

Différenciation des mots de fonction et des mots de contenu par la prosodie : analyse d'un corpus trilingue de langage adressé à l'enfant et à l'adulte

Christelle Dodane*, Jean-Marc Blanc** & Peter Ford Dominey**

*Laboratoire Dynamique du Langage, UMR CNRS 5596
14 avenue Berthelot, 69 363 LYON Cedex 07 - France
dodane@isc.cnrs.fr

**Institut des Sciences Cognitives, UMR CNRS 5015
67 Bd Pinel, 69675 BRON Cedex – France
dominey@isc.cnrs.fr ; blanc@isc.cnrs.fr

ABSTRACT

This research investigated the role of salient prosodic cues in the first approximant assignment in function and content words by infants. In order to discover these cues and to see if they could vary cross-linguistically, infant-directed speech was compared to adult-directed speech in three different languages, French, English and Japanese. The same story was successively read by 15 mothers to their infant and to an adult (5 mothers per language). The acoustic analyses reveal that non-final syllable duration, Fo peaks and relative amplitude and amplitude peaks are particularly relevant in the three languages to allow categorization in function and content items, but they have a different relative weight across languages because of the specific prosodic organization of these languages.

1. INTRODUCTION

Afin de pouvoir commencer à élaborer le savoir syntaxique de sa langue maternelle, le bébé doit d'abord être capable de segmenter le signal de parole en constituants. Or, comment l'enfant peut-il commencer à segmenter le flux de parole alors que les différents paramètres acoustiques varient de façon continue et qu'il existe une très grande variabilité intra et interlocuteurs ? Les théories d'initialisation prosodique supposent que le signal de parole contient des signaux prosodiques réguliers et suffisamment saillants pour permettre à l'enfant d'accéder à une information grammaticale de type rudimentaire sur les principales catégories linguistiques de sa langue. La distinction entre mots de fonction et mots de contenu pourrait notamment guider l'enfant sur la route de l'initialisation syntaxique (Gleitman, [1]) et sémantique (Pinker [2]). Une segmentation de ce type serait très utile car, bien que les mots de fonction soient en nombre restreint, ils ont une très grande fréquence dans la parole. Leur repérage permettrait d'accéder à un grand nombre de frontières de mots (Christophe & al., [3] ; Ramus [4]). Il se trouve que les mots de fonction sont minimalisés dans leur forme parlée (Culter [5] ; Shi [6] ; Selkirk [7] ; Morgan & al. [8] ; Shi & al. [9]), et en particulier, au niveau prosodique (Selkirk [7]). Les deux classes de mots ont donc un poids perceptuel très différent, les mots de contenu étant prosodiquement plus saillants. Les bébés

sont sensibles à ces différences puisque dès la naissance, ils sont capables de discriminer les mots de contenu des mots de fonction (Shi & al. [10]) et à 6 mois, ils marquent une préférence pour les mots de contenu (Shi & Werker [11]) et se focalisent sur les mots ayant un fort poids sémantique (acquisition du lexique). La sensibilité aux mots de fonction apparaît plus tardivement, entre 8 et 13 mois (Shi & al. [12]). Si cette distinction entre mots de contenu et mots de fonction est importante pour l'enfant, il est probable qu'elle soit exagérée par les parents lorsqu'ils s'adressent à leurs enfants (langage adressé à l'enfant, désormais L.A.E.) dans un processus de « pédagogie inconsciente ». Le L.A.E. étant une modalité de langage où le niveau prosodique est largement exagéré (Papousek & Papousek [13] ; Fernald [14]), il est probable que les indices prosodiques pertinents pour une telle distinction soient également exagérés. Afin de déterminer quels indices prosodiques pourraient permettre aux bébés de différencier les mots de contenu des mots de fonction, il serait donc intéressant de comparer deux modalités de langage, le LAE et le langage adressé à l'adulte (désormais LAA). Par ailleurs, ces indices devraient diverger en fonction des caractéristiques du système prosodique de la langue étudiée. Nous proposons donc de comparer deux modalités de langage (LAA vs LAE), et ce, dans trois langues appartenant à des catégories prosodiques différentes, l'anglais (tendance à la rythmicité accentuelle), le français (tendance à la rythmicité syllabique) et le japonais (langue à rythmicité moraique). Nous établirons un relevé contrastif permettant d'évaluer le rôle de chaque indice au sein 1°) de chaque langue 2°) de toutes les langues étudiées 3°) en fonction des deux modalités étudiées.

2. MÉTHODOLOGIE

Participants

15 mères volontaires ont participé à cette étude, 5 mères françaises, 5 mères japonaises et 5 mères anglaises. Si toutes les mères ont été enregistrées à Lyon, les mères japonaises et les mères anglaises ne s'adressaient à leur enfant que dans leur langue maternelle. L'âge de leur enfant allait de 6 mois à 22 mois.

Enregistrements

Les mères ont toutes été enregistrées à leur domicile, par le père ou une personne familière afin d'obtenir les enregistrements les plus naturels possibles. Il était demandé aux mères de lire la même histoire à un adulte familier (modalité LAA), puis à leur enfant (modalité LAE). Les enregistrements étaient dissociés dans le temps afin que les modalités soient nettement différenciées. L'histoire était extraite d'un livre pour enfant français (Cousins [15]). Elle a été traduite en anglais et en japonais par des locuteurs natifs. Les enregistrements ont été réalisés avec un enregistreur minidisque (Sony MZN-910S) et un micro unidirectionnel Philips SBC-MD695. Chaque mot, chaque syllabe et chaque phonème a été étiqueté en fonction de leur appartenance à une catégorie grammaticale générique (fonction vs contenu) et une catégorie grammaticale spécifique (nom, verbe, adverbe, adjectif, déterminant, auxiliaire, conjonction, pronom et enclitique uniquement pour le japonais). Pour le japonais, le codage a été vérifié par un linguiste et locuteur natif du japonais.

Analyses acoustiques

Les phrases ont été échantillonnées à 22 kHz, 16 bits en mono. Les analyses acoustiques ont été réalisées avec le logiciel Praat. Le contour de Fo a été extrait, puis post-traité (suppression des sauts d'octave, lissage, interpolation). Un prosogramme a été édité (Mertens [16]) à partir du contour de Fo. Tous les événements mélodiques affichant une montée suivie d'une descente sont considérés comme des accents de hauteur (pics de Fo). La totalité des accents relevés grâce à cette méthode ont été vérifiés pour les trois langues par un phonéticien et musicien entraîné. La durée de la totalité des mots, des syllabes et des phonèmes (en ms) a été extraite automatiquement à partir des grilles de segmentation de Praat. Le contour d'amplitude a été extrait avec Praat et l'amplitude relative de chaque syllabe a été déterminée en calculant le rapport entre son énergie (RMS, root-mean square en dB) et l'énergie de la syllabe la plus intense de l'énoncé (Shi & al. [9]).

	Mots				Syllabes			
	F type	F tok.	C type	C tok.	F type	F tok.	C type	C tok.
Anglais	19	42	44	73	23	44	59	107
Français	21	44	42	66	22	33	71	128
Japonais	17	48	42	68	22	52	67	149

Table 1 : Nombres de types et de token (abrégié en tok.), de mots et de syllabes pour les mots de contenu (C) et les mots de fonction (F) en anglais, français et japonais.

Ces mesures ont permis l'extraction de 33 indices acoustiques différents, construits à partir des mesures de durée (mots, syllabes finales et non finales, voyelles, consonnes, durée moyenne des syllabes portant les pics de Fo), de Fo (Fo moyen, minimum et maximum en Hz, intervalles en 1/2 tons et en 1/2 tons par seconde sur les syllabes finales et non finales et les mots - Fo moyen des

pics de Fo, relevé des pics de Fo) et d'amplitude (amplitude relative, amplitude moyenne, minimum et maximum en dB, déviation standard sur les syllabes finales et non finales et les mots - relevé des pics d'amplitude).

Analyses statistiques

Des analyses de variance à deux facteurs et à trois facteurs (ANOVA à mesures répétées) ont été réalisées dans le but de déterminer l'influence de la modalité (1^{er} facteur : LAA - LAE), de la catégorie lexicale (2^{ème} facteur : mot de contenu - mot de fonction) et de la langue (3^{ème} facteur : français - anglais - japonais).

3. RÉSULTATS

3.1. Résultats et discussion

Résultats généraux

Parmi les 33 indices prosodiques étudiés, 26 se montrent significativement pertinents (effet de catégorie) pour différencier les items de contenu (désormais C) des items de fonction (désormais F). Les résultats convergent dans les trois langues pour montrer que les C sont prosodiquement plus saillants que les F. Cependant, certains indices jouent un rôle plus important dans cette différenciation. C'est notamment le cas des indices de durée, des pics de Fo, des pics d'amplitude et de l'amplitude relative. Pour cette raison, nous nous limiterons à la présentation de ces indices. Par ailleurs, le rôle respectif de ces indices varie en fonction de la langue et notamment, de son organisation prosodique. Enfin, on relève un effet de la modalité (LAA vs LAE) dans les trois langues, même si cet effet diffère là aussi, en fonction de la langue.

	Anglais	Français	Japonais
Durée des mots	p<,0000*	p<,0002*	p<,0000*
Durée SNF	p<,0004*	p<,0030*	p<,0811
Pics de Fo	p<,0003*	p<,0000*	p<,0002*
Amplitude relative	p<,0020*	p<,0003*	p<,0018*
Pics d'amplitude	p<,0002*	p<,0002*	p<,0000*

Table 2 : Significativité des indices les plus saillants dans les trois langues étudiées en fonction de la catégorie (F/C)

Durée des syllabes

En anglais et en français, la durée des syllabes appartenant aux mots de C est significativement plus grande que celles appartenant aux mots de F. Ainsi, en anglais, on relève une augmentation de 127 ms de F à C en LAA et de 154 ms de F à C en LAE (effet de catégorie : p<,0000* ; effet de modalité : p<,0032*). En français, on relève également une augmentation de durée, mais bien moins marquée qu'en anglais avec une augmentation de 44 ms de F à C en LAA et de 62 ms de F à C en LAE (effet de catégorie : p<,0009* ; effet de modalité : p<,0142*). En revanche, pour le japonais, les syllabes de F sont plus longues que

les syllabes de C ; en effet, on relève une baisse de 15 ms de F à C en LAA et de 39 ms de F à C en LAE (pas d'effet de catégorie). Ces résultats s'expliquent par le fait que la majorité des énoncés japonais dans notre corpus se termine par une particule enclitique, c'est-à-dire un mot de F. Par ailleurs, l'allongement final (AF) est spécialement marqué en LAE pour les trois langues de notre corpus. Ainsi, en anglais, la durée moyenne des syllabes finales (SF : syllabes précédant une pause) est de 484 ms en LAA et de 619 ms en LAE, soit une augmentation de 132 ms d'une modalité à l'autre. En français, on relève une durée moyenne de 312 ms en LAA et de 442 ms en LAE, soit une augmentation de 130 ms et en japonais, une durée moyenne de 272 ms en LAA et de 398 ms en LAE, soit une augmentation de 126 ms. Du point de vue perceptif, l'AF est particulièrement utile pour le bébé car il lui permet de segmenter le signal de parole en constituants de grande taille (Koponen & Lacerda [17]), les propositions et les phrases. Cependant, les différences de durée sont également importantes pour repérer des constituants plus petits, tels que les mots F ou de C. L'AF venant perturber la mesure de la durée des syllabes, nous avons choisi de ne conserver que les syllabes non finales (SNF), c'est-à-dire celles qui ne portent pas l'AF (figure n°1).

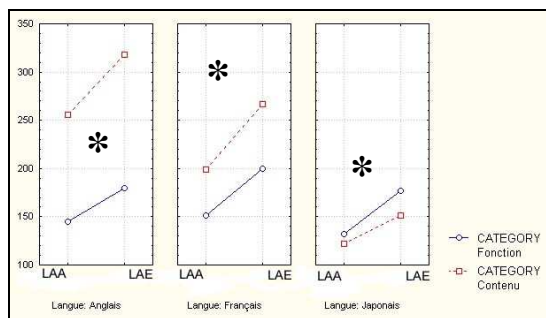


Figure 1 : Durée des finales non finales (en ms) des mots de F et des mots de C en LAA et LAE avec de gauche à droite, l'anglais, le français et le japonais.

Ainsi, dans les trois langues, les SNF des C sont significativement plus longues que dans les F. Cette différence est beaucoup plus marquée en anglais (augmentation de 63 ms en LAA et de 77 ms en LAE), qu'en français (augmentation de 33 ms en LAA et de 25 ms en LAE) et en japonais (augmentation de 15 ms en LAA et de 19 ms en LAE). Cette grande différence de durée en anglais s'explique par le fait que les accents sont localisés sur les mots de C. Tandis que les syllabes portant l'accent sont dilatées au niveau temporel, les syllabes F, qui sont atones, sont compressées. Si l'on pouvait prévoir ce type de résultat en anglais en raison de la rythmicité accentuelle de cette langue, les résultats sont plus surprenants en français. En effet, bien que le français soit décrit comme une langue dont les syllabes non finales ont une tendance à l'isochronie, les SNF des C sont plus longues (+ 33 ms en LAA et + 25 ms en LAE). En ce qui concerne l'effet de la modalité, il est significatif en français ($p < .0173^*$) et en anglais ($p < .0055^*$), mais pas en japonais ($p < .0786$).

Pics de Fo

Les mots de C sont mis en relief par la présence très fréquente d'une montée intonative ou d'un pic de Fo (qu'on relève en moyenne dans plus de 60% des mots de C en position non finale). Cet indice se révèle hautement significatif dans les trois langues (voir figure n°2) et il semble particulièrement important en français ($p < .0000^*$) et en japonais ($p < .0002^*$). Pour le français, ces résultats rejoignent ceux de Blanc & al. [18] qui montrent qu'un réseau de neurones est capable de reconnaître 83,6 % (corpus LSCP) et 70,3 % (corpus Multext) des mots de C à partir du seul contour de Fo. En revanche, on ne relève aucun effet de modalité dans les trois langues. La proportion de pics de Fo est la même en LAA et en LAE, leur répartition dépendant de la structure prosodique et linguistique des énoncés. S'il y a bien un effet de modalité, il concerne la hauteur des pics de Fo. Ainsi, les pics sont significativement plus élevés en LAE qu'en LAA en français (337 Hz en LAA ; 481 Hz en LAE) et en anglais (324 Hz en LAA ; 400 Hz en LAE). En revanche, l'effet de modalité n'est pas significatif en japonais (225 Hz en LAA ; 284 Hz en LAE).

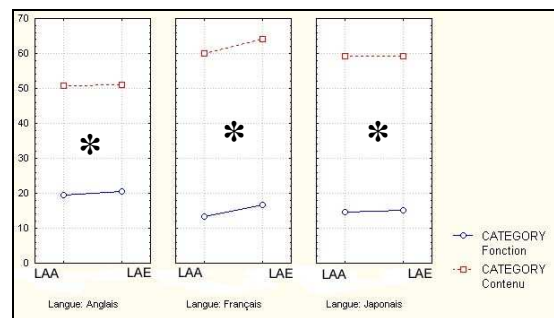


Figure 2 : Proportion (en %) de pics de Fo correspondant aux mots de C et aux mots de F en LAA et LAE, avec de gauche à droite, l'anglais, le français et le japonais.

Amplitude relative et pics d'amplitude

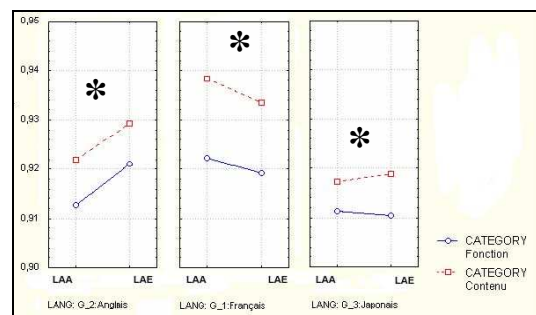


Figure 3 : Amplitude relative des mots de C et des mots de F en LAA et LAE, avec de gauche à droite, l'anglais, le français et le japonais.

L'amplitude relative des SNF est significativement plus élevée sur les C que sur les F en anglais, en japonais et spécialement en français. Ces résultats vont dans le même sens que ceux de Shi & al. [9] pour le Mandarin et le Turc. En revanche, l'effet de la modalité est différent en fonction des langues (figure n°3). En français, l'amplitude décroît pour les F et les C en passant du LAA au LAE. En

anglais, c'est le contraire puisque l'amplitude croît pour les F et les C en passant du LAA au LAE. Enfin, en japonais, il n'y a pas de différence marquée entre le LAA et le LAE. Par ailleurs, la majorité des pics d'amplitude sont localisés sur les mots de C. En effet, plus de 70% des pics sont localisés sur les mots de C dans les trois langues, contre moins de 30% sur les F. Cet indice est spécialement marqué en japonais et en français.

4. CONCLUSION

Les indices étudiés révèlent que les mots de C sont prosodiquement beaucoup plus saillants que les mots de F en français, en anglais et en japonais. Ainsi, la durée des SNF est plus longue pour les C, en particulier pour l'anglais. Les mots de C sont mis en relief par la présente très fréquente de pics de Fo, spécialement en français et en japonais. Enfin, l'amplitude relative est plus élevée sur les syllabes de C et les pics d'amplitude sont le plus souvent localisés sur les mots de C. Dans notre corpus, la « minimalisation » acoustique des F relevée en français, en anglais et en japonais rejoint les résultats de Shi & al. [9] pour l'anglais, le turc et le mandarin. Cependant, il serait nécessaire de compléter cette étude par une analyse

fine et détaillée des phénomènes accentuels tels que l'accent secondaire en français et des phénomènes d'alignement tonal dans les trois langues. Il serait également intéressant de tester chacun de ces indices avec un réseau de neurones, comme nous l'avons déjà fait pour les pics de Fo [18], afin de pouvoir hiérarchiser leur importance au sein de chaque langue, ainsi que pour toutes les langues étudiées. Par ailleurs, nos résultats font apparaître une augmentation des paramètres prosodiques lorsque l'on passe de la modalité LAA à la modalité LAE et ce, dans les trois langues étudiées. Cependant, tandis qu'on observe une exagération prosodique beaucoup plus forte en anglais, le LAE est beaucoup plus « modéré » en japonais, le français se situant dans une position intermédiaire. Ces variations trouvent sans doute leur origine dans des différences culturelles liées à l'utilisation du LAE dans les langues étudiées.

Nous tenons à remercier Reiko Vacheret pour son aide précieuse concernant l'analyse du japonais, ainsi que toutes les mamans et les bébés qui ont participé à notre étude. Cette étude a été financée par le programme « Human Frontier Science Program ».

BIBLIOGRAPHIE

- [1] L. Gleitman. The structural sources of verb meanings. *Language Acquisition*, 1, 3-55, 1990.
- [2] S. Pinker. *Language learnability and language development*. Cambridge: Harvard Univ. Press, 1984.
- [3] A. Christophe, T. Guasti, M. Nespor, E. Dupoux, and B. Van Ooyen. Reflections on phonological bootstrapping: its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, 12, (8/6), 585-612, 1997.
- [4] F. Ramus. *Rythmes des langues et Acquisition du Langage*. Thèse de Doctorat, Paris, EHESS, 1999.
- [5] A. Cutler. Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistics Research*, 22, 109-131, 1993.
- [6] R. Shi. Perceptual correlates of content words and function words in early language input. PhD. Brown University, Providence, 1995.
- [7] E. Selkirk. The prosodic structure of function words. In Morgan & Demuth, *Signal to Syntax*, 187-213, 1996.
- [8] J. Morgan, R. Shi & P. Allopenna, P. Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In Morgan & Demuth, *Signal to Syntax*, 263-283, 1996.
- [9] R. Shi, J. Morgan & P. Allopenna. Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, 25, 169-201, 1998.
- [10] R. Shi, J. Weker & J. Morgan. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72, B11-B21, 1999.
- [11] R. Shi & J. Werker. Six-month-old infants' preference for lexical words. *Psychological Science*, 12/1, 70-75, 2001.
- [12] R. Shi, J. Werker & A. Cutler, A. Function words in early speech perception. Proc. of the XVth ICPhs, Barcelona, 3309-3012, 2003.
- [13] M. Papousek & H. Papousek. Musical elements in the infant's vocalizations: Their significance for communication, cognition and creativity. In Lipsitt & Rovee-Collier, 163-224, 1981.
- [14] A. Fernald. Intonation and communication intent in mother's speech to infants: is the melody the message? *Child Development*, 60, 1497-1510, 1989.
- [15] L. Cousins. En route avec Mimi. Albin Michel, 2001.
- [16] P. Mertens. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des Langues*, 45(2), 109-130, 2004.
- [17] E. Koponen & F. Lacerda. Final lengthening in infant-directed speech may function as a cue to phrase constituents. *Phonum*, 9, 9-12, 2003.
- [18] Blanc, J.M., Dodane, C. & Dominey, P.F. Temporal Processing for Syntax Acquisition: A simulation Study. Proc. XXVth Ann. Meet. Cognitive Science Society, 2003.