academic Press, pp. 77-111.
- Issatchenko A. V. et Schädlich H. J., 1966, Untersuchungen über die deutsche Satzintonation, *Studia Grammatica*, 7, pp. 7-64.
- Kochanski G. P. et Shih C., soumis pour publication en 2003, A Quasi-glottogram signal for voicing and power estimation, *Journal of the Acoustical Society of America*.
- Kohler K. J., 1991, Terminal intonation patterns in single accent utterances of German : phonetics, phonology, and semantics, in *Studies in German Intonation [Arbeitsberichte des*

*Instituts für Phonetik der Universität Kiel (AIPUK) n°25*, Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK), University of Kiel (Germany).

Kohler K. J., 2003, Neglected Categories in the Modelling of Prosody: Pitch Timing and Non-Pitch Accents, International Congress of Phonetic Sciences, Barcelona.

Ladefoged P., Draper M. H. et Whitteridge D., 1958, Syllables and Stress, *Miscellanea Phonetica*, III, pp. 1-14.

Launey M., 1994, *Une grammaire omniprédicative: essai sur la morphosyntaxe du nahuatl classique*, Sciences du Langage, Paris, CNRS Editions.

Mazaudon M., 2003, From discourse to grammar in Tamang: topic, focus, intensifiers and subordination, in *Language Variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, D. Bradley, R. LaPolla, B. Michailovsky et G. Thurgood, Pacific Linguistics, Canberra, A.N.U., pp. 145-157.

Mazaudon M., sous presse, From discourse to grammar in Tamang: topic, focus, intensifiers and subordination, in *Language Variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, G. Thurgood, Pacific Linguistics, Canberra, A.N.U.

Michaud A., 2003, The Rising tone of Naxi: from syntax and intonation to the lexicon?, Prosodic Interfaces, Nantes, France.

Mol H. G. et Uhlenbeck G. M., 1956, The linguistic relevance of intensity in stress, *Lingua*, 5, pp. 205-213.

Mooshammer C. et Fuchs S., 2002, Stress distinction in German: simulating kinematic parameters of tongue-tip gesture, *Journal of Phonetics*, 30(3), pp. 337-355.

Ohala J., 1978, The Production of Tone, in *Tone: a linguistic survey*, V. A. Fromkin, New York/San Francisco/London, Academic Press.

Ohala J. et Ewan W. G., 1973, Speed of pitch change, 84th Meeting of the Acoustical Society of America, abstract published in the Journal of the Acoustical Society of America 53(1):345.

Orlikoff R. F., 1998, Scrambled EGG: The Uses and Abuses of Electroglottography, *Phonoscope*, 1(1).

Paris M.-C., 1979, *Nominalization in Mandarin Chinese: the morpheme 'de' and the 'shi...de' constructions*, Département de recherches linguistiques, Paris, Université Paris VII.

Pike K. L., 1948, *Tone Languages. A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*, Ann Arbor, University of Michigan Press.

Rialland A., 1997, Le parcours du "downstep", ou l'évolution d'une notion, *Bulletin de la Société de Linguistique*, XCII(1), pp. 207-243.

Rossi M., 1971, L'intensité spécifique des voyelles, *Phonetica*, 24, pp. 129-161.

Rossi M., 1999, *L'intonation, le système du français: description et modélisation*, Gap/Paris, Ophrys.

Rothenberg M. et Mahshie J. J., 1988, Monitoring vocal fold abduction through vocal fold contact area, *Journal of Speech and Hearing Research*, 31, pp. 338-51.

Roubeau B., Chevrie-Muller C. et Arabia-Guidet C., 1987, Electroglottographic study of the changes of voice registers, *Folia Phoniatrica*, 39, pp. 280-89.

Roubeau B., Chevrie-Muller C. et Arabia-Guidet C., 1991, Control of laryngeal vibration in register change, in *Vocal Fold Physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms*, J. Gauffin et B. Hammarberg, San Diego, Singular Publishing Group, pp. 279-86.

- Shih C., 2000, A Declination Model of Mandarin Chinese, in *Intonation: Analysis, Modelling and Technology*, A. Botinis, Kluwer Academic Publishers, pp. 243-268.
- Shih C. et Kochanski G. P., 2000, Chinese Tone Modeling with Stem-ML, ICSLP 2000, Beijing.
- Shih C., Kochanski G. P., Fosler-Lussier E., Chan M. et Yuan J., 2001, Implications of Prosody Modeling for Prosody Recognition, ISCA Workshop on Prosody in Speech Recognition and Understanding, Red Bank, New Jersey.
- Sluijter A., Van Heuven V. J. et Pacilly J. J. A., 1997, Spectral balance as a cue in the perception of linguistic stress, *Journal of the Acoustical Society of America*, 101, pp. 503-513.
- Strik H. et Boves L., 1992, A physiological model of intonation, AFN 1992, University of Nijmegen.
- Sundberg J., 1979, Maximum speed of pitch changes in singers and untrained subjects, *Journal of Phonetics*, 7, pp. 71-79.
- Titze I. R., 1988, Regulation of Vocal Power and Efficiency by Subglottal Pressure and Glottal Width, in *Vocal fold physiology: voice production, mechanisms and functions*, O. Fujimura, New York, Raven Press, pp. 227-237.
- Titze I. R., 1990, Interpretation of the Electroglottographic signal, *Journal of Voice*, 4(1), pp. 1-9.
- Titze I. R., 1994, *Principles of voice production*, Englewood Cliffs, Prentice-Hall.
- Vaissière J., 1977, Premiers essais de l'utilisation de la durée pour la segmentation en mots dans un système de reconnaissance, 8èmes Journées d'Etudes sur la Parole, Aix-en-Provence.
- Vaissière J., 1983, Language-independent Prosodic Features, in *Prosody: Models and Measurements*, A. Cutler et R. Ladd, Berlin, Springer Verlag, pp. 53-66.
- Vaissière J., 1986, Variance and Invariance at the Word Level, in *Invariance and Variability in Speech Process*, J. S. Perkell et D. Klatt, Lawrence Erlbaum Associates, pp. 534-539.
- Vaissière J., 1995, Phonetic explanations for cross-linguistic similarities, *Phonetica*, 52, pp. 123-130.
- Vaissière J., 2004, The Perception of Intonation, in *Handbook of Speech Perception*, D. B. Pisoni et R. E. Remez, Blackwell Textbooks in Linguistics, Oxford, U.K. & Cambridge, Massachusetts, Blackwell.
- Vu Ngoc T., d'Alessandro C. et Rosset S., 2002, A phonetic study of Vietnamese tones : acoustic and electroglottographic measurements, International Conference on Speech and Language Processing, Boulder, Colorado.
- Xu Y., 1997, Contextual tonal variations in Mandarin, *Journal of Phonetics*, 25, pp. 61-83.
- Chao Yuen-ren, 1929, Beiping intonation [en chinois], in *A.A. Milne: The Camberley Triangle*, Shanghai, Zhonghua Bookstore, pp. (appendice).
- Chao Yuen-ren, 1932, A preliminary study of English intonation (with American variants) and its Chinese equivalent, *Bulletin of the Institute of History and Philology*, BIHP Supplement: Studies presented to Cai Yuanpei on his 65th birthday, pp. 105-156.
- Chao Yuen-ren, 1933, Tone and intonation in Chinese, *Bulletin of the Institute of History and Philology*, 4:3, pp. 121-134.
- Chao Yuen-ren, 1935, Intonation of National Language [en chinois], *National Language Weekly*, 214.