

Identification des corrélats visibles de la focalisation contrastive en français

Marion Dohen et Hélène Lævenbruck

Institut de la Communication Parlée, UMR CNRS 5009 / INPG / Université Stendhal
46 avenue Félix Viallet – 38031 Grenoble Cedex 1, France
Tél.: ++33 (0)4 76 57 45 34 - Fax: ++33 (0)4 76 57 47 10
Mél: {dohen ; loeven}@icp.inpg.fr - http://www.icp.inpg.fr

ABSTRACT

This study aims at determining whether there are visual cues to contrastive focus in French. An audiovisual corpus was recorded from a male native speaker of French consisting of sentences with a subject-verb-object (SVO) structure. Four conditions were recorded: contrastive focus on each phrase (SVO) and broad focus. It was first checked whether the utterances displayed a typical focused intonation. Then the lip area and jaw opening were extracted from the video. The analysis of the data enabled us to extract a set of visible correlates of contrastive focus in French: a) increase in lip area and jaw opening on the focused item b) lengthening of the prefocal syllable and of the focal syllables (particularly on the first phoneme of the focused phrase) c) hypo-articulation of the post focus sequence. Thus there are visual cues to contrastive focus that may be used in communication.

1. INTRODUCTION

1.1. Objectif

La focalisation contrastive permet de mettre en relief un mot ou un groupe de mots dans un énoncé par opposition à un autre. En français, elle peut être syntaxique (« C'est x qui a mangé la pomme. ») ou prosodique (« X_F a mangé la pomme. »). Nous étudierons ici la focalisation contrastive prosodique en français.

Jusqu'à maintenant les études ont surtout mis en évidence les corrélats intonatifs et temporels de la focalisation contrastive en français [1-5]. Ces corrélats interviennent sans doute dans la perception auditive de la focalisation. Cependant d'un point de vue audiovisuel, d'autres corrélats semblent importants. En effet, des études ont montré que les variations de F0 pouvaient avoir des conséquences articulatoires potentiellement visibles, telles qu'une plus grande ouverture de la bouche [6-9]. Dans cette étude nous allons tenter d'identifier précisément différents paramètres visibles possibles du bas du visage pour la focalisation.

1.2. Contexte

Nous avons utilisé ici le modèle de Jun & Fougeron [3,4] qui est en accord avec la plupart des descriptions de l'intonation du français et utilise un système de transcription qui a montré son utilité pour de nombreuses

langues. Ce modèle comporte 2 unités hiérarchiques. La plus basse est le Syntagme Accentuel (SA) marqué à droite par l'accent primaire (H*, ton haut) et parfois marqué par une séquence initiale : l'accent secondaire LHi (Low-High initial, succession d'un ton bas et d'un ton haut). La plus élevée est le Syntagme Intonatif (SI) qui peut moduler le SA (e.g. si un SA est à la fin d'un SI, H* est remplacé par le ton de frontière du SI : L% ou H% comme c'est le cas pour le dernier SA de la figure 1a.).

Dans ce modèle, la focalisation contrastive est marquée par un ton Hf élevé suivi d'un plateau tonal bas : voir figure 1b. Hf remplace souvent Hi (figure 1b) mais il peut aussi remplacer à la fois Hi et H*.

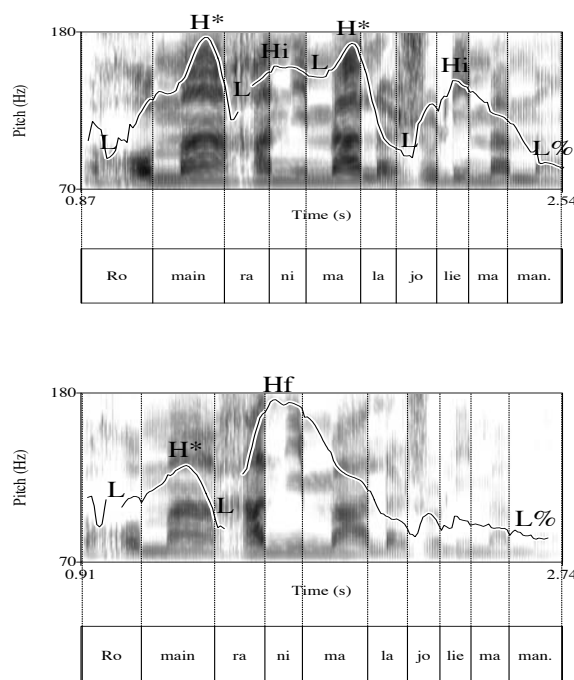


Figure 1 : Spectrogramme et suivi de F0 pour un SI comprenant 3 SA. a) (haut) cas de focus large. b) (bas) focus sur le SA verbal. L'énoncé est {[Romain]_{SA}[ranima]_{SA}[la jolie maman]_{SA}}_{SI}.

2. METHODE EXPERIMENTALE

2.1. Corpus et enregistrement audiovisuel

Le corpus est composé de 8 phrases ayant une structure sujet-verbe-objet (SVO), des syllabes CV (Consonne-Voyelle) et des sonores pour faciliter la détection

automatique de F0. Les phrases ont été conçues pour que l'énoncé soit produit en un seul SI composé de trois SA. Le patron tonal par défaut est donc dans le cas neutre : { [LHiLH*]_S [LHiLH*]_V [LHiLL%]_O } ([3]). Ci-dessous, les huit phrases composant le corpus (les nombres à côté de S, V et O correspondent au nombre de syllabes de chaque groupe) :

- (1) [Jean]_{S1} [veut ménager]_{V4} [nos jolis nouveaux navets]_{O7}.
- (2) [Romain]_{S2} [ranima]_{V3} [la jolie maman]_{O5}.
- (3) [Mélanie]_{S3} [vit]_{V1} [les mauvais loups malheureux]_{O7}.
- (4) [Véronique]_{S4} [mangeait]_{V2} [les mauvais melons]_{O5}.
- (5) [Les mauvais loups]_{S4} [mangeront]_{V3} [Jean]_{O1}.
- (6) [Mon mari]_{S3} [veut ranimer]_{V4} [Romain]_{O2}.
- (7) [Les loups]_{S2} [suivaient]_{V2} [Marilou]_{O3}.
- (8) [Le beau marin]_{S4} [vit]_{V1} [Véronique]_{O4}.

C'est un locuteur francophone masculin qui a enregistré ce corpus. Quatre modes ont été enregistrés : focalisation contrastive étroite sur S, V et O et focalisation large (version « neutre » de l'énoncé). La focalisation a été produite grâce à une tâche de correction comme dans l'exemple ci-dessous (majuscules : focalisation) :

Le locuteur lit : Les loups suivaient Marilou.

Prompt audio : Les loups suivaient Aurélie ?

Le locuteur dit : Les loups suivaient MARILOU.

Le locuteur n'a reçu aucune consigne sur la manière de produire la focalisation (i.e. quelle(s) syllabe(s) devaient être accentuée(s) par exemple). Quatre types de parole ont été enregistrés : parole lexicalisée, délexicalisée, chuchotée et délexicalisée chuchotée. Cependant deux modes seront étudiés ici : délexicalisé et lexicalisé. La parole délexicalisée consiste à remplacer toutes les syllabes des mots par une seule syllabe (ici /ma/). On sait ainsi que les différences mesurées entre les syllabes seront dues uniquement à des variations prosodiques puisque les paramètres articulatoires inhérents sont les mêmes. Ainsi si l'on n'avait mesuré aucune différence liée à la prosodie sur les paramètres visibles, on aurait su qu'il n'y en avait sûrement pas non plus en parole lexicalisée. D'un autre côté, s'il existe des paramètres prosodiques visibles on aura une idée plus précise de ce qu'ils sont et on pourra étudier plus précisément la parole lexicalisée. En tout, 256 énoncés ont été enregistrés (8 phrases, 4 focalisations, 4 types de parole et deux répétitions pour chaque énoncé).

2.2. Validation acoustique

Le but est ici de vérifier que les indices acoustiques des énoncés focalisés sont conformes à ceux décrits dans les études précédentes. Or le maximum de F0 pour chaque énoncé était bien sur une syllabe focalisée et quand ce n'était pas le cas, nous avons vérifié que cela était dû à la déclinaison. En effet, un syntagme objet focalisé (fin d'énoncé) peut présenter des pics de F0 d'amplitudes égales (ou plus faibles) que ceux du syntagme sujet. Néanmoins, les auditeurs compensent ce phénomène lors de la perception [10]. La première syllabe du premier mot de contenu du syntagme focalisé était bien marquée par un accent Hf [3]. En tenant compte de la déclinaison, F0 était toujours plus important sur le syntagme focalisé. Les

énoncés produits comportaient donc bien les indices acoustiques décrits dans d'autres études : [1,2,3,4,5].

2.3. Techniques de mesure

La figure 2 donne un exemple des images enregistrées et des mesures faites. Un logiciel conçu à l'Institut de la Communication Parlée (ICP) ([11,12]) permet d'extraire automatiquement le contour des lèvres et les coordonnées de certains points du visage grâce à un maquillage bleu. On obtient ainsi les données décrivant la forme et la protrusion des lèvres, ainsi que la position de la mâchoire.

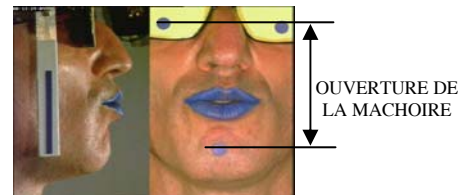


Figure 2 : Signal vidéo : mesure des paramètres.

3. ETUDE PRELIMINAIRE : PAROLE DELEXICALISEE

Une étude préliminaire a été menée sur la parole délexicalisée [13] afin de voir s'il y avait des paramètres visibles de la focalisation contrastive d'une part et d'avoir une idée de quels étaient ces paramètres d'autre part. Cette étude nous a permis de conclure que la focalisation contrastive s'accompagnait d'une **ouverture de la mâchoire plus grande et plus rapide** sur les syllabes focalisées. De plus, nous avons mesuré une **fermeture initiale des lèvres significativement plus longue** au début du SA focalisé ainsi qu'un **allongement pré-focal** et une **hypo-articulation post-focale**.

4. ANALYSE DES CORRELATS VISIBLES POTENTIELS DU FOCUS CONTRASTIF POUR LA PAROLE LEXICALISEE

4.1. Analyse préliminaire du problème

Les corrélats visuels peuvent être de deux types : mouvements articulatoires et variations de durée. Dans l'étude préliminaire ce sont la hauteur des lèvres et l'ouverture de la mâchoire qui variaient le plus du cas de focalisation large au cas de focalisation étroite pour la syllabe en question ([ma]). Pour la parole lexicalisée ces paramètres doivent être adaptés. Il faut en effet identifier des corrélats qui varient significativement pour toutes les configurations articulatoires présentes dans le corpus. Les paramètres articulatoires varieront en effet non seulement à cause des variations prosodiques mais aussi à cause des différences entre syllabes.

Corrélats articulatoires potentiels

Le but est ici de déterminer les paramètres articulatoires dont les variations seront les plus significatives quelle que soit la syllabe. Notre étude préliminaire ([13]) avait montré que la focalisation contrastive induisait une hyper

articulation. Or l'hyper-articulation peut être réalisée de manières différentes : pour un /a/ l'ouverture des lèvres et de la mâchoire seront plus grandes mais pour un /i/ ce sera plutôt la largeur des lèvres qui augmentera et pour un /u/ la protrusion. Notre corpus ne présentant que très peu de cas où la protrusion serait le paramètre le plus affecté,

nous ne l'avons pas étudié ici. Nous avons donc étudié : la hauteur des lèvres (HL), la largeur des lèvres (LL) et l'ouverture de la mâchoire (OM). Le paramètre d'aire aux lèvres (AL) tient compte à la fois des variations de HL et LL. Nous étudierons donc OM et AL.

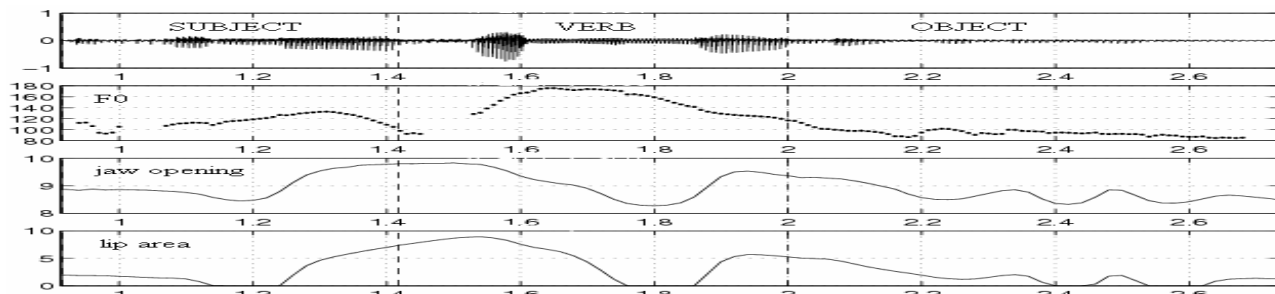


Figure 4 : Signal acoustique, ouverture de la mandibule (cm) et aire aux lèvres (cm²) en fonction du temps (s). L'énoncé était [Romain]_s [RANIMA]_v[la jolie maman]_o. Le constituant focalisé est délimité par les traits en pointillés.

Corrélat de durée potentiels

Nous mesurerons les allongements focal et pré-focal potentiels. L'allongement significatif de la fermeture initiale des lèvres pour le premier segment du syntagme focalisé mesuré dans l'étude préliminaire n'était peut-être qu'un artefact de la syllabe choisie. En effet, pour la parole lexicalisée on n'observera pas forcément une fermeture initiale des lèvres (comme pour le /m/). Pour vérifier ceci, nous avons donc mesuré la durée du premier phonème du syntagme focalisé qu'on nommera ci-après « durée du premier segment ».

4.2.Résultats

Corrélat articulaires

Nous avons calculé pour chaque paramètre les maxima sur chaque syllabe puis les pourcentages d'augmentation par rapport au cas de focalisation large. Enfin nous avons calculé la moyenne des pourcentages d'augmentation sur tous les groupes de toutes les phrases.

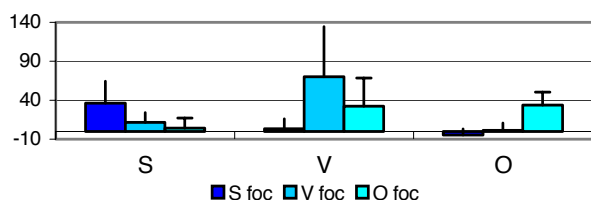


Figure 3 : Moyenne des moyennes des pourcentages d'augmentation de AL (cm²) pour S, V et O.

Nous avons remarqué qu'en moyenne l'aire aux lèvres augmentait de 36,4% lors de la focalisation étroite. Cette valeur peut sembler peu importante mais compte tenu de la diversité des voyelles du corpus et donc de la diversité des stratégies mises en œuvre pour focaliser, cette valeur est tout à fait intéressante. La Figure 3 présente la moyenne des pourcentages moyens

d'augmentation de AL pour chaque type de groupe syntaxique (S, V ou O). Des test ANOVA permettent de rejeter dans les trois cas de focalisation étroite l'hypothèse d'égalité des trois moyennes (S : F(2,42) = 12,21 p<0,01, V : F(2,42)= 9,48 p<0,01 O : F(2,42) = 50,19 p<0,01). On constate de plus (tests de Student) que les moyennes des augmentations moyennes de S, V et O lors de la focalisation étroite sont significativement supérieures à 0 (cas de focalisation large). On mesure donc une augmentation significative de l'aire aux lèvres lorsqu'il y a focalisation contrastive étroite.

L'augmentation moyenne de l'ouverture de la mâchoire du cas de focalisation large vers la focalisation étroite est de 51.72%. Nous avons observé que les syntagmes après la focalisation avaient une aire aux lèvres et une ouverture de la mâchoire réduites par rapport au cas de focalisation large : **hypo-articulation post-focale**.

Corrélat de durée

Nous avons mesuré la durée moyenne de chaque syllabe focalisée et l'avons comparée à la durée moyenne de la même syllabe dans le cas de la focalisation large. Nous avons trouvé que la durée moyenne des syllabes focalisées était significativement plus grande (test epsilon p < 0,01) que dans le cas de la focalisation large. L'allongement focal moyen est de 33,71%. Nous avons également comparé la durée de la syllabe pré-focale avec la durée de cette même syllabe dans le cas de la focalisation large. Nous avons trouvé un allongement pré-focal significatif de 19,63% (test epsilon p<0,01). Nous avons aussi mesuré la durée du premier segment du syntagme focalisé et l'avons comparée à la durée de ce même segment dans le cas de la focalisation large. Nous avons trouvé un allongement du premier segment focalisé significatif de 59% (test epsilon p< 0,01). Cet allongement est donc plus important que l'allongement focal (33,71%).

Bilan des corrélats visibles potentiels de la focalisation contrastive en français

La focalisation contrastive en français est caractérisée par une augmentation significative de l'aire aux lèvres des syllabes focalisées et des durées de la syllabe pré-focale, des syllabes focalisées et, encore plus significativement, du 1^{er} segment du syntagme focalisé. La séquence post-focale est hypo-articulée (figure 4).

5. CONCLUSION

Une étude préliminaire [13] sur la parole délexicalisée avait montré que la focalisation impliquait a) une amplitude et une vitesse d'ouverture de la mâchoire plus grandes b) une plus longue fermeture initiale des lèvres c) un allongement des syllabes pré focale et focales d) une hypo-articulation de la séquence post-focale. Pour la parole lexicalisée nous avons mesuré des augmentations significatives de l'aire aux lèvres et de l'ouverture de la mâchoire lors de la focalisation étroite. Ces résultats sont en accord avec ceux obtenus pour la parole lexicalisée. Pour les /ma/ l'hyper-articulation consistait en une plus grande ouverture de la mâchoire mais pour la parole lexicalisée, on pouvait soit observer une augmentation de l'ouverture des lèvres et de la mâchoire, soit une augmentation de l'étirement des lèvres. Or les variations de l'aire aux lèvres comportent à la fois ces deux types de variations. Nous avons également mesuré un allongement de la syllabe pré-focale et des syllabes focales et une augmentation très significative de la durée du premier phonème du syntagme focalisé. Ce corrélat semble correspondre à la longue fermeture des lèvres pour la syllabe initiale du syntagme focalisé. L'étude décrite dans [14] montre que les corrélats trouvés pour la parole délexicalisée sont très bien perçus (86% de bonnes réponses en perception visuelle seule pour un niveau de hasard de 25%). Une étude perceptive a également été menée pour la parole lexicalisée (71,45% de bonnes réponses) et a fait l'objet d'une autre soumission à cette conférence [15].

6. REMERCIEMENTS

Nous remercions M.-A. Cathiard et J.-L. Schwartz pour leurs conseils scientifiques, G. Rolland pour l'enregistrement du corpus, C. Savariaux et A. Arnal pour leur aide technique, P. Welby pour ses commentaires et les participants aux tests perceptifs.

BIBLIOGRAPHIE

- [1] A. Clech-Darbon, G. Rebuschi & A. Rialland. Are there Cleft Sentences in French?. In *The Grammar of Focus*. Tuller L. & Rebuschi G. (Eds), Amsterdam: Benjamins, 83-118, 1999.
- [2] A. Di Cristo. Intonation in French. In *Intonation systems: a survey of twenty languages*, Hirst D. & Di Cristo A. (Eds.). Cambridge University Press, 195-218, 1998.
- [3] S.-A. Jun, C. Fougeron. A Phonological Model of French Intonation. In *Intonation: Analysis, modelling and technology*, A. Botinis (Ed.). Dordrecht: KAP, 209-242, 2000.
- [4] S.-A. Jun, C. Fougeron. Realizations of Accentual Phrases in French Intonation. *Probus* 14, 147-172, 2002 .
- [5] P. Touati. Structures prosodiques du suédois et du français. *Working Paper 21*. Lund University Press, 1987.
- [6] H. Loevenbruck. An Investigation of Articulatory Correlates of the Accentual Phrase in French. *Proceedings of ICPhS'99*, San Francisco, 1, 667-670, 1999.
- [7] M. Tabain. Effects of prosodic boundary on /aC/ sequences: articulatory results. *JASA*, 113(5) : 2834-2849, 2003.
- [8] K. De Jong. The supraglottal articulation of prominence in English: linguistic stress as localized hyper-articulation. *Journal of the Acoustical Society of America*, 97 : 491-504, 1995.
- [9] P. Keating, M. Baroni, S. Mattys, R. Scarborough, A. Alwan, E.T. Auer & L.E. Bernstein. Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English. *Proceedings of 15th ICPhS*, Barcelona, 2071-2074, 2003.
- [10] M. Liberman & J. Pierrehumbert. Intonational invariance under changes in pitch range and length. In *Language sound to structure: studies in phonology presented to Morris Halle by his teacher and students*. Aronoff M. & Oehrle R. (eds.). MIT Press, 157-233, 1984.
- [11] M.-T. Lallouache. Un poste Visage-Parole couleur. Acquisition et traitement automatique des contours de lèvres. PhD Thesis, Institut National Polytechnique de Grenoble, 1991.
- [12] M. Audouy. *Traitement d'images vidéo pour la capture des mouvements labiaux*. Final engineering report, Institut National Polytechnique de Grenoble, 2000.
- [13] M. Dohen, H. Loevenbruck, M.A. Cathiard, J.L. Schwartz. Potential Audiovisual Correlates of Contrastive Focus in French. *Proceedings of Eurospeech'03*. Geneva, 145-148, 2003.
- [14] M. Dohen, H. Loevenbruck, M.A. Cathiard, J.L. Schwartz. Audivisual Perception of Contrastive Focus in French. *Proceedings of the AVSP'03 Conference*. St Jorioz, France, 245-250, 2003.
- [15] M. Dohen & H. Loevenbruck. (soumis). La focalisation contrastive est-elle visible? Une étude perceptive visuelle de la focalisation contrastive en français. Actes des XXVèmes JEP, avril 2004.