

Calcul rythmique et quantification de la vitesse d'élocution

Michela Russo / William John Barry

Université Paris 8/UMR 7023, CNRS (2, rue de la Liberté - F-93526 SAINT-DENIS cedex 02)

Saarland University (FR 4.7 Phonetics Building 17.2 P.O. Box 151150 D-66041)

mrusso@univ-paris8.fr

wbarry@COLI.Uni-SB.de

ABSTRACT

A comparison of Italian and German is carried out using acoustic measures related to structural properties: Vowel-interval and consonantal inter-vowel-interval durations are used to obtain rhythmic measures based on a number of approach [14], [7], [1], [2], [3], [15], [16]. The resulting distributions of rhythm scores for the two languages revealed a strong overlap. These fluctuations in measurement can be shown to relate to the lexical and intonational structure of the individual utterances, thus placing the concept of rhythm typology in the same sphere as other perceptually based generalisations on human behaviour. Statistical analysis shows, however, that the factor 'language' is still a significant differentiator.

Keywords: typology, rhythmic measures, textual structure, speaking style, production.

1. INTRODUCTION

Il est communément admis que l'italien et l'allemand appartiennent à des groupes rythmiques différents: isosyllabique vs. isoaccentuel. S'il paraît clair que le rythme dépend de propriétés phonologiques (cf. en particulier [4], [5]), en revanche les méthodes de quantification de ces propriétés paraissent moins claires.

Dans le passé, les mesures expérimentales inspirées de la théorie de l'isochronie n'ont eu que peu de succès. Pour la tradition de la typologie rythmique, la classification des langues constitue une donnée primitive, c'est-à-dire une propriété inhérente à chaque langue. Par contre, des travaux plus récents considèrent la classification rythmique comme une propriété émergente, c'est-à-dire comme le produit de processus phonotactiques et phonétiques de production du langage (cf. [4], [11]). La différence de classification entre l'italien et l'allemand est plausible étant donné les différences structurales de ces deux langues: l'italien a une structure syllabique relativement simple (CV- dominant). Il ne présente pas d'opposition phonologique de longueur vocalique, ni de réduction des voyelles au niveau phonologique. D'autre part, il est caractérisé par l'opposition phonologique de la longueur consonantique et de l'allongement allophonique de la voyelle accentuée en syllabe ouverte. L'allemand, au contraire, est caractérisé par des structures syllabiques complexes (groupes consonantiques fréquents) et par la distinction phonémique de la longueur vocalique. Il n'a pas de réduction vocalique au niveau phonologique, bien qu'il ait un schwa lexical, et il présente un abrègement

vocalique remarquable au niveau phonétique dans les contextes [-accent]. En outre, on trouve fréquemment en allemand la réduction des groupes consonantiques et des géminées en sandhi externe. Cette combinaison de propriétés phonologiques et d'observations phonétiques suggèrent que les identités rythmiques 'émergentes' de l'italien et de l'allemand sont probablement plus ou moins distinctes, selon le niveau de réalisation des énoncés produits. La comparaison proposée ici entre l'italien et l'allemand s'articule autour des mesures acoustiques en corrélation avec les propriétés structurales mentionnées. La complexité syllabique, les distinctions de longueur vocalique et consonantique, la réduction vocalique dépendant de l'accent et la propension aux processus de réduction vocalique et consonantique (affaiblissement, élision, 'schwa-isation') ont été considérés comme des facteurs qui contribuent au rythme d'un énoncé et, par conséquent, à l'impression rythmique d'une langue. De manière cruciale, non seulement le rythme devient mesurable ([14], [7], [1], [2], [3], [15], [16]) comme phénomène de la parole plutôt que comme propriété inhérente à une langue, mais il devient aussi nécessairement un *continuum* et non une propriété catégorielle.

2. LES MESURES RYTHMIQUES ET LE TEMPO RYTHMIQUE

Les critères structuraux de différenciation de Dauer ainsi que les récentes approches instrumentales (quantitatives) basées sur la logique des critères de Dauer fournissent un cadre pour une différenciation plus fine que la dichotomie traditionnelle isosyllabique vs. isoaccentuelle. A la fois structurellement et quantitativement, l'italien et l'allemand semblent toujours appartenir à deux catégories différentes (cf. [5], [14], [7]).

Ce déplacement de l'interprétation du 'rythme' est cependant problématique:

- a) la façon dont Dauer considère les propriétés structurales est abstraite, et très éloignée des phénomènes de surface de la parole, donc du niveau de l'expérience rythmique.
- b) Ramus, Grabe et Low réduisent le complexe des propriétés structurales de Dauer à des mesures de durée.
- c) Ces deux approches ne prennent pas en compte que les différences entre les langues dépendent de manière critique du matériel examiné (*message-dependent selection*), qui est

représentatif de la langue. Les mesures quantitatives seront nécessairement sujettes à des effets dus au style de parole, au locuteur et au texte.

L'utilisation de mesures de durée, en contraste avec les critères structuraux différenciés proposés par Dauer, est largement défendable. On sait que de nombreux facteurs sont liés à la durée. Les traits de longueur sont par définition en corrélation avec les mesures de durée; les traits de qualité dans les voyelles et les consonnes sont également liés à la durée réalisée puisque la réduction spectrale et articulatoire peut être vue comme un produit de réduction temporelle. Il a été même démontré que le changement de F0 a une influence sur la perception de la longueur. En tous les cas, réduire le rythme à des paramètres de durée a une plausibilité primitive, l'essence du rythme étant de l'ordre d'un modèle temporel.

Ces approches sur le rythme (cf. [12], [13], [14], [7]) démontrent que si l'on utilise une classification appropriée il n'est alors pas nécessaire de recourir à la mesure des intervalles interaccentuels et intersyllabiques. Ces études fournissent des mesures de la proportion vocalique de l'énoncé et de la durée des intervalles vocaliques et consonantiques, et montrent comment les différences rythmiques traditionnelles peuvent être mise en évidence par les éléments acoustiques qui en constituent la base expérimentale.

Cependant, la notion de vitesse d'élocution (*speech rate*) demeure problématique si on compare les résultats des différentes mesures. Avec leur *Pairwise Variability Index* (PVI) normalisé, Grabe et Low affirment pouvoir corriger les fluctuations du *speech rate*, et rejettent les mesures de Ramus en partie sur la base de leur vulnérabilité à capturer les changements de *tempo*. En fait, bien que l'on traite d'un sujet très discuté dans la littérature ([7], [13], [14]), la relation entre *articulation rate* et mesures rythmiques (*changing tempo*) n'a pas encore été examinée empiriquement de manière systématique (mais cf. [1], [2], [3], [15] et [16]).

Nous avons essayé de lier le *tempo* à la quantification rythmique. Nous fondons notre analyse sur l'*articulation rate*, c'est-à-dire le nombre de syllabes ou phonèmes réalisés par seconde à l'exclusion des pauses, plutôt que sur le *speech rate*, c'est-à-dire le nombre de syllabes ou phonèmes par seconde, pauses comprises. Le *tempo*, de même que les mesures rythmiques, est un produit de la structure syllabique et de la durée. En termes de perception, le rythme est constitué de séquences plus lentes et plus rapides. Ainsi, le *tempo* mesuré et le rythme mesuré ne peuvent pas être indépendants.

Nous avons appliqué les deux approches de la quantification du rythme qui ont été établies par les travaux récents [12], [13], [14] et [7]; nous avons examiné en particulier la notion d'*articulation rate* en relation avec les mesures rythmiques (l'écart-type des durées d'intervalles vocaliques et consonantiques ΔV et ΔC ; la proportion des intervalles vocaliques %V et le PVI = *Pairwise variability Index*). La parole montre des réductions indépendamment du type rythmique et les

langues à structure syllabique complexe simplifient leur structure dans la parole spontanée. Pour cette raison, nous avons examiné les paramètres rythmiques de l'italien et de l'allemand dans la parole spontanée.

2.1. Objectifs et hypothèses

Les analyses ont été effectuées pour l'allemand sur 4 locuteurs du corpus de Kiel [8], [9] et pour l'italien sur 6 locuteurs de Naples et 7 de Pise du corpus AVIP/API (<ftp://ftp.cirass.unina.it>).

Notre hypothèse de départ, basée sur les propriétés phonético-phonologiques de l'italien et de l'allemand, est que les corpus d'italien et d'allemand diffèrent de manière significative à la fois dans les mesures consonantiques et vocaliques. Nous avons exploré, en outre, la position de l'italien et de l'allemand le long de l'espace rythmique isosyllabique/isoaccentuel en relation avec les autres langues étudiées précédemment par les auteurs. Une hypothèse à vérifier est de savoir si les relations entre l'italien et l'allemand que nous avons mises en évidence confirmaient au moins celles tracées par les études précédentes [12], [14].

2.1.1 Méthodes d'analyse

L' 'ipu' (*inter-pause unit* ou unité interpausale) est l'unité de l'énoncé utilisée pour calculer les mesures rythmiques individuelles, qui est groupé et moyenné en fonction des locuteurs et des langues (et aussi en fonction des groupes régionaux pour l'italien). Etant donné que le rythme requiert une séquence de syllabes pour pouvoir être perçu, une limite inférieure de 4 syllabes pour un 'ipu' doit être posé. Tout 'ipu' plus court est donc exclu de l'analyse. Cette limite inférieure est en somme arbitraire, mais ce choix s'appuie sur le fait qu'en deçà de cette limite les 'ipu' se manifestent avec des valeurs de *tempo* et des mesures rythmiques extrêmes que l'on ne trouve pas pour des 'ipu' de 4 syllabes ou plus.

Les mesures rythmiques sont calculées en s'appuyant sur les travaux [12], [13], [14] et [7], [3].

Les mesures de Ramus sont (i) la proportion de voyelles dans les 'ipu' (%V), (ii) la déviation standard ou écart-type de la durée vocalique dans les 'ipu' (ΔV) et (iii) la déviation standard de l'intervalle consonantique intervocalique (ΔC).

Les mesures de Grabe et Low correspondent par essence aux mesures de variabilité de Ramus, mais sont calculées par paires (*pairwise*) à travers les 'ipu' plutôt que globalement sur les 'ipu'. Elles sont donc appelées 'Pairwise Variability Indices' (PVI). La différence (i) entre les voyelles consécutives et (ii) les intervalles intervocaliques consécutifs est moyennée selon les 'ipu', ce qui donne une mesure de la variabilité vocalique et consonantique. Dans le cas d'intervalles vocaliques, la différence est liée à la somme des deux voyelles. Cette 'normalisation' est supposée nécessaire (et possible) pour les intervalles vocaliques de façon à contrebalancer les variations de *tempo* parce que les voyelles varient plus que les consonnes pour ce qui est du *speech tempo*, et il

n'y a jamais plus d'une voyelle dans un intervalle vocalique. Les deux formules de PVI sont les suivantes ($r = raw$, $n = normalisé$):

(i) PVI consonantique non normalisé

$$r PVI = \left[\sum_{k=1}^{m-1} d_k - d_{k+1} \right] / (m-1)$$

(ii) PVI vocalique normalisé

$$n PVI = 100 \times \left[\sum_{k=1}^{m-1} \frac{d_k - d_{k+1}}{(d_k + d_{k+1}) / 2} \right] / (m-1)$$

d = durée de la voyelle ou de l'intervalle consonantique, m = nombre d'intervalles.

Un certain nombre d'autres mesures sont calculées, en ce qu'elles sont censées offrir une meilleure visibilité de la relation entre rythme et tempo. Parmi celles-ci, on a le ratio nombre de consonnes/nombre de voyelles (en tant que mesure brute de la complexité syllabique dans un 'ipu' ou dans un corpus) et le ratio durée vocalique /durée consonantique (en tant que mesure de la structuration temporelle de la syllabe)

Les corpus ont été segmentés et étiquetés, en fournissant les identités segmentales et les durées qui forment la base des mesures rythmiques et les mesures de tempo. Les pauses, les hésitations et autres interruptions sont également annotées, afin d'identifier prosodiquement les 'ipu' ininterrompus.

3. RÉSULTATS DE CETTE MÉTHODE

Les deux mesures de tempo (syllabes et phones par seconde) sont toujours fortement liées quel que soit le corpus, et il est souvent conclu de façon prématurée qu'elles permettent de capturer essentiellement le même phénomène.

À partir de la base de données de parole spontanée du corpus allemand de Kiel et à partir des enregistrements italiens de l'AVIP/API, les deux mesures de tempo sont corrélées jusqu'à $R = 0.894$ pour les données italiennes et 0.815 pour les données allemandes. Ces degrés de corrélation signifient que 79.9% et 66.3%, respectivement, de la variation de tempo dans chaque mesure est prédictible à partir de l'autre mesure. Réciproquement, cela signifie que dans les données italiennes 20.1%, et dans les données allemandes 33.6% de la variation est indépendante de l'autre mesure. Aussi, il y a clairement la place pour pousser plus avant les investigations dans cette voie de la variabilité de la parole.

3.1 Différences entre langues

Nous examinons la première hypothèse selon laquelle l'italien et l'allemand diffèrent de manière significative dans leurs mesures rythmiques. Or toutes les mesures de rythme ne montrent pas de différence.

Avec ANOVAS *one-way* pour tester l'influence du contexte linguistique (*language background*) avec les mesures de Ramus (ΔV , ΔC et %V), les deux mesures vocaliques montrent un effet de langue/variante régionale clair ($F_{2, 1475} = 155.8$, $p < 0.001$ pour ΔV ; $F = 458.8$, $p <$

0.001 for %V), mais la mesure ΔC n'est pas significativement différente ($F_{2, 1475} = 0.255$, $p = 0.77$). Un test 'Sheffé post-hoc' montre que les deux mesures vocaliques permettent de distinguer les deux variantes italiennes de l'allemand, et dans ΔV (mais pas dans %V), les groupes de locuteurs de Pise et Naples sont également différents ($p < 0.05$).

Les mesures PVI de Grabe et Low (PVI-V et PVI-C) montrent toutes les deux un effet très significatif en fonction du groupe de locuteurs ($F_{2, 1475} = 10.43$, $p < 0.001$ mais PVI-C n'achève pas le seuil de 5%, $F_{2, 1475} = 2.89$, $p = 0.056$ pour PVI-C). Cependant, la différenciation des groupes, comme indiquée par les tests Scheffé post-hoc, est vue comme déviante par rapport à ce que l'on trouve dans les valeurs de Ramus: l'italien régional du groupe de locuteurs napolitains se distingue de manière significative dans sa mesure de la variabilité vocalique (PVI-V) à la fois par rapport au groupe pisan et au groupe allemand ($p < 0,001$), alors que les groupes allemands et pisans ne se distinguent pas. La variabilité consonantique (PVI-C), d'autre part, ne permet pas de distinguer les groupes.

Ce constat suggère que les valeurs rythmiques ne constituent pas des indicateurs fiables du statut rythmique, en termes typologiques, des langues ou même de leurs variantes régionales, mais sont plutôt le reflet du matériel linguistique qui identifie les énoncés produits, et du style dans lequel l'énoncé est produit (cf. deux locuteurs lisant le même texte donnent lieu à des variations significatives, mais dans des dimensions différentes pour des différents textes).

4. DIFFÉRENCES ENTRE LOCUTEURS

La seule mesure qui offre une séparation indépendante du matériel entre les locuteurs italiens et allemands est la proportion vocaliques en tant que simple pourcentage des durées des 'ipu' (%V). Ceci n'a rien de surprenant, vu que la structure syllabique de l'italien est, en moyenne, plus simple que la structure syllabique de l'allemand; c'est-à-dire, il y a moins de consonnes par syllabe, et de fait proportionnellement moins de voyelles. Ceci peut être quantifié avec deux mesures: 1. Le ratio moyen des consonnes par rapport aux voyelles (C_n/V_n) dans les 'ipu' est 1.181 pour l'italien (1.179 for Naples, 1.183 for Pise) et 1.52 pour l'allemand; 2. traduit en ratio de durée (*durational ratio*): V_{dur}/C_{dur} est 1.217 pour l'italien (1.201 pour Naples, 1.232 pour Pise) et 0.738 pour l'allemand.

5. CONCLUSION

Les mesures rythmiques montrent une variation considérable dans les deux langues en fonction (i) du locuteur, (ii) de la structure textuelle, (iii) du style de parole (les locuteurs italiens sont systématiquement plus lents que les locuteurs allemands). La distribution des *scores* rythmiques pour les deux langues révèle une superposition partielle. Ces fluctuations dans les mesures

apparaissent liées à la structure lexicale et intonative des énoncés individuels, positionnant donc le concept de typologie rythmique dans la sphère des autres généralisations perceptives basées sur l'expérience du comportement humain.

Les analyses statistiques montrent, en fait, que le facteur 'langue' est encore un facteur de différenciation significatif. Cependant, en fonction de la mesure rythmique utilisée, les différences entre l'italien et l'allemand diminuent en même temps qu'augmente le *tempo* (ΔV et ΔC), ou restent faibles mais relativement constantes à travers le *tempo* (PVI). %V, d'autre part, semble être un facteur différentielle à la fois sensible aux différences de structures syllabiques et résistant au *tempo*, en ce qui concerne l'italien et l'allemand.

Les résultats de nos analyses jettent une lumière intéressante sur le comportement des mesures 'rythmiques' récemment proposées, mais appliquées à des

quantités de parole lue plutôt restreintes (et plus ou moins contrôlées en ce qui concerne le *speech rate*).

En ce qui concerne le sujet de la normalisation, qui a alimenté de nombreuses discussions dans la littérature, nous soutenons que la normalisation locale effectuée par le PVI vocalique ne corrige pas les différences de durée sur le long terme, comme celles mises en évidence entre la parole de l'allemand articulée de manière rapide et la parole de l'italien articulée de manière lente. En termes relatifs (d'indice PVI), par conséquent, la variabilité vocalique de l'italien et la variabilité de l'allemand sont apparues identiques, alors qu'en termes absolus de ΔV , et naturellement aussi en termes de %V, elles sont apparues très différentes. D'autre part, le PVI semble normaliser les différences entre les syllabes consécutives qui sont structurellement (et par conséquent probablement rythmiquement) importantes, comme les voyelles brèves ou longues, accentuées ou non.

BIBLIOGRAPHIE

- [1] W. J. Barry and M. Russo. Measuring rhythm. Is it separable from speech rate?. In *AAI Workshop, Prosodic Interfaces*, Nantes, Université de Nantes, AAI, pages 15-20, 2003.
- [2] W. J. Barry and M. Russo. Isocronia oggettiva o soggettiva? Relazioni tra tempo articolatorio e quantificazione ritmica. In *Il parlato Italiano*, Napoli, D'Auria, cdrom A02, 2004.
- [3] W. J. Barry, B. Andreeva, M. Russo, S. Dimitrova and T. Kostadinova. Do Rhythm Measures Tell us Anything about Language Type?. In *Proc. 15th Intl. Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003, Barcelona: Causal Productions Pty Ltd, pages 2693-2696, 2003.
- [4] R. M. Dauer. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* pages 11, 51-62, 1983.
- [5] R. M. Dauer. Phonetic and phonological components of language rhythm. In *Proc. 11th Intl. Congress of Phonetic Sciences*, Tallinn, Estonia, U.S.S.R., volume 5, pages 447-450, 1987.
- [6] D. Gibbon and U. Gut. Measuring speech rhythm. In *Proc. Eurospeech 2001-Scandinavia*, Aalborg, Kommunik Grafiske Losninger A/S, volume 1, pages 95-98, 2001.
- [7] E. Grabe and E. L. Low, Durational Variability in Speech and the Rhythm Class Hypothesis. In *Papers in Laboratory Phonology VII*, The Hague, Mouton de Gruyter, pages 515-546, 2002.
- [8] IPDS. *The Kiel Corpus of Read Speech*. Kiel, Institut für Phonetik und digitale Sprachverarbeitung, volume 1, CD-ROM #1, 1994.
- [9] IPDS. *The Kiel Corpus of Spontaneous Speech*. Kiel, Institut für Phonetik und digitale Sprachverarbeitung, vol. 1-3, CD-ROM #2-4, 1994-1997.
- [10] E. L. Low. *Prosodic prominence in Singapore English*. Doctoral dissertation, University of Cambridge, 1998.
- [11] M. Nespôr. On the rhythm parameter in phonology. In *Logical Issues in Language Acquisition*, Dordrecht, Foris, pages 157-175, 1990.
- [12] F. Ramus. *Rythme des langues et acquisition du langage*. Doctoral dissertation, Paris, EHESS, 1999.
- [13] F. Ramus. Acoustic correlates of linguistic rhythm: Perspectives. In *Proc. 1st Intl. Conference on Speech Prosody*, Aix-en-Provence, Laboratoire Parole et Langage, pages 115-120, 2002.
- [14] F. Ramus, M. Nespôr and J. Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, pages 265-292, 1999.
- [15] M. Russo and W. J. Barry. In che misura l'italiano è 'iso-sillabico'? Una comparazione quantitativa tra l'italiano e il tedesco. In *Generi, Architetture e forme testuali*, Firenze, Cesati, pages 387-401, 2004.
- [16] M. Russo and W. J. Barry. Isochrony reconsidered. Objectifying relations between Rhythm Measures and Speech Tempo. *Speech Prosody 2008, 4th Conf. on Speech Prosody*, Campinas, 2008, Brazil, to appear.