Les phonostyles: une description prosodique des styles de parole en français[1]

Anne Catherine Simon (Université de Louvain)
Antoine Auchlin (Université de Genève)
Mathieu Avanzi (Université de Neuchâtel et de Paris X Nanterre)
Jean-Philippe Goldman (Université de Genève – Université de Louvain)

1. Introduction

Comment les paramètres prosodiques varient-ils en fonction des styles de parole ? L'étude de la variation stylistique part du constat qu'un locuteur adapte sa manière de parler en fonction de certains paramètres de la situation de communication. Notre étude vise à découvrir si certains paramètres prosodiques (le débit, l'intonation, les accents, etc.) varient systématiquement en fonction de facteurs situationnels et si cette variation permet de définir des styles de parole spécifiques (ou phonostyles). À cette fin, nous analysons un corpus de 70 minutes de parole échantillonné en 7 activités de communication.

2. La variation stylistique

La variabilité phonétique est une caractéristique majeure des productions orales, qui sont irreproductibles (Moreau 1997 : 272). Les phonéticiens se sont attachés à tracer le domaine de la variabilité, tout autant qu'à identifier les invariants phonologiques qui permettent de catégoriser les réalisations phonétiques (Meunier 2005). La variation de la parole n'est pas pour autant aléatoire: outre les facteurs contextuels phonétiques (dont les effets sont prévisibles), la variation est partiellement reliée à caractéristiques géographiques, historiques ou sociales du locuteur, ou à des caractéristiques de la situation. Pierre Léon, lorsqu'il définit les phonostyles, tient compte de ces différents facteurs : les phonostyles sont des styles sonores « tels qu'ils sont perçus en tant que caractéristiques d'un **individu** (jeune, vieux, homme, femme), d'un **groupe social** (prolétaire, bourgeois), ou d'une **circonstance particulière** (discours politique, sermon, etc.) » (1993: 3). C'est à ce dernier aspect que nous nous intéressons. Dès lors qu'il n'existe pas de locuteur monostyle, quels sont les facteurs du contexte immédiat de la production de parole qui sont liés à la variation intra-locuteur (Moreau 1997 : 275) ? À la suite des études de Labov (1972), trois ordres de facteurs sont mis en avant : le type de relation entre le locuteur et son interlocuteur, le contexte social et la thématique de la communication.

> Style reflects the action of the environment upon the individual and the individual upon the environment. It is his perception of the various "status" levels of his listener and of the type of situation in which he finds himself. It is also a projection of himself, his background, and [it] is a setting of the type of tone of conversation he whishes to have. All of this is a mixture of conscious and unconscious

[1] Cette recherche a bénéficié de l'appui du projet Expressive financé par la Région wallonne dans le cadre du Programme WIST2 (convention n° 616422).

(voluntary/involuntary) effort on his part and is not always perceived in the same way as it was intended. (Eskenazi 1993: 502)

La plupart des auteurs ayant élaboré des modèles de classification des styles se fondent sur une échelle de familiarité. Tel est le cas de Joos (1972) qui classe les styles sur une échelle à 5 degrés : intime (*intimate*), familier (*casual*), consultatif[2] (*consultative*), formel (*formal*) et figé (*frozen*). Ce type d'échelle présente en une dimension (le degré de formalité et, partant, le degré supposé d'attention portée par le locuteur à sa production) différents ordres de facteurs, comme le caractère public ou privé de la communication, le degré de familiarité entre les interlocuteurs ou le contexte social.

Concernant notre corpus (cfr infra), il ne nous a pas semblé pertinent de chercher à classer ces situations très diverses sur un axe allant du plus intime vers le plus figé. Nous avons plutôt cherché à prendre en compte de manière spécifique chaque dimension constitutive d'un « style ».

3. Définir les styles de parole

Les enregistrements de notre corpus relèvent de situations contrastées qui sont susceptibles de faire émerger, chez les locuteurs, des styles de parole variés « perçus en tant que caractéristiques de circonstances particulières » (Léon 1993). Comme notre corpus ne comporte pas suffisamment d'échantillons de chaque style (cfr §4, infra) pour tenter de les discriminer sur une base statistique (dans une approche *data driven*), nous avons établi une catégorisation a priori des échantillons, sur la base de facteurs situationnels. Cette catégorisation est multidimensionnelle : nous classons les échantillons selon plusieurs axes, en fonction des différents paramètres observables de la situation de communication.

Labov (1972) retient trois ordres de facteurs : la relation entre le locuteur et son interlocuteur, le contexte social et la thématique de la communication ; Eskénazi (1993 : 502) définit les styles de parole selon trois dimensions : le degré d'intelligibilité, le degré de familiarité et le statut social (des locuteurs et de la situation) ; le modèle conceptionnel (Koch et Oesterreicher 2001), que nous avons retenu pour notre étude, se fonde sur une dizaine de critères pour « catégoriser le comportement communicatif des interlocuteurs par rapport aux déterminants situationnels et contextuels[3] » (2001 : 586). Selon ce modèle, le style informel (« langue de proximité ») se caractérise par une relation intime entre les interlocuteurs, une grande liberté thématique, de l'expressivité, une certaine improvisation (qui va de pair avec des hésitations et des formulations décondensées), une coprésence spatio-temporelle (qui autorise l'ancrage de l'interprétation sur d'autres paramètres que le verbal), etc. À l'opposé, le style formel (« langue de la distance ») se caractérise par le fait que les interlocuteurs sont inconnus l'un pour l'autre, ne sont pas en présence, ne peuvent recourir qu'à un contenu verbal pour interpréter le discours (ce qui

[2] Le style consultatif correspond au discours surveillé (*careful*) que l'on produit lorsqu'on répond à des questions dans le cadre d'une interview.
[3] Cette catégorisation est dite « conceptionnelle ». Elle fonctionne indépendamment du médium, phonique ou graphique.

exige un caractère plus explicite), etc. Entre ces deux extrêmes, existe une gradation très fine des différents types conceptionnels.

Pour des raisons pratiques, nous appelons désormais « styles » les types de production langagière qui correspondent aux sept situations communicatives représentées dans notre corpus, sans présumer de leur spécificité. Chaque style a été noté par trois juges selon les dix critères de Koch & Oesterreicher (2001). Pour chaque critère, on a attribué une valeur entière comprise entre 1 et 5 selon que le style est typique de la « langue de la distance » ou de la « langue de la proximité ». Les styles sont rarement homogènes du point de vue dont ils se caractérisent. Par exemple, l'interview et le journal radiophoniques sont deux types de communication médiatique. Le journal parlé est proche de la langue de la distance (non-présence des interlocuteurs, contrainte thématique forte, discours préparé) tandis que l'interview radiophonique est intermédiaire entre la langue de la distance (absence d'intimité entre les interlocuteurs, détachement actionnel et situationnel, communication relativement préparée) et la langue de la proximité (coprésence, dialogue, présence d'une certaine émotivité, etc.).

[Figure 1]

Pour chaque critère, nous avons fait la moyenne des 3 évaluations. Nous avons ensuite additionné les scores des 10 critères pour classer les styles sur un continuum allant des meilleurs représentants de la langue de la distance à ceux de la langue de la proximité. Il résulte un profil conceptionnel pour chaque style. L'analyse des données montrera que l'addition pure et simple des critères n'est pas nécessairement pertinente : certains des paramètres sont interdépendants, et on peut les réduire à des grandes dimensions comme les paramètres liés à l'interaction (et à la coopérativité), au contenu thématique, et au caractère (non) préparé du discours. On verra qu'un critère semble jouer un rôle crucial sur la variation stylistique : le caractère public ou privé de la parole.

[Figure 2]

4. Données et outils d'analyse prosodique

Le tableau 1 présente le corpus sur lequel se fonde cette étude. Ce corpus comprend des données représentant 7 activités de parole : nouvelles radiophoniques[4] (JPA), discours de chefs d'états à l'occasion d'une fête nationale (POL), réponse à une demande d'itinéraire en milieu urbain (ITI), récit conversationnel[5] (RCV), lecture à haute voix[6] (LEC), conférence dans le cadre d'un colloque universitaire (CFU) et interview radiophonique (INT). Au total, 24 échantillons de 3 minutes en moyenne représentent 70 minutes de corpus, soit environ 10 minutes par style. Les 29 locuteurs sont des hommes et des femmes originaires de Belgique francophone (BE), de France métropolitaine (FR), ou de Suisse romande (CH). Cinq des enregistrements mettent en scène deux locuteurs, les

[4] Enregistrées sur les trois chaînes nationales belge francophone, française et suisse romande.
[5] Un locuteur raconte une mésaventure qui lui est arrivée à une personne intime (ami, frère, etc.).
[6] Il s'agit d'une lecture faite pour une enquête linguistique (données élicitées). L'informateur ne connaît pas le texte à lire (un article de journal).

autres sont monologaux. La façon dont ce corpus a été échantillonné ne permet pas de comparer différents événements de parole d'un même locuteur.

[Tableau 1]

L'annotation automatique du corpus est réalisée grâce à plusieurs outils informatiques, développés comme des scripts du logiciel Praat (Boersma et Weenink 2008). Ces outils sont utilisés pour extraire les mesures et informations caractérisant le phonostyle de chaque enregistrement. Dans un premier temps, l'outil EasyAlign (Goldman 2008) opère une segmentation automatique en repérant, au moyen d'un moteur de reconnaissance de la parole, les frontières de phonèmes et de mots, puis les frontières de syllabes. L'outil ProsoProm, basé sur Prosogramme (Mertens 2004), déduit de la courbe intonative une représentation stylisée de la mélodie réduite aux noyaux syllabiques, c'est-à-dire la partie voisée et intense des syllabes. Ensuite, sur la base d'une série de mesures acoustiques pour chaque syllabe (durée relative, hauteur relative, mouvement mélodique, etc.), l'outil estime le caractère proéminent ou non de celle-ci dans un contexte local. En dernier lieu, l'outil ProsoReport déduit de ces annotations détaillées une série de mesures moyennes pour chaque enregistrement, résumant non seulement les variables temporelles et mélodiques globales, mais aussi des estimations de la proportion de syllabes proéminentes. Une annotation des hésitations a été réalisée manuellement[7]. Toutes ces mesures sont exploitées dans notre étude.

5. Paramètres prosodiques de la variation stylistique

Affirmer qu'un locuteur varie sa manière de parler en fonction de la situation de communication suppose qu'on établisse un lien entre facteurs de la situation et indicateurs prosodiques. Les étiquettes pour qualifier les styles sont aussi nombreuses que les paramètres prosodiques susceptibles de varier selon la situation. Outre les paramètres prosodiques (fréquence fondamentale, durée, intensité), il a été démontré que la qualité vocale et les variations segmentales jouent un rôle important (Eskénazi 1993 : 503-504).

Si l'on s'en tient aux variables prosodiques classiques, la liste est longue des paramètres observables (Llisteri 1992 en dresse la liste ; voir aussi Hirschberg 2000). En accord avec notre méthode d'annotation semi-automatique, nous avons retenu les variables suivantes :

- variables temporelles (taux d'articulation, débit de parole et débit d'articulation) ;
- variables mélodiques (étendue du registre et agitation mélodique) ;
- accentuation (proportion de syllabes proéminentes) ;
- pauses pleines (marquant des hésitations).

Comment relier les paramètres prosodiques aux facteurs contextuels ? Une variante phonétique ne se relie pas d'une manière univoque à *une* signification sociale ou stylistique (Hambye et Simon 2005) : les variables ont potentiellement plusieurs significations (Eckert 2003) et ces significations sont contextuelles. Afin de formuler des hypothèses quant à nos observations, nous nous inspirons cependant de l'hypothèse H&H

[7] Pour une présentation détaillée des processus d'annotation, voir Simon et al. (2008) et Goldman et al. (2008).

de Lindblom (1990), selon laquelle la production du discours est adaptative. Le locuteur adapte son articulation aux exigences communicatives et situationnelles et varie entre *hyper* (sur-articulation) et *hypo* (sous-articulation). Cette adaptation reflète la prise en compte tacite par le locuteur des sources d'information contextuelles indépendantes du signal lui-même et du besoin de marquage qui en découle.

Nos hypothèses concernent la manière dont les différents paramètres prosodiques se relient à certains aspects de la situation de communication.

1. Aspects temporels : on s'attend à ce que le débit de parole soit plus élevé en lecture (journal parlé, lecture de texte) qu'en conversation (interview, récit conversationnel) (Hirschberg 2000 : 336) ; on s'attend à trouver un débit lent dans les situations solennelles (discours officiels ; voir Léon 1993 : 169) ; on s'attend à ce qu'un faible taux d'articulation (silences fréquents) puisse caractériser une grande formalité (p.ex. discours officiels) ou au contraire un discours hésitant (p. ex. récit conversationnel).

2. Aspects mélodiques : la mélodicité (étendue du registre utilisé) peut être un indice d'implication, utilisée pour captiver l'audience (Meunier et Peraya 2004), ou un indice d'émotion (Léon 1993) lié à des situations d'intimité.

3. La densité accentuelle correspond à la proportion de syllabes proéminentes par rapport aux syllabes non proéminentes[8]. Plus les syllabes proéminentes sont nombreuses, plus courtes sont les unités accentuées découpées dans le flux de parole. Il est difficile de prévoir selon quels paramètres la longueur des groupes accentuels pourrait varier.

4. Marques d'hésitation (allongements vocaliques et pauses remplies) : on s'attend à ce que le nombre de marques d'hésitation croisse avec le caractère improvisé du discours.

6. Résultats

6.1. Paramètres temporels

Le *taux d'articulation* correspond au temps occupé par de la parole articulée (le reste étant rempli par du silence). Par conséquent, le *débit de parole* indique le nombre de syllabes prononcées par seconde en tenant compte des silences, tandis que le *débit d'articulation* indique le nombre de syllabes par seconde en excluant les silences.

Trois styles se caractérisent par des mesures extrêmes de débit d'articulation : la lecture et les journaux parlés montrent les débits d'articulation les plus élevés (respectivement 6,13 et 5,87 syl./s), tandis que le discours politique a le débit le plus lent[9] (4,85 syl./s). Les autres styles ont des valeurs intermédiaires assez comparables. La variation de débit

[8] On se base sur une détection automatique des syllabes proéminentes (Simon et al. 2008) et on présuppose que les syllabes proéminentes sont accentuées.

[9] Pour le discours politique (POL), l'écart-type est important. Un des trois échantillons (pol-be) se caractérise par des valeurs extrêmes (2,27 syl./sec de débit de parole) à cause d'un taux d'articulation particulièrement bas (56,8% de parole articulée, le reste étant occupé par du silence).

ne peut pas être rapportée à un paramètre contextuel en particulier : ces trois styles sont lus, donc préparés ; l'un est privé (la lecture), les deux autres sont publics, etc.

[tableau 2]

Il apparaît que le taux d'articulation montre une corrélation significative avec le degré de préparation du discours : moins un discours est préparé, plus il y a de pauses. Quant au débit de parole, on voit qu'il est plus rapide dans les cas de styles formels.

[tableau 3]

6.2. Paramètres intonatifs[10]

L'*étendue de la fréquence fondamentale* est mesurée en demi-tons, afin de permettre la comparaison entre les locuteurs. Elle correspond à la différence entre le minimum et le maximum de fréquence utilisé – en retirant les 5% supérieurs et inférieurs. La *proportion de tons dynamiques* correspond au nombre de noyaux syllabiques qui contiennent un mouvement mélodique. Cette mesure donne une indication de la mélodicité, ou agitation mélodique, de la parole analysée.

[tableau 4]

Contrairement à nos attentes concernant la parole dite émotive (voir 5), la mélodicité semble fortement conditionnée par le caractère public ou privé de l'adressage de la parole. La mesure de l'étendue de f0 sépare nettement deux populations : les discours publics (INT, POL, JPA et CFU) montrent une grande étendue comprise entre 10,3 et 13,8 ST, tandis que les discours privés (ITI, RCV, LEC) montrent un registre compressé, compris entre 6,3 et 8,6 ST. L'aspect public ou privé de la parole est le seul facteur ayant une incidence mesurée.

[tableau 5]

6.3. Syllabes proéminentes et densité accentuelle

Notre annotation des données (cfr § 4, supra) prévoit une détection automatique des syllabes proéminentes. Une syllabe est détectée comme proéminente si elle se distingue des syllabes environnante par son allongement, par un mouvement mélodique montant, par une hauteur relative supérieure ou par la présence d'une pause. Il est abusif de considérer que toute syllabe proéminente acoustiquement est linguistiquement accentuée. Toutefois, on peut considérer que cette mesure donne une bonne approximation de la longueur des groupes intonatifs : une proportion élevée de syllabe proéminente indique des groupes prosodiques court, et inversement.

[10] Les mesures des enregistrements avec plus d'un locuteur ne sont pas fiables pour l'étendue de la f0, puisqu'on mesure en fait le registre créé par les voix des deux locuteurs. Cependant, nos interviews impliquent chaque fois des locuteurs du même sexe, ce qui diminue cet inconvénient. Dans une des demandes d'itinéraire également, le temps de parole est occupé par moins de 10% par un autre locuteur.

[tableau 6]

Les unités prosodiques les plus longues sont produites en situation de lecture et de récit conversationnel ; les plus courtes se retrouvent dans les interviews radiophoniques, dans les discours politiques et dans les explications d'itinéraires.

6.4. Pauses pleines d'hésitation

Les pauses pleines (annotées manuellement) sont des syllabes fortement allongées avec un contour mélodique plat. Le plus souvent, elles correspondent à la particule d'hésitation « euh » ; dans les autres cas, l'hésitation affecte un déterminant, une conjonction ou un mot interrompu. Leur nombre varie sensiblement d'un style à l'autre.

[tableau 7]

Sans surprise, le nombre d'hésitation est directement dépendant du degré de préparation du discours. Les hésitations sont plus fréquentes dans les discours privés que publics.

[tableau 8]

7. Conclusion

Notre étude confirme-t-elle l'existence de phonostyles caractéristiques de certaines situations de parole ? Oui et non. Certains styles sont très homogènes, dans le sens que les différents échantillons d'un même style présentent le même comportement prosodique (pour le rythme, la mélodicité, les proéminences et les hésitations). C'est essentiellement le cas de la parole des journalistes radio et des lectures neutres à voix haute et, dans une moindre mesure, des discours politique.

En outre, ces trois styles présentent certaines valeurs prosodiques extrêmes qui les distinguent fortement des autres styles (voir tableau 9) : le journal parlé se caractérise par le débit de parole le plus rapide et par le taux d'articulation le plus élevé ; la lecture se caractérise par le débit d'articulation le plus élevé et par l'étendue de f0 la plus réduite ; les discours politiques présentent les débits les plus lents. Ces trois styles ne comptent presqu'aucune hésitation. On peut conclure que ces trois styles – journalistique, lecture neutre et discours politique – sont relativement normés et qu'ils induisent des comportements vocaux stéréotypés, et donc identifiables. Ils correspondent aux phonostyles de Léon (1993).

Les autres « styles » de notre corpus ont un comportement à la fois moins homogène (plus de variation entre les locuteurs dans un même style) et plus moyen[11] (des valeurs prosodiques non spécifiques, sans effet distinctif).

[Tableau 9]

[11] Si l'on excepte que le récit conversationnel présente le nombre le plus élevé d'hésitations.

À côté de la caractérisation a priori des styles en fonction de l'activité communicative du locuteur, nous avons proposé une méthode de classification selon des facteurs contextuels. En effet, l'utilisation d'une seule dimension pour mesurer la variation stylistique (par ex. une échelle de familiarité) ne permet pas de rendre compte des variations prosodiques. Nous avons donc utilisé les dix critères de Koch et Oesterreicher (2001) et nous les avons regroupés en trois dimensions (discours préparé ou non, public ou privé, formel ou informel). Il ressort que certaines variations prosodiques semblent fortement corrélées à une dimension particulière (ainsi, l'étendue du registre intonatif augmente avec le caractère public du discours).

De cette étude ressort une vision plurielle des styles, dans leurs déterminations situationnelles et dans leurs manifestations prosodiques. Observations à l'appui, il apparaît que certains styles sont dotés de caractéristiques prosodiques stables et distinctives. Ceux-ci pourraient être considérés comme des « genres » de discours, professionnels ou non, relativement normés et stabilisés dans les pratiques (la lecture, les journaux parlés, les discours politiques). Par ailleurs, certains aspects prosodiques sont en phase avec certains aspects de la situation, d'une manière transversale à différentes activités communicatives. La prosodie est donc aussi réglée par les conditions particulières dans lesquelles la parole est conçue.

Références bibliographiques

BOERSMA, P., WEENINK, D. 2008, "Praat: doing phonetics by computer (Version 5.0.35)" [Programme informatique]. Page consultée le 23/09/2008, http://www.praat.org/

ECKERT, P. 2003, "The meaning of style", in W.-F. CHIANG, E. CHUN, L. MAHALINGAPPA, S. MEHUS (éd.). *Texas Linguistics Forum* 47.

ESKÉNAZI, M. 1993. "Trends in Speaking Styles Research", *Proceedings Eurospeech'93*, Berlin, 501-509.

GOLDMAN, J.-Ph. 2008, "EasyAlign: a semi-automatic phonetic alignment tool under Praat", [Programme informatique], Page consultée le 01/10/2008, http://latlcui.unige.ch/phonetique/easyalign.

GOLDMAN, J.-Ph., AUCHLIN, A., SIMON, A.C., AVANZI, M. 2008, "Phonostylographe : un outil de description prosodique. Comparaison du style radiophonique et lu", *Cahiers de Linguistique française* 28, 219-237.

HAMBYE, Ph., Simon A. C. 2005, "The production of social meaning through the association between varieties and styles: a small case study on vowel lengthening in Belgium", in *Defining linguistic varieties and styles: Objective and subjective perspectives, special issue of the Canadian Journal of Linguistics* 49, no. 3-4, 1001-1025.

HIRSCHBERG, J. 2000. "A corpus-based approach to the study of speaking style", in M. HORNE (éd.), *Prosody: Theory and Experiments,* Kluwer Academic Publisher, The Netherlands, 335-350.

JOOS, M. 1972. "The isolation of styles", in J. FISHMAN (éd.) *Readings in the Sociology of Language,* Mouton, La Haye, 185-191.

KOCH, P., OESTERREICHER, W. 2001, "Langage parlé et langage écrit", in G. HOLTUS, M. METZELTIN, Ch. SCHMITT (éd.). *Lexikon der Romanistischen Linguistik*, I/2, Niemeyer, Tübingen, 584-627. .

LABOV, W. 1972, *Sociolinguistic Patterns*. U. of Pennsylvania Press, Pennsylvania.

LEON, P. 1993, *Précis de phonostylistique, Parole et expressivité*, Nathan Université, Paris.

LINDBLOM, B. 1990, "Explaining phonetic variation: a sketch of the H and H theory", in MARCHAL A., HARDCASTLE, W. J.(éd.), *Speech Production and Speech Modelling*, Kluwer Academic Publishers, NATO ASI Series, Dordrecht, 403-439.

LLISTERI, J. 1992. "Speaking styles in speech research", *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*, Dublin, Ireland, 15-17 juillet 1992, http://liceu.uab.es/~joaquim/publicacions/SpeakingStyles_92.pdf [consulté en août 2008]

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.

MEUNIER, J.-P., PERAYA, D. 2004, *Introduction aux théories de la communication, Analyse sémio-pragmatique de la communication médiatique*, De Boeck, Bruxelles.

MOREAU, M.-L. 1997. *Sociolinguistique. Concepts de base*. Mardaga, Liège.

OBIN, N., LACHERET-DUJOUR, A., VEAUX, Ch., RODET, X., SIMON, A. C. 2008, "A Method for Automatic and Dynamic Estimation of Discourse Genre Typology with Prosodic Features", in *Proceedings of Interspeech 2008*, Brisbane.

SIMON, A. C., AVANZI, M., GOLDMAN, J.-Ph. 2008, "La détection des proéminences syllabiques. Un aller-retour entre l'annotation manuelle et le traitement automatique", Actes du Congrès Mondial de Linguistique française, papier n° 151, Paris. [doi:10.1051/cmlf08256].