

Le focus prosodique n'est pas que déictique : le modèle VID (Valence-Intensité-Domaine)

Aubergé, V. & Rilliard, A.

Institut de la Communication Parlée UMR 5009 CNRS-INPG- Université Stendhal, Grenoble, France
{auberge, rilliard}@icp.inpg.fr

ABSTRACT

This paper summarizes several perception experiments showing that the morphology of the prosodic focus conveys more information than the only deictic information: (1) the binary valence - yes/no focus – which is perceptively quite categorical (a magnet effect is clear on the basis of an identification and a discrimination experiment [1]), (2) the intensity information, used by the speaker to give his preference for one of two focused elements, (3) the information of the focus domain, that are some segmentation cues about the focused element (phonological unit or word unit), which are perceptively identified by listeners. The morphological cues revealing Valence-Intensity-Domain are observed in particular in morphing procedure making clear the thresholds of quite-categorical behaviors.

1. INTRODUCTION

La focalisation dans la chaîne verbale peut être définie sommairement comme une fonction utilisant des matériaux communicatifs dont la prosodie et la syntaxe pour faire émerger une information nouvelle ou préciser un élément mis en contraste. Ainsi, Rossi [9] considère la focalisation comme un ensemble d'outils qui permettent au locuteur de hiérarchiser l'information et de faire émerger un élément spécifique sur lequel le locuteur veut attirer l'attention de l'auditeur. La fonction d'emphase et de focus lexical sont présentées comme deux fonctions exclusive dans la littérature [5]. La fonction d'emphase est surtout reliée à des traits expressifs comme le degré d'intérêt et est décrite plutôt comme une fonction gradiente, tandis que la fonction de focalisation est binaire (présence/absence de focus). Le focus prosodique est très robuste : un focus lexical est réalisé en français (et peut-être universellement [3]) par une proéminence acoustique, principalement tonale, sur la première syllabe du mot (sauf dans des cas particuliers de conflits ou d'effet stylistique). Pourtant un mot contenant un tel ton sur n'importe laquelle de ses syllabes, est néanmoins perçu comme focalisé [2]. Jackendoff [4] a montré que le domaine du focus peut être méta-linguistique quand il renvoie à des valeurs communicatives différentes du focus lexical : l'« ordinary focus » se distingue du focus méta-linguistique dont le domaine est la syllabe ou le phonème et qui pointe sur la forme phonologique du signifiant. Dans des études précédentes, nous avons montré, sur la base d'une analyse acoustique de F0, de l'intensité et de la durée, que les réalisations prosodiques du focus ordinaire (nouveau ou contraste sur un élément lexical) et du

focus méta-linguistique (focus syllabique) sont similaires. Les hauteurs de F0 et d'intensité sont identiques sur la première syllabe (pour un focus sur le mot ou directement sur la première syllabe), la pente qui ramène F0 de la première syllabe au niveau du F0 du mot, ont des décours différents [2]. Il s'agit donc de vérifier si cette apparente dissimilitude a une pertinence cognitive. Nous avons essayé de montrer par différentes expériences perceptives que trois types d'informations fonctionnelles sont véhiculées par la morphologie prosodique du focus¹ : (1) sa *Valence*, information grammaticale binaire sur la présence ou l'absence du focus sur l'élément, nous montrons qu'il est réalisé par un processus tonal (indice statique), (2) son *Intensité*, information pragmatique de quantité d'insistance, par exemple ou délivrer une valeur d'emphase ou encore pour instancier un choix entre deux éléments focalisés ; l'intensité est réalisé grâce une compétence psycho-acoustique de perception gradiente des variations de hauteurs tonales, montrée en particulier par Ladd, mais nous proposons que cette intensité fonctionnelle ne soit cognitivement pertinente qu'après que la valence positive de présence focus aie été instanciée, et non pas directement à partir d'un non focus comme le suggère l'expérience de Ladd et (3) son *Domaine*, information linguistique sur le segment qui constitue l'élément focalisé (la syllabe pour un focus méta-linguistique, le mot pour un focus de contraste ou de nouveauté), qui est réalisée par la valeur de la pente de F0 en fin de ton de focus (indice dynamique).

Nous rappelons ainsi brièvement les résultats d'une expérience de perception catégorielle, menée sur des stimuli « morphés » progressivement des contours sans focus aux contours avec focus, produits par un locuteur, et qui montre clairement une décision binaire par un effet magnet sur absence/présence de focus. Dans cette expérience, la valeur « frontière » de la présence de focus donne la base de la gradience qui opère au-dessus de la perception de présence de focus. Enfin, nous résumons deux expériences dont la première montre la compétence perceptive de discrimination du domaine (syllabe vs. mot) et dont la seconde, basée sur un morphing de la pente de F0, met en évidence un seuil de pente dans certains stimuli pour la perception du domaine de l'élément focalisé. Ces expériences nous permettent de proposer le modèle VID qui attribue à la morphologie prosodique du focus une fonction à trois valeurs informationnelles : Valence, Intensité, Domaine.

¹ « morphologie » de la prosodie est ici employé au sens générique, i.e. organisation de la forme prosodique

2. INDICES DE LA VALENCE DE LA FONCTION

Le corpus utilisé pour les expériences présentées ici (voir [2] et [1] pour plus de détails) est basé sur une même structure syntaxique porteuse, dans laquelle les items lexicaux varient uniquement selon la dimension phonotactique. Les phrases françaises enregistrées ont une longueur de 6 à 8 syllabes, et chaque item lexical comporte de 1 à 3 syllabes. Chaque phrase est enregistrée avec le focus réalisé alternativement sur chaque item lexical pour le focus sur le mot, sur chaque syllabe pour le focus métalinguistique et sans focus. Différents tests de perception [2] ont permis de valider la bonne perception du focus sur l'item visé. Le protocole de ces expériences est détaillé dans [1].

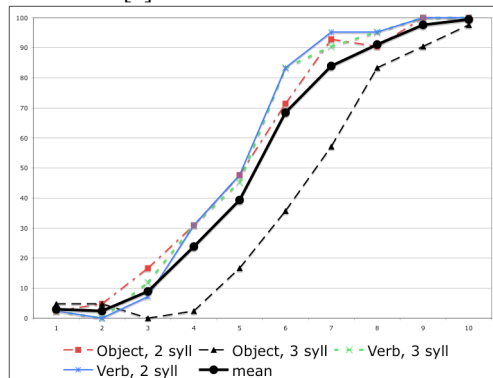


Figure 1. Pourcentages d'identification du focus pour les 4 continums et pourcentage moyen. Les ordonnées représentent les 10 itérations du morphing depuis le stimulus neutre jusqu'au focalisé.

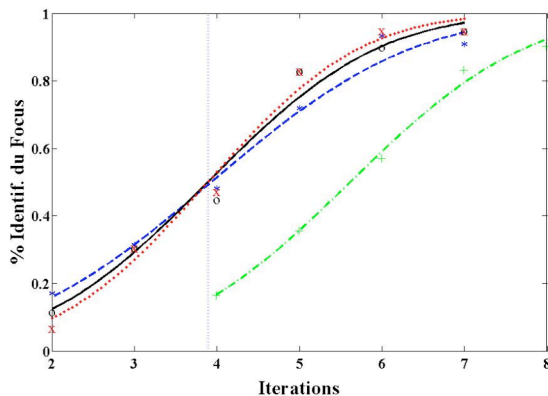


Figure 2. Interpolation Probit des résultats pour chaque continuum. Légende : (*) et tirets bleus : focus sur l'objet de 2 syllabes ; (+) et tirets-points verts : focus sur l'objet de 3 syl ; (o) et ligne continue : verbe de 3 syl ; (x) et pointillés : verbe 2 syl. La ligne verticale représente le seuil de perception du focus. En ordonnée l'itération et en indice le % d'identification du focus.

4 continums acoustiques en 10 pas, partant de la phrase sans focus pour aboutir à la même phrase portant un focus sur un mot, ont été créés. Les pas de F0, d'intensité et de durée utilisés pour ces morphings sont tous inférieurs au seuil de perception, tel qu'il est décrit par Rossi [8]. Une tâche d'identification cherche à mettre en évidence un comportement quasi catégoriel pour la perception de ce continuum prosodique : un effet magnet [6]. Les résultats (cf. fig. 1) obtenus sont comparables à ceux décrits par Ladd [7]. Une régression logistique (analyse Probit) sur ces résultats (cf. fig 2) met en évidence le seuil de

perception du focus aux alentours de la quatrième itération sur les 10.

Afin de compléter cette tâche d'identification, et de vérifier si le comportement est vraiment catégoriel, une expérience de discrimination a été réalisée, dont les résultats ne mettent en lumière aucun pic de discrimination aux alentours du seuil de perception. Nous devons donc conclure à un comportement de type effet magnet, avec une valeur de F0 (et aussi d'intensité) statique au-delà de laquelle les sujets perçoivent une focalisation (cf. fig. 3).

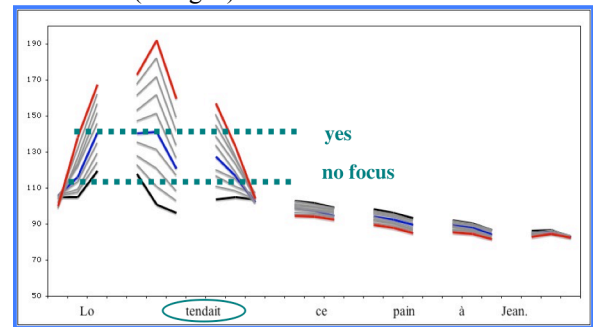


Figure 3. Morphologie de F0 dans le cas d'un verbe de 2 syllabes, pour les deux valences du focus. Les extrêmes correspondent aux deux références naturelles

3. INDICES DE L'INTENSITE DE LA FONCTION

Ladd [7] a montré qu'un auditeur en condition psycho-acoustique perçoit les variations de proéminence (F0 seulement) de manière gradiente, alors que les mêmes stimuli dans une tâche de perception linguistique sont perçus de manière catégorielle, conformément aux résultats présentés précédemment. Lors d'une expérience préliminaire, nous avons observé que les auditeurs utilisent ces capacités psychoacoustiques par exemple pour choisir un élément parmi deux éléments lexicaux focalisé au sein d'un même énoncé. Cette fonction, que nous appellerons fonction de préférence, suppose (1) une fonction déictique sur les deux éléments lexicaux, c'est-à-dire une valence positive de la fonction de focus sur les deux domaines lexicaux sur lesquels (2) va opérer l'intensité, par une gradience supérieure de l'un des deux pour implémenter la préférence. Ce genre de distinction est utilisé typiquement lors de dialogues homme-machine, pour exprimer la préférence du locuteur à propos d'un choix entre deux items. Les auditeurs écoutaient des phrases du type : « Préférez-vous une correspondance à Paris ou à Londres cette semaine ? », dans laquelle Paris ou Londres recevraient une gradience supérieure, les deux étant obligatoirement au-dessus du seuil de perception de la valence positive de focus. Pour être interprétable cette expérience demande à être reproduite avec des stimuli plus qui contrôlent systématiquement l'ordre de présentation, leur place dans l'énoncé (une même valeur variation de hauteur n'est pas perçue identiquement en début ou fin d'énoncé [3], et le domaine du focus, c'est-à-dire l'implémentation de la fonction de pointage lexicale vs. pointage méta-linguistique.

4. INDICES DU DOMAINE DE LA FONCTION

4.1. Expérience de discrimination

Lors de cette expérience, les auditeurs doivent juger si le locuteur parle plus spécifiquement d'une personne, d'une action ou d'un objet (en contraste avec un autre), ou s'il tâche de désambigüiser une syllabe mal comprise d'un mot particulier (focus métalinguistique). 25 sujets ont écouté tous les stimuli une fois seulement, et donné leur réponse grâce à l'interface présentée à la figure 4.

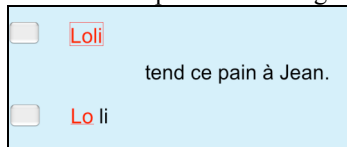


Figure 4. Interface du test de discrimination des focus sur le mot ou sur la première syllabe.

Les stimuli sélectionnés présentent toutes les longueurs et les positions possibles de focus sur un mot dans notre corpus, ainsi que leur pendant focalisé sur la première syllabe du même mot. Cela car en français, les mots focalisés sont réalisés avec une importante proéminence sur la première syllabe, et doivent donc être comparés aux stimuli ne portant un focus que sur cette syllabe-ci.

Table 1. Résultats de la tâche de discrimination.

	% bonnes réponses	
	brut	corrigé / hasard
Focus sur le mot	79,7	59,3
Focus 1 ^{ère} syllabe	83,7	67
Total	81,7	63,3

Les auditeurs ont le sentiment de répondre au hasard, mais les résultats (cf. table 1) montrent un taux de discrimination largement supérieur au hasard, pour les deux types de focus.

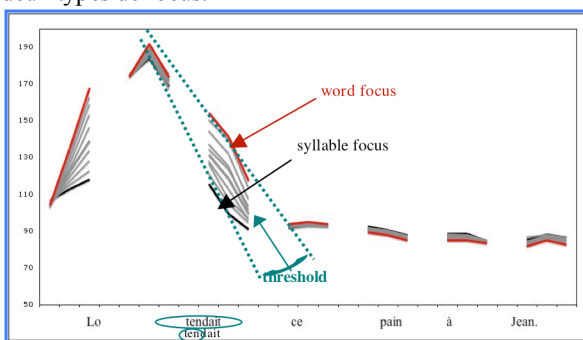


Figure 5. Morphologie de la F0 dans le cas des verbes de 2 syllabes, pour les deux domaines de focus. Les deux contours extrêmes correspondent aux deux références naturelles.

Il faut noter que l'analyse acoustique (cf. [2]) réalisée sur ces stimuli montre des niveaux de F0 et d'intensité similaire pour la première syllabe des mots réalisés avec un focus sur le mot ou sur la syllabe. Ceci est indirectement confirmé par l'expérience d'identification du focus présentée ci-après, pendant laquelle un seuil est relevé pour le focus sur le mot aussi bien que sur la syllabe. Cela pourrait indiquer que les indices permettant

l'identification du domaine sont à rechercher dans la dynamique du contour du focus, entre la première syllabe et les suivantes (cf. figure 5). L'expérience suivante cherche à savoir si l'identification du domaine correspond à un traitement catégoriel ou continu, et à déterminer la frontière morphologique dans le cas d'un traitement pseudo-catégoriel.

4.2. Expérience d'identification

Les stimuli utilisés pour cette expérience sont basés sur 6 phrases différentes, toutes construites sur la même structure syntaxique Sujet-Verbe-Objet. Un mot de chaque phrase porte le focus, soit le sujet, le verbe ou l'objet. La longueur des mots focalisés varie de 2 à 3 syllabes, tous les autres mots de la phrase étant monosyllabiques. Les phrases ont été enregistrées par un locuteur masculin francophone qui a produit le focus sur le mot en entier, ou bien seulement sur sa première syllabe. Un continuum est ensuite construit par analyse-synthèse, qui passe de l'une des phrases de chaque paire à la seconde grâce à un morphing des paramètres de F0, intensité et durée réalisé sur 10 pas (en utilisant le logiciel Praat pour modifier F0 et durée, et un script Matlab pour modifier l'intensité). Les différences introduites entre chacun de ces pas sont toutes en dessous du seuil de perception du glissando décrit par Rossi [8]. Cela donne finalement 6 continums de 10 stimuli différents.

Table 2. ANOVA réalisée sur les réponses des sujets au test d'identification du focus sur le mot ou sur la syllabe. Les facteurs pris en compte sont le pas du morphing (Itération), la longueur du mot focalisé, la position de ce mot dans la phrase et les 3 répétitions.

Facteur	ddl	F	p	sig.
Itération	9	64,129	,000	*
Longueur	1	4,790	,053	
Position	2	9,981	,001	*
Répétition	2	1,067	,363	
Itération * Long	9	14,345	,000	*
Itération * Position	18	7,168	,000	*
Long * Position	2	1,785	,194	
Iter.*Long* Pos	18	3,599	,000	*

Les 11 sujets écoutent chaque stimulus, dans un ordre aléatoire. L'ensemble des stimuli est présenté trois fois à chaque sujet. Ils doivent répondre si le focus qu'ils perçoivent correspond à un focus sur le mot ou sur la première syllabe du mot.

La cohérence des résultats inter-sujets est validée par un alpha de Cronbach significatif ($\alpha=0.84$). Ensuite, une ANOVA permet de mesurer l'influence relative des différents paramètres de ce test de perception sur les réponses des sujets : les différents pas du morphing, la longueur du mot focalisé, sa place dans la phrase (reliée à la fonction syntaxique du mot), et les trois répétitions. Les

effets principaux qui en ressortent (cf. table 2) permettent de souligner (1) l'effet primordial du morphing et (2) celui de la position du mot dans la phrase. Par ailleurs, (3) des interactions significatives sont relevées entre l'itération, la longueur du mot et sa position. Les trois répétitions n'ont aucun effet sur les réponses. Ces résultats confirment la capacité des auditeurs à percevoir le domaine du focus et donc à distinguer les deux fonctions sous-jacentes.

Cependant, l'effet important de la position du mot dans la phrase soulève d'autres questionnements sur l'influence de facteurs qu'il n'est pas possible de tester dans cette expérience, et nous conduit à vouloir poursuivre ces expériences plus avant. En particulier la nature phonologique des phonèmes (consonne voisée ou non) constituant la syllabe focalisée doit être contrôlée, et la position du mot focalisé découplée de sa fonction syntaxique. En particulier, on peut se demander si l'interaction significative entre la longueur du mot focalisé et l'itération est le reflet d'une plus grande facilité des sujets à traiter les mots quand ils sont plus longs. Mais comme la longueur n'est pas significative seule, et que les objets de deux syllabes reçoivent de bons résultats, cela pourrait aussi être dû à l'un des facteurs listés auparavant.

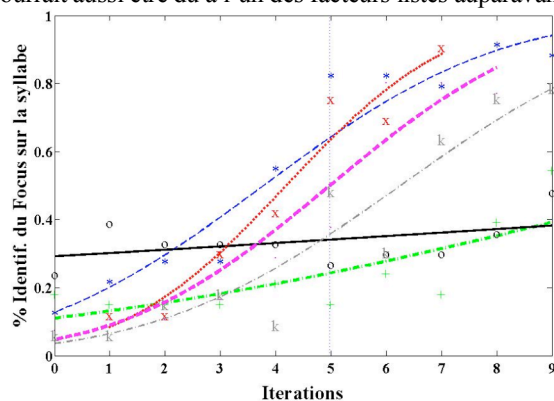


Figure 7. Résultats de l'analyse Probit. Les 6 courbes sont le résultat d'une interpolation des résultats moyens obtenus pour chaque continuum à chaque itération. Légende : (+) et tirés-points verts : sujet de 2 syllabes ; (*) et tirés bleus : objet de 2 syl ; (o) et ligne noire : verbe de 2 syl ; (k) et tirés-points gris : sujet de 3 syl ; (x) et pointillés rouge : objet de 3 syl ; (.) et tirés magenta : verbe de 3 syl. La ligne verticale indique la position de la frontière perceptive pour l'ensemble des courbes – sauf (b) et (c). En ordonnée l'itération et en indice le % d'identification du focus syllabique.

Afin de tester si les réponses des sujets sont de nature catégorielle ou non, une analyse Probit a été menée sur les résultats (cf. fig. 7). Cette analyse montre que 4 continums sur les 6 testés montrent un passage abrupt des réponses de focalisation sur le mot à celles sur la syllabe, autour de la cinquième itération. Pour les deux autres, la focalisation syllabique n'a pas, ou mal, été reconnue. Ces résultats sont cohérents avec ceux de l'ANOVA, et nous poussent à poursuivre ces recherches afin de mieux comprendre les paramètres qui peuvent faire que certains stimuli sont bien ou mal reconnus. Le résultat important de cette étude reste cependant que pour 4 phrases, les sujets répondent à un continuum d'un focus

sur le mot à un focus sur la syllabe comme s'il existait une frontière perceptive entre les deux.

5. CONCLUSIONS

Nous avons résumé ici plusieurs expériences perceptives qui confirment, après Ladd [7] que la prosodie du focus véhicule bien, à travers des indices statiques tonals traités dans une perception « quasi » catégorielle, la binarité présence ou absence de focus (Valence), mais ne peut cependant pas être réduite à cette seule valeur fonctionnelle, puisque des indices de gradience (traités seulement quand la catégorie également présence de focus est établie) informent sur l'Intensité du focus, qui ne peut pas exister en deçà d'une valence positive (qui est utilisable dans des fonctions communicatives comme la préférence), tandis que des indices dynamiques informent sur le segment focalisé, à travers un effet « quasi » catégoriel, dont le seuil devra cependant être établi plus systématiquement. Le modèle VID, qui reprend les seuils établis dans les expériences perceptives, est en cours d'implémentation et de validation dans le système de synthèse, pour une mise en œuvre dans le système de dialogue de FT R&D.

REMERCIEMENTS

Nous remercions particulièrement Philippe Bretier et Franck Panaget de France Telecom R&D, pour les discussions fructueuses qui nous ont permis en particulier de cerner les fonctions communicatives du focus prosodique.

BIBLIOGRAPHIE

- [1] Aubergé, V. 2001. Modélisation de la prosodie par formes globales : amont ou aval de la phonologie tonale ? 23rd JEP, France, 281-284.
- [2] Brichet, C., Aubergé V., 2002. La prosodie de la focalisation en français : faits perceptifs. 94-99, 24es JEP, Nancy.
- [3] Gussenhoven C (2003). Perceiving paralinguistic meaning. *Proceedings of Prosodic Interfaces 2003*, edited by Amina Mettouchi and Gâelle Ferré (eds). Université de Nantes. 47-49.
- [4] Jackendoff, R. 2002. *Foundations of language*, Oxford: Oxford University Press.
- [5] Kohler K. (2006). What is Emphasis and how is it coded ? Int Conf of Speech Prosody, Dresden.
- [6] Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories; Monkeys do not. *Perception & Psychophysics*, 50, 93-107.
- [7] Ladd, D.R. & Morton, R. 1997. The perception of intonational emphasis : continuous or categorical ? *Journal of Phonetics*, 25, 313-342.
- [8] Rossi, M. 1978. La perception des glissandos descendants dans les contours prosodiques. *Phonetica*, 35, 11-40.
- [9] Rossi, M. 1985. L'intonation et l'organisation de l'énoncé, *Phonetica*, 42, 135-153.