Sur le statut de l'unité intonative dans une tâche métalinguistique : exploitation expérimentale d'un concept linguistique

Irina Nesterenko

Département de Phonétique, Université d'Etat de Saint-Pétersbourg
et Laboratoire Parole et Langage
Université de Provence, 29 av. R. Schuman – 13621 Aix-en-Provence Cedex 1, France
Tél.: ++33 (0)4 42 95 36 39 - Fax: ++33 (0)4 42 59 50 96
Mél: Irina.Nesterenko@lpl.univ-aix.fr

ABSTRACT

The present paper deals with the data obtained in a metalinguistic perception experiment exploring the notion of the Intonation Unit (IU). Seven Russian speaking subjects were asked to segment a proposed set of stimuli into IU, and this under three conditions: the amount of prosodic/linguistic information available augmenting from one condition to another. We analyse the agreement observed between listeners and the weight of the individual strategies applied. We project the application of the results of the present experiment as well as other analyses of the speech material in NLP systems.

1. INTRODUCTION

La présente étude propose une exploitation métalinguistique d'une des notions centrales dans les recherches en prosodie – celle de l'unité intonative. Dans la tradition moderne, à commencer par les théories métriques et auto-segmentales, la prosodie est conçue comme un système organisateur (organisation hiérarchique des constituants prosodiques et de leurs éléments nucléaires, Beckman [1]). Dans notre étude nous adoptons le modèle proposé dans Di Cristo [3] : on considère que le système prosodique d'une langue est une construction complexe, comportant trois sous-systèmes qui gèrent l'organisation tonale, l'organisation métrique et l'organisation temporelle du message. Pourtant, les composants prosodiques ne sont pas privés de la dimension fonctionnelle : l'une de leurs fonctions principales est de signaler l'organisation structurelle de l'énoncé, assurant la compréhension du message par l'auditeur. Or, la plurifonctionnalité de la prosodie y introduit une grande variabilité, surtout quand on étudie les corpora de parole spontanée.

La segmentation du discours en unités intonatives, le choix d'un contour mélodique approprié, la reconnaissance et la reconstruction d'une telle organisation structurale, sont des questions de grande importance dans le domaine du traitement automatique des langues. La dimension de variabilité que nous avons évoquée plus haut est aussi bien un obstacle qu' un avantage : d'une part, la variabilité crée le bruit qui abaisse la performance des systèmes de reconnaissance ; d'autre part, elle augmente, si incluse, le coté naturel et l'intelligibilité de la parole synthétisée.

Dans le cadre de notre recherche, nous travaillons sur la description et la représentation formelle de l'intonation du parler russe spontané. Nous envisageons une description applicable à la synthèse vocale qui assurera l'intégration des informations issues des probabilités sur la variabilité des contours mélodiques observée dans le discours humain. Ce projet côtoie la problématique générale d'annotation et d'extraction automatiques des unités intonatives à partir des grands corpora oraux. Nous nous sommes posée la question du choix de l'unité de description de base appropriée, étant donnée la diversité des approches et des définitions dans le domaine de la phonologie prosodique, ainsi que notre souci d'intégrer l'approche pluri-linéaire à l'intonation et au discours, développé par le groupe Prodige [4] et une approche de compétition de la génération du sens avec les informations apportées par différentes composantes linguistiques (Blache et Di Cristo [2])

Nous présentons dans la suite une étude métalinguistique du concept de l'unité intonative : nous commençons par la description de notre corpus de travail ; nous présentons ensuite le test de perception appliqué et nous proposons plusieurs analyses des résultats obtenus, en étudiant, entre autres, le degré d'accord entre les auditeurs.

2. PRÉSENTATION DE L'EXPÉRIENCE ET DU CORPUS

Dans le cadre de nos recherches, nous travaillons sur un corpus de parole spontanée en langue russe, constitué au Département de Phonétique de l'Université d'Etat de Saint-Pétersbourg (projet de recherche international INTAS-915). Les enregistrements ont été effectués dans la chambre anéchoïque du Département de Phonétique à l'aide des microphones-casques individuels. Nous avons travaillé sur les propos d'une locutrice âgée de 20 ans, extraits d'un dialogue informel entre elle et une de ses amies. Le corpus analysé comporte 8 minutes de parole.

La quête d'une unité appropriée de la description formelle de l'intonation côtoie la problématique des constituants, qui est largement traitée dans les travaux de la phonologie prosodique actuelle ; or, Shattuck-Hufnagel et Turk [10] ont démontré les désaccords et la divergence des concepts dans les travaux de différents auteurs. Nous avons choisi de proposer une tâche métalinguistique de découpage des énoncés en unités intonatives à des spécialistes – phonéticiens confrontés à ce problème lors de leurs activités d'enseignement et de recherche.

2.1. Stimuli et Sujets

Nous avons choisi dans le corpus 25 « unités de performance », correspondant à des interventions ou parties d'une intervention de la locutrice, délimitées par des pauses et comportant au moins une unité intonative. Travaillant avec le corpus de parole spontanée, nous avons été confrontée au problème des discontinuités propres à ce style de parole (les hésitations et les allongements syllabiques imprédictibles, liés au processus de la construction en ligne du message). Nous avons choisi d'inclure les extraits avec ces discontinuités dans l'ensemble de nos stimuli, étant donnée la perspective globale d'annotation prosodique automatique des corpus oraux.

7 Sujets (6 femmes et 1 homme) de langue maternelle russe ont participé à l'expérience. Tous sont des doctorants et des enseignants-chercheurs au Département de Phonétique de l'Université d'Etat de Saint-Pétersbourg.

2.2. Procédure

Traditionnellement, l'unité intonative est définie par rapport à ses marques de frontières et à son organisation interne : c'est une unité comportant au moins un accent nucléaire et délimitée du reste du discours par un ton de frontière Haut ou Bas. Or, nous considérons que de nombreuses études ont été influencées par le parallélisme établi entre les unités intonatives et les constituants syntaxiques. Cette réflexion nous a mené à proposer à nos auditeurs d'effectuer la tâche de segmentation sous trois conditions différentes :

- **Condition 1 :** les stimuli ont une fréquence fondamentale monotone (ce ton correspondant à la valeur moyenne de la fréquence fondamentale dans le stimulus). On a appliqué par la suite un filtre passe-bas (seuil = 500 Hz), ce qui a supprimé l'information segmentale.
- **Condition 2 :** les stimuli sont modifiés par l'application du filtre passe-bas (les énoncés ont été désémantisés, tout en préservant leur courbe mélodique initiale).
- **Condition 3 :** les stimuli sont de la parole normale, sans modification.

Les sujets ont passé le test individuellement. Les stimuli étaient affichés dans le programme de traitement du signal Praat. Les sujets pouvaient écouter l'énoncé-stimulus autant de fois qu'ils en avaient besoin. Ils avaient pour tâche d'indiquer, dans les stimuli présentés, les frontières des unités intonatives qu'ils marquaient sur une ligne spécialement réservée de Praat.

3. RÉSULTATS

Au niveau de l'analyse linguistique des données obtenues, nous avons constaté une large variabilité dans les réponses de nos auditeurs ; nous sommes bien consciente que la tâche proposée aux sujets en est en partie responsable. Pourtant, même en Condition 3, on a retrouvé les cas où les jugements des auditeurs divergeaient beaucoup ; et dans ce cas, la variabilité peut être attribuée à des facteurs linguistiques influençant la performance des auditeurs. Nos hypothèses de base portaient sur la différence de performance des auditeurs sous les trois conditions, ainsi que sur les divergences des stratégies individuelles. Par la suite, nous envisageons de mettre en relation les données subjectives et les paramètres acoustiques afin de dégager les stratégies particulières appliquées par les auditeurs.

3.1. Accord entre les auditeurs

Dans les études qui portent sur l'évaluation d'un système de transcription (notamment le système ToBI [5, 8, 11]), on mesure, entre autres, le degré d'accord entre plusieurs experts. Nous avons appliqué cette méthode à nos données, en testant séparément pour les trois conditions le degré de consentement observé entre les auditeurs sur la présence versus l'absence d'une frontière à un endroit précis dans le signal. Comme mesure d'accord, nous avons calculé le coefficient kappa de Cohen, qui permet de tester le degré d'accord observé par rapport au degré d'accord dû au hasard. Les valeurs de K obtenues sont classifiées selon Landis et Koch [6].

Ainsi, pour calculer les coefficients du consentement K, nous avons choisi de regrouper les frontières marquées par les différents auditeurs, lorsque leur position respective variait de plus ou moins une syllabe, en vérifiant, pourtant, que l'écart ne dépasse pas ±50 ms. Nous avons obtenu les valeurs de K suivantes :

- **Condition 3** : K=0,45 (le consentement «modéré) ;
- **Condition 2** : K=0,2 (le consentement à la frontière entre « mauvais » et « médiocre ») ;
- **Condition 1** : K=14,7 (le « mauvais » consentement).

Si nous comparons le K obtenu pour la Condition 3 avec ceux rapportés dans les travaux cités plus haut, le consentement entre les auditeurs observé dans notre expérience apparaît moins bon (K=0,65 et 0,62 pour les sujets masculins et feminins respectivement dans l'étude

de Syrdal & McGory [11]) ; notons, pourtant, que travaillant avec le système de transcription ToBI, les juges évaluent le degré de liaison entre deux mots, sans faire explicitement appel à des constituants prosodiques. Le faible accord entre les auditeurs observé pour les deux autres conditions peut s'expliquer par la nature de la tâche : la variabilité entre les auditeurs peut être due à l'absence d'un repère unifié, par exemple, la syllabation ou la fin du mot.

3.2. Facteur Condition et stratégies individuelles des auditeurs

Pour l'aperçu global des résultats, nous avons travaillé sur le nombre moyen de frontières identifiées par stimulus et par condition, ce choix méthodologique nous permettant de niveler, à cette étape, l'influence du facteur Sujet. Le test statistique de Kruskal-Wallis donne une valeur de p assez élevée (H corrigé pour ex-aequo=2,07, p=0,355), ce qui nous oblige à réfuter l'hypothèse zéro selon laquelle les trois distributions des moyennes proviennent d'une même distribution sous-jacente. Cette différence peut être attribuée à l'influence du facteur Condition.

Les phénomènes linguistiques et phonétiques associés à la notion de frontière proviennent de niveaux linguistiques différents, selon les modèles pluridimensionnels de traitement du langage. Nous avons décidé d'examiner de plus près les interactions entre les trois conditions.

Dans cet objectif, nous avons examiné nos données à l'aide d'un modèle linéaire avec effets mixtes [7, 9], afin d'observer la variation du nombre de frontières indiquées par les auditeurs en fonction des trois conditions.

Le test sur la totalité des données prouve l'effet net du facteur Condition ($F(2, 347) = 5{,}837$; $p = 0{,}003$). Simultanément, les régresseurs de l'effet fixe examinés par ce modèle, montrent que la première condition se distingue nettement de la deuxième, et à fortiori de la troisième, étant données les distributions des moyennes et des médianes pour les trois conditions. Or, pour les deux conditions préservant les informations mélodiques, les différences ne sont pas significatives pour la variable dépendante choisie (cf. la table infra et le graphique de la figure 1 qui représente les rapports entre le nombre moyen de frontières par stimulus sous les trois conditions):

	Valeur t	Valeur p
Condition 1 vs. 2	t (347) = 2,924	p = 0,0037
Condition 3 vs. 2	t (347) = -0,071	p = 0,943

En effet, quand les auditeurs sont confrontés aux stimuli de la Condition 1, c'est l'information rythmique qui ressort. Ainsi, dans les modèles des constituants prosodiques, qui regroupent les unités intonatives et rythmiques, ces dernières sont généralement plus petites, ce qui pourrait provoquer la tendance observée chez nos auditeurs de mettre plus de frontières dans la première condition.

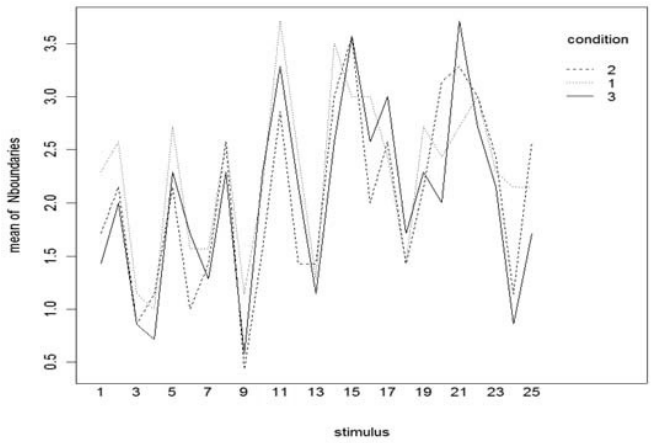

Figure 1 : Comparaison des moyennes du Nombre de frontières selon la Condition par stimulus

En même temps, après une analyse linguistique des données obtenues, l'absence de contraste entre les Conditions 2 et 3 est assez cohérent et s'explique par la nature de la variable dépendante choisie : il s'agit du fait que le nombre de frontières que les auditeurs perçoivent sous les deux conditions est le même ; pourtant, la position des frontières dans les énoncés-stimuli diffère considérablement, ce que nous illustrons avec un exemple. Le stimulus 25 dans notre corpus correspond à l'énoncé « Non, moi, malheureusement, je n'ai pas participé à ces compétitions ». Sous la Condition 3 les auditeurs à l'unanimité mettent une frontière après « non », bien qu'il n'y ait aucune marque prosodique associée. Pourtant, dans la Condition 2, cette frontière disparaît chez tous les auditeurs; or, une autre frontière avant l'accent nucléaire sur le verbe surgit chez 5 auditeurs sur 7.

Nous aurons tendance à interpréter ces résultats dans le contexte de relations entre les constituants prosodiques et syntaxiques : les modules de génération de l'intonation dans la plupart des synthétiseurs présupposent une forte congruence entre la syntaxe et la prosodie. Cette hypothèse, facilitant l'interprétation au niveau de l'analyse prosodique, présuppose l'application de plusieurs analyses linguistiques d'une grande complexité : l'analyse syntaxique détaillée, l'algorithme de génération du contour mélodique à partir de la structure syntaxique donnée, la liste des exceptions d'autant plus grande si on travaille avec du discours spontané. Or, nous nous confortons dans la démarche choisie, à savoir : trouver une unité de description appropriée à l'approche plurilinéaire de la prosodie, ainsi qu'au modèle de compétition dans son application à l'analyse du discours.

Dans les analyses statistiques appliquées nous avons traité le facteur Sujet comme un effet random. Pourtant,

nous supposons que les différences éventuelles entre les stratégies appliquées par les différents auditeurs sous trois conditions pourraient s'annuler mutuellement dans ce traitement global. Ainsi nous avons constaté que les stratégies mises en œuvre lors de la tâche de perception, peuvent varier d'un auditeur à l'autre (cf. figure 2). Figure 2 représente la distribution du nombre moyen de frontières par condition selon l'auditeur : on s'aperçoit que le rapport entre les moyennes selon la condition ne varie pas dans le même sens pour tous les auditeurs.

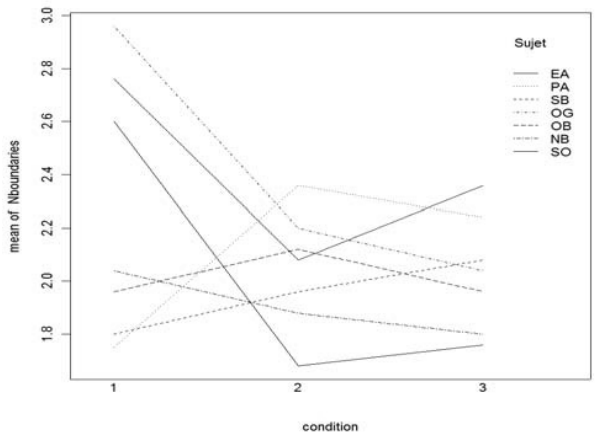

Figure 2 : Variabilité inter-Sujet du nombre moyen de frontières par Condition.

Dans l'analyse linguistique des données nous avons constaté, également, les différentes stratégies des auditeurs dans le cas du traitement de l'accent nucléaire : la frontière pourrait être posée soit avant, soit après, soit avant et après le pic mélodique qui forme dans ce cas une unité séparée. Nous envisageons une étude plus approfondie des ces stratégies individuelles.

4. CONCLUSIONS

L'accord « modéré » obtenu dans la tâche métalinguistique de découpage des énoncés en unités intonatives nous conforme dans notre recherche d'une unité appropriée pour la description et la modélisation de l'intonation du parler russe spontané dans le cadre d'un modèle pluridimensionnel de l'activité langagière de l'homme. En analysant les résultats obtenus, nous avons pu constater la variabilité des stratégies mises en place par les différentes auditeurs ainsi que les discontinuités dans leur performance dues à l'influence du postulat de congruence entre les unités prosodiques et les unités syntaxiques. Nous envisageons de poursuivre l'exploitation des données obtenues selon les pistes indiquées dans cet article afin de mettre en relation les stratégies utilisées par les auditeurs et la structuration informationnelle du message (par intermédiaire du concept de focalisation et du concept prosodique associé – celui de l'accent nucléaire).

Remerciements : Je remercie tous les auditeurs qui ont participé à cette expérience ; Mr Daniel Hirst et Mr Robert Espesser pour leurs conseils en traitement statistique des données ; Mr Cyril Auran pour les scripts qui ont permis l'exploitation automatique des données.

BIBLIOGRAPHIE

[1] M.E. Beckman. The parsing of prosody. *Language and Cognitive Processes*, 11:17-67, 1996.

[2] P. Blache et A. Di Cristo. Variabilité et dépendence des composants linguistiques. In *TALN 2002,* Nancy, 24-27 Juin 2002.

[3] A. Di Cristo. Interpréter la prosodie. In *Actes de 23-7mes Journées d'Etudes sur la Parole*, pages 13-29, Aussois, 21-23 juin 2000.

[4] A. Di Cristo *et al*. An integrative approach to the relations of prosody to discourse: towards a multilinear representation of an interface network. In *Prosodic Interfaces,* International AAI Workshop, Nantes, March 27-29, 2003.

[5] M. Grice *et al*. Consistency in transcription and labelling of German intonation with GToBI. In *Proc. 4th Internat. Conf. Spoken Language Processing*, volume 3, pages 1716-1719, Philadelphia, 1996. ICSLP.

[6] J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174, 1977.

[7] J.C. Pinheiro, D.M. Bates. *Mixed-Effects Models in S and S-PLUS*. Coll. Statistics and Computing. New York, NY, USA : Springer, 2000.

[8] J. Pitrelli, M. Beckman, and J. Hirshberg. Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proc. 3rd International Conf. Spoken Language Processing*, volume 2, pages 123-126, Yokohama, 1994.

[9] R Development Core Team (2003). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

[10] S. Shattuck-Hufnagel and A. Turk. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2): 193-247, 1996.

[11] A.K. Syrdal and J. McGory. Inter-trinscriber reliability of ToBI prosodic labelling. In *Proceedings of ICPhS 2003*, Barcelone, August, 3-8, 2003.