



**HAL**  
open science

# Usons de la dimension vocale jusqu'à la corde : la voix du locuteur enchâssé dans le discours rapporté direct à l'oral

Bertrand Verine

► **To cite this version:**

Bertrand Verine. Usons de la dimension vocale jusqu'à la corde : la voix du locuteur enchâssé dans le discours rapporté direct à l'oral. *Les cahiers de praxématique*, 2007, 49, pp.159-181. hal-00700352

**HAL Id: hal-00700352**

**<https://hal.science/hal-00700352>**

Submitted on 22 May 2012

**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.

USONS DE LA DIMENSION VOCALE JUSQU'À LA CORDE :  
LA VOIX DU LOCUTEUR ENCHASSE DANS LE DISCOURS RAPPORTÉ DIRECT À L'ORAL

Il est (...) possible de considérer que certaines formes prosodiques se caractérisent par un sens stable très général, tout en fonctionnant comme des indices de contextualisation.

Grobet, 2001, *in* Roulet E., L. Filliettaz & A. Grobet, 234.

Dans l'appel à contribution de ce numéro, Jeanne-Marie Barbéris a pointé un double paradoxe : d'une part, le fait que le terme *voix* ait été emprunté au champ du discours oral pour rendre compte des hétérogénéités énonciatives avant tout dans les textes écrits, et d'autre part, le fait que, dans les analyses dialogiques ou polyphoniques, la *voix* soit le plus souvent mise en relation non avec les instances de production et de réception matérielles des énoncés, mais avec les instances de prise en charge de leur contenu sémantique et pragmatique. Cela pose une question toujours angoissante quand on s'efforce de faire de la recherche scientifique : la *voix*, que de nombreux chercheurs tentent de constituer en concept opératoire, voire en concept plein et entier, ne serait-elle qu'une métaphore ?

Pour proposer quelques éléments de réponse à cette interrogation, je me consacrerai prioritairement ici aux locuteurs tels que nous les avons définis dans Bres et Verine (2002). Le but de ce travail était de formaliser un cadre unifié pour l'approche praxématique des marqueurs de dialogisme et, en particulier, d'articuler le discours rapporté (désormais DR) aux autres phénomènes dialogiques, tels que la négation, la confirmation, le détournement ou l'ironie. Je me limiterai ici au discours direct (désormais DD) qui, de manière presque toujours explicite, fait apparaître la hiérarchisation de deux énoncés renvoyant à deux situations d'énonciation différentes : l'énoncé enchâssant, que nous symbolisons par E, correspond à l'interaction en cours ou au texte considéré, comportant en principe une forme du type *X dire (à Z) : « Y »*, qui désigne un autre acte d'énonciation ; dans le segment bivocal « Y », l'énoncé enchâssant fait entendre, selon des formes et des proportions très diverses, le fruit de cet autre acte d'énonciation, soit l'énoncé enchâssé, que nous symbolisons par e. À chacun de ces deux énoncés nous faisons correspondre quatre instances :

- l'énonciateur, instance d'actualisation modale et/ou de repérage déictique du côté de la production, que nous symbolisons par E<sub>1</sub> pour l'énoncé enchâssant et par e<sub>1</sub> pour l'énoncé enchâssé ;
- l'énonciataire, instance modale et déictique du côté de la réception, que nous symbolisons par E<sub>2</sub> ou par e<sub>2</sub> selon le niveau d'enchâssement ;
- le locuteur, instance de profération sonore ou d'inscription graphique, que nous symbolisons par L<sub>1</sub> pour l'énoncé enchâssant et par l<sub>1</sub> pour l'énoncé enchâssé ;
- l'interlocuteur, instance de réception auditive et/ou visuelle, que nous symbolisons par L<sub>2</sub> ou par l<sub>2</sub> selon le niveau d'enchâssement.

Si l'on centre l'attention sur l'instance du locuteur dans le discours rapporté direct (désormais DRD), on constate que les genres écrits, en particulier dans le discours journalistique où ces formes abondent, ont développé des conventions typographiques de plus en plus nombreuses pour marquer et délimiter matériellement l'altérité du segment bivocal, tendant à l'identifier à l'énoncé enchâssé : alinéas, guillemets ou tirets, alternance entre caractères droits et gras ou italiques, chevrons en marge et parfois désormais changements de couleur... Comment, dans les genres oraux, un locuteur enchâssant L<sub>1</sub> utilise-t-il sa voix, au sens propre d'instrument de

profération, quand il rapporte en DD les mots d'un locuteur enchâssé  $l_1$  ? Marque-t-il toujours phoniquement la différence entre l'énoncé enchâssant et le segment bivocal ? Cette bivocalité, que nous posons au niveau énonciatif de toutes les formes dialogiques – dans la mesure où l'énonciateur enchâssant prend nécessairement position sur la validité, ou non, de l'énoncé enchâssé – s'exerce-t-elle également au niveau locutoire, et dans quelles proportions ? Afin de défricher cette problématique, je passerai d'abord en revue quelques travaux sur la question (1), avant d'illustrer le caractère non prédictible du marquage vocal (2) et de dégager quelques contraintes à prendre en considération pour progresser dans sa description (3).

## **1. L'état d'un art naissant**

Je partirai du constat posé par Roxane Bertrand et Robert Espesser (1998 : 48), selon qui, « de cette hétérogénéité montrée relative aux DRD, seules des marques de nature verbale (pour l'essentiel syntaxiques : temps des verbes utilisés, pronoms, etc.) ont été abordées. Si certains auteurs (Goffman, Gumperz, Jacob) ont abordé les questions du marquage prosodique dans le rapport de parole, leurs considérations sont souvent demeurées au stade théorique ». J'ajouterai que lors des deux premiers colloques du groupe Ci-Dit, exclusivement consacrés au DR (López Muñoz, Marnette & Rosier éd., 2004 et 2005), très peu de communications ont porté sur ses occurrences orales, et aucune sur leur dimension prosodique. Je présenterai synthétiquement les travaux auxquels j'ai eu accès, en distinguant ceux qui s'appuient sur des mesures physiques et ceux qui se basent sur des observations auditives. Je précise au préalable que je me suis moi-même limité jusqu'ici à quelques notations sur l'inflexion riieuse (Fauré et Verine 2004), sur l'intensité renforcée ou affaiblie et sur l'autoparodie (Verine 2005 et 2006). Le présent travail se situera dans la lignée des recherches auriculaires, non sans tenter de tirer profit des conclusions auxquelles parviennent les recherches instrumentées.

### **1.1. Travaux instrumentés**

Les textes auxquels j'ai eu accès s'attachent à la dimension vocale dans l'ensemble de la catégorie du DD en distinguant ses deux composantes fondamentales que sont l'énoncé enchâssant et l'énoncé enchâssé, mais sans jamais poser explicitement la question du locuteur cité, de sa représentation ou d'une éventuelle bivocalité.

Ainsi, Bertrand et Espesser ont comparé, sur un corpus de 188 occurrences, cinq paramètres dans la zone de transition initiale constituée par les derniers segments de l'énoncé enchâssant et les premiers segments bivocaux faisant entendre l'énoncé enchâssé. Trois de ces paramètres ne donnent pas de résultats significatifs, à savoir qu'on ne constate d'écart interprétable ni sur l'intensité en décibels, ni sur le nombre de syllabes par seconde, ni sur le taux de présence de tenues voisées. Les deux types de segments s'opposent en revanche par deux traits. Pour ce qui est de la ligne mélodique, d'une part, la fin des énoncés enchâssants présente une « structure plus aléatoire », alors que le début des segments bivocaux offre des « mouvements plus continus » (*op. cit.* : 54). Les auteurs expliquent cette différence en termes de moindre hésitation et de meilleure planification de la locutrice étudiée sur le discours autre que sur son dire en cours : « on peut supposer en effet que celle-ci, lorsqu'elle s'apprête à produire du DRD, *sait ce qu'elle va dire* (du moins sur la durée de l'empan envisagé, – paires entre 1 et 4 secondes) » (*ibidem*).

En ce qui concerne la dynamique tonale, d'autre part, le début des segments bivocaux se caractérise par un élargissement vers les valeurs élevées de la fréquence fondamentale  $F_0$ , que les auteurs expliquent en termes d'emphatisation : « le DRD, réputé pour son statut d'énoncé pertinent, notamment d'un point de vue informatif, revêt, selon nous, une part d'implication plus grande chez la locutrice qui le produit » et, réciproquement, « les DRD sont précédés d'éléments verbaux de préparation (introductions) fonctionnant à la manière de signaux d'annonce, mais sans doute caractérisés par une moindre implication énonciative des sujets »

(*ibidem*). Bertrand et Espesser (2002) ont approfondi l'étude de ce dernier paramètre, mais sur la question spécifique de l'opposition entre DD d'origine autophonique, dans lequel le locuteur enchâssant cite un de ses propres énoncés, et DD d'origine hétérophonique, qui cite un autre locuteur : je n'entrerai pas ici dans le détail de cette distinction.

Les résultats de 1998 me paraissent appeler au moins deux remarques et deux questions. La première remarque tient très généralement à leur représentativité, car les 188 occurrences analysées ont été produites par une seule locutrice à l'occasion d'une même conversation ; or nous observerons *infra* (notamment exemples [4] et [5]), à la simple écoute, de notables différences selon les locuteurs et/ou selon les circonstances. La seconde remarque a trait à l'absence d'exemple transcrit dans cet article : cela présente évidemment l'avantage d'empêcher de corrélérer *a priori* tel phénomène vocal avec un cotexte sémantique ou une contrainte pragmatique trop spécifiques. Cela conduit, cependant, à une première question : la dynamique tonale étendue vers le haut dans les segments bivocaux ne pourrait-elle pas, assez souvent, s'expliquer par la fréquence des modalités interrogative et exclamative dans les discours représentés ? Ma seconde question concerne l'explication des différences mélodiques par l'opposition entre le discours propre, à planifier, et l'autre discours, souvent déjà connu<sup>1</sup>, opposition dont nous observerons *infra* (notamment exemple [2]) qu'elle est loin de s'avérer systématique : ne peut-on pas, au contraire, faire l'hypothèse que la profération moins assurée de l'énoncé enchâssant dans le temps du dire provienne d'une planification plus coûteuse de l'énoncé autre dans le temps de l'à-dire, cet effort supérieur de programmation aboutissant à une élocution plus continue du segment bivocal ? J'argumenterai très rapidement cette hypothèse en rappelant que si, à l'écrit, il peut suffire de recopier l'énoncé déjà produit que l'on souhaite enchâsser, à l'oral, « on ne se souvient que très partiellement de paroles qui viennent juste d'être prononcées » (Vincent et Dubois, 1997 : 20) : la production d'un DD a donc ceci de coûteux qu'elle implique de (re)construire une approximation de paroles dites, avec pour contrainte (plus ou moins forte selon les contextes) que cette approximation paraisse convaincante.

Quoi qu'il en soit, ces résultats semblent avérer un changement de voix entre l'énoncé enchâssant et le segment bivocal, donc présupposer un marquage prosodique du locuteur cité  $I_1$ . Il n'en va pas de même chez Mary-Annick Morel (1998), qui reprend un article de 1996 paru dans les *Cahiers du français contemporain* 3. S'attachant à un seul élément, la dynamique tonale, elle l'étudie sur un corpus très diversifié, ce qui permet d'affiner et d'élargir le critère de Bertrand et Espesser. Celui-ci se trouve décomposé en trois indices : hauteur du verbe introducteur par rapport à l'amont de l'énoncé enchâssant, hauteur initiale du segment bivocal, et variation ou non de la mélodie au cours du segment bivocal. Morel ne fournit aucun chiffre et n'envisage pas toutes les combinaisons possibles entre les trois indices, mais distingue cinq configurations typiques qu'elle relie au degré d'implication du locuteur enchâssant et à son degré d'adhésion au contenu de l'énoncé enchâssé. Sur ces cinq configurations, deux seulement semblent coïncider avec les résultats précédents : (i) celle où le segment bivocal est proféré en plage haute et sans modulation, qui marquerait le caractère inattendu de la situation, et (ii) celle où un verbe en plage haute introduit un discours également en plage haute, qui indiquerait l'implication maximale du locuteur citant par rapport à un énoncé constituant une surprise ou un tournant. Les autres configurations retenues sont (iii) l'introducteur en plage haute validant un énoncé enchâssé qui appuie la thèse enchâssante, (iv) l'introducteur en plage basse corrélé à l'autonomie du locuteur cité ou à la réticence envers l'autre discours, et (v) le segment bivocal en plage basse sans modulation associé au total désengagement du locuteur citant. Selon ces données, et sur le seul critère de la dynamique tonale, le marquage vocal du locuteur cité  $I_1$  et, même, le changement de voix

---

<sup>1</sup> Aucun des travaux consultés ne prend en compte la distinction entre les énoncés enchâssés préexistants et ceux dont l'énonciateur enchâssant suppose l'existence (cf. Verine 2005).

de l'énoncé enchâssant au segment bivocal semblent donc loin de constituer la configuration la plus typique.

Le travail de Jacqueline Léon (1988) paraît aller dans le même sens, mais pour des raisons bien différentes. En effet, alors que Morel présuppose l'étanchéité des énoncés E et e, Léon commente de nombreux exemples où la différence entre les deux segments s'avère abolie ou subvertie. Elle prend notamment en compte les DD non canoniques, introduits de manière implicite ou non univoque, et les discours directs libres (DDL), sans aucun élément recteur. Bien qu'elle ne produise aucun chiffre, elle démontre ainsi qu'assez souvent « les phénomènes prosodiques, au lieu de contribuer à assurer la rupture entre deux actes (ou pseudo-actes) d'énonciation séparés, ont pour effet de condenser deux voix en une » (*op. cit.* : 116). De fait, « on peut (...) observer de nombreux cas de recouvrement, total ou partiel, entre l'intonation de l'introduction et l'intonation de l'énoncé rapporté » (*op. cit.* : 112) : l'exemple 6 (*ibidem*) illustre même le changement implicite de locuteur cité, c'est-à-dire la juxtaposition, sans introducteur ni variation intonative de deux énoncés enchâssés que seul le contexte permet d'imputer à deux voix différentes. Dans d'autres cas, « la discordance introduite par l'intonation met L en position de distance évaluative des paroles de l. Ce mode de prise en charge énonciatif, considéré par J. Authier et J. Rey-Debove comme incompatible avec une forme de DD, s'apparente plutôt au DIL [discours indirect libre] : deux voix, celles de L et l, se font entendre simultanément » (*op. cit.* : 115).

Je ne saurais rendre compte de toute la palette de nuances qu'illustrent Gabriele Klewitz et Elizabeth Couper-Kuhlen (1999) puisque, partant d'exemples où le marquage prosodique fonctionne comme un équivalent strict du marquage typographique, elles en viennent graduellement à parler de *cadrage* puis de simple *balisage* pour conclure : « the flags (...) are not iconic but indexical of a deictic shift in reported speech » (*op. cit.* : 26) et « the prosodic marking (...) is a stylistic device rather than a norm : it may be used to signal reported speech or not, depending on speakers' local goals and strategic choices » (*op. cit.* : 30). Elles confirment donc, sur un corpus anglophone, que le changement de voix n'a pas de caractère systématique en DD, au moins au regard des paramètres fondamentaux de la phonologie (cf. *infra* exemple [2]), que les cas de parodie où L<sub>1</sub> oppose sa voix propre au simulacre de la voix de l<sub>1</sub> doivent de ce fait être considérés comme des spectacularisations exceptionnelles (cf. *infra* exemple [3]) et que, symétriquement, il arrive que L<sub>1</sub> instille son intonation évaluative dans l'énoncé imputé à l<sub>1</sub> (cf. *infra* exemple [7]).

## 1.2. Travaux auriculaires

Revenant au domaine francophone, je ne signalerai que pour mémoire l'article de Hanne Leth Andersen (2000), car le rôle de la prosodie s'y trouve seulement mentionné en introduction. De même, le travail de Diane Vincent et Sylvie Dubois, qui constitue la seule monographie existante sur le DR en français oral, inclut le changement du timbre, la qualité de la voix et l'intonation parmi les indices pertinents de discordance énonciative (1997 : 86), mais constate : « on trouve, dans nos données, assez peu de traces des accents, des voix et des " tics " de l'énonciateur. Les propos sont mis en évidence, mais l'auteur perd ses caractéristiques individuelles et physiques, sauf si le locuteur veut mettre l'accent précisément sur ces caractéristiques, ce qui se fait alors au détriment du contenu » (*op. cit.* : 20). De fait, le seul exemple commenté est une occurrence de liaison intempestive.

Deux articles consacrés à des corpus germanophones fournissent, en revanche, des remarques transférables pour tenter de s'orienter dans la recherche du marquage vocal du locuteur enchâssé. Celui de Klaus Müller (1992) a beaucoup plus généralement pour objet le marquage prosodique du caractère joyeux ou dramatique des étapes d'un récit parmi lesquelles figurent de nombreux DD. Son hypothèse d'ensemble est que si les éléments lexicaux et paraverbaux se conjuguent parfois pour renforcer leurs effets, leur combinatoire s'avère non prédictible, ce

qu'attestent deux constats. D'une part, « even dramatic episodes are potential laughables, if the narrator does contextualize this fact through her own laughing after the episod. This " strong invitation " can compensate a rather sad content and/or a dramatic rhythm and/or dark and angry voice » (*op. cit.* : 217). D'autre part, en rapportant une longue séquence dialogale, la narratrice étudiée s'impute systématiquement un timbre grave, coïncidant avec sa préoccupation au moment des faits racontés, alors qu'elle associe toujours le timbre aigu à l'enjouement de son conjoint masculin : la hauteur ne serait donc pas, en l'occurrence, imitative du type physique des locuteurs cités, mais représentative de leur état d'esprit (cf. *infra* exemple [5]). La très grande précision de la transcription permet par ailleurs de repérer qu'au moins deux énoncés enchâssés en DD ne se trouvent ni rythmiquement ni intonativement démarqués de l'énoncé enchâssant. En somme, comme l'écrit Susanne Günthner dans le commentaire de ce travail, il semble que « there is no stable one-to-one relationship between a specific voice and a specific character » (1992 : 226).

Elle ajoute une remarque qui me paraît capitale, à savoir qu'à l'oral les locuteurs enchâssants commentent beaucoup plus rarement qu'à l'écrit la façon de parler des locuteurs qu'ils citent, et ce en raison de la possibilité d'intégrer à leur propre diction tel ou tel trait prosodique ou paralinguistique qu'ils jugent pertinent. Je préciserai que, selon moi, le facteur déterminant dans les occurrences marquées n'est pas la fidélité envers le référent, donc la restitution vraisemblable de la voix du locuteur enchâssé, mais l'efficacité à l'égard du destinataire actuel : c'est seulement lorsqu'il lui paraît pertinent de contextualiser davantage l'autre discours que le locuteur enchâssant recourt au marquage vocal. Cela me semble également montrer que, sur ce point, l'oral peut utiliser simultanément des ressources polysémotiques là où l'écrit est condamné à la linéarité monosémotique : si l'on continue parfois à jauger négativement la moindre précision et la moindre densité du discours oral, c'est qu'il nous manque encore certaines unités et certains instruments de mesure pour rendre les marqueurs prosodiques quantifiables et évaluables.

À partir d'un corpus de trente-huit conversations entre parents ou amis, à table ou au téléphone, Günthner (1999) illustre la diversité des utilisations par plusieurs locuteurs enchâssants des indices intonatifs mentionnés jusqu'ici, mais aussi de propriétés vocales telles que les chuchotements, les soufflements ou les inspirations. Elle attribue à l'ensemble de ces traits cinq fonctions principales dont, à nouveau, la combinatoire n'est pas prédictible. La première peut être de distinguer les énoncés enchâssants et bivocaux, en particulier quand le locuteur principal rapporte ses propres discours (*infra* exemple [5]). La seconde peut être de discriminer plusieurs locuteurs enchâssés, notamment en l'absence de formule rectrice (*infra* exemple [4]), ce qui est fréquent lorsque les tours de parole rapportés se multiplient. La troisième peut être de caractériser le type d'acte discursif auquel contribue l'énoncé enchâssé : par exemple, le reproche (*infra* exemple [4g]), l'excuse, la plainte, etc. La quatrième, que Günthner associe tantôt à la précédente tantôt à la suivante, peut être de contextualiser l'état émotionnel du locuteur cité (*infra* exemple [4f]), dont l'auteure prend pour exemple l'hystérie, le calme, l'insistance ou la détermination. Par le choix de ces catégories, on se doute que la dernière fonction des indices intonatifs et des propriétés vocales peut être d'évaluer, du point de vue actuel du locuteur enchâssant, l'autre discours et l'autre locuteur (*infra* exemple [7], cf. aussi Günthner 2001 : 8-12).

Günthner développe ce cadre d'analyse en partant des exemples les plus théâtraux pour aller progressivement vers des cas de plus en plus complexes, voire paradoxaux. C'est ce même type de cheminement que je voudrais maintenant suivre.

## **2. Les degrés d'audibilité du locuteur enchâssé**

Dans les cas très majoritaires où le contexte de l'énonciation rapportée est coupé de l'interaction en cours, tout le problème est de déterminer dans quelle proportion tel ou tel

élément prosodique renvoie à l'une ou l'autre des situations d'énonciation et de leurs nombreux paramètres. Pour tenter de progresser dans le balisage de cette épineuse question, je commencerai par articuler la catégorie peu étudiée du *discours convoqué* avec les formes habituelles de DR que plusieurs auteurs, comme Roulet (2001) notamment, préfèrent dénommer *discours représenté* plutôt que *rapporté*.

## 2.1. Discours convoqué dans les médias audiovisuels

Dans une note infrapaginale, Manuel Fernandez (2001 : 72) définit le *discours convoqué* comme « la forme particulière du discours rapporté des médias audiovisuels qui consiste à donner la parole à des intervenants extérieurs au dispositif énonciatif proprement médiatique (interview, micro-trottoir...) ». Il s'agit donc de toutes les occurrences où les techniques d'enregistrement permettent d'insérer dans le dire en cours un fragment d'une autre énonciation en reproduisant sa matérialité sonore, voire une partie de son contexte visuel. Il convient d'ailleurs d'étendre cette dénomination à d'autres secteurs d'activité que les médias, à commencer par certaines conférences scientifiques, mais aussi de nombreuses interactions instrumentées par l'ordinateur, dans lesquelles un insert peut également reproduire la matérialité graphique originelle d'un texte écrit. L'articulation du DR spontané et du DR instrumenté est rendue concrète par l'exemple [1]<sup>2</sup> :

(1) (Deux jeunes viennent de mourir électrocutés en tentant d'échapper à une interpellation policière.)

(L<sub>1</sub>) Nicolas Sarkozy<sup>3</sup> qui promet que **tout' la lumière s'ra fait' sur cette affaire'** / qui dit vouloir **s'attaquer / aux violenc's urbain's en décrétant la toléranc' zéro** / et en annonçant qu'il ira / chaqu' semaine' dans un quartier difficile / il était à nouveau c' matin en Sein'-Saint-Denis

(l<sub>1</sub>) Clichy-sous-Bois / on a eu d'abord ce drame avec ces deux jeun's / c'est un drame / (*malheureus'ment pour les parents*) et / je reçois les parents cet après-midi (*hmhm*) (*haha*) / j' les reçois pour les assurer d'une chos' dans / l'épreuv' qui est la leur / c'est qu'ils auront accès à tous les documents pour savoir quelle est exactement la vérité / premier point / deuxièm' point / j'ai mis les moyens nécessair's à Clichy-sous-Bois / pour que vous puissiez vivr' tranquill'ment / y a aucun' raison qu' Clichy-sous-Bois connaiss' des nuits d'émeute / comm' cela / troisièm'ment je vais voir dans un instant l'imam / pour parler avec lui d' c' qui s'est passé / dans la mosquée de manière à c' que le calme revienne / parc' qu'on n' peut rien traiter et rien fair' quand y a pas du calme / en l'état de me::s informations / je confirm' que **c'est bien un' grenad' lacrymogène** euh qui vient d'une:: / -fin / qui est en dotation / dans euh les compagnies euh d'intervention qui étaient / sur place en Sein'-Saint-D'nis cett' nuit-là ce qui n' veut pas dir' / que c'est un tir qui a été vou- euh / fait par un policier hein ça c'est l'enquêt' qui le dira

(L<sub>1</sub>) Nicolas Sarkozy au micro de: Sophie Parmentier vous l'avez entendu c'est **donc bien / un' grenad' lacrymogèn'** qui a été euh tirée dans la mosquée (« Le 13-14 » d'Yves Decamp, *France Inter*, 31 octobre 2005).

Je ne détaillerai pas ici le fait que le journaliste a choisi de redoubler chacun des trois points du discours de N. Sarkozy en le rapportant deux fois, d'une part sous la forme de l'enregistrement audio, donc du discours convoqué, d'autre part sous l'une des formes habituelles du DR (segments graissés) : aux plans conceptuel et terminologique, cela tend, en tout cas, à prouver que le *discours rapporté* fonctionne comme une catégorie englobante dont le *discours convoqué* et le *discours représenté* constituent deux sous-classes (cf. Verine 2008).

---

<sup>2</sup> Conventions de transcription : [:] allongement vocalique ; [/] pause ; [+] syllabe inaudible ; ['] élision de e muet ; [-] troncation de syllabe ; [↑] intonation ascendante ; l'encadrement d'un segment par [(1)] indique une inflexion rieuse, par [(2)] un ton élevé, par [(3)] un ton affaibli, par (4) une stylisation vocale. Tous les noms désignant des personnes privées ont été remplacés par des pseudonymes. Pour respecter l'absence de marqueur univoque de clôture de l'énoncé enchâssé à l'oral, je bannis de mes transcriptions tout guillemet et les tirets autres que celui indiquant le début d'un tour de parole mais j'italique les segments bivocaux. Les segments italiqués entre parenthèses transcrivent les chevauchements par un autre locuteur. En [1], je graisse les DR par lesquels le journaliste reformule les trois points principaux de l'enregistrement de N. Sarkozy.

<sup>3</sup> Alors ministre de l'intérieur.

J'insisterai en revanche sur le fait que cette occurrence permet d'entendre trois types de contraste sonore entre discours représenté et discours convoqué, lorsque ce dernier n'a pas été retouché par des moyens techniques. Elle rend d'abord audible le changement d'environnement acoustique entre les situations d'énonciation enchâssante et enchâssée : en l'espèce, on passe du son totalement aseptisé du studio où la voix enchâssante du journaliste L<sub>1</sub>, absolument nette, occupe tout l'espace sonore, à l'atmosphère beaucoup plus complexe de Clichy-sous-Bois dans laquelle la voix enchâssée de N. Sarkozy I<sub>1</sub>, légèrement étouffée, se détache sur le bruit de fond d'une ou plusieurs autres voix à proximité, et sur les échos plus lointains de cris d'enfants puis de moteurs de voitures. On remarque ensuite l'absence de filtrage de trois segments dont la décontextualisation empêche de déterminer s'ils constituent les seuls fragments audibles d'une conversation parallèle, ou des évaluations des destinataires originels I<sub>2</sub>, les clichois et les reporters présents lors de la déclaration, ou encore une suggestion (*malheureus'ment pour les parents*) puis deux régulateurs (*hmhm* et *haha*) adressés à N. Sarkozy par ses conseillers. On observe enfin l'absence de coupe au montage de deux lapsus potentiellement lourds de signification politique dans la parole même du ministre : *qui vient d'une:: / -fin / qui est en dotation / dans euh les compagnies euh d'intervention*, d'une part, *un tir qui a été vou- euh / fait par un policier hein*, d'autre part. On perçoit ainsi que le discours convoqué, lorsqu'il n'est pas retouché, recèle des indices pertinents pour contextualiser l'énoncé enchâssé (environnement sonore), mais que l'incomplétude de cette contextualisation peut rendre indécidable la pertinence de certains constituants (interventions d'arrière-plan), et même rendre signifiants des segments qui, en interaction, étaient susceptibles de passer inaperçus (lapsus). Cette part incompressible de décontextualisation et la fréquence des enregistrements retouchés dans les médias prouvent que le discours convoqué ne saurait constituer un quelconque horizon idéal du discours rapporté : la matérialité de la voix de I<sub>1</sub> contribue, plus encore que dans le texte écrit, à l'illusion de littéralité et d'étanchéité des discours, mais elle ne modifie pas fondamentalement le statut de l'énoncé enchâssé, que l'énonciation enchâssante intègre à sa perspective propre.

En comparaison, le discours représenté, fût-il direct, fournira un nombre encore moindre d'éléments du contexte, pour trois raisons au moins. D'abord, ils ne peuvent pas être tous mémorisés, selon l'argument de Vincent et Dubois (*supra*). Ensuite, ils ne peuvent pas tous être restitués de manière synthétique par le locuteur enchâssant, ce qui va notamment de soi pour les bruits parasites, ou encore pour les chevauchements. Enfin, tous ne sauraient apparaître comme pertinents, puisque le DR spontané n'a pas, le plus souvent, pour fonction d'informer sur l'énonciation enchâssée en tant que telle, mais de contribuer au dire en cours du locuteur enchâssant. À telle enseigne que, dans un nombre non négligeable de cas, aucun de ces éléments contextuels, pas même les spécificités de la voix du locuteur enchâssé, ne seront ni catégorisés lexicalement, ni stylisés vocalement.

## 2.2. Absence de marquage de la locution enchâssée dans l'interaction orale

L'exemple [2] montre que, lorsqu'il ne converge pas avec d'autres indices, chaque paramètre vocal pris isolément s'avère à peu près ininterprétable :

(2) (Alain Sérane est en train de répondre à la question : *si je te demandai:s le souvenir le plus pénible de ces treize mois le le plus mauvais.*)

ça: bon euh c'-tait pénible / pénible: / eu:h / d'une part psychologiquement pasqu'y avait d'autre:s euh travailleurs y avait d'autres mineurs / aors chacun son niveau / e::t / dans ma tête je disais euh / (3) *fan des purges* (3) *les les autres qui sont* pasqu'on pouvait pas tellement parler entre nous / (3) je disais *fan des purges là les autres qui sont à côté de nous là is vont euh* (3) / y en a que ça va les leur foutre un (2) coup (2) au moral (*Ladrecht*<sup>4</sup>, Sérane II, 65A).

<sup>4</sup> Interviews recueillies dans le bassin minier d'Alès par Jacques Bres et Françoise Madray de 1982 à 1984.



En prolongement des observations de Léon (1988), on constate que le débit aussi bien que l'intensité présentent des à-coups parfois décalés entre eux et qui, surtout, ne coïncident pas avec la distinction entre énoncés enchâssants et segments bivocaux. Jusqu'à la première proposition rectrice de DD (*dans ma tête je disais*), l'intensité est moyenne et le débit lent, avec de multiples pauses et allongements vocaliques. Puis le débit s'accélère fortement sur le premier segment bivocal, l'incidente explicative du locuteur enchâssant et tout le début de la reprise du DD, avant de devenir moyen à partir de *y en a que*. De son côté, l'intensité est très affaiblie sur le juron *fan des purges*, redevient moyenne sur la fin du premier DD et sur l'incidente explicative, avant de se réaffaiblir aussi bien sur la proposition rectrice *je disais* que sur la reprise du segment bivocal, cet affaiblissement englobant les éléments encore non proférés à *côté de nous là is vont*. Mais encore, redevenue moyenne à partir de *y en a que*, l'intensité connaît un pic spectaculaire sur le mot *coup*. On l'entend, seul ce dernier trait paraît réellement cohérent en termes d'emphatisation eu égard au sens produit par l'ensemble de l'extrait.

### 2.3. Discours direct parodique

À l'opposé, il peut arriver, de façon très exceptionnelle, que le locuteur enchâssant fasse entendre, au moyen de quelques traits idiolectaux, une approximation convaincante non seulement de l'expression, mais de la voix du locuteur enchâssé. L'occurrence [3] est d'autant plus remarquable qu'elle a été produite par un député européen et professeur à Sciences Po Paris, en situation de parole publique et en position d'expert de la vie politique française<sup>5</sup> :

(3) c'est là où le discours de vérité / euh qui est nécessaire doit être prononcé et / si grand' que soit mon admiration pour Raymond Barr' je crois qu'i n' suffit pas / de dir' (4) *oui oui oui / i faut dir' la vérité* (4) *aux Français / ça n' suffit pas car la vérité tell' qu'elle est entendue ainsi c'est simplement / l'idée euh Français vous allez en baver* (4) *d'ailleurs je n' suis pas là pour fair' plaisir* (4) *aux gens bon / et on a eu (i faut r'tenir son bilan) on a eu c' discours qui est très sympathiqu' pasque y a y a quand mêm' un' un' un' un' fraîcheur dans l' discours d' Raymond Barr' que nous d'vons saluer* (Jean-Louis Bourlanges, « L'esprit public », *France Culture*, 2 avril 2006).

Aussitôt après la proposition rectrice *i n' suffit pas / de dir'*, L<sub>1</sub> Jean-Louis Bourlanges adopte soudainement une voix de gorge et aiguë à la fois, ainsi que deux traits phonologiques absents du reste de son discours, l'extrême fermeture de presque tous les [i] et la prononciation alvéolaire (à l'anglaise) du [d] de *dire*, puis du [t] de *vérité* : la combinaison impromptue de ces trois paramètres conduit nécessairement les récepteurs d'un certain âge à reconnaître dans ce segment un écho de la voix de Raymond Barre, désigné juste avant comme le locuteur-énonciateur enchâssé de ce DD. Plus encore, le premier paramètre, associé cette fois à la prononciation dorsale (à l'espagnole) du [s] de *suis* et, moins nettement, du [z] de *plaisir*, permet aux destinataires entrés dans la connivence d'attribuer au même L<sub>1</sub> Raymond Barre le DDL *d'ailleurs je n' suis pas là pour fair' plaisir aux gens*.

Il convient de souligner le caractère à la fois partiel et très économique du procédé. Partiel, car la parodie n'utilise ici que deux ou trois traits idiolectalisants (Détrie 2005) et n'inclut pas le dernier syntagme de chaque segment bivocal, *aux Français* et *aux gens*, sur lesquels reparaît la diction ordinaire de Bourlanges. Économique, car la simulation vocale rend possibles de manière très elliptique la discrimination énonciative et la disjonction argumentative du DDL par rapport au DD immédiatement antérieur *c'est simplement / l'idée euh Français vous allez en baver*. Sans entrer dans la complexité du mouvement argumentatif de cette séquence, je dirai que la parodie de la voix de Raymond Barre, contrebalancée par la *fraîcheur* et le caractère *sympathique* que L<sub>1</sub> *salue* avec *admiration*, raillent par l'humour le

<sup>5</sup> Je remercie Catherine Détrie de m'avoir procuré cet exemple.

fonctionnement du discours enchâssé comme vœu pieux qui *ne suffit pas*, et que Bourlanges propose de mettre en actes jusque dans ses ultimes conséquences. Le changement de voix contribue donc au réglage de la distance entre le locuteur enchâssant et l'énoncé enchâssé dont il partage l'esprit mais dont il discute l'application. L'exemple [4] nous fait entendre une fraîcheur plus spontanée :

(4) (Pendant les vacances, A et B sont partis en excursion, tandis que C et D restaient à la maison avec leurs enfants patraques. Le soir, Marianne, notée C, narre sa journée en présence de Lambert, noté D. On entend en fond des éclats de voix d'enfants et/ou de télévision. Pour la commodité de l'analyse, je détache du tour de parole 21, les huit DD sous-numérotés [4a] à [4h].)

19C – (+++) (1) alors j' te finis notre *disput'* (rire) c'est intéressant (1) (++)

20D – ça intéress' peut-être pas tout l' monde hein (rire B, C et D)

21C – si c'est toujours marrant d' voir la merd' chez les autres (rire A, B et C) // et donc euh / (*vous voulez quelque chose non* ↑ +++ D) (1) on a mangé tristement une assiett' de riz (rire C) et puis (1) /

(4a) je dis à Lambert *écout' vraiment ça je je / je me sens pas bien / je vais aller faire un' siest' réveill'-moi dans un' demi-heure / euh j' vais aller m'allonger un peu // et euh j' f'rai la vaissell' j'ui dis j' f'rai la vaissell' je f'rai la vaissell' mais réveill'-moi dans un' demi-heur' / et j'émerge d'un sommeil / péguieux<sup>6</sup> lourd atroc' tu vois d'un seul coup en entendant l' bruit du chauffe-eau / j' me précipite à la cuisine*

(4b) en disant *mais non mais je je non non mais laiss' je vais la fair' la vaissell' /*

(4c) (4) *bien je m'ennuyais tell'ment / que j'avais plus que ça à faire (4) / euh /*

(4d) (4) *ah bon mais euh pasque mais (4) (y avait pas de film à la télé A) il est tard* ↑ *mais il est tard alors* ↑ *c'est pasque /*

(4<sup>e</sup>) (4) *il est six heur's je pense (4) /*

(4f) (4) *mais / mais je / mais qu'est-c' que (4) / (rire A) mais vous avez fait quoi / et les enfants ont fait quoi* ↑ /

(4g) (4) *ils ont regardé tout' la la télé tout' l'après-midi (4) / et alors là j'ai explosé*

(4h) en disant (2) *merd' (2) mais j' t'avais dit d' me réveiller (+++ D) c'est pas la pein' de m' fair' des reproch's* (rire A et B) (*Conil*<sup>7</sup> 12, 19-21).

En première approximation, on perçoit sur toute la durée des énoncés réactifs [4c], [4e] et [4g] la parodie de la voix de L<sub>1</sub> Lambert, dans la mesure où L<sub>1</sub> Marianne reprend le registre grave de son conjoint de manière « absolue », en outrepassant fortement vers le bas sa propre tessiture (Couper-Kuhlen 1996 : 367)<sup>8</sup>. On retrouve, *mutatis mutandis*, les deux fonctionnalités observées en [3] et pointées par Günthner (1999) : d'un côté, le changement de voix accompagne l'absence de propositions rectrices en facilitant l'interprétation de l'alternance des locuteurs enchâssés ; de l'autre, le contrat posé étant que *c'est toujours marrant d' voir la merd' chez les autres*, la parodie confère au récit de la dispute une dimension ludique qui tempère ce que la communication des faits a d'agressif pour la face de Lambert. En revanche, les énoncés initiatifs que Marianne s'impute à elle-même présentent une gradualité du marquage vocal qui pose la question du seuil à partir duquel on doit considérer un segment comme simulé. De fait, si [4a] n'offre à l'oreille aucune spécificité vocale par rapport à la narration enchâssante, [4b] symbolise le déphasage de l'actant réveillé en sursaut avant tout par le tâtonnement désordonné des morphèmes sans suite *mais non mais je je non non*, qu'accompagne une très discrète inflexion aiguë sur les trois *non*, tandis qu'en [4d] et [4f] les tâtonnements *ah bon mais euh pasque mais* et *mais / mais je / mais qu'est-c' que* sont prononcés d'une manière à la fois aiguë et affaiblie qui confère à la voix de la locutrice enchâssée un caractère très enfantin, ce trait d'ingénuité n'affectant pas le reste des deux énoncés. Enfin, l'*explosion* de [4h] est soulignée par une intensité beaucoup plus forte et un accent d'insistance sur *merd'*, ainsi qu'un fort accent sur la dernière syllabe de *réveiller* et

<sup>6</sup> *Poissex* en français méridional.

<sup>7</sup> Échanges familiers que j'ai enregistrés à micro caché en 1998.

<sup>8</sup> Cet article traite de cas plus spécifiques : les reprises immédiates de paroles de l'interlocuteur, à la limite entre DR, écho et parodie.

de *reproch's*, sans qu'on puisse pour autant parler de ton colérique ou de simulation de la dispute.

La série des exemples [4b] à [4h] me paraît ainsi montrer qu'entre les deux extrêmes que constituent l'absence de marquage (exemples [2] et [4a]), d'une part, et la parodie, nécessairement partielle mais parfois frappante (exemple [3]), d'autre part, la majorité des DD repérables à l'oreille comme vocalement marqués, que nous signalons conventionnellement en les encadrant par (4) dans nos transcriptions, s'avèrent des stylisations forcément très fragmentaires, aussi bien par leur durée que par le nombre de traits pertinents qu'ils mettent en jeu (Günthner 2001 : 11 parle de *stylisation prosodique*). Tout comme pour l'expression verbale des énoncés enchâssés, leur intérêt ne réside donc pas dans l'exactitude de la représentation qu'ils donnent de la voix de l'autre locuteur : cet intérêt doit être recherché dans l'interaction des marqueurs vocaux entre eux et avec les autres marqueurs syntaxiques et lexicaux de l'hétérogénéité énonciative montrée. Faute de pouvoir actuellement inventorier ces marqueurs vocaux dans leur multiplicité et dans la subtilité de leurs combinaisons, je voudrais en dernier lieu pointer quelques contraintes à prendre en compte pour mener à bien cet inventaire.

### 3. L'alternance et l'entrelacement des marqueurs vocaux

#### 3.1. Stylisation contrastive du même et de l'autre

Une première contrainte tient, pour l'analyste, au caractère essentiellement contextuel de la stylisation vocale. De fait, le registre ultragrave que la locutrice enchâssante Marianne impute en [4] à la voix enchâssée de son conjoint ne peut plus être considéré comme un trait idiolectalisant du locuteur Lambert dès lors qu'on connaît l'occurrence [5] :

(5) (1) alors j' me suis installée dans l' bus / j'ai croisé mes p'tits bras j'ai fait la moue et ça a duré comm' ça trois heur's / et alors (1) y a la copin' qui v'nait *euh / ça va t'as pas mal au cœ<sup>u</sup>r* ↑ (4) *non non ça va non non non* (4) / (1) ça m'a trop fait chier chier chier chier (1) / (Conil 5, 1C).

Un registre auditivement semblable à celui de [4c], [4e] et [4g] apparaît ici associé à la représentation que Marianne donne de sa propre voix enchâssée dans le tour réactif sans introducteur : *non non ça va non non non*. Ce marqueur vocal appartient donc en réalité à l'idiolecte de Marianne elle-même, qui l'utilise de façon contrastive pour rendre audible l'état émotionnel de renfrognement caractérisant des actants différents dans les situations d'énonciation représentées – ce qui constitue la quatrième fonction assignée par Günthner (1999) au marquage vocal. Je voudrais, pour ma part, souligner le fonctionnement contrastif de ce marquage et, par delà, sa variabilité contextuelle. De fait, dans l'occurrence [5], la voix habituelle de la locutrice Marianne n'est associée ni à sa position de narratrice L<sub>1</sub>, ni à sa position d'actant I<sub>1</sub>, mais au DD (proposition rectrice + segment bivocal) de l'autre locutrice citée, *la copin'*. Par opposition, deux marqueurs différents distinguent, dans ce contexte particulier, les voix propres à Marianne : celui du renfrognement corrélé à sa voix de locutrice enchâssée, et l'inflexion rieuse associée à sa voix de narratrice.

Une seconde contrainte concerne l'interférence des représentations groupales avec les singularités individuelles : dans certains cas, la voix d'un locuteur présentant des traits idio- ou sociolectaux pourtant faciles à approcher apparaît stylisée par un stéréotype auquel elle ne correspond pas. Ainsi en [6], l'interviewé Sérane cite un de ses amis, Recani, dont les enquêteurs connaissent par ailleurs l'accent corse, en remplaçant à deux reprises le phonème [y] par le phonème [u], ce que la transcription note en remplaçant le graphème *u* de *tu* et *tuer* par le graphème *ou* :

(6) (Alain Sérane narre une algarade entre Georges Recani, mineur gréviste, et un notable socialiste.)

alors il a commencé à lui dire: // *eu:h / ah tou veux nous touer toi / espèce de: je sais pas qu'est-c' qu'i lui a dit (Ladrecht, Sérane I, 1, 6A).*

Or, dans sa propre interview, la diction de Recani ne se caractérise pas par cette particularité phonétique, mais par l'alternance de *r* grasseyés et roulés, et par certaines courbes plus chantantes que celles des Français continentaux. En l'absence de [R] dans le segment bivocal, le locuteur enchâssant a-t-il craint d'échouer à produire une courbe chantante ? ou de teinter de comique une narration sérieuse illustrant la virulence des mineurs grévistes envers les responsables socialistes ? Quelle qu'en soit la raison, il a préféré un marquage peu convaincant à l'absence de marquage : la parodie ne vise donc manifestement pas ici à convaincre, mais à démarquer l'énoncé enchâssé en soulignant son caractère hyperbolique, qui met contrastivement en valeur la pondération de Sérane et son rôle de médiateur dans le reste du récit.

### 3.2. Locuteur enchâssant, le retour

Une dernière contrainte a trait au caractère non discret des indices vocaux et au fait qu'ils peuvent fonctionner comme marqueurs sur trois plans : nous avons observé jusqu'ici, presque toujours conjointement, (i) celui de l'interaction représentée, où se situe la stylisation de l'autre locuteur, et (ii) celui de la relation interdiscursive de  $L_1$  à  $l_1$ , où se situe l'évaluation de l'autre discours par le locuteur enchâssant ; or il convient de ne pas omettre (iii) celui de l'interaction en cours, où se situe la relation interlocutive de  $L_1$  à son destinataire actuel  $L_2$ . J'en prendrai deux exemples concernant l'inflexion rieuse qui, dans les occurrences de l'extrait [7], ne paraît pertinente que pour la seule relation interlocutive en cours :

(7) (Dans une location d'été, Marianne, notée C, commente une visite de la propriétaire qui, au vu de son embonpoint abdominal, a cru qu'elle était enceinte. Lors d'une précédente visite, la bailleresse s'était permis des remarques sur l'entretien de la maison.)

1D – mais elle a rien dit cett' fois-ci ↑

2C – non / sauf que j'étais enceinte ha salope

3A – mais elle a p't-êtr' pensé beaucoup / elle a p't-êtr' beaucoup pensé

4C – évidemment

5D – boh moi ell' peut penser c' qu'ell' veut -fin

4C – évidemment dans ton état tu peux pas (1) beaucoup nettoyer la terrasse (1) (rire C et B)

6B – alors là en fait c'était pour tester pourquoi vous fait's rien le ménag' là c'est (1) ça manqu' ça (1) / (Conil 4).

Explicitons d'abord le contexte. Les cinq premiers tours de parole commentent l'indiscrétion de la bailleresse sur un mode sérieux : en 2C, la construction détachée *ha salope* implique que la locutrice a reçu l'argument de sa grossesse supposée comme une grave atteinte à sa face, tandis que *rien dit cett' fois-ci* en 1D et *beaucoup pensé* en 3A font allusion à la sortie précédente sur la propreté douteuse de la maison. La répétition de *penser* en 3A et 5D facilite l'interprétation de la seconde partie du tour 4C comme enchâssant sans régie le discours intérieur supposé de la propriétaire, prolongé en 6B par un DD non canonique qu'introduit *c'était pour tester*. Si nous nous centrons maintenant sur le segment bivocal de 4C, au niveau verbal, le concessif *évidemment* et la litote *tu peux pas beaucoup* (= {tu arrêtes complètement}) soulignent que *nettoyer la terrasse* constitue un prérequis de la ménagère dont le non-respect n'est concevable que si son *état* l'en empêche.

Au niveau paraverbal, l'inflexion rieuse ne saurait donc être pertinente ni pour marquer un quelconque enjouement de la locutrice enchâssée, ni pour indiquer une connivence quelconque de la locutrice enchâssante envers ce discours. Elle ne prend sens que du plaisir de se dédommager de ce quiproquo désagréable en le mettant en scène pour / avec ses interlocuteurs du moment, et en faisant jouer l'un contre l'autre les deux stéréotypes machistes prêtés à la propriétaire andalouse : le sociotype de la parfaite ménagère espagnole

dont le logis est nécessairement toujours impeccable, et le préconstruit *dans ton état*, assimilant la grosseur à une maladie. La locutrice B réagit positivement à ce persiflage ludique, puisqu'elle imagine un autre discours intérieur de la propriétaire dont elle affecte semblablement la dernière proposition, *ça manqu' ça*, d'une inflexion rieuse qui n'a de pertinence qu'au niveau de l'interlocution enchâssante (pour des cas similaires, cf. Günthner 2001 : 10, Auchlin *et al.* 2004 : 232-233).

De telles occurrences rendent audible la bivocalité locutoire du DD que pointait déjà Léon (1988). Elles tendent à prouver que, contrairement à l'illusion d'étanchéité construite par les objectivations typographiques et par la tradition didactique, l'énoncé enchâssant ne s'interrompt pas pour laisser place nette à l'énoncé enchâssé : en contrepoint de la rupture énonciative qu'il marque le plus souvent, le DD fait interagir E et e de manière continue, mais non nécessairement explicite, à l'instar des autres structures dialogiques. À l'oral comme à l'écrit et comme dans les occurrences instrumentées par les technologies audiovisuelles, sa spécificité consiste à exhiber (plus ou moins fortement) un pseudo-effacement de E au profit de e, alors que les autres formes de dialogisme maintiennent la primauté de l'énoncé enchâssant, au point de rendre parfois problématique la récupération de l'énoncé enchâssé.

Ces quelques configurations sont évidemment très loin d'épuiser la diversité des combinaisons possibles. Elles me semblent cependant montrer que, sauf imitateur au sens fort du terme, ce n'est presque jamais la globalité de la voix du locuteur enchâssé que l'on entend dans le discours rapporté direct à l'oral mais, et dans certains cas seulement, un ou quelques traits plus ou moins suggestifs de cette voix : je proposerai donc de répondre à la question initiale de Jeanne-Marie Barbéris que le concept de voix n'est pas une métaphore, mais une synecdoque généralisante du tout pour la partie, une vaste enquête statistique demeurant nécessaire pour préciser les proportions effectives et les contextes d'apparition du marquage vocal.

Plus largement, je démarquerai ma conclusion de celle de Günthner (1992 : 230) qui pose successivement que « contextualization cues have no decontextualized meaning of the kind we find in lexical entries » et que « instead of trying to find context-free meanings of contextualization cues we have to concentrate on the ongoing interpretation of interactional activities of which such contextualization cues are part ». Je reprendrai telle quelle la seconde assertion puisque, à l'exception très relative de l'exemple [3], toutes les occurrences considérées ici me paraissent prouver que les marqueurs vocaux prennent sens avant tout de leur interaction contextuelle entre eux et avec les autres marqueurs syntaxiques et lexicaux de l'hétérogénéité énonciative montrée. Quant à la première assertion, je dirai en revanche que, puisque la praxématique ne reconnaît pas de signification décontextualisée aux entrées lexicales, après avoir progressé depuis trente ans sur la question du réglage du sens, il me semble que nous devons, aujourd'hui, ouvrir l'immense chantier du réglage du son dans l'interaction verbale afin d'étudier par quels processus les marqueurs vocaux contribuent, selon les mots de Siblot (2001), à la « mise en relation de la représentation du monde en langage avec le réel ».

## Références

- Andersen H. L., 2000, « Discours rapporté en français parlé : rection du verbe de citation et éléments délimitant la citation directe », in Andersen H. L. & A. B. Hansen (éd.), *Le Français parlé : corpus et résultats*, Copenhague : Museum Tusulanum Press, 143-155.
- Auchlin A. *et al.*, 2004, « (Én)action, expérientiation du discours et prosodie », *Cahiers de linguistique française* 26, 217-249.
- Bertrand R. & R. Espesser, 1998, « Prosodie et discours rapporté : la mise en scène des voix », in Verschueren J. (éd.), *Pragmatics in 1998 : Selected Papers from the 6th International Pragmatics Conference*, vol. 2, Anvers : International Pragmatics Association, 45-56.

- Bertrand R. & R. Espesser, 2002, « Voice diversity in conversation : a case study », in Bel B. & I. Marlien (éd.), *Proceedings of SP2002*, Aix-en-Provence : Laboratoire Parole et Langage, 171-174.
- Bres J. & B. Verine, 2002, « Le bruissement des voix dans le discours : dialogisme et discours rapporté », *Faits de langues* 19, 159-169.
- Couper-Kuhlen E., 1996, « The prosody of repetition : on quoting and mimicry », in Couper-Kuhlen E. & M. Selting, *Prosody in Conversation*, Cambridge : C.U.P., 366-405.
- Détrie C., 2005, « La dynamique idiolectalisante, entre singularisation et répétition », *Cahiers de praxématique* 44, 51-75.
- Fauré L. & B. Verine, 2004, « Authentifier un discours autre en y mettant du sien : les vocalisations *ah* et *oh* en frontière de discours rapporté direct à l'oral », in López Muñoz J. M., S. Marnette & L. Rosier (éd.), 317-327.
- Fernandez M., 2001, « Un agresseur invisible, des victimes innocentes », in Charaudeau P. et al., *La Télévision et la guerre*, Bruxelles : De Boeck, 71-80.
- Günthner S., 1992, « Comments on : Klaus Müller : theatrical moments », in Auer P. & A. Di Luzio (éd.), *The Contextualization of Language*, Amsterdam / Philadelphie : John Benjamins, 223-230.
- Günthner S., 1999, « Polyphony and the " layering of voices " in reported dialogues : an analysis of the use of prosodic devices in everyday reported speech », *Journal of pragmatics* 31, 685-708.
- Günthner S., 2001, « Constructing scenic moments : grammatical and rhetoric-stylistic devices for staging past events in everyday narratives », *Interaction and linguistic structures* 22, 3-22.
- Klewitz G. & E. Couper-Kuhlen, 1999, « Quote – unquote ? The role of prosody in the contextualization of reported speech sequences », *Interaction and linguistic structures* 12, 4-33.
- Léon J., 1988, « Formes de discours direct dans des récits oraux », *Linx* 18, 107-124.
- López Muñoz J. M., S. Marnette & L. Rosier (éd.), 2004, *Le discours rapporté dans tous ses états*, Paris : L'Harmattan.
- López Muñoz J. M., S. Marnette & L. Rosier (éd.), 2005, *Dans la jungle des discours : genres de discours et discours rapporté*, Cadix : Servicio de Publicaciones de la Universidad.
- Morel M.-A., 1998, « Le discours rapporté direct », in Morel M.-A. & L. Danon-Boileau, *Grammaire de l'intonation : l'exemple du français*, Gap ; Paris : Ophrys, 130-134.
- Müller K., 1992, « Theatrical moments : on contextualizing funny and dramatic moods in the course of telling a story in conversation », in Auer P. & A. Di Luzio (éd.), *The Contextualization of Language*, Amsterdam / Philadelphie : John Benjamins, 199-221.
- Roulet E., 2001, « L'organisation énonciative et l'organisation polyphonique », in Roulet E., L. Fillietaz & A. Grobet, *Un modèle et un instrument d'analyse de l'organisation du discours*, Berne : Peter Lang, 277-306.
- Siblot P., 2001, « Réglage du sens », in Détrie C., P. Siblot & B. Verine (éd.), *Termes et concepts pour l'analyse du discours. Une approche praxématique*, Paris : Champion, 295.
- Verine B., 2005, « Comment les interjections vocaliques modalisent l'insertion argumentative des discours rapportés directs non-véridiques à l'oral », in Lopez-Muñoz J.-M., S. Marnette & L. Rosier (éd.), 497-506.
- Verine B., 2006, « Marquages et fonctionnements dialogiques du discours intérieur autophonique de forme directe à l'oral », *Travaux de linguistique* 52, 57-70.
- Verine B., 2008, « Une variante contextuelle du discours rapporté avec mention : le discours convoqué dans le journal et le magazine d'information radiophoniques », communication au colloque *Représentations du sens linguistique IV*, Helsinki.
- Vincent D. & S. Dubois, 1997, *Le Discours rapporté au quotidien*, Québec : Nuit blanche.