

Marquage acoustique du focus contrastif non codé syntaxiquement en français

Corine Astésano¹, Cyrille Magne¹, Michel Morel², Annelise Coquillon³, Robert Espesser³,
Mireille Besson¹, Anne Lacheret-Dujour²

¹Institut de Neurosciences Cognitives de la Méditerranée, CNRS, 31 Ch. J. Aiguier, 13402 Marseille Cedex 20, France
Tél. : ++33 (0)4 91 16 43 94 – Fax : ++33 (0)4 91 77 49 69

Mél : {astesano ; cmagne ; mbesson}@lnf.cnrs-mrs.fr

²CRISCO, Université de Caen, place de la Paix, 14032 Caen cedex, France

Tél. : ++33 (0)2 31 56 56 27 – Fax : ++33 (0)2 31 56 54 27

Mél : {anne.lacheret ; michel.morel}@crisco.unicaen.fr - <http://www.crisco.unicaen.fr>

³Laboratoire Parole & Langage, Université de Provence, 29 Av. R. Schuman, 13621 Aix-en-Provence Cedex 1, France

Tél. : ++33 (0)4 42 95 36 19 – Fax : ++33 (0)4 42 59 50 96

Mél : {annelise.coquillon ; robert.espesser}@lpl.univ-aix.fr

ABSTRACT

The present work is dedicated to contrastive focus in French. It is part of a collaboration between the INCM in Marseille, the CRISCO laboratory in Caen and the LPL in Aix. The present paper proposes an analysis of the acoustic realisation of contrastive focus in French, where both syllabic- and word-level features are investigated with regards to focus position (middle or end) in the sentence. This phonetic work is aimed at providing a linguistic interpretation for the results of a perceptual experiment on the same material, which tackles the issue of on-line processing of pragmatic focus, using the Event Related Potentials (ERP) method (Magne et al, these proceedings).

1. INTRODUCTION

Notre travail s'inscrit dans le cadre d'un projet TTT (Terrains, Techniques et Théories, S. Robert (dir.)) financé par le ministère de la recherche, et centré sur la thématique de la complexité linguistique. Une des questions majeures abordée dans ce projet est celle de l'interaction entre les différents niveaux de traitement du langage. Dans ce contexte, notre objectif étant de mieux comprendre l'effet des contraintes pragmatiques liées à la matérialisation linguistique de la structure informationnelle dans le message, nous avons choisi d'amorcer ce travail par l'étude du marquage prosodique du focus contrastif en français en utilisant les méthodes comportementale et électrophysiologique (PEs). Sous l'angle neurophysiologique, l'expérience a été conduite dans l'optique de déterminer si un indice électrophysiologique est associé au traitement en temps réel du focus contrastif (Magne et al, ces actes, [1]). En aval de ce travail, une analyse acoustique des patrons accentuels à valeur focale s'avérait indispensable pour, d'une part, éclairer les résultats PEs, d'autre part, enrichir nos connaissances sur les corrélats phonétiques de l'accent focal en français. C'est ce second point qui fait l'objet de notre communication. Un rappel définitoire de la notion de focus sous l'angle fonctionnel et phonétique, nous permet de préciser nos hypothèses sur les marqueurs

acoustiques de l'accent focal. Nous présentons ensuite le matériel expérimental développé pour l'étude. Enfin, une analyse acoustique exhaustive du corpus nous permet de faire le point sur nos hypothèses initiales.

2. ELÉMENTS DE DÉFINITION

Nous envisageons la notion de focus à la fois sous l'angle informationnel (pour être considéré comme focus, un élément doit avant tout être porteur d'informations), et du point de vue de son marquage phonétique.

2.1. Définition fonctionnelle du focus contrastif

La notion d'information nouvelle telle que nous la concevons ici utilise le cadre théorique proposé par Lambrecht [2]. Parmi les différents types de focus recensés dans la littérature, nous avons centré notre étude sur le focus contrastif. Un tel type d'opération, fondé sur la sélection exclusive d'une unité dans une classe paradigmatique posée (explicitement ou non), est la plupart du temps utilisé comme mécanisme d'identification référentielle (ex. *C'est les cigognes qui partent en Afrique en hiver*) ou circonstancielle (ex. *c'est en hiver que les cigognes partent en Afrique*). Concernant le domaine de la projection focale, les unités manipulées ici sont des focus étroits : l'accent frappe toujours un et un seul item lexical, en en faisant ainsi un îlot de saillance perceptive.

2.2. Les marqueurs phonétiques du focus

Sous l'angle phonétique, la focalisation d'une unité se manifeste par un accent d'intensité et de hauteur mélodique sur la première syllabe à attaque consonantique de l'unité à focaliser [3, 4], parfois combinés à l'utilisation du trait de durée. Le focus contrastif plus précisément serait caractérisé par un patron intonatif montant/descendant sur le mot, dont le pic serait généralement aligné sur la première syllabe [5]. Dans ce cas, le focus est marqué également par un allongement significatif de l'attaque consonantique [6, 7]. De plus, une augmentation de la durée globale du mot est également

observée [8]. Cette rapide description des caractéristiques acoustiques du focus contrastif indique une tendance à un marquage global du mot lexical focalisé. Ces tendances ont également été observées dans le cas du focus d'intensification dans des travaux récents en psycho-acoustique, montrant que le marquage focal était souvent déclencheur de stratégies holistiques entraînant une modification globale du contour intonatif porté par l'unité focalisée [9]. Enfin, l'opération de focalisation se construit également autour de contrastes syntagmatiques faisant alterner figures et fond. En conséquence le marquage accentuel d'un focus va de pair avec une reconfiguration prosodique des éléments environnants : réduction de l'intensité autour du focus [10], aplatissement des contours mélodiques et disparition des accents de pitch en position post-focale [11, 12], et également réduction des pics accentuels en position pré-focale [13]. Notre étude a donc pour but de tester empiriquement d'une part, l'hypothèse holistique du marquage du focus contrastif, et d'autre part, l'hypothèse selon laquelle la figure de focus est mise en valeur par les caractéristiques prosodiques des éléments environnants.

3. MATÉRIEL ET PROCÉDURE

Notre objectif dans cette étude était d'évaluer la participation relative des 3 paramètres prosodiques (fréquence fondamentale F0, Intensité I, Durée D) et leurs points de localisation dans le marquage focal. Par ailleurs, il s'agissait de conduire notre observation sur des données aussi écologiques que possible. C'est pourquoi nous nous sommes fixés un certain nombre de contraintes pour la mise en place du matériel linguistique et les modalités d'enregistrement.

3.1. Sélection linguistique du matériel

Le matériel est constitué d'une série de couples question-réponse, la réponse venant saturer un paradigme informationnel posé dans la question. La moitié des questions induisait la réalisation d'un accent focal sur un mot en position médiane de la phrase-réponse (ex. *Où vas-tu à vélo, au village ou à la mer? Je vais au **village** à vélo*), l'autre moitié induisant la réalisation d'un accent focal en position terminale de la phrase (ex. *comment vas-tu au village, à vélo ou en bus? Je vais au village à **vélo***). Ainsi, la même phrase-réponse est réalisée avec deux types de focus différents, en fonction du contexte discursif introduit par la question. Sous l'angle cognitif, ce corpus nous permet de tester, dans le cadre d'une étude électrophysiologique, l'effet perceptif d'une violation d'accent focal dans différentes positions de la chaîne et de le comparer à l'effet produit par un focus prosodique congruent dans les mêmes positions (cf. [1], pour une description détaillée de cette étude). Sous l'angle linguistique, l'intérêt de ce corpus consiste à analyser, dans les phrases-réponses, la réalisation acoustique du focus contrastif en français en fonction du contexte prosodique (effet de position du mot focalisé dans la phrase).

Dans la suite du texte, nous désignerons les mots-cibles focalisés en fonction de leur distribution (mot en position médiane ou terminale de phrase) par MM et MT respectivement. Notons enfin que les mots-cibles sont soit bisyllabiques (pour la majorité, 75%), soit trisyllabiques (25%).

3.2. Modalités d'enregistrement

240 couples question-réponse ont été construits. De manière à introduire de la variation, les questions ont été posées par deux locuteurs hommes et un enfant. En revanche, les 240 réponses (120 réponses où le focus porte sur MM, et 120 où il porte sur MT) ont été produites par la même locutrice afin de produire un corpus perceptivement homogène. Les enregistrements ont été réalisés en chambre sourde avec microphone AP-850 Galaxy et carte-son SoundBlaster PCI 128, à 22050 Hz et 16 bits. Le logiciel d'enregistrement et de traitement du signal Kparole développé dans le cadre du projet de synthèse vocale Kali [14] a été utilisé pour l'enregistrement et le nettoyage des signaux. Les stimuli ont été validés par 4 sujets en fonction de 3 critères : qualité de l'enregistrement, stimulus linguistiquement plausible, patron intonatif acceptable.

3.3. Analyses acoustiques

Les extraits ont été numérisés et déséchantillonnés à 16 kHz. La F0 a été modélisée par une courbe spline quadratique, interpolant des points cibles <ms ; Hz>, correspondant aux variations de F0 locales pertinentes [15]. Les mots cibles ont été étiquetés manuellement en syllabes. Les paramètres acoustiques suivants ont été extraits automatiquement : (i) au niveau du mot : durée D (ms) globale du mot, pics de F0 (Hz) et d'intensité I (dB), localisation (en pourcentage de la durée du mot) des pics de F0 et d'intensité par rapport au début du mot (ii) au niveau syllabique (identification des syllabes portant les pics de F0 et de I et le maximum de durée).

Pour chaque phrase, les valeurs des pics de F0 et de I ont ensuite été normalisées sur la base de leur valeur moyenne respective dans la phrase, afin d'obtenir un indice de *saillance prosodique* locale. Pour le paramètre de I, la normalisation permet de s'affranchir des fluctuations d'énergie RMS (distance micro/lèvres, notamment).

4. RESULTATS

Nous avons soumis nos données à des Analyses de Variance (Anova), afin de mieux cerner l'influence du contexte prosodique sur le focus, d'une part, et de déterminer les proportions et l'empan de son marquage acoustique, d'autre part.

Concernant le premier point, sur l'ensemble des mots-cibles du corpus, des Anova simples ont été menées pour lesquelles la variable indépendante était l'information de position du mot-cible MM et MT, les variables

dépendantes étant les *écarts* des valeurs de durée de mots, de pics de F0 et de I entre le mot-cible focalisé et le même mot-cible dans sa version non focalisée. Ce type d'Anova nous renseigne sur les éventuelles différences de saillance prosodique du Focus contrastif en position médiane ou terminale de phrase (MM vs. MT), et donc sur l'influence du contexte prosodique du focus en fonction de sa distribution dans la phrase. Ainsi, MM dans sa version non focalisé se trouve dans un contexte de pré-focus, MT étant focalisé ; Inversement, MT non focalisé se trouve dans un contexte de post-focus, puisque MM est focalisé. On s'attend donc à ce que les *écarts* de valeurs (D, F0 et I) soient plus importantes pour MT que pour MM, puisque la position post-focale de MT non focalisé implique, d'après la littérature, une réduction des patrons prosodiques, notamment de F0 et de I.

Pour ce qui est de l'empan du focus, sur l'ensemble des mots-cibles du corpus, des Anovas simples ont été menées séparément pour MM et MT, avec l'information Focalisé/Non Focalisé comme variable indépendante et les valeurs de durée globale, des pics de F0 et de I, ainsi que de localisation (ms) de ces pics par rapport au début du mot, comme variables dépendantes. Afin d'affiner l'information de localisation des pics, nous avons réalisé des Anovas factorielles sur les mots bisyllabiques (75% du corpus), pour lesquelles la variable indépendante supplémentaire était l'information de la syllabe (S1 initiale ou S2 terminale) portant le pic de F0, le pic de I ou le maximum de durée.

4.1. Influence du contexte prosodique

Nos résultats montrent qu'effectivement, le marquage du focus mobilise conjointement les paramètres F0, I et D (cf. 4.2. infra, pour les détails). Cependant, le renforcement acoustique par la focalisation est différent selon la position médiane ou terminale de mot :

- l'allongement de la durée globale du mot est plus important pour MM que pour MT, l'écart de durée entre les MM focalisés et non focalisés étant significativement plus important que celui des MT ($F(1, 238) = 29.37, p < .001$) ;
- l'augmentation de l'amplitude du pic de F0 est moindre pour MM que pour MT, l'écart des valeurs de F0 pour MM étant significativement moins important que pour MT ($F(1, 238) = 336.92, p < .001$).
- On n'observe pas de différence significative pour le pic de I.

Ces résultats s'interprètent ainsi : les MM non focalisés sont, dans notre dessin expérimental, nécessairement en position pré-focale. Par conséquent, le locuteur prépare le focus MT en affaiblissant les indices prosodiques précédents (notamment ici, la durée globale du mot). C'est pourquoi l'écart de durée entre MM focalisés et non focalisés est renforcé. Les patrons de F0 sont en revanche conservés. A l'inverse, l'écart de durée est moindre et celui des valeurs de F0 est plus important entre les mots focalisés et non focalisés pour MT, car en position post-focale, les patrons de F0 seraient aplatis (augmentant le

contraste avec les mots focalisés), les patrons de durée étant conservés [11, 12].

4.2. Proportions et empan du marquage prosodique du focus

D'une manière générale, les mots-cibles focalisés sont significativement plus allongés que leurs pendants non focalisés ($F(1, 238) = 76.35, p < .001$ pour MM, et $F(1, 238) = 17.7, p < .001$ pour MT). Les pics de F0 sont significativement plus élevés pour les mots-cibles focalisés (MM : $F(1, 238) = 108.82, p < .001$; MT : $F(1, 238) = 305.2, p < .001$), ainsi que les pics de I (MM : $F(1, 238) = 54.14, p < .001$; MT = $F(1, 238) = 47.32, p < .001$). La localisation du pic de F0 est significativement plus précoce dans les MM focalisés que dans les mêmes mots non focalisés (38% vs. 83% du mot ; $F(1, 238) = 433.62, p < .001$). Bien que la tendance soit la même pour les MT focalisés, ils ne sont pas significativement différents des MT non focalisés (38% vs. 41% du mot ; $F(1, 238) = .06, p = .81$). La localisation du pic de I est significativement plus précoce dans les MM focalisés (39% vs. 58% ; $F(1, 238) = 42.55, p < .001$), la même tendance étant vraie pour MT sans atteindre le seuil de signification (35% vs. 40% ; $F(1, 238) = 3.01, p = .08$).

L'observation plus précise de la localisation des pics de F0 et I, et des maxima de durée sur MM et MT mots focalisés bisyllabiques indiquent des tendances tout à fait comparables. Bien que l'allongement porte sur les deux syllabes, le maximum de durée se trouve sur la syllabe S2 (LSD de Fisher à $p < .001$ pour MM et MT). Le pic de F0 en revanche est davantage porté par S1, mais cette tendance n'est pas significative (LSD à $p = .25$ et $p = .07$ pour MM et MT respectivement). En effet, dans les deux positions, la localisation du pic est soit tardive dans S1 (80% de la syllabe), soit précoce dans S2 (30% de la syllabe), indiquant une tendance à un ancrage central du patron montant/descendant de F0 dans le mot bisyllabique. Le pic de I est clairement situé sur S1 pour MM et MT (LSD à $p < .001$ et $p = .007$ respectivement), et il est localisé tardivement dans S1 (70%) ou précocement dans la syllabe s'il est réalisé dans S2 (37%), indiquant là encore une tendance à un ancrage central du pic de I sur les mots bisyllabiques.

Des tendances similaires ont été observées pour les mots trisyllabiques (maximum de durée sur la dernière syllabe du mot, et pics de F0 et de I davantage sur la première syllabe), mais la trop faible proportion de trisyllabiques dans le corpus (25%) ne permet pas de valider ces résultats statistiquement.

5. DISCUSSION ET PERSPECTIVES

Ces résultats nous ont permis de caractériser plus précisément le focus contrastif en français sur le plan prosodique. Tout d'abord, nos résultats sur les *écarts* de valeurs des paramètres prosodiques entre les mots focalisés et non focalisés dans les deux positions médiane (MM) et terminale (MT) de phrases, confirment

expérimentalement les tendances observées dans la littérature sur le français, i.e. une réduction des paramètres prosodiques en position pré-focale [13], et essentiellement un aplatissement des contours de F0 en position post-focale [11, 12]. Nos résultats révèlent en outre que les contours d'intensité sont comparables en position pré-focale et post-focale, à la différence de Dahan & Bernard [10].

En ce qui concerne plus spécifiquement le marquage prosodique du focus contrastif en français, nos résultats indiquent tout d'abord que le focus est marqué globalement de manière identique sur les mots terminaux et médians (cf. [16] pour des résultats comparables en anglais). En outre, nos résultats démontrent quantitativement que les différents paramètres (D, F0 et I) se combinent pour marquer localement le mot entier : le paramètre de durée marque proportionnellement davantage la fin des mots, alors que les pics de F0 et d'intensité se situent préférentiellement sur le début du mot. La diffusion des indices prosodiques sur l'ensemble du mot, ainsi que la localisation des pics de F0 et de I vers le milieu des mots bisyllabiques, plaident ainsi pour un marquage holistique du focus contrastif en français.

Nous envisageons de prolonger cette étude dans 4 directions. (i) Étendre notre corpus en enregistrant les mêmes phrases avec une prosodie déclarative neutre, afin notamment d'affiner la description des mots en position pré- et post-focale. (ii) Augmenter le nombre de mots trisyllabiques, afin de confirmer statistiquement les tendances de marquage holistique du focus. (iii) Adapter la méthodologie de Bricchet & Aubergé [9] à notre corpus, en validant perceptivement le marquage holistique du focus contrastif par une expérience de 'gating' (présentation auditive en pas croissant (en ms) du mot). (iv) intégrer les informations ici fournies (nature et distribution des paramètres prosodiques D, I, et F0) dans un modèle statistique de régression multiple, afin de déterminer quels sont les paramètres les plus pertinents dans la caractérisation du focus contrastif en français. Ces résultats devraient nous permettre de disposer d'informations quantitatives solides, à implémenter dans les modèles de synthèse du français [14].

6. BIBLIOGRAPHIE

- [1] C. Magne, C. Astésano, A. Lacheret-Dujour, M. Morel & M. Besson. Traitement de l'accent de focalisation en français: approche électrophysiologique. Actes des XXV JEP, Fès, 2004.
- [2] K. Lambrecht. *Information Structure and Sentence Form*. Cambridge, CUP. 1994.
- [3] A. Lacheret and F. Beaugendre. *La prosodie du français*. Editions du CNRS, Paris.
- [4] V. Padeloup. *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. Thèse de Doctorat, Université de Provence. 1990.
- [5] A. Di Cristo. French. In D. Hirst, A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge University Press, Cambridge, 195–218, 1998.
- [6] A. Séguinot. L'accent d'insistance en français standard. *Studia Phonetica*, 12, 1-58, 1976.
- [7] C. Astésano. *Rythme et accentuation en français. Invariance et variabilité stylistique*. Editions L'Harmattan, Collection Langue & Parole, 2001.
- [8] A. Di Cristo. *De la microprosodie à l'intonosyntaxe*. Publications de l'Université de Provence. 1985.
- [9] C. Bricchet & V. Aubergé. La focalisation en français : morphologie de la prosodie. *Actes des Journées Prosodie 2001*, V. Aubergé, A. Lacheret, H. Loevenbruck (eds), 67-70, 2003.
- [10] D. Dahan & J.M. Bernard. Interspeaker variability in emphatic accent production in French. *Language and Speech*, 39 (4), 341-374. 1996.
- [11] A. Di Cristo,; L. Jankowski. Prosodic Organisation and Phrasing after Focus in French. *ICPhS'99*, 1565-1568.
- [12] E. Delais-Roussarie, A. Rialland, J. Doetjes & J.M. Marandin. The prosody of post-focus sequences in French. *Speech Prosody*, D. Hirst (ed.), 2002.
- [13] P. Touati. *Structures prosodiques du suédois et du français*. Lund University Press. 1987.
- [14] M. Morel & A. Lacheret. Kali, synthèse vocale à partir du texte : de la conception à la mise en oeuvre. *Traitement automatique des langues, synthèse de la parole à partir du texte*, 42, Ch. D'Alessandro (ed.) Hermès, Paris, 193-221, 2001.
- [15] D. Hirst, & R. Espesser. Automatic modelling of fundamental frequency with a quadratic spline function. *T.I.P.A*, 15, 71-85. 1993.
- [16] E. Selkirk. Contrastive Focus vs. presentational focus: Prosodic evidence from right-node raising in English. *Speech Prosody*, D. Hirst (ed.), 2002.