

La focalisation contrastive est-elle visible ?

Une étude perceptive visuelle 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

The purpose of this study was to determine whether the visual modality is useful for the perception of prosody. An audio-visual corpus consisting of four focus conditions (subject, verb, object focus and broad focus) was recorded from a male native speaker of French. A preliminary production study showed visible correlates of contrastive focus in French are a) increase in lip area and jaw opening on the focused syllables b) lengthening of the prefocal syllable and the focal syllables (with a considerably higher lengthening for the first segment of the focused phrase). The present perceptual study showed that a) contrastive focus was well perceived visually; b) no practice was necessary and c) subject focus was slightly easier to identify than the other focus conditions. We also found that the presence and salience of the visual cues enhances perception.

1. INTRODUCTION

La prosodie est souvent conçue comme un ensemble de commandes glottiques et subglottiques résultant en des variations acoustiques (fréquence fondamentale (F0), l'intensité) ou en des variations temporelles (durée syllabique). Par conséquent, les études sur la perception de la prosodie, sont pour la plupart centrées sur la modalité auditive. Du point de vue de la modalité visuelle, les gestes glottiques et subglottiques sont peu visibles. Toutefois, il a été suggéré, pour certaines langues, que la focalisation prosodique, par exemple, puisse être perçue visuellement. Parmi les indices visuels étudiés, les mouvements des sourcils [1] et de la bouche [2,3] semblent être influencés par la focalisation prosodique.

La focalisation contrastive permet de mettre en relief un mot ou groupe de mots 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. »). A notre connaissance, il n'existe qu'une étude de la perception visuelle de la focalisation contrastive, et elle concerne uniquement l'anglais [4]. Elle laisse suggérer qu'il existe des corrélats faciaux de la focalisation, qui permettent sa détection visuelle (sans signal audio).

Nous étudions ici la focalisation contrastive prosodique en français. Nous donnons les résultats d'une étude

annexe sur ses caractéristiques visuelles puis présentons les résultats de tests perceptifs visuels.

2. METHODE EXPERIMENTALE

2.1. Corpus

Nous utilisons ici le modèle de Jun & Fougeron [5,6] qui est en accord avec la plupart des descriptions de l'intonation du français et s'appuie sur un système de transcription employé 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 LHi (*Low-High initial*, succession de tons bas et haut), appelée accent secondaire. La plus élevée est le Syntagme Intonatif (SI) qui peut moduler le SA (e.g. si un SA est en position finale d'un SI, H* est remplacé par le ton de frontière du SI, L% ou H%). Dans ce modèle, la focalisation contrastive est marquée par un ton Hf élevé suivi d'un plateau tonal bas[5-8].

Le corpus est composé de 8 phrases ayant une structure sujet-verbe-objet (SVO), des syllabes CV (Consonne-Voyelle), et une majorité de sonores pour faciliter la détection de F0. Les phrases ont été conçues de manière à être prononcées comme un SI composé de trois SA. Des exemples sont fournis en 4.

2.2. Enregistrement audio-visuel

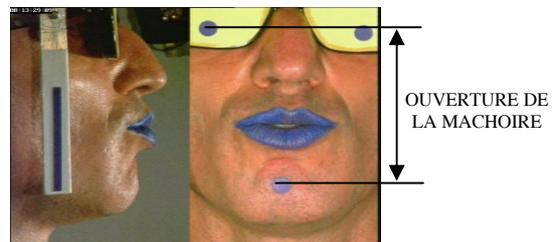


Figure 1 : Signal vidéo : profil et face.

Un locuteur francophone masculin a été enregistré (face et profil : figure 1) pour quatre conditions : focalisation contrastive étroite sur S, V ou O et focalisation large sur l'énoncé. Pour obtenir des focalisations étroites, le locuteur entendait un énoncé déclencheur dans lequel S, V ou O était erroné. Il devait produire la phrase correcte en contrastant le syntagme erroné dans le déclencheur :

Enoncé déclencheur : Denis ranima la jolie maman.

Production du locuteur : ROMAIN ranima la jolie maman.

Quatre modes ont été enregistrés : lexicalisé, délexicalisé, chuchoté et délexicalisé-chuchoté. Seuls les deux premiers ont été étudiés. Le mode délexicalisé a été prononcé en remplaçant toutes les syllabes par [ma]. Son objectif est de permettre la comparaison acoustique et articulatoire des syllabes. 256 énoncés ont été produits (8 phrases, 4 focalisations, 4 modes d'élocution, 2 répétitions), 128 ont été étudiés.

2.3. Analyse annexe des productions

Les analyses acoustique et articulatoire sont décrites en détail dans [7]. L'analyse acoustique a montré que les enregistrements contiennent des indices insonorifs de la structure focale, cohérents avec d'autres observations [5,11-12]. L'analyse articulatoire et temporelle nous a permis d'identifier un ensemble de **corrélats visibles** (i.e. des corrélats articulatoires ou duratifs visibles par le (les) interlocuteur(s) et qui pourraient être utilisés pour la perception) de la focalisation contrastive en français : a) augmentation de l'**aire aux lèvres** et de l'**ouverture mandibulaire** sur la séquence focale, b) **hypo-articulation de la séquence post-focale** (aire aux lèvres et ouverture mandibulaire réduites), c) **allongement de la syllabe pré-focale et des syllabes focales** (en particulier de l'attaque de la séquence focale).

3. ETUDE PERCEPTIVE PRELIMINAIRE : PAROLE DELEXICALISEE

Dans une étude préliminaire [13], nous avons étudié le mode délexicalisé. Nous avons mené une étude perceptive dans laquelle on présentait aux participants des stimuli visuels. Les participants devaient identifier la condition de focalisation (étroite sur S, V, O ou large). Les résultats ont montré que, pour ce mode, les sujets perçoivent correctement la focalisation *via* la modalité visuelle seule (86% de bonnes réponses, seuil de hasard à 25%). Les méprises sont plus fréquentes dans le sens d'une focalisation étroite perçue comme large que l'inverse. La focalisation sur le sujet est significativement mieux détectée que les autres. Lorsqu'elle est mal détectée, elle est prise pour une focalisation large. Ce résultat peut s'expliquer par le fait que les syntagmes verbal et objet qui suivent un sujet focalisé sont hypo-articulés : ils ne peuvent donc pas être classés comme focalisés. Ainsi si la focalisation sur le sujet n'a pas été perçue, la réponse la plus probable est la focalisation large. De plus, les énoncés à taux d'erreur important sont ceux pour lesquels les corrélats visibles sont peu marqués et les énoncés à faible taux d'erreur ont des corrélats visibles bien marqués. L'ensemble de ces résultats indique qu'il existe bien des indices visuels utiles pour la perception de la focalisation contrastive en mode délexicalisé en français et que ces indices incluent probablement les paramètres articulatoires et temporels que nous avons mesurés.

4. EXPERIENCE PERCEPTIVE : METHODE EXPERIMENTALE

Les corrélats identifiés dans l'étude annexe [7] sont-ils

perçus visuellement par des observateurs francophones ?

Les 4 phrases équilibrées du corpus (nombre de syllabes proche dans chaque syntagme S, V, O) ont été étudiées :

- (1) [Romain]_{S2} [ranima]_{V3} [la jolie maman]_{O5}.
- (2) [Véronique]_{S3} [mangeait]_{V2} [les mauvais melons]_{O5}.
- (3) [Mon mari]_{S3} [veut ranimer]_{V4} [Romain]_{O2}.
- (4) [Les loups]_{S2} [suivaient]_{V2} [Marilou]_{O3}.

Nous avons indiqué aux participants qu'ils suivraient une conversation entre 2 locuteurs. Ils entendraient d'abord un locuteur prononçant une phrase (déclencheur audio). Un des éléments de cette phrase (S, V ou O) serait mal compris par le deuxième locuteur, qui produirait la phrase sur un mode interrogatif (les participants n'entendraient pas cette question). Le premier locuteur répéterait en insistant sur l'élément mal compris. Les participants verraient une vidéo de profil et de face de ce locuteur, sans son. L'expérience se déroulait ainsi :

- Locuteur 1 (audio seul) : Romain ranima la jolie maman.*
Locuteur 2 (rien) : Denis ranima la jolie maman ?
Locuteur 1 (vidéo seul) : ROMAIN ranima la jolie maman.

On prévenait les participants que le deuxième locuteur comprenait parfois bien la phrase (la focalisation était large). Leur tâche consistait à déterminer s'il y avait eu une mauvaise compréhension et quel élément avait été mal compris, i.e. à déterminer la condition de focalisation. Ils utilisaient un surligneur pour marquer le constituant perçu comme focalisé sur une grille de réponse ; en cas de focalisation large, ils surlignaient la case vide :

Romain	ranima	la jolie maman.	
--------	--------	-----------------	--

Pour le déclencheur audio, qui permet de connaître la phrase sur laquelle portera la focalisation, nous avons utilisé les enregistrements en focalisation large et pour les répétitions vidéo, nous avons utilisé toutes les conditions. Nous disposions de 32 paires de phrases (1 paire : un énoncé audio seul avec focalisation large et un énoncé vidéo seul) au total (4 phrases, 4 conditions de focalisation, 2 répétitions) dont 8 à focalisation large. Les participants ont passé 5 tests correspondant à 5 combinaisons aléatoires des 32 paires mais dans un ordre différent). Chaque participant a donc évalué 160 paires de phrases. 33 locuteurs francophones (8 hommes et 25 femmes) âgés de 18 à 52 ans ont participé.

5. RÉSULTATS

5.1. Résultats généraux

La moyenne des pourcentages de bonnes réponses (condition de focalisation bien identifiée) pour les 33 participants s'élève à 71,45%. Ce score étant bien supérieur au hasard (25%), il permet de conclure que les participants sont sensibles aux informations visuelles sur la focalisation contrastive. Ce score élevé était surprenant, étant donné que la plupart des participants a rapporté que le test était difficile. Ceci suggère que les indices visuels seraient utilisés de façon non-explicite. Nous avons également vérifié que les scores étaient indépendants de l'ordre de présentation des stimuli.

5.2. Influence de l'entraînement

Les participants ont tous passé les mêmes tests, mais dans des ordres différents. Nous avons donc vérifié les résultats de chaque test en fonction de sa position au cours de l'expérience. Il se pourrait en effet que les performances des participants augmentent au fur et à mesure de l'expérience. Une analyse de la variance à un facteur (ANOVA) montre que les 5 moyennes ne sont pas significativement différentes ($F(4,160) = 0.55, p = 0.7$). Nous concluons donc que les scores des participants ne sont pas améliorés par l'entraînement.

5.3. Analyse par constituant syntaxique

Cette analyse a pour but d'examiner si la position de la focalisation étroite (S, V ou O) a une influence sur la performance des participants. Une analyse de la variance à un facteur (ANOVA) montre que les résultats des 4 conditions de focalisation sont significativement différents ($F(3,96)=11.3, p<0.01$). Des comparaisons multiples indiquent que le score moyen est significativement plus élevé pour la condition de focalisation sur le sujet que sur le verbe ou l'objet ($p<0.001$). Les scores des conditions de focalisation sur le verbe, l'objet et la focalisation large ne sont pas significativement différents, pas plus que ceux de la focalisation sur le sujet et la focalisation large. Il semble donc que, pour ce corpus, la condition de focalisation sur le sujet soit plus facile à détecter. Pourtant, on aurait pu s'attendre à ce que les différences articulatoires dues à la focalisation soient plus saillantes au milieu de la phrase (sur le verbe) plutôt qu'au début. Les données articulatoires vont d'ailleurs dans ce sens, car les maxima d'aire au lèvres et d'ouverture mandibulaire sont plus élevés pour la focalisation sur le verbe. Cependant, comme nous l'avons expliqué en 2.3, lorsqu'un syntagme est focalisé, les syntagmes qui le suivent sont hypo-articulés (ouverture mandibulaire et aire aux lèvres réduites). Lorsque le sujet est focalisé, cette hypo-articulation est observée pour tout le reste de la phrase. Si l'observateur fait du syntagme sujet une référence en termes d'étendue des mouvements articulatoires, alors les deux syntagmes qui suivent doivent lui paraître nettement moins articulés. La différence entre l'hyper-articulation du sujet et l'hypo-articulation des syntagmes suivants est probablement un indice du lieu de la focalisation.

5.4. Analyse des types d'erreur

Le but est ici d'examiner les confusions faites par les participants. La figure 2 fournit la matrice de confusion des erreurs observées. Il semble que les sujets aient le plus souvent confondu une focalisation étroite avec une focalisation large (15,8% des focalisations sur le sujet, 26,1% sur le verbe, 9% sur l'objet). Ceci était prévisible car les corrélats visibles de la focalisation contrastive sont plus ou moins marqués selon les stimuli. Toutefois, on observe aussi des confusions entre focalisation large et focalisation étroite sur le sujet (9%) ou le verbe (11,5%).

réponse stimulus	S	V	O	LARGE
S	80.2	3.9	0.2	15.8
V	7.4	65.7	0.8	26.1
O	0.6	25.4	65	9
LARGE	9	11.5	4.5	74.9

Figure 2 : Pourcentages de réponses des 4 types pour les 4 types de stimulus. 80,2% des stimuli avec focalisation sur le sujet ont bien été identifiés comme tels.

La figure 2 indique également un taux de confusion important entre stimuli à focalisation sur l'objet et ceux à focalisation sur le verbe (25,4%). Ce résultat est intéressant car il avait également été observé pour la parole délexicalisée [13]. Nous avons alors expliqué ceci par le fait que lorsqu'un syntagme est focalisé, la focalisation est distribuée sur l'ensemble de ses syllabes. Ainsi, plus le syntagme focalisé est long, moins son hyper-articulation serait significative. Dans le corpus étudié, certains syntagmes étaient relativement longs par rapport à d'autres. Les objets du corpus sont parfois assez longs (jusqu'à 7 syllabes) et pour ces objets là, l'hyper-articulation n'est peut-être pas aussi marquée que pour les plus courts. Les données articulatoires confirment cette hypothèse : en moyenne dans ce corpus, quand l'objet est focalisé, il n'est pas aussi hyper-articulé que sont les autres syntagmes. La notion d'effort articulatoire pourrait expliquer le fait que l'hyper-articulation soit distribuée. Quand les syntagmes sont courts, l'augmentation de l'aire aux lèvres, l'ouverture et la durée nécessite probablement moins d'effort que quand ils sont longs. Toutefois cette notion d'effort n'explique pas à elle seule pourquoi la confusion s'effectue vers une focalisation sur le verbe. Dans ce corpus, lorsque le nombre de syllabes de l'objet est élevé, celui du verbe est faible. Ainsi, lorsque l'objet est long, la focalisation est uniformément répartie (e légère hyper-articulation de toutes les syllabes). Or la syllabe qui précède cet objet long focalisé porte les corrélats visibles du H* d'un verbe court. Cette syllabe accentuée peut ainsi paraître plus hyper-articulée que les syllabes suivantes et conduire à l'identification d'une focalisation sur le verbe.

5.5. Analyse approfondie par stimulus

Dans cette section, les stimuli bien et mal identifiés ont été analysés de façon approfondie. Sur les 24 stimuli à focalisation étroite (parmi les 32 stimuli, 8 étaient à focalisation large), 4 ont été *mal perçus* (pourcentage de réponses correctes inférieur ou proche de 25 %), 9 ont été *bien perçus* (entre 60% et 80%) et 11 ont été *très bien perçus* (plus de 80%). Pour les stimuli *mal perçus*, les paramètres visibles mesurés ne sont pas conformes à la description du 2.3. En effet, en moyenne, l'aire aux lèvres sur les constituants focalisés (respectivement l'ouverture mandibulaire) est bien plus faible que pour les stimuli très bien perçus : 3,15 vs. 5,53 cm² (3,34 vs. 5,76 cm). De même l'allongement pré-focal est moins important (durée

syllabique de 193 vs. 221 ms). Il est également intéressant de noter que trois des stimuli *bien perçus* et cinq stimuli *très bien perçus* sont associés à des corrélats visibles complémentaires : si l'un ou plusieurs des paramètres sont atypiques (par rapport au 2.3), l'un des autres au moins est très marqué. Parmi ces stimuli atypiques, trois stimuli *bien perçus* et deux stimuli *très bien perçus* correspondent à des corrélats temporels non marqués, mais à un corrélat articulatoire bien marqué. Pour un troisième stimulus *très bien perçu*, les corrélats articulatoires ne sont pas marqués, mais les corrélats temporels le sont. Seuls deux stimuli atypiques ne présentent pas ce compromis. Leurs corrélats sont tous « moyennement marqués ». Or ces stimuli correspondent aux scores les plus élevés (98,2% et 98,8%). Il se peut donc que la focalisation contrastive soit mieux détectée si tous les corrélats sont présents, même s'ils ne sont pas particulièrement marqués. Mais il se peut également (comme cela a été suggéré par [4]) qu'il existe des corrélats plus subtils que ceux qui ont été évalués.

6. DISCUSSION ET CONCLUSION

Notre étude préliminaire sur la parole délexicalisée suggère qu'il existe des indices visuels de la focalisation contrastive en français et que ces indices sont bien perçus par les observateurs. L'objectif de l'étude présentée ici était de vérifier que ces résultats tiennent pour la parole lexicalisée. Le corpus contenait des phrases lexicalisées dans 4 conditions de focalisation (étroite sur le sujet, le verbe ou l'objet et large). Des corrélats visibles ont été identifiés dans une étude annexe sur la production. Nous avons mené une expérience perceptive afin de tester si ces corrélats visibles constituent des indices pour la perception de la focalisation contrastive. Nous avons présenté aux participants des stimuli purement visuels. Les résultats montrent que les observateurs perçoivent avec succès la focalisation, sans entraînement. La focalisation sur le sujet est significativement plus facile à détecter que les autres conditions et lorsqu'elle n'est pas détectée elle est confondue avec une focalisation large. Le fait que les stimuli les plus mal perçus correspondent à des corrélats visibles peu marqués indiquent que les corrélats identifiés sont probablement effectivement utilisés par les observateurs. Cependant, certains stimuli sont bien perçus alors que leurs corrélats visibles sont uniformément peu marqués. Ceci pourrait signifier que tous les corrélats doivent être présents, même s'ils ne sont pas très marqués, pour une identification correcte de la focalisation. Il est également possible que d'autres corrélats existent. D'autre part, il a été montré que les stratégies articulatoires utilisées pour marquer la focalisation dépendent du locuteur [2, 4]. En perspective, nous envisageons d'enregistrer d'autres locuteurs pour vérifier si les résultats peuvent être généralisés.

7. REMERCIEMENTS

Nous remercions Marie-Agnès Cathiard et Jean-Luc Schwartz pour leurs conseils scientifiques précieux, Guillaume Rolland pour l'enregistrement du corpus, Christophe Savariaux et Alain Arnal pour leur aide sur les

données vidéo, Pauline Welby pour ses commentaires sur cet article et tous les participants aux tests perceptifs.

BIBLIOGRAPHIE

- [1] B. Granström, D. House & M. Lundeberg. Prosodic Cues in Multimodal Speech Perception. *Proceedings of ICSLP'99*. San Francisco, 1, 655-658, 1999.
- [2] K. De Jong. The supraglottal articulation of prominence in English: linguistic stress as localized hyper-articulation. *J. Acoust. Soc. A.m.* 97, 491-504, 1995.
- [3] L.E. Bernstein, S.P. Eberhardt & M.E. Demorest. Single-channel vibrotactile supplements to visual perception of intonation and stress. *J. Acoust. Soc. A.m.* 85, 397-405, 1989.
- [4] 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.
- [5] S.-A. Jun & C. Fougeron. A Phonological Model of French Intonation. In *Intonation: Analysis, modelling and technology*, A. Botinis (ed.). Dordrecht, 209-242, 2000.
- [6] S.-A. Jun & C. Fougeron. Realizations of Accentual Phrases in French Intonation. *Probus* 14, 147-172, 2003.
- [7] E. Delais-Roussarie, A. Rialland, J. Doetjes & J.M. Marandin. The Prosody of Post Focus Sequences in French. *Proceedings of Speech Prosody 2002*, Aix en Provence, 2002.
- [8] A. Di Cristo & L. Jankowski. Prosodic Organisation and Phrasing after Focus in French. *Proceedings of ICPhS 1999*, 1565-1568, 1999.
- [9] M. Dohen & H. Lœvenbruck. Identification des corrélats visibles de la focalisation contrastive en français. Actes des XXVèmes JEP, 19-22 avril 2004.
- [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] 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.
- [12] 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.
- [13] M. Dohen, H. Lœvenbruck, M.A. Cathiard, J.L. Schwartz. Audivisual Perception of Contrastive Focus in French. *Proceedings of the AVSP 2003 Conference*. St Jorioz, France, 245-250, 2003..