

Étude diachronique de l'accent initial au travers d'archives audio

Philippe Boula de Mareuil¹, Albert Rilliard¹ & Alexandre Allauzen^{1,2}

¹ LIMSI-CNRS, BP 133, 91403 Orsay CEDEX, France

² Univ. Paris-Sud, 91405, Orsay, France

{mareuil;rilliard;allauzen}@limsi.fr

<http://www.limsi.fr>

ABSTRACT

This study makes use of advances in automatic speech processing to analyse French audiovisual archives. A 10-hour corpus spanning over five decades of broadcast news is investigated from the angle of prosody evolution. This paper focuses on word-initial stress, which may give an impression of emphatic style. Our measurements suggest that the following features as well as mean pitch have decreased since the forties: pitch rise and vowel duration associated with initial stress. In the meantime, the onsets of supposedly stressed initial syllables have become longer while average speech rate has not changed. This puzzling outcome rises interesting questions for research on French prosody.

Keywords: prosody, diachrony, phonostylistics

1. INTRODUCTION

Il semble que nous soyons capables de reconnaître un enregistrement fait il y a quelques décennies et de le distinguer d'un enregistrement contemporain. Les conditions techniques telles que la distance du micro et son type ont certes évolué. Elles sont au moins en partie responsables d'une qualité de voix particulière qui est immédiatement perceptible et peut facilement être caricaturée. Des traits linguistiques plus ponctuels, peut-être plus subtils mais néanmoins saillants peuvent également avoir changé. Nous nous concentrons ici sur la prosodie du français. L'objectif général de cette étude est de démêler les paramètres qui permettent de caractériser un style d'annonceur qui date des années 1940, par exemple, et de l'opposer au style journalistique actuel. On dispose maintenant d'archives couvrant plus de cinquante ans d'informations audiovisuelles qui rendent possible une telle recherche.

Bien que le français soit traditionnellement décrit comme possédant un accent de fin de groupe, de nombreuses études théoriques, expérimentales et appliquées ont mis en évidence l'émergence d'un accent initial de mot en français contemporain [6][9][17][8][15]. Cet accent initial est complémentaire de l'accent final (qui reste une propriété essentielle du système accentuel français), mais se répand dans les styles journalistique et didactique, dans les informations radio/télédiffusées, les conférences publiques et les salles de classe. Cette mutation, dont l'origine est difficile à retracer avec précision, remonterait à la fin du XIX^e siècle voire plus tôt [7]. En accord avec elle, les modèles phonologiques récents de

la prosodie du français [7][18][13] ont intégré un double marquage des mots pleins par un accent initial et un accent final, le premier étant essentiellement mélodique et le dernier caractérisé par l'allongement [18][3]. Dans certains contextes, l'accent initial sous-jacent peut être réalisé en surface par un accent emphatique (*accent d'insistance*) avec des patrons mélodiques plus dynamiques et un allongement plus accusé de l'attaque de la syllabe proéminente. Tandis que l'accent emphatique a une fonction pragmatique ou paralinguistique propre, l'accent initial non-emphatique a plutôt une fonction rythmique motivée par des contraintes d'eurythmie favorisant des alternances d'accents bien équilibrées [12] : des suites trop longues de syllabes inaccentuées sont ainsi évitées. Cependant, les descriptions phonétiques proposent rarement des corrélats acoustiques clairs qui permettraient de distinguer les syllabes portant un accent initial de leurs contreparties inaccentuées, de façon satisfaisante et sans recourir à l'interprétation. Il est admis que la distinction peut être dépendante de l'auditeur et du locuteur [20][10], et de nombreux facteurs influents demandent à être examinés.

Suivant des critères proposés par [19] notamment, notre propre approche de l'accent initial concerne principalement les séquences clitique non-clitique. Une telle séquence correspond au tronçon de mots (*chunk*) le plus fréquent, et constitue un bon candidat pour un accent initial sur le mot non-clitique. Nous avons considéré les suites clitique non-clitique de manière extensive dans le but de faire apparaître des différences prosodiques avec le temps. Cette étude empirique vise d'abord à évaluer si oui ou non l'accent initial gagne du terrain dans le style journalistique, à travers des archives audiovisuelles de 1940 à 1995. D'un côté, une façon de parler emphatique est typique des enregistrements anciens [14] ; de l'autre, l'accent initial est patent dans la diction des personnalités journalistiques et politiques d'aujourd'hui. Notre but est double : vérifier si un changement linguistique est en cours (avec une tendance croissante ou décroissante à l'accent initial) et dégager des traits de prononciation discriminants entre styles.

Après une description du corpus (section 2), la section 3 donne des résultats sur l'accent initial de mot à travers la montée mélodique et l'allongement de l'attaque ou du noyau. Enfin, la section 4 tire quelques conclusions et envisage de possibles perspectives.

2. CORPUS ET MÉTHODE

Le corpus utilisé a été collecté lors du projet européen ECHO [4]. Une partie des données provient de documents d'actualités cinématographiques diffusés pendant la Seconde Guerre Mondiale. La plus grande partie provient du projet EURODELPHES et couvre la période de 1945 à 1995. Ce second projet avait pour but de créer un manuel scolaire d'histoire multimédia sur les évolutions de l'Europe depuis la fin de la Guerre. L'ensemble du corpus est constitué de 189 documents audiovisuels, chacun contenant au moins un locuteur et durant de 20 secondes à 20 minutes. Les documents ont été transcrits manuellement au niveau orthographique, puis segmentés automatiquement au niveau phonétique au moyen de l'alignement. Celui-ci utilisait des modèles acoustiques indépendants du contexte entraînés de façon extensive et un dictionnaire de prononciation spécifiquement adapté au corpus. La méthode, décrite dans [1], a été validée dans plusieurs publications depuis lors [2]. Des valeurs de pitch ont été attribuées à chaque phonème en moyennant les mesures de fréquence fondamentale (F_0) prélevées toutes les 10 ms par le logiciel PRAAT (www.praat.org) avec les options standard — les valeurs de F_0 en dessous de 75 Hz apparaissent comme non-définies. Une représentation élémentaire mais robuste de la prosodie est ainsi obtenue, où chaque phonème est caractérisé par sa durée, son pitch moyen, le mot associé et d'autres informations facilitant les traitements ultérieurs. En complément, d'autres stylisations de la mélodie sont envisageables et même souhaitables pour mieux prendre en compte la forme des contours de F_0 [19][16], mais elles n'ont pas été appliquées ici.

Pour cet article, les données ont été regroupées en 4 périodes afin d'avoir une répartition équilibrée : 1940–1959, 1960–1969, 1970–1979 et 1980–1995. Les années 60 et 70 sont en effet surreprésentées dans notre corpus. La table 1 fournit pour chaque période la durée totale des séquences audio, la durée moyenne des phonèmes qui est assez comparable entre les périodes retenues, les pourcentages de voyelles détectées comme non-voisées (lesquels, fait rassurant, sont relativement faibles) et le pitch moyen. Ce dernier est restreint aux locuteurs masculins en raison de la faible présence féminine dans les documents d'archives.

Table 1 : Durée des données, durée moyenne des phonèmes, pourcentage de voyelles non-voisées et pitch moyen (restreint aux hommes).

	durée (s)	dur.phon (ms)	V non-voisées (%)	Pitch moyen (Hz)
1940–1959	7434	79	4.8	175
1960–1969	11552	81	3.2	144
1970–1979	6053	83	4.7	141
1980–1995	9403	80	3.5	137

Il est notable que le pitch des hommes est plus élevé dans les années 40 et 50 et qu'il ne cesse de décroître après ces années. Ceci a également pour effet que les triangles vocaliques des années 40 et 50 sont plus

grands que ceux des autres périodes, lesquels sont très similaires (l'estimation des formants a été effectuée avec PRAAT comme dans [10]). Le pitch plus élevé peut s'expliquer par les conditions d'enregistrement qui obligeaient les journalistes à parler fort, debout face au micro. Cela ne signifie pas nécessairement que les journalistes parlaient plus haut. Le support original des enregistrements (disque ou cassette) peut avoir un impact sur des paramètres comme le pitch moyen ; mais c'est ce que nous recevons quand nous écoutons des documents d'archives anciens. En outre, les paramètres plus locaux tels que ceux que nous allons expliciter dans ce qui suit ne devraient pas être affectés.

3. RÉSULTATS SUR L'ACCENT INITIAL

3.1. Montée mélodique

La variation de hauteur (ΔF_0 en demi-tons) a été mesurée entre la dernière voyelle d'un clitique et la première voyelle du non-clitique subséquent. Une question se pose d'emblée : que met-on sous l'étiquette clitique ? Sur la base de connaissances antérieures, un premier ensemble de quelque 300 mots outils a été établi, incluant des formes de verbes auxiliaires comme *être* et *avoir*. Sur la base des mots les plus fréquents de notre corpus, une seconde liste de 30 mots outils a été élaborée : *le, la, les, un, une, du, des, de, à, pour, en, dans, et, que, qui, est*, etc. Les négations *pas* et *plus* ont été exclues car il est difficile de les considérer comme des clitiques. On obtient 23 000 séquences clitique non-clitique avec le premier ensemble et 21 000 avec le second. C'est ce dernier ensemble qui a été conservé car, mieux contrôlé, il n'en offre pas moins une large couverture. L'inconvénient est que des séquences telles que *ne plus* se voient dès lors rangées parmi les séquences clitique non-clitique. Il est donc important de se concentrer sur les séquences de mots de type clitique polysyllabe, ce qui de plus évite la confusion entre accents initial et final dans les monosyllabes.

La table 2 donne le nombre de contextes correspondant à notre sélection ainsi que le pourcentage cumulé des cas où la différence de F_0 (polysyllabe moins clitique) est supérieure à 1, 2, 3 ou 4 demi-tons (dt). La distribution non-cumulée est représentée figure 1.

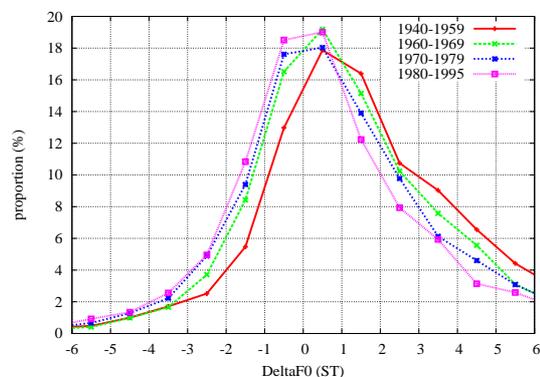


Figure 1 : ΔF_0 entre la voyelle de la voyelle initiale du polysyllabe et du clitique qui précède (en demi-tons).

Bien sûr, une différence de F_0 plus importante en valeur absolue conduit à un pourcentage d'occurrences

plus bas. Cependant, pour chaque seuil, la hiérarchie est la même. La figure 1 montre que plus le document est récent, plus la courbe est basse pour les valeurs positives (droite) et plus elle est haute pour les valeurs négatives (gauche).

Dans la table 2, le pourcentage de ΔF_0 supérieur à 1 dt, par exemple, est de plus de 50 % pour les années 40 et 50, et décroît avec le temps. À la suite de [19] et d'études plus récentes sur le français [11], le seuil de 3 dt est considéré comme une bonne estimation des corrélats acoustiques de la proéminence prosodique. D'après cette interprétation, plus d'une séquence clitique non-clitique polysyllabique sur quatre donne lieu à une proéminence prosodique dans les années 40 et 50, contre moins d'une sur cinq dans les années 80 et 90. Cette proportion augmente sensiblement avec la taille du polysyllabe dans les années 40 et 50 : pour 26 % des dissyllabes, 31 % des trisyllabes et 32 % des tétrasyllabes, la différence de F_0 avec le clitique précédent excède 3 dt. Cependant, cette tendance ne se confirme pas pour les décennies suivantes.

Table 2 : nombre de contextes pour les séquences clitique non-clitique polysyllabique avec la proportion des cas où la différence de F_0 excède 1, 2, 3 et 4 dt.

	#cntxts	%>1dt	%>2dt	%>3dt	%>4dt
1940–1959	2104	55	39	28	19
1960–1969	4154	47	32	21	14
1970–1979	2107	43	29	20	14
1980–1995	3793	38	26	18	12

Nous n'avons observé aucune tendance claire à l'allongement des voyelles initiales de polysyllabes par rapport au clitique précédent. Par exemple, les pourcentages de différences de durée supérieures à 20 ms sont respectivement de 61 %, 53 %, 55 % et 57 % pour les périodes 1940–1959, 1960–1969, 1970–1979 et 1980–1995. Cependant, d'autres patrons de durée peuvent avoir évolué depuis la Seconde Guerre Mondiale, et c'est ce sur quoi nous allons nous pencher à présent.

3.2. Allongement attaque / noyau

La durée des attaques et voyelles initiales de polysyllabes précédés d'un clitique a été calculée dans les mêmes contextes, en appliquant les règles de syllabation proposées par [1]. Par exemple dans la séquence *de puissants*, l'attaque est [pɥ] et le noyau vocalique est [i]. La table 3 montre que la durée des attaques augmente avec le temps. L'augmentation est même plus régulière si on restreint l'analyse aux attaques simples (constituées d'une seule consonne). Les années 40 et 50 diffèrent particulièrement des décennies suivantes, pour lesquelles plus de 50 % des attaques durent plus de 70 ms. D'autres résultats pour les seuils de 60 ms et 80 ms sont rapportés dans la table 3 (ces seuils sont centrés autour des durées moyennes). Cet allongement de l'attaque contredit la tendance suggérée précédemment : on aurait plutôt prédit l'inverse si l'allongement est un corrélat de l'accent initial [15] [12] [3].

Table 3 : durée des attaques des polysyllabes précédés d'un clitique (valeur moyenne et pourcentage d'occurrences supérieures à un seuil donné).

	Toutes les attaques			Attaques simples		
	durée moyenne	%>70 ms	%>80 ms	durée moyenne	%>60 ms	%>70 ms
1940–1959	70ms	49	37	63ms	58	42
1960–1969	77ms	60	47	71ms	69	54
1970–1979	80ms	62	49	72ms	70	56
1980–1995	79ms	63	50	73ms	71	58

Une autre interprétation est que l'importance relative des corrélats de l'accent initial a pu changer en un demi-siècle. Quoi qu'il en soit, un changement est notable, alors que les débits de parole sont quasiment égaux tout au long de cette période (cf. table 1). Ce changement est comparé à l'allongement vocalique dans ce qui suit.

Comme le montre la table 4, la durée des voyelles dans les contextes clitique non-clitique polysyllabique décroît de 1940 à 1995. L'évolution est parallèle si on considère les seules voyelles plus hautes d'au moins 3 dt que la voyelle du clitique précédent. Les durées moyennes et les pourcentages sont alors naturellement plus élevés que si l'ensemble des voyelles est pris en compte, et le fait que les voyelles accentuées sont un peu plus courtes dans les années 60 que dans les années 70 est plus marqué.

Table 4 : durée des voyelles initiales de polysyllabes précédés d'un clitique (moyenne et pourcentage d'occurrences supérieures à un seuil donné). La partie droite est limitée aux voyelles plus hautes d'au moins 3 dt que la voyelle du clitique précédent.

	Toutes les voyelles			Voyelles > 3dt		
	durée moyenne	%>70 ms	%>80 ms	durée moyenne	%>70 ms	%>80 ms
1940–1959	79ms	61	48	85ms	66	55
1960–1969	73ms	51	40	75ms	56	42
1970–1979	73ms	52	38	79ms	58	47
1980–1995	71ms	47	34	78ms	54	39

Finalement, l'augmentation de la durée des attaques et la diminution de la durée des noyaux, sur la période étudiée, rendent stable la durée des syllabes supposées accentuées. Ceci laisse plus de temps pour des excursions de pitch dans les données de 1940–1959. Même si d'autres interprétations restent ouvertes, la nôtre est que l'accentuation initiale a diminué depuis les années 40.

4. CONCLUSION ET PERSPECTIVES

Ainsi, le traitement de la parole (l'alignement automatique et l'extraction de F_0) a permis une approche dirigée par les données des changements linguistiques, en particulier des changements prosodiques dans le style journalistique français de la fin du XX^e siècle. Dans ce contexte assez bien circonscrit, nous avons pu identifier des formes spécifiques aux différentes époques, et mener une étude

quantitative qui va au-delà des descriptions impressionnistes habituelles. Les résultats montrent une diminution du pitch moyen et de l'accent initial de mot (au moins en ce qui concerne ses corrélats mélodiques), accent qui était plus marqué dans les années 40 et 50. Ces deux décennies sont celles qui diffèrent le plus des autres.

Alors que la vitesse d'élocution n'a pas évolué au fil des années, l'augmentation observée de la durée de l'attaque de ce qui peut être considéré comme des syllabes portant un accent initial pose d'intéressantes questions sur la prosodie du français. Malgré la quantité de données manipulées, des analyses statistiques sont probablement nécessaires pour comprendre cette évolution en apparence paradoxale des corrélats temporels liés à l'attaque et des corrélats mélodiques de syllabes initiales supposées accentuées. En tout état de cause, des différences moyennes de durée de 10 ms ne sont pas négligeables, mais aboutissent certainement à moins de cas audibles que des différences de F_0 de 3 demi-tons. En comparaison, le seuil minimum pour une différence perceptible de durée est estimé à 20 ms [5].

Le traitement automatique de la parole permet de sélectionner des passages comportant des accents initiaux bien audibles. Des tests d'écoute informels utilisant la modification/resynthèse de prosodie peuvent révéler combien les paramètres de durée et de F_0 contribuent au caractère daté de certains enregistrements. Des expériences perceptives plus poussées sont en cours afin d'évaluer l'importance relative de la prosodie dans la caractérisation de styles de parole plus ou moins anciens et démodés. Nous espérons qu'elles permettront de démêler la contribution de l'allongement de l'attaque et celle de la montée mélodique à la perception de l'accent initial en français.

5. REMERCIEMENTS

Nous sommes très reconnaissant envers Laurent Vinet et l'Institut National de l'Audiovisuel (INA) pour le corpus audiovisuel et sa transcription, rendus disponibles dans le cadre des projets EURODELPHES et ECHO. Merci aussi à Martine Adda-Decker et Cécile Woehrling pour leur aide précieuse à cette recherche.

BIBLIOGRAPHIE

- [1] M. Adda-Decker, P. Boula de Mareüil, G. Adda, L. Lamel. Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, 46(2):119–139, 2005.
- [2] M. Adda-Decker, De la reconnaissance automatique de la parole à l'analyse linguistique de corpus oraux, *JEP*, Dinard, pages 389–400, 2006
- [3] C. Astésano, E.G. Bard, A. Turk, A. Functions of the French initial accent: a preliminary study. *Speech Prosody*. Aix-en-Provence, pages 139–142, 2002.
- [4] C. Barras, A. Allauzen, L. Lamel, J.-L. Gauvain. Transcribing audio-video archives. *ICASSP*. Orlando, pages 13–16, 2002.
- [5] K. Bartkova, C. Sorin. A model of segmental duration for speech synthesis in French. *Speech Communication*, 6(3):245–260, 1987.
- [6] F. Carton, A. Marchal, D. Hirst, A. Séguinot. *L'accent d'insistance (Emphatic stress)*. Didier, Montréal, 1977.
- [7] A. Di Cristo. Le cadre accentuel du français contemporain : essai de modélisation. *Langues*, 2(3):184–205, 1999.
- [8] G. Fant, A. Kruckenberg, L. Nord. Durational correlates of stress in Swedish, French and English. *Journal of Phonetics*, 19:351–365, 1991.
- [9] I. Fónagy, P. Léon. *L'accent en français contemporain*. Didier, Montréal, 1980.
- [10] C. Gendrot. Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français. Thèse de doctorat. Université de Paris III, 2006.
- [11] J.-P. Goldman, M. Avanzi, A.-C. Simon, A. Lacheret, A. Auchlin. A methodology for the automatic detection of perceived prominent syllables in spoken French. *Interspeech*. Anvers, pages 98–101, 2007.
- [12] L. Jankowski, C. Astésano, A. Di Cristo. The initial rhythmic accent in French: Acoustical and perceptual prosodic cues. *ICPhS*. San Francisco, pages 257–260, 1999.
- [13] S.-A. Jun, C. Fougeron. A phonological model of French intonation. In *Intonation: Theories and Models*, A. Botinis (ed.). Kluwer Academic Publishers, Dordrecht, pages 209–242, 2000.
- [14] C. Méadel. *Histoire de la radio des années trente*. INA/ANTHROPOS, Paris, 1994.
- [15] P. Mertens. Accentuation, intonation et morpho-syntaxe. *Travaux de linguistique*, 26:21–69, 1993.
- [16] P. Mertens. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. *Speech Prosody*. Nara, pages 23–26, 2004.
- [17] V. Padeloup. Modèle de règles rythmiques du français appliqué à la synthèse de la parole. Thèse de doctorat. Université de Provence, 1990.
- [18] M. Rossi. L'intonation, le système du français : description et modélisation. Ophrys, Paris, 1999.
- [19] J. 't Hart, R. Collier, A. Cohen. A perceptual study of intonation: an experimental-phonetic approach to speech melody. Cambridge University Press, Cambridge, 1991.
- [20] J. Vaissière. Language-independent prosodic features. In *Prosody: Models and measurements*, A. Cutler & D.R. Ladd (eds). Springer-Verlag, Berlin, pages 53–66, 1983.