

# ***Phonostylographe* : un outil de description prosodique. Comparaison du style radiophonique et lu<sup>1</sup>**

Jean-Philippe Goldman<sup>1</sup>, Antoine Auchlin<sup>1</sup>,  
Anne Catherine Simon<sup>2</sup>, Mathieu Avanzi<sup>3</sup>

<sup>1</sup>Département de Linguistique (Université de Genève), <sup>2</sup>Département d'Études romanes (UCLouvain), <sup>3</sup>Institut de philologie Romane et de Linguistique Française (Université de Neuchâtel)  
<goldman@lettres.unige.ch, auchlin@lettres.unige.ch,  
anne-catherine.simon@uclouvain.be, mathieu.avanzi@unine.ch>

## **Résumé**

*The paper presents a software for prosodic and phonostylistic description and its application to a study comparing radiophonic and read speech. This quasi-automatic and modular tool consists of a set of Praat-based acoustic tools like phonetic segmentation, melodic stylisation and prominence detection. A phonostylistic voice report – called phonostylographe – is produced on the basis of an audio file and its orthographic transcription. In this paper, we use this tool to identify phonostylistic properties of chronicles of a French public radio: three two-minutes-long recordings are compared with plain neutral reading of the same texts. Results confirm our initial hypotheses about phonostylistic distinctive features of Radio-France's chronicles – leaving out questions open to further study.*

## **1. Méthodes et outils d'analyse prosodique**

L'analyse de la prosodie concerne plusieurs dimensions (intonation, accentuation, rythme, qualité vocale). Chaque dimension requiert une représentation ou une transcription spécifique (contours intonatifs, types d'accents, patrons rythmiques). Une représentation proche de la substance acoustique mesure des paramètres (f0, durée, intensité...) et les interprète en fonction de variables externes<sup>2</sup>. Une transcription symbolique se limite aux variations prosodiques fonctionnelles qu'elle représente au moyen d'un ensemble fermé de symboles (tons,

<sup>1</sup> Nous remercions le Fonds National Suisse de la Recherche Scientifique (subside n° 100012-113726/1) ainsi que le Fonds national belge de la Recherche scientifique (FRFC n° 2.4523.07).

<sup>2</sup> Par exemple, comparer le registre tonal moyen des présentateurs de journaux télévisés en France et au Québec (Ménard 2003).

contours ou accents). Chaque type de transcription peut-être produite manuellement (par un expert) ou (semi)automatiquement (à l'aide de logiciels).

Nous présentons des outils pour analyser la prosodie de la manière la plus automatique et la plus fiable possible. Ces outils, développés sous Praat (Boersma & Weenink 2007), permettent de :

- styliser la courbe de  $f_0$  (Mertens 2004), c'est-à-dire en donner une représentation simplifiée qui correspond à la perception ;
- aligner et segmenter le signal sonore en phonèmes, syllabes et mots graphiques (Goldman 2007) pour interpréter les variations prosodiques en fonction de leur ancrage dans la chaîne parlée ;
- détecter de manière automatique les syllabes proéminentes (Avanzi et al. 2007) ;
- calculer automatiquement des paramètres comme le temps de parole et le temps d'articulation, le débit, la durée syllabique moyenne, l'extension et la moyenne du registre tonal, la proportion de syllabes proéminentes... et les présenter dans un rapport que l'on nomme *phonostylographe* (cf. infra §2).

Dans notre approche, ces outils peuvent se combiner. Chaque chercheur décide quelle information linguistique il souhaite utiliser : alors que la mesure du débit ou la détection de proéminences présupposent une segmentation en syllabes, le calcul du registre tonal moyen peut s'effectuer sans annotation linguistique du signal de parole.

### 1.1. Stylisation de la $f_0$

La stylisation de la fréquence fondamentale est une procédure qui,

« by eliminating all details of the pitch contour that play no communicative role, those perceptual properties of the pitch contour become apparent that are essential constituents patterns of the intonation patterns of the utterance. » (Hermes 2006, 29)

La stylisation automatique présente l'avantage d'éliminer non seulement des micro-variations non communicatives dues à des phénomènes de coarticulation mais aussi, dans notre cas, la plupart des erreurs de détection de la fréquence fondamentale, en sélectionnant les parties stables et intenses du signal, qui correspondent le plus souvent aux voyelles et dans lesquelles la  $f_0$  est généralement estimée correctement.

Le système de stylisation prosogramme (Mertens 2004) fournit une représentation de la variation mélodique qui correspond aux variations perçues. Il ne s'agit pas d'une opération mathématique de lis-

sage de la courbe telle qu'elle est générée par l'algorithme MOMEL (Hirst & Espesser 1993), par exemple.

Le résultat d'une stylisation peut prendre la forme d'une courbe continue, si l'on choisit d'interpoler des segments entre les parties voisées du signal, ou d'une suite de segments, si l'on choisit de ne styliser que les parties voisées. Prosogramme s'inscrit dans la seconde option et fournit une suite de « segments » stylisés sous la forme de morceaux de droites. En résumé,

« [le prosogramme] vise une transcription prosodique lisible, objective, quantifiée, semi-automatique, perceptuellement motivée, indépendante de telle ou telle théorie de l'intonation. Il s'agit d'une stylisation de la courbe de  $f_0$  basée sur un modèle de la perception tonale existant, qui est appliquée ici aux voyelles. Les variations infraliminaires apparaissent comme des traits plats, les glissandos comme des courbes. » (Mertens 2004, 109)

La stylisation réalisée par prosogramme opère une segmentation du continuum sonore (les traits noirs correspondent aux segments stylisés de la  $f_0$ , voir figure 2 ci-dessous). Cette segmentation peut être obtenue sur une base strictement acoustique, à partir des pics de sonie, ou sur la base d'un étiquetage phonétique qui contraint les noyaux syllabiques à styliser aux voyelles. La première technique permet de styliser la  $f_0$  indépendamment de toute annotation alors que la seconde requiert une annotation à fournir dans un TextGrid séparé.

Nous recourons à une version modifiée du prosogramme (Avanzi et al. 2007) qui autorise le noyau (normalement limité à la partie vocale de la syllabe) à s'étendre à toute la partie voisée de la syllabe (incluant les semi-voyelles et les consonnes sonores au-dessus d'un certain seuil d'intensité).

### 1.2. *Alignement et segmentation syllabique*

Les exploitations d'un fichier sonore dépourvu de toute annotation se limitent à des mesures très générales :  $f_0$  moyenne et étendue du registre tonal, ratio signal / bruit... Dès que l'on cherche à mesurer le débit ou que l'on souhaite représenter les contours intonatifs, une segmentation en syllabes s'impose.

L'outil *EasyAlign* (Goldman 2007) permet de réaliser rapidement, avec une intervention manuelle minimale, un alignement phonétique du signal de parole à partir d'une transcription orthographique au format texte. *EasyAlign* crée un TextGrid dans lequel la transcription orthographique, alignée en unités de 5 à 10 secondes, est phonétisée à l'aide d'un dictionnaire. L'utilisateur vérifie cette phonétisation (réalisation des schwas et des liaisons, élisions...) avant qu'elle soit alignée

au niveau des phonèmes, puis des syllabes. Un modèle de détection des silences permet d'annoter également ces derniers.

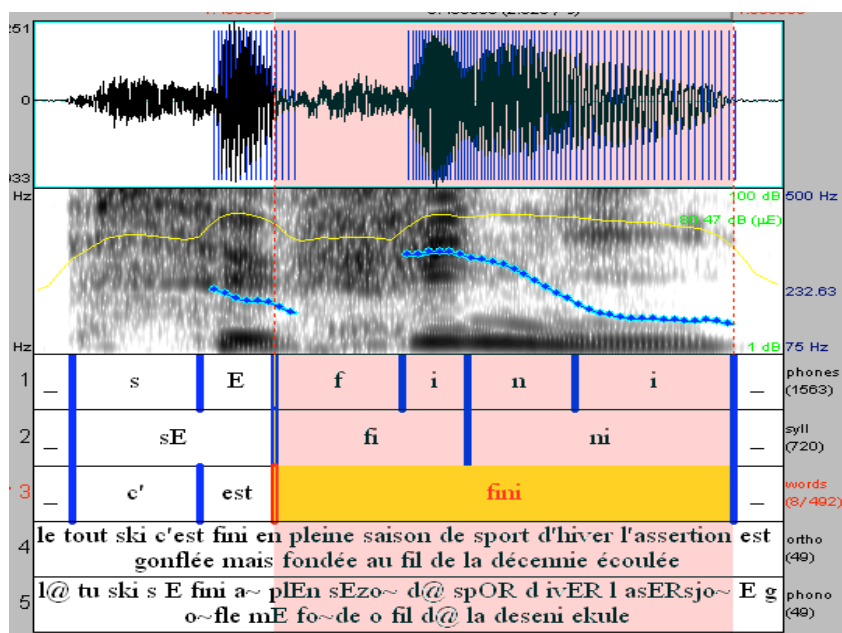


Figure 1 : TextGrid à cinq tires généré par *EasyAlign*

### 1.3. Stylisation et détection automatique des proéminences

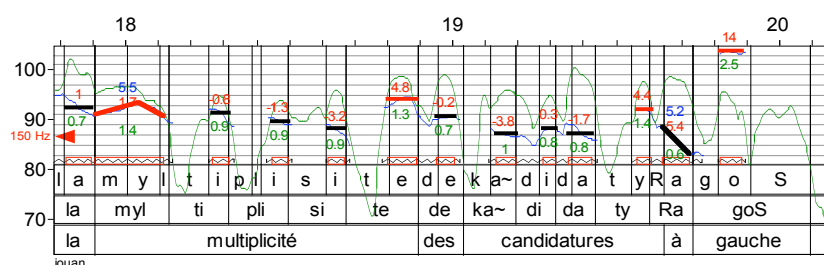
La détection des syllabes proéminentes est une étape obligée dans une étude prosodique, à cause du rôle fondamental des proéminences dans l'accentuation ou le rythme. Propriété graduelle de la syllabe, la proéminence se définit comme « a quantitative parameter of a syllable [...] that describes markedness relative to surrounding syllables » (Portele & Heuft 1997, 63, cités par Hermes 2006, 30).

L'outil *ProsoProm* (Goldman et al. 2007 ; Avanzi et al. 2007), basé sur la stylisation du prosogramme et sur l'alignement syllabique, a été développé pour obtenir une détection automatique des syllabes proéminentes qui corresponde au mieux à une détection auditive.

Actuellement, *ProsoProm* ajoute au TextGrid d'annotation une tire supplémentaire (un intervalle par syllabe) dans laquelle un jugement binaire (et non graduel) de proéminence est attribué à chaque syllabe en fonction de la durée et de la hauteur relatives d'une syllabe par rapport aux syllabes environnantes et de son mouvement mélodique.

Plus précisément, si l'un des deux paramètres acoustiques relatifs dépasse un certain seuil, alors la syllabe sera notée proéminente.

Le résultat de la détection se présente sous la forme d'un prosogramme augmenté de plusieurs informations (voir figure 2). Les segments stylisés détectés comme proéminents sont tracés en rouge.



**Figure 2 : Prosogramme enrichi.** La courbe de f0 stylisée est représentée par les traits épais. Les valeurs des paramètres acoustiques apparaissent pour chaque noyau stylisé (de bas en haut : durée relative, hauteur relative de f0 et mouvement relatif intrasyllabique, ce troisième chiffre apparaît uniquement pour les tons dynamiques)

Dans la figure 2, la proéminence de la première syllabe du mot « multiplicité » est détectée en partie grâce à sa hauteur relative, mais surtout grâce à son mouvement interne. De même, la proéminence de la dernière syllabe du mot « candidatures » est détectée, et ce malgré la présence d'un segment mélodiquement beaucoup plus haut deux syllabes plus tard (« gauche »). Quant à la syllabe intermédiaire « à », elle est certes dynamique mais sa hauteur relative est trop faible pour la considérer comme proéminente.

#### 1.4. Annotation morphosyntaxique

Si l'on considère qu'une syllabe proéminente est potentiellement une syllabe accentuée, le recours à l'information grammaticale est indispensable pour identifier le type d'accent (primaire ou secondaire, Lacheret-Dujour & Beaugendre 1999 ; initial ou final, Mertens 1987, etc.).

Il n'y a pas de critère qui fasse l'unanimité pour distinguer, en français, les mots lexicaux des mots grammaticaux. Nous nous sommes limités à une annotation grossière de type « clitique » vs « non clitique » afin de distinguer les syllabes finales et initiales des mots lexicaux, annotation basée sur Mertens (2001) et Mertens (2006). Sur cette base, une tire d'annotation syllabique est créée dans laquelle on distingue les syllabes initiales (« i ») et les syllabes finales (« f »). Les mots non clitiques monosyllabiques sont également notés « f ». Cette tâche d'annotation a été faite manuellement, mais l'utilisation d'un étiqueteur grammatical pour l'automatiser est envisagée.

## 2. Le phonostylographe

Le phonostylographe fournit un *rapport prosodique* détaillé pour enregistrement, c'est-à-dire un ensemble exhaustif de descriptions phonostylistiques, en tirant parti des outils précédents. Le degré de détail et de profondeur de ce rapport dépend des annotations et des traitements utilisés. Nous décrivons qualitativement le contenu du phonostylographe dans sa version complète. Après une présentation globale des mesures prises sur un enregistrement complet, nous montrons comment le phonostylographe permet de réaliser des comparaisons ciblées qui tiennent compte, par exemple, du caractère proéminent ou non des syllabes.

### 2.1 *Phonostylographe global*

Sur la base d'un **alignement syllabique** et d'une **stylisation mélodique**, le phonostylographe fournit des mesures globales et fiables des paramètres prosodiques suivants<sup>3</sup> :

<sup>3</sup> Certaines mesures omises ici sont disponibles pour l'utilisateur avancé qui souhaite une description exhaustive. Il existe également une version allégée du phonostylographe.

<i>Alignement syllabique et stylisation mélodique</i>		
<b>syllabes</b>	o nombre de syllabes	
<b>durée</b>	o durée de parole	s. (avec pauses)
	o durée d'articulation	s. (sans pauses)
	o taux d'articulation	% (articul./parole)
	o débit de parole	syll./s. (pour parole)
	o débit d'articulation	syll./s. (pour articul.)
	o durée moyenne syllabes + déviation std	s.
	o durée moyenne nuclei <sup>4</sup> syl. + dév. std	s.
	o proportion des nuclei pdt l'articulation	%
<b>f0 globale</b>	o moyenne + déviation standard	ST (relatif à 1 Hz)
	o étendue (inter quantile range)	ST
<b>f0 tons</b>	o proportion des tons statiques,	%
	o prop. des tons dynamiques montants,	%
	o prop. des tons dynamiques descendants	%
	o prop. des tons dynamiques complexes	%
<b>f0 mvt</b>	mouvement intra-nuclei des tons	
	o dynamiques (excluant les statiques)	ST/syll. et ST/sec.
	o mouvement absolu intra-nuclei (pour tous les tons)	ST/syll. et ST/sec.
	o mouvement absolu inter-nuclei (i.e mvt moyen entre les tons)	ST/syll. et ST/sec.
	o agitation mélodique	ST/syll. et ST/sec.
<b>intensité</b>	o intensité moyenne dans les nuclei	dB
	o intensité moyenne hors des nuclei	dB
	o différence d'intensité entre nuclei et non-nuclei	dB

## 2.2 Phonostylographe ciblé

L'outil de **détection automatique des proéminences** et l'**annotation morphosyntaxique** permettent d'aller plus loin en détaillant ces mêmes paramètres prosodiques pour un sous-ensemble de syllabes sélectionnés selon des critères précis. Par exemple, les mesures de durée, de hauteur et d'intensité des syllabes proéminentes peuvent être comparées à celles des syllabes non proéminentes, ou bien, les syllabes initiales aux syllabes finales. Ces deux tires d'annotation supplémentaires<sup>5</sup> font office de filtres syllabiques.

Plus précisément, si l'outil de **détection de proéminence** est utilisé, le phonostylographe donne le nombre ou le pourcentage de syllabes proéminentes et une description des paramètres prosodiques identi-

<sup>4</sup> Lorsque la stylisation est effectuée sur la base d'un alignement syllabique, le nucleus correspond à la partie voisée de la syllabe qui a fait l'objet d'une stylisation perceptive. Ce nucleus peut s'étendre au-delà de la voyelle.

<sup>5</sup> La tire d'annotation des proéminences est obtenue automatiquement par *ProsoProm* et la tire d'annotation morphosyntaxique est réalisée manuellement.

ques à ceux présentés en 2.1, mais en distinguant les syllabes proéminentes des syllabes non proéminentes.

<i>Détection de proéminences</i>		
syllabes	proportion de syllabes proéminentes	%
	proportion de syllabes non proéminentes	%
durée, f0, intensité, ... des syllabes proéminentes vs non proéminentes		

De manière identique, l'**annotation morphosyntaxique** permet de distinguer les mesures prises sur les syllabes initiales de mots pleins des mesures qui décrivent les syllabes finales.

<i>Annotation morphosyntaxique</i>		
syllabes	proportion de syllabes en position initiale de mot lexical	%
	proportion de syllabes en position finale de mot lexical	%
durée, f0, intensité, ... des syllabes initiales vs finales		

Enfin, ces **deux critères** peuvent être **combinés** pour sélectionner les syllabes en fonction de leur localisation (initiale/finale de mots lexicaux) et de leur caractère proéminent.

<i>Annotation morphosyntaxique et détection de proéminences</i>		
syllabes	proportion de syllabes proéminentes en position initiale de mot lexical	%
	proportion de syllabes proéminentes en position finale de mot lexical	%
durée, f0, intensité, ... des syllabes initiales / finales proéminentes / non proéminentes		

De là, il est possible d'interpréter le taux de syllabes proéminentes par seconde (que l'on pourrait associer au « tempo »), ou encore la proportion de syllabes proéminentes en position initiale (assimilable au taux d'accents d'insistance produits le locuteur). De manière plus générale, l'usage combiné d'annotations acoustique et linguistique permet d'envisager un repérage automatique d'unités prosodiques majeures grâce aux syllabes proéminentes finales.

Avant d'aller plus loin, il faut tout de même signaler deux limitations : d'une part la syllabe étant choisie comme unité prosodique de base, il est impossible à l'heure actuelle de caractériser et de quantifier des phénomènes polysyllabiques comme des contours (par exemple une insistance étalée sur tout un mot ou un contour d'appel). Il en va de même pour des modifications temporaires de paramètres prosodiques globaux (par exemple une incise basse). D'autre part, il y a un danger de circularité dans la méthode pour certains aspects : les syllabes proéminentes étant détectées automatiquement par leur hauteur ou leur durée relatives, il est évident qu'elles seront en



moyenne plus hautes et/ou plus longues que les autres syllabes. Cela dit, les autres critères non corrélés comme leur nombre et leur position restent tout à fait valables.

### 3. Comparaison du style radiophonique et du style lu

Nous analysons dans le reste de cet article différentes mesures établies à partir de trois chroniques radiophoniques et des lectures neutres des textes correspondants. Le phonostylographe comparé de ces six enregistrements est présenté en annexe.

#### 3.1. Présentation du corpus

Burger & Auchlin (2007) font l'hypothèse qu'il existe un phonostyle spécifique au genre radiophonique de la chronique sur France Info ; ils le qualifient de *pan-francophone didactique* en raison de caractéristiques perçues décrites de façon informelle par les aspects suivants : surarticulation ; sursegmentation ; hyperactivité mélodique et sur-exploitation de l'espace dynamique en particulier par la récurrence d'accents initiaux.

Cette hypothèse est approfondie, étayée et détaillée dans Goldman & Auchlin (2006) par des mesures chiffrées semi-automatiques, comparant deux journalistes (un homme et une femme) de France Info à deux journalistes de France Inter (même distribution H/F), traçant en contraste les contours d'un phonostyle (hypothétique) avec un autre.

Les résultats attestent, si l'on compare France Info à France Inter : un plus grand parcours tonal (variations tonales cumulées), une plus grande durée des consonnes sur le temps total d'articulation, un plus grand nombre d'accents initiaux par groupe accentuel ainsi que, de façon parallèle homme/femme, un plus grand ambitus mélodique global, particulièrement sur les accents initiaux (1.3 à 2.6 st de différence pour les A.I.).

Le corpus à la base de notre étude est constitué de six enregistrements : deux chroniques radiophoniques de France Info (un homme et une femme) et une de France Inter (une femme choisie pour une certaine ressemblance phonostylistique avec les deux autres), respectivement FI-g, FI-a, FR-j (total 6'38), et la lecture des trois textes de ces chroniques réalisée par une locutrice unique, sans changement des conditions durant la tâche de lecture : L-g, L-a, L-j (total 7'03 ; durée totale du corpus : 13'41).

Le tableau 1 présente la durée des enregistrements et le nombre de noyaux syllabiques détectés. Les colonnes du milieu « L-total » et

« R-total » indiquent la somme par type de corpus, lecture ou radio<sup>6</sup>. Les différences de nombre de syllabes pour les mêmes textes s'expliquent par des choix phonologiques variant selon les locuteurs.

Tableau 1 : Durées syllabiques et durées d'enregistrement totales (en sec.)

	L-j	L-a	L-g	L-total	R-total	FR-j	FI-a	FI-g
Nombre de syllabes	772	644	525	1941	1901	755	638	508
Durée de parole	161,295	138,925	122,19	422,41	384,40	149,675	123,21	111,51

### 3.2. Différences entre phonostyles

L'analyse quantitative (voir annexe) permet différentes comparaisons, à la fois intra-locuteur (les trois lectures L attestent une variabilité intra-locuteur), entre style lu et radio (les colonnes du milieu entre elles), mais aussi entre les styles radio, France Info et France Inter. Pour déterminer les contours d'un phonostyle, nous avons posé les hypothèses suivantes (voir Burger & Auchlin 2007 ; Goldman & Auchlin 2006) : le style radiophonique se caractérise, par rapport à la lecture neutre,

- H1 : par une plus grande « agitation mélodique » ;
- H2 : par une surarticulation, massivement localisée en position initiale de mots lexicaux (*i.a.* énergie répartie sur les consonnes) ;
- H3 : par une quantité d'accents initiaux plus élevée qu'en condition de lecture.

### 3.3. Mesures qui confirment nos hypothèses

Le taux d'articulation (proportion de temps d'articulation par rapport au temps de parole total incluant les pauses) varie de manière significative entre le style lu et le style radio. Cette différence n'est pas due à une différence de débit durant le temps d'articulation (le débit est globalement identique dans les deux styles) mais elle s'explique par le fait que, en lecture, la locutrice insère plus de pauses silencieuses.

<sup>6</sup> Nous entendons ici par « style lu » la relecture neutre, et par « style radio », les chroniques originales pour lesquelles les journalistes devaient à coup sûr être en situation de lecture, tout au moins en production préparée.

**Tableau 2** : Durées de parole, durées d'articulation (en sec.)  
et taux d'articulation (en %)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Durée de parole</b>	161,2	138,9	122,2	<b>140,8</b>	<b>128,1</b>	149,6	123,2	111,5
<b>Durée d'art.</b>	127,8	110,9	99,9	<b>113</b>	<b>112,4</b>	132,8	110,6	93,7
<b>Taux d'art.</b>	79,2	79,8	81,8	<b>80,2</b>	<b>87,5</b>	88,8	89,8	84

La hauteur moyenne de f<sub>0</sub> est difficilement comparable, puisque les locuteurs sont aussi bien des hommes que des femmes. En revanche, le fait que la déviation standard de f<sub>0</sub> soit plus élevée dans la parole radiophonique montre que cette dernière fait globalement l'objet de plus d'agitation mélodique que la parole lue.

**Tableau 3** : Moyenne et déviation standard de f<sub>0</sub> (en demi-tons relatifs à 1 Hz)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>f<sub>0</sub> moy.</b>	93,98	94,75	94,71	<b>94,48</b>	<b>90,37</b>	92,09	91,74	87,29
<b>Dév. std</b>	8,07	2,73	5,4	<b>5,40</b>	<b>7,21</b>	9,21	3,7	8,73

L'étendue de f<sub>0</sub> correspond à l'espace du registre tonal utilisé par chaque locuteur. D'un style à l'autre, cette mesure ne varie pas de manière significative (voir Tableau 4). Par contre, si l'on excepte les 5% de syllabes les plus hautes et les 5% les plus basses (cf. §4 cependant), l'étendue du registre tonal exploitée en parole lue est significativement plus réduite que celle exploitée par les journalistes professionnels (style radio). La différence se marque sur les mesures moyennes lu vs radio (8,57 ST d'étendue tonale en lu contre 12,9 ST en radio) ; il faut remarquer qu'on a deux populations bien séparées. En outre, en lecture, l'étendue est très stable dans la partie moyenne (95%-5%) – ce qui justifie la validité de cette mesure, alors qu'en radio, elle varie de manière asymétrique et présente plus de variation dans la partie haute du registre mélodique (i.e. les excursions mélodiques vers les tons hauts sont plus amples en style radiophonique).

**Tableau 4** : Étendue du registre tonal : étendue totale et étendue calculée en exceptant les 5% de syllabes les plus hautes et les 5% les plus basses  
(en demi-tons relatifs à 1 Hz)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Étendue f<sub>0</sub> max-min</b>	15,8	10,9	23	<b>16,57</b>	<b>18,87</b>	23	15,4	18,2
<b>95%-5%</b>	8,6	8,4	8,7	<b>8,57</b>	<b>12,90</b>	13,8	12,2	12,7

Si l'on compare les types de « tons<sup>7</sup> » (rappelons que le ton porté par une syllabe peut être plat ou dynamique), on constate une proportion plus importante de tons dynamiques dans le style radio. En outre, il y a plus de tons dynamiques descendants en radio. Cette différence n'est pas régulièrement répartie : elle est surtout due à la proportion élevée de tons descendants dans la parole de France Info.

Tableau 5 : Tons (statiques, montants et descendants) : pourcentage et occurrences

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Statiques</b>	89,2 (689/772)	85,2 (549/644)	85,7 (450/525)	<b>86,70</b>	<b>82,87</b>	82,8 (625/755)	79,8 (509/638)	86 (437/508)
<b>Dynami- ques mon- tants</b>	3,5 (27/772)	7,6 (49/644)	7,2 (38/525)	<b>6,10</b>	<b>6,73</b>	9,4 (71/755)	6,7 (43/638)	4,1 (21/508)
<b>Dynami- ques des- cendants</b>	7,3 (56/772)	7,1 (46/644)	7 (37/525)	<b>7,13</b>	<b>10,37</b>	7,8 (59/755)	13,5 (86/638)	9,8 (50/508)

La mesure du mouvement mélodique moyen (en ST/seconde) correspond au « chemin mélodique » total parcouru par un locuteur au cours d'une seconde d'articulation (c'est-à-dire en excluant les pauses). Ce chemin mélodique, dû aux mouvements mélodiques intrasyllabiques (tons dynamiques) et aux écarts mélodiques entre les syllabes, est plus important dans la parole radiophonique. Il traduit une impression d'agitation mélodique accrue.

Tableau 6 : Mouvements mélodiques (en ST/sec.)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Mouvements intra-nuclei des tons</b>	18,2	20	20	<b>19,40</b>	<b>25,67</b>	26,3	26,1	24,6
<b>Intra-mvt</b>	2	3	2,9	<b>2,63</b>	<b>4,40</b>	4,5	5,3	3,4
<b>Inter-mvt</b>	13,3	14,5	14,9	<b>14,23</b>	<b>17,23</b>	18,4	15	18,3
<b>Agitation</b>	15,3	17,5	17,8	<b>16,87</b>	<b>21,63</b>	22,9	20,3	21,7

Enfin, la proportion de syllabes proéminentes varie de manière significative entre les deux styles de parole, en faveur du style radio. Plus encore, le nombre de syllabes proéminentes en position initiale de mot lexical est plus important en radio. Ceci est renforcé par le fait que le nombre de proéminences en position finale de mots lexicaux est identique pour les deux styles. Le pourcentage de syllabes proéminen-

<sup>7</sup> Sans faire référence aux langues à tons, on parle de ton pour désigner la partie stylisée de chaque syllabe. Sur le prosogramme, chaque syllabe est représentée par un trait gras noir, qui peut être plat (ton dit statique), montant ou descendant (ton dit dynamique).

tes peut paraître élevé, il dépend fortement des seuils de détection. Cela dit, quelle que soit la valeur de ces seuils, la comparaison entre corpus reste tout à fait valable.

**Tableau 7 :** Syllabes proéminentes : pourcentage de syllabes proéminentes par rapport aux syllabes non proéminentes ; pourcentage de syllabes proéminentes en position d'accent initial (syllabe initiale de mot lexical polysyllabique) ; pourcentage de syllabes proéminentes en position d'accent final (dernière syllabe de mot lexical)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Prom.</b>	29,5	33,2	34,5	<b>32,40</b>	<b>37,03</b>	35,6	35,9	39,6
	(228/772)	(214/644)	(181/525)			(269/755)	(229/638)	(201/508)
<b>Prom/i</b>	19,7	18,1	26,3	<b>21,37</b>	<b>31,43</b>	30,1	30,3	33,9
<b>Prom/f</b>	58,6	64,5	58,8	<b>60,63</b>	<b>59,6</b>	59,3	59,4	60,1

Cette première série de calculs confirme certaines de nos intuitions, et montre qu'il est parfaitement envisageable de cerner quantitativement les différences entre les deux phonostyles de notre corpus.

### 3.3. Mesures qui infirment nos hypothèses ou sont difficilement interprétables

Pourtant, d'autres mesures (voir annexe) font état d'une similarité entre les deux styles de parole et infirment donc nos hypothèses. Ainsi, la durée moyenne des syllabes ne présente pas de différence. Par contre, on constate que ce paramètre est sujet à des variations intralocuteur importantes : la locutrice ayant réalisé les trois lectures présente une variation de 166 à 190 ms pour la durée moyenne des syllabes (syllabes 15% plus longues dans une des trois lectures).

**Tableau 8 :** Durée des syllabes (en sec.)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Durée moy.</b>	0,166	0,172	0,19	<b>0,18</b>	<b>0,18</b>	0,176	0,173	0,184
<b>Dév. std.</b>	0,067	0,067	0,085	<b>0,07</b>	<b>0,08</b>	0,079	0,062	0,09

L'hypothèse qu'une plus grande partie d'énergie articulatoire est répartie davantage sur les consonnes que sur les voyelles dans la parole radiophonique ne se vérifie pas non plus. La mesure prise en compte ici est l'intensité moyenne des noyaux syllabiques par rapport à l'intensité moyenne sur le reste de la parole articulée (ligne « Diff. »). La différence moindre chez la locutrice en lecture suggère que ses portions consonantiques sont plus intenses qu'en style radiophonique.

**Tableau 9 :** Intensité dans les noyaux – parties stylisées (en dB)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Nuclei</b>	75,3	74	74,8	<b>74,70</b>	<b>74,60</b>	71,3	81	71,5

<b>Non-nuclei</b>	66,2	64,7	64,9	<b>65,27</b>	<b>62,40</b>	59,5	67,7	60
<b>Diff.</b>	9,1	9,3	10	<b>9,47</b>	<b>12,20</b>	11,8	13,3	11,5

Par contre, sur le temps d'articulation, la proportion de noyaux syllabiques par rapport aux « non-noyaux » (i.e. des consonnes) varie significativement. Cela signifie que si les consonnes ne sont pas plus énergiques en parole radiophonique, elles sont cependant plus longues, et la quantité globale d'énergie qui leur est consacrée reste supérieure à celle de la lecture (en tout cas pour la locutrice ayant participé à notre étude).

**Tableau 10** : Durée des noyaux (en secondes) et proportion des noyaux sur la durée totale d'articulation (sans les pauses) en %

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
<b>Durée noy.</b>	60,09	51,767	41,747	<b>153,6</b>	<b>138,5</b>	56,588	47,342	34,574
<b>% noy/artic.</b>	47	46,7	41,8	<b>45,17</b>	<b>40,77</b>	42,6	42,8	36,9

### 3.4. Conclusion de l'analyse du corpus

Au final, cette analyse sur un corpus de taille limitée aura permis de donner du crédit à certaines intuitions concernant les paramètres prosodiques pertinents pour discriminer des phonostyles (en l'occurrence parole lue vs parole radio), et à en infirmer d'autres. Le phonostylographe a permis de tester trois hypothèses.

La première concernait le taux de modulation mélodique. La mise en contraste des mesures (1) de f0 moyenne et déviation standard, (2) d'étendue de f0 et (3) du parcours tonal total conforte l'idée que la parole radio fait l'objet de davantage d'« agitation mélodique » que la parole lue.

Notre seconde hypothèse (qui avait trait à l'hyperarticulation) est en partie confirmée par les mesures portant sur (4) la durée totale des consonnes par rapport à la durée d'articulation et (5) la proportion de prééminences en position initiale des mots lexicaux. Cependant, elle est infirmée par la mesure de (6) l'intensité relative sur les consonnes, qui est plus basse dans le style radio, contrairement à ce qu'on aurait pu croire.

Quant à notre troisième hypothèse, qui stipulait que la parole radiophonique comporte une quantité d'accents initiaux plus élevée que la parole lue, elle s'avère fondée puisque la parole radio en compte 10% de plus que la parole lue (voir tableau 7).

#### 4. Conclusion générale

Dans l'utilisation présentée ici, le phonostylographe contredit l'idée qu'un phonostyle peut être marqué par un petit nombre de traits disséminés, l'*émaillage* de Ch. Bally (Léon 1993 ; Burger & Auchlin 2007) : l'outil proposé calcule des moyennes de manifestations spécifiques, permettant regroupements et différenciations globales.

Certaines variations (e.a. nuclei ratio, qui varie proportionnellement entre chaque chronique et sa lecture) semblent liées au matériau linguistique (structures syllabiques, longueur des mots, nombre de clitiques par rapport au nombre de mots lexicaux, contenu...).

On peut aussi utiliser ces mesures pour chercher des phénomènes isolés remarquables : par exemple les éléments écartés des moyennes (les extrêmes des statistiques) peuvent être retenus, et filtrés par tel ou tel ensemble de paramètres, linguistiques et prosodiques, désignant des cibles linguistiques ou discursives.

La distribution temporelle des manifestations pourrait également être représentée pour donner une dimension évolutive aux mesures, en les représentant dans leur progression interne à un corpus d'un locuteur, afin de repérer des temps et des durées particuliers, et leurs effets de sens aux niveaux linguistique et discursif<sup>8</sup>.

Le phonostylographe doit encore prendre en considération les facteurs spectraux, notamment dans les segments non voisés, les moins bien explorés pour l'instant (voir cependant d'Alessandro 2006).

L'esquisse présentée nous semble ouvrir la voie, à plus long terme, à une description systématique, sur un large corpus, de variétés phonostylistiques du français contemporain.

#### Bibliographie

- AVANZI M., J.-P. GOLDMAN, A. LACHERET-DUJOUR, A.-C. SIMON & A. AUCHLIN (2007), « Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé », *Cahiers of French Language Studies, Actes du Symposium « Les français des corpus »*, Bristol, 20-21 janvier 2007.
- BOERSMA P. & D. WEENINK (2007), *Praat: doing phonetics by computer* (version 4.6.09), disponible à <http://www.praat.org>
- BURGER, M. & A. AUCHLIN (2007), « Quand le parler radio dérange : remarques sur le phono-style de France Info », in Broth M., Forsgren M., Norén C. & Sullet-Nylander, F. (éds), *Le Français parlé des médias. Actes du colloque de*

<sup>8</sup> Cela permettrait par exemple, plutôt que de mesurer un débit d'articulation moyen sur un enregistrement complet, de représenter les accélérations et les ralentissements de débit au niveau local. Les comparaisons internes à un corpus deviendraient pertinentes.

- Stockholm 8-12 juin 2005, Stockholm, Acta Universitatis Stockholmiensis, 97-111.
- D'ALESSANDRO, Ch. (2006), «Voice Source Parameters and Prosodic Analysis», in Sudhoff S. et al. (eds), *Methods in Empirical Prosody Research*, Berlin-New York, Walter de Gruyter, 63-87.
- GOLDMAN, J.-P. (2007), *EasyAligner: a semi-automatic phonetic alignment tool under Praat*, available at <http://latcui.unige.ch/phonetique>
- GOLDMAN J.-PH, M. AVANZI, A. LACHERET-DUJOUR, A.-C. SIMON & A. AUCLIN (2007), «A Methodology for the Automatic Detection of Perceived Prominent Syllables in Spoken French», *Proceedings of Interspeech07*, Antwerpen, August 27-31, 2007.
- GOLDMAN J.P. & A. AUCLIN (2006), « Quelques observations intuitives et mesurées sur le phonostyle de France Info », Communication au Colloque international Phonologie du Français Contemporain 2006, *Approches phonologiques et prosodiques de la variation sociolinguistique: le cas du français*, Louvain-la-Neuve, 6-8 juillet 2006.
- HERMES D.J. (2006), « Stylization of Pitch Contours », in Sudhoff S. et al. (eds). *Methods in Empirical Prosody Research*, Berlin-New York, Walter de Gruyter, 29-61.
- HIRST D. & R. ESPESSER (1993), «Automatic modelling of fundamental frequency using a quadratic spline function», *Travaux de l'Institut de Phonétique d'Aix* 15, 75-85.
- LACHERET-DUJOUR A. & F. BEAUGENDRE (1999), *La prosodie du français*, Paris, Éditions du CNRS.
- LÉON P. (1993), *Précis de phonostylistique. Parole et expressivité*, Paris, Nathan.
- MÉNARD L. (2003), « Le registre en tant que marqueur prosodique de dialectalité : le cas du français du Québec et de la France », in M. Demers (éd.), *Registre et voix sociale*, Québec, Nota bene, 35-58.
- MERTENS P. (1987), *L'intonation du français. De la description linguistique à la reconnaissance automatique*, Thèse de doctorat, Université de Leuven.
- MERTENS P. (2004), « Le prosogramme : une transcription semi-automatique de la prosodie », *Cahiers de l'Institut de Linguistique de Louvain* 30/1-3, 7-25.
- MERTENS P. (2006), «A Predictive Approach to the Analysis of Intonation in Discourse in French», in Y. Kawaguchi, I. Fonagy. & T. Moriguchi (eds), *Prosody and Syntax*, Amsterdam, John Benjamins, 64-101.
- MERTENS P., J.-P. GOLDMAN, E. WEHRLI, A. GAUDINAT (2001), « La synthèse de l'intonation à partir de structures syntaxiques riches », *Traitement Automatique des Langues* 42/1, 142-195.



## Annexe

Voice report	L-j	L-a	L-g	LU	RAD.	FR-j	FI-a	FI-g	
Number of nuclei:	772	644	525			755	638	508	
Speech duration :	161.295	138.925	122.19			149.675	123.21	111.51	s (including pauses)
Articulation dur:	127.773	110.9	99.944			132.853	110.62	93.693	s (excluding pauses)
Articulat. ratio:	79.2	79.8	81.8	<b>80.27</b>	<b>87.53</b>	88.8	89.8	84	% (articul / speech)
Speech rate:	4.8	4.6	4.3	<b>4.57</b>	<b>4.93</b>	5	5.2	4.6	nuc/s (over speech)
Articulation rate:	6	5.8	5.3	<b>5.70</b>	<b>5.63</b>	5.7	5.8	5.4	nuc/s (over articulation)
Syllable duration mean:	0.166	0.172	0.19	<b>0.18</b>	<b>0.18</b>	0.176	0.173	0.184	s
Syllable duration dev:	0.067	0.067	0.085	<b>0.07</b>	<b>0.08</b>	0.079	0.062	0.09	s
Nuclei duration mean:	0.078	0.08	0.08	<b>0.08</b>	<b>0.07</b>	0.075	0.074	0.068	s
Nuclei duration dev:	0.037	0.04	0.039	<b>0.04</b>	<b>0.04</b>	0.038	0.045	0.039	s
Nuclei duration :	60.09	51.767	41.747			56.588	47.342	34.574	s (only nuclei)
Nuclei ratio:	47	46.7	41.8	<b>45.17</b>	<b>40.77</b>	42.6	42.8	36.9	% (nuclei / articulation)
<b>About nuclei f0</b>									
f0 mean:	93.98	94.75	94.71	<b>94.48</b>	<b>90.37</b>	92.09	91.74	87.29	ST (above 1Hz)
f0 stddev:	8.07	2.73	5.4	<b>5.40</b>	<b>7.21</b>	9.21	3.7	8.73	ST (standard deviation)
Estimated quantiles: (in semitones above 1Hz)									
1.%	85	89.6	78.2			79.9	85.3	79.1	(~min)
5.%	90.4	90.6	90.5			86.6	86.5	82	
50.%	94.7	94.7	95.3			92.6	91.2	87.6	(median)
95.%	98.9	99	99.2			100.4	98.7	94.6	
99.%	100.8	100.5	101.2			102.9	100.7	97.3	(~max)
Estimated spreading:									
max-min	15.8	10.9	23	<b>16.57</b>	<b>18.87</b>	23	15.4	18.2	ST (range)
95%-5%	8.6	8.4	8.7	<b>8.57</b>	<b>12.90</b>	13.8	12.2	12.7	ST (narrow range)
median-5%	4.2	4.3	3.9	<b>4.13</b>	<b>7.53</b>	7.9	7.6	7.1	ST (left range)
95%-median	4.3	4.1	4.8	<b>4.40</b>	<b>5.40</b>	5.9	4.7	5.6	ST (right range)



5.%	0.1	0.3	0.6		0.7	0.3	0.6	ST	
50.%	3.1	3.5	3.7		4.4	4.3	4.5	ST	
95.%	5.8	6.6	6.7		9.4	9	8.5	ST	
<b>About prom</b>									
<b>nonProm</b>	70.5 (544/772)	66.8 (430/644)	65.5 (344/525)	<b>67.60</b>	<b>62.97</b>	64.4 (486/755)	64.1 (409/638)	60.4 (307/508)	% of syll
<b>prom</b>	29.5 (228/772)	33.2 (214/644)	34.5 (181/525)	<b>32.40</b>	<b>37.03</b>	35.6 (269/755)	35.9 (229/638)	39.6 (201/508)	% of syll
<b>i/prom</b>	16.2	13.6	19.3	<b>16.37</b>	<b>12.70</b>	19.7	0	18.4	% of prom are i
<b>f/prom</b>	67.1	68.7	60.8	<b>65.53</b>	<b>42.93</b>	60.6	0	68.2	% of prom are f
<b>prom/i</b>	19.7	18.1	26.3	<b>21.37</b>	<b>31.43</b>	30.1	30.3	33.9	% of i are prom
<b>prom/f</b>	58.6	64.5	58.8	<b>60.63</b>	<b>59.60</b>	59.3	59.4	60.1	% of f are prom
<b>About intensity in nuclei</b>									
<b>Nuclei:</b>	75.3	74	74.8	<b>74.70</b>	<b>74.60</b>	71.3	81	71.5	dB (mean intensity in nuclei)
<b>Non-nuclei:</b>	66.2	64.7	64.9	<b>65.27</b>	<b>62.40</b>	59.5	67.7	60	dB (mean intensity outside nuclei)
<b>Diff:</b>	9.1	9.3	10	<b>9.47</b>	<b>12.20</b>	11.8	13.3	11.5	dB

