

Analyse quantitative comparée de la prosodie des langues à accent fixe et à accent libre.

Adam Pawłowski

Université de Wrocław, pl.Nankiera 15, 54/140 Wrocław, Pologne

Abstract

The prosodic structure of a language with a constant position of accent and of a language with a free and moving position of accent is analysed with the ARIMA method. Although the latter possesses a great potential of building any rhythmical sequence, the former turns out to be more rhythmical because of a regular repetition of rhythmical groups of the same length.

Résumé

La structure prosodique des langues à accent fixe et des langues à accent libre et mobile est analysée à l'aide de la méthode ARIMA. Les langues à accent libre et mobile possèdent un potentiel permettant de construire toute séquence rythmique mais les langues à accent fixe s'avèrent plus rythmiques en raison de la répétition régulière de groupes rythmiques de même longueur.

Mots-clés: time-series analysis, prosody, accent, sequential modelling

1. Linéarité du langage – aperçu général

L'étude de la *structure linéaire du texte* est une nouvelle direction de la recherche quantitative sur le langage. La revue des nombreux travaux en LQ démontre que les échantillons textuels ont été généralement considérés comme des *populations statistiques* par excellence, remplissant toutes les exigences imposées à des collections de ce genre. Cette assumption est souvent abusive¹ et découle surtout de la tradition des recherches statistiques en sciences naturelles et sociales. La conséquence de cette optique unilatérale est, d'un côté, la découverte de nombreuses lois quantitatives décrivant le langage en tant que population d'unités relevant des textes ou du système² et, de l'autre, des lacunes importantes dans les connaissances sur les structures linéaires du langage³. Les structures séquentielles existent sur tous les niveaux du texte, mais les hypothèses les plus riches en contenu concernent le *niveau prosodique* du texte.

2. Notions de base d'analyse prosodique

La prosodie est une branche de la phonologie qui étudie les traits phoniques suprasegmentaux, caractérisant les unités phoniques composées, telles que la syllabe, le groupe rythmique et la phrase. L'analyse prosodique consiste en l'étude, dans la ligne du texte, des relations entre les *tons*, les *accents d'intensité* et les *quantités*. Dans la description de ces attributs, on n'applique pas de catégories physiques, mais les échelles ordinales, composées de deux ou plusieurs niveaux (p. ex. voyelle brève/longue). Le continu phonique est ainsi réduit à une échelle discrète et le caractère composé de certains traits prosodiques est ignoré⁴. Pourtant, cette réduction

¹ Les conditions d'homogénéité des échantillons textuels et d'indépendance des unités sont rarement remplies; en plus, on a du mal à définir la population générale à distribution connue, à laquelle les tests statistiques se réfèrent.

² Parmi les lois les mieux testées il y a celles de Zipf et de Menzerath ainsi que les lois décrivant la structure quantitative du lexique (Hammerl et Sambor, 1993).

³ La plus récente bibliographie de la LQ (Köhler, 1995) ne cite qu'une vingtaine de titres caractérisés par les mots-clés *séquentiel*, *linéaire*, *syntagmatique* etc. sur la totalité de *ca* six mille publications recensées.

⁴ L'accent d'intensité est souvent accompagné d'un rallongement de la voyelle et de la hausse du ton.

n'entraîne aucune perturbation de la communication linguistique. C'est précisément cette organisation prosodique qui forme le tissu rythmique du texte et facilite sa compréhension et mémorisation.

3. Systèmes d'accentuation

Du point de vue de la position de l'accent d'intensité dans le mot ou le groupe rythmique, les langues alphabétiques se répartissent en trois groupes:

- langues à *accent fixe* qui frappe toujours la même syllabe du mot ou du groupe rythmique (en polonais la pénultième, en français la dernière etc.);
- langues à *accent libre* qui peut frapper toute syllabe du mot ou du groupe rythmique mais garde toujours sa position dans les forme dérivés (l'allemand);
- langues à *accent libre et mobile* qui peut frapper toute syllabe du mot ou du groupe rythmique et peut changer de position dans le paradigme du même mot (russe);

4. Objectif

L'objectif que nous visons est la comparaison de la structuration rythmique des textes en langues à *accent fixe* et à *accent libre et mobile*. En ce qui concerne ces dernières, on peut s'attendre à ce que dans les textes où la fonction communicative domine sur la fonction esthétique, la distribution des accents dans la ligne du texte soit moins régulière par rapport aux langues à accent fixe. La situation serait différente dans les textes où la fonction esthétique domine sur d'autres fonctions du langage. En fait, si la prosodie des langues à accent libre et mobile paraît moins régulière dans les textes « peu soignés », elle offre d'énormes possibilités dans la création des schémas rythmiques et l'auteur désireux de les exploiter peut façonner le texte à son gré et, pratiquement, sans aucune limite (Levý, 1965). Dans cette étude, nous allons nous pencher sur le premier volet de cette hypothèse, concernant les textes sans versification⁵.

5. Données

L'hypothèse sera testée sur les fragments des récits fictionnels et des textes journalistiques en polonais et en russe. Le premier est une langue à accent fixe paroxytonique et le deuxième une langue à accent libre et mobile⁶. Les deux langues ont la même origine et présentent beaucoup de similitudes au niveau de la structure morphosyntaxique et de la structure quantitative du lexique. Une des caractéristiques qui les oppose d'une façon nette est la prosodie. On a choisi un texte d'auteur russe avec sa traduction et un texte d'auteur polonais avec sa traduction. Le nombre d'échantillons (2*20 fragments des textes de presse et 4*15 fragments du récit fictionnel) et la longueur de chaque fragment dépassant 150 syllabes semblent largement suffisant pour valider ou invalider l'hypothèse avancée.

6. Quantification

On analyse l'ordre d'apparition dans le texte des syllabes accentuées et non-accentuées. Trois niveaux d'intensité de l'accent sont distingués – a. principal (codé comme 1), a. secondaire (codé comme 0,5) et manque d'accent (codé comme 0). Par exemple la phrase polonaise « *Krzyki i gromkie śmiechy dobiegały także z innego miejsca* » a été codée comme {1 0 0 1 0 1 0 0,5 0 1 0 0,5 0 0 1 0 1 0}. La même phrase en russe – « *Крики и ревуций хохот донеслись и из другово места* » – a été codée comme {1 0 0 0 1 0 1 0 0 0 1 0 0 0 1 0 1 0}. Ce type de codage provient de l'analyse métrique et a été utilisé avec succès dans d'autres études quantitatives de la

⁵ Une structure rythmique particulière caractérise non seulement les textes à organisation formelle explicite, mais aussi les textes privés des contraintes rythmiques imposées par les règles de versification (Bratley et Ross, 1981; Skinner, 1941; Pawłowski, 1997; Pawłowski, 1999).

⁶ La liste des textes analysés est présentée dans l'Appendice.

prosodie⁷. Les tests ont démontré que l'application d'une échelle cardinale n'est pas un obstacle dans le traitement mathématique des séries numériques ainsi générées.

7. Méthode

Les séries numériques correspondant à la structure prosodique (rythmique) des textes sont traitées avec la méthode ARIMA de Box et Jenkins⁸. On obtient une description quantifiée de cette structure sous forme des formules des processus stochastiques du type ARMA(p, q).

8. Résultats

Dans tous les échantillons analysés on a découvert les processus MA(q). En polonais, on a observé les processus MA(1) et MA(2) (respectivement 64% et 36% des cas). En russe, on a observé uniquement les processus MA(1). Malgré le manque de contraintes rythmiques extérieures, les coefficients des processus sont élevés et démontrent qu'aussi les textes en prose possèdent un contour rythmique prononcé. Le tableau 1 présente les pourcentages moyens de la variance originale des séries, expliqués par les modèles proposés. Cette valeur synthétise le degré de structuration du texte: plus elle est élevée, plus le texte est rythmique.

	Polonais (accent fixe)	Russe (accent libre et mobile)
Récit fictionnel (M.B.)	35%	22%
Récit fictionnel (J.I.)	34%	21%
Texte journalistique	32%	18%

Tab.1 Pourcentage de la variance expliquée par les modèles

La structuration rythmique de la langue à accent fixe est nettement plus régulière que celle d'une langue à accent libre et mobile (Tab.1). Cette caractéristique est de nature linguistique et semble indépendante du style et de la réalisation individuelle. Elle est la conséquence de la longueur plus ou moins constante des mots et des groupes rythmiques qui, avec la forte tendance à placer l'accent sur la même (ici pénultième) syllabe du groupe, impose au texte un rythme régulier.

Apparemment, le potentiel d'une langue à accent libre qui permet de construire toute séquence rythmique en profitant des grandes parties du lexique (Levý, 1965) ne se réalise que dans les textes versifiés ou très soignés, où la fonction esthétique domine sur la fonction communicative.

Le résultat obtenu confirme aussi l'observation générale que les textes journalistiques, dont la fonction primaire est d'informer, sont moins rythmiques que la prose artistique qui, en plus de communiquer, devrait plaire. Cette caractéristique est indépendante de la langue.

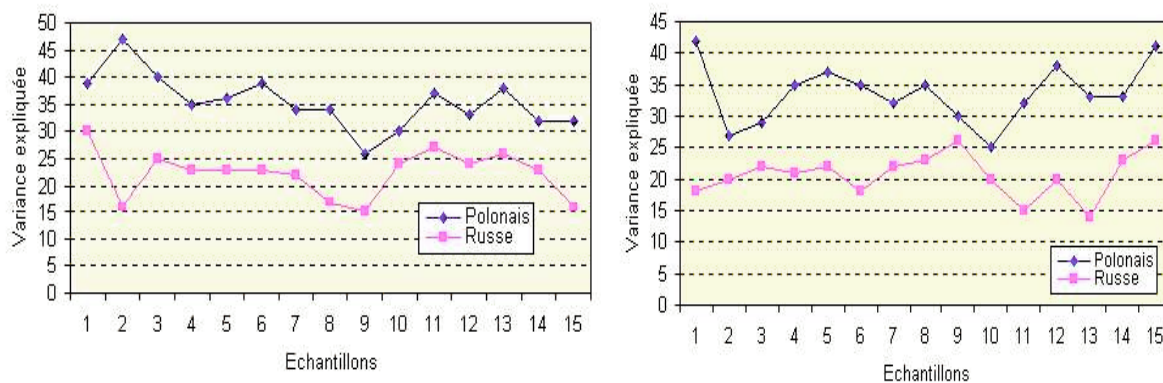


Fig. 1 Pourcentages de la variance expliquée dans les mêmes fragments de Boulgakov (gauche) et d'Iwaszkiewicz (droite)

⁷ Azar et Kedem, 1979; Bratley et Ross 1981; Pawłowski, 1997.

⁸ Glass et al, 1975; McCleary et Hay, 1980; Whiteley, 1980.

En utilisant les valeurs de la variance expliquée par les modèles, nous avons aussi comparé la structuration rythmique dans les mêmes fragments du récit fictionnel en polonais et en russe. On trouve une très faible corrélation, au niveau de la prosodie, entre les fragments correspondants (Fig.1). Dans le cas de Boulgakov, le coefficient de corrélation est de 0,23. Pourtant, si le premier échantillon « anormal » est éliminé, le r_{xy} augmente jusqu'à 0,69. Dans le cas d'Iwaszkiewicz, il n'y a pas de corrélation ($r_{xy} = 0,06$). Vraisemblablement, la relation entre le texte source et le texte cible au niveau de la prosodie dépend de la structure des langues respectives, mais aussi de la qualité et du style de la traduction.

Références

- Azar M. et Kedem B. (1979), Some Time Series in the Phonetics of Biblical Hebrew. *Bulletin of the ALLC* 7(2): 111-129.
- Bratley P. et Ross D. (1981), Syllabic Spectra. *ALLC Journal* 2(2), 41–50.
- Glass G., Wilson V. et Gottman J. (1975), *Design and Analysis of Time-Series Experiments*. Colorado: Colorado Associated University Press.
- Hammerl R. et Sambor J. (1993), *O statystycznych prawach językowych* (Sur les lois statistiques du langage). Warszawa: Zakład Semiotyki Logicznej Uniwersytetu Warszawskiego.
- Köhler R. (1995), *Bibliography of Quantitative Linguistics*. Amsterdam: John Benjamins.
- Levý J. (1965), *W sprawie ścisłych metod analizy wiersza* (A propos des méthodes exactes d'analyse du vers) [in:] M.R. Mayenowa (1965), *Poetyka i matematyka* (Poétique et mathématiques), Warszawa: PIW.
- McCleary R./Hay R. (1980), *Applied Time Series Analysis for the Social Sciences*. Beverly Hills: Sage.
- Pawłowski A. (1997), Time-Series Analysis in Linguistics. Application of the ARIMA Method to Some Cases of Spoken Polish. *Journal of Quantitative Linguistics* 4(1-3), 203-221.
- Pawłowski A. (1999), Language in the line vs. language in the mass: On the efficiency of sequential modelling in the analysis of rhythm. *Journal of Quantitative Linguistics* 6(1), 70–77.
- Skinner B.F. (1941), A quantitative estimate of certain types of sound-patterning in poetry. *American Journal of Psychology* 54, 64-79.
- Whiteley P. (1980), Time Series Analysis. *Quality and Quantity* 14, 225-247.

Appendice

Récit fictionnel

- Михаил Булгаков, *Мастер и Маргарита*, Москва: Художественная Литература, 1998.
Pages: 14, 50, 55, 65, 75, 78, 81, 151, 154, 182, 207. 254, 255, 302, 344;
- Michał Bułhakow, *Mistrz i Małgorzata*, Warszawa: Czytelnik, 1988.
Pages: 12, 58, 64, 78, 90, 94, 98, 193, 197, 236, 268, 333, 336, 396, 452;
- Jarosław Iwaszkiewicz, *Sława i chwala*, Warszawa: PIW, v.1, 2, 1973.
Pages: 28, 59, 97, 146, 198, 251. 261, 305, 316, 380, 428, 439, 481 (t.1), 22, 96 (t.2);
- Ярослав Ивашкевич, *Собрание сочинений т.6, Хвала и слава*, Москва: Художественная Литература, 1975.
Pages: 28, 55, 93, 139, 188, 240, 251, 293, 304, 368, 414, 426, 468, 504, 575;

Presse

- Polityka* no 41/1999 (2214) – 10 fragments;
- Rzeczpospolita* no 237/1999 (5402) – 10 fragments;
- Известия* nos 25381–25384/1999;