Corpus du parler russe spontané : annotations et observations sur la distribution des frontières prosodiques

Irina Nesterenko

▶ **To cite this version:**

Irina Nesterenko. Corpus du parler russe spontané : annotations et observations sur la distribution des frontières prosodiques. Travaux interdisciplinaires du Laboratoire Parole et Langage, 2006, 25, pp.151-172. hal-00315590

HAL Id: hal-00315590
https://hal.science/hal-00315590

Submitted on 29 Aug 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CORPUS DU PARLER RUSSE SPONTANÉ : ANNOTATIONS ET OBSERVATIONS SUR LA DISTRIBUTION DES FRONTIÈRES PROSODIQUES

Irina Nesterenko

1. Introduction

La présente contribution a pour objectif de démontrer l'intérêt qu'il y a à mener un travail de description prosodique à partir d'un corpus doté d'une annotation élaborée.

Nous insisterons tout particulièrement sur le fait que la méthode de corpus permet de conjuguer les démarches analytiques descendante *(top-down)* et ascendante *(bottom-up) :* une fois les entités prosodiques définies sur la base des critères formels, la tâche est de mettre en évidence des contraintes de haut niveau (contraintes linguistiques), influant sur leur formation et la variabilité observée. En d'autres termes, les unités prosodiques formelles doivent être mises en relation avec des unités linguistiques et discursives.

Le travail sur une formalisation des analyses prosodiques s'impose d'autant plus que nous travaillons sur un corpus en langue russe. Les paradigmes de la phonologie intonative et de la phonologie des domaines, standards dans les études prosodiques actuelles, n'ont pas été appliqués jusqu'à présent dans les descriptions de la prosodie du russe.

Le présent article sera organisé en deux parties. Nous allons d'abord proposer une description de notre corpus d'études ; dans cet objectif, nous discuterons des modalités de recueil de données avant de détailler les annotations que nous avons effectuées pour en préparer des traitements postérieurs. Dans un deuxième temps, nous nous servirons des annotations établies dans le but d'aborder l'épineuse question du phrasé prosodique.

2. Travailler sur la parole spontanée

Nous avons fait le choix de travailler sur la parole spontanée. Il importe de remarquer qu'une grande partie des recherches antérieures en prosodie (quelle que soit leur finalité) se sont fondées principalement sur de la parole lue, et ceci également pour la langue russe. En effet, en lecture, les

NESTERENKO, Irina (2006), Corpus du parler russe spontané : annotations et observations sur la distribution des frontières prosodiques, *Travaux Interdisciplinaires du Laboratoire Parole et Langage,* vol. 25, p. 151-172.

variations prosodiques sont contrôlées et restreintes, ce qui permet d'appréhender au mieux les rapports entre les dimensions et les structures prosodique et syntaxique de l'énoncé. Or, même dans ce cadre simplifié, il apparaît que la gestion de certains phénomènes (tels l'emplacement des frontières prosodiques et le degré de saillance des éléments structurels de l'énoncé) n'est pas aussi aisée qu'il y paraît. Dans la suite des recherches prosodiques appliquées, il a été attesté que le transfert direct des connaissances acquises pour le style lu à de la parole non-contrainte se trouve être inadéquat, surtout dans le contexte d'une prise de conscience sur les mécanismes régissant la production de la parole spontanée.

Notre deuxième choix méthodologique a été de nous positionner dans la perspective analytique guidée par les données. Dans le contexte d'une confrontation *(cf.* Lickley *et al.,* 2005) entre les méthodes de la phonologie de laboratoire (Pierrehumbert *et al.,* 2001 ; D'Imperio, 2005) et de la linguistique de corpus, que, pour notre part, nous considérons comme complémentaires, nous jugeons souhaitable, à l'étape initiale, de fonder notre étude sur l'observation des usages que font des locuteurs des moyens mis à leur disposition par le système prosodique de la langue.

Pour argumenter notre position, il nous semble nécessaire de fournir quelques précisions sur la dichotomie établie entre la parole de laboratoire et la parole spontanée.

Par parole de laboratoire, on entend généralement des enregistrements effectués dans une chambre anéchoïque, ce qui assure une bonne qualité du signal : le matériau linguistique représente des listes de mots ou de phrases, construites au préalable par l'expérimentateur, avec un contrôle systématique de diverses sources de variabilité. En plus d'une bonne qualité de l'enregistrement, cruciale dans les études acoustiques, un atout majeur de ce type de corpus réside dans la possibilité d'élaborer des stimuli en fonction des hypothèses testées. En revanche, en travaillant à partir d'un corpus authentique, on pourrait avoir besoin de nombreuses heures d'enregistrement pour aboutir à une quantité représentative d'exemples du phénomène étudié ; dans un corpus construit ce phénomène est présent de manière condensée.

Pourtant, le désavantage des corpus construits réside dans la restriction stylistique qui leur est propre : très fréquemment, il s'agit de textes/stimuli/listes lues, et non pas de parole produite naturellement. Précisons également que ces corpus représentent de la parole produite dans un but non communicatif. Qui plus est, dans le travail à partir des corpus construits, chacun est amené à adopter un raisonnement descendant (théorie ~ empirie) sur les conséquences des principes sous-jacents d'une théorie linguistique.

À l'opposé, les corpus de parole attestée relèvent d'une parole authentique. Ici, la parole n'est pas préparée à l'avance par l'expérimentateur, la parole est produite dans l'intention de communiquer un message et, selon la situation, les processus de planification et de production se réalisent

conjointement (et on en trouve des marques dans le signal sous la forme de différents phénomènes de disfluences).

Dans ce débat sur le statut et la validité des données dans les études phonétiques expérimentales, il est légitime de poser la question des objectifs auxquels se rapportent ces données. Ainsi, de manière générale, tout investigateur se penche sur les données observables soit pour construire une hypothèse de recherche, soit pour valider une hypothèse issue de l'analyse d'autres types de données et/ou de réflexions théoriques. On distingue donc les objets d'observation et les objets de validation : tous deux peuvent provenir de corpus fabriqués comme de corpus de parole attestée. Le seul contexte dans lequel la parole naturelle est préférable est le cadre d'une étude discursive et conversationnelle avec l'accent mis sur les aspects interactionnels. Ainsi Beckman (1997) conclut-il son plaidoyer en faveur de la parole spontanée :

> *« I have spent so much time motivating why we want to look at spontaneous speech for a better modelling of how focus domain relates to accentuation in English because this is a good example of how lab speech has served us well and of how it fails to serve our needs completely. »*

Dans la même lignée, Lickley *et al.* (2005) défendent les approches de la phonologie de laboratoire :

> *« Doubts are sometimes raised about the "ecological validity" of conclusions based on the analysis of intonation in controlled sentences read aloud under laboratory conditions, and more generally about the intonation of "decontextualised" sentences. We argue that these concerns, though by no means entirely misplaced, are not sufficiently serious to discredit laboratory phonology methods. »*

Les auteurs remarquent, entre autres, que les différences entre la parole de laboratoire et la parole naturelle, spontanée, sont levées dès que l'on atteint les niveaux d'implémentation phonologique et phonétique de la parole *(cf.* le modèle de Levelt, 1989). Cette conclusion intrinsèquement liée à l'approche modulaire entre en conflit avec les observations formalisées sous la forme de la règle de variabilité *(variability rule)*, formulée par Cedegren et Sankoff (1974). Selon cette règle, la probabilité d'une forme phonétique, donc d'une réalisation concrète, est une fonction (dans le sens mathématique du terme) d'un ensemble de contextes.

Pour notre part, nous considérons que l'attention portée aux données contextualisées, à la parole spontanée, ne remet pas en cause les principes fondateurs de l'approche de la phonologie de laboratoire. Dans l'analyse du corpus, on valide les mêmes hypothèses phonologiques (ou autres) à travers l'observation d'une masse considérable de productions langagières, nous pouvons même nous faire une impression sur la fréquence d'un phénomène donné, et, donc, sur sa productivité et sa valeur cognitive (c'est-à-dire le poids des connaissances probabilistes dont disposent les locuteurs), points omis dans les travaux à partir des données construites. Le point méthodologique

crucial consiste donc à disposer de corpus suffisamment grands, dotés d'une annotation multi-niveaux et d'outils automatisés d'annotation et d'analyse : l'exploitation de cette piste permettra de lever la controverse entre parole de laboratoire et parole naturelle. Nous aborderons donc la question de l'annotation adoptée après avoir fourni des détails sur le corpus de travail.

3. Présentation du corpus

L'intégralité du corpus, dont une partie a été utilisée dans le cadre du présent travail, a été recueillie au laboratoire de phonétique de l'université d'État de Saint-Pétersbourg lors d'un travail sur un projet de recherche international *« Spontaneous Speech of Typologically Inrelated Languages (Russian, Finnish and Dutch) : Comparison of Phonetic Properties »* (projet subventionné par le programme INTAS # 00-0915). Ce projet a été conçu dans une perspective comparative des propriétés phonétiques de la parole spontanée et de la lecture : cette orientation du projet a déterminé les modalités de recueil des données orales.

Le recueil des données comportait deux phases : d'abord, des dialogues spontanés ont été déclenchés chez dix locuteurs (5 hommes et 5 femmes) et enregistrés dans une chambre anéchoïque. Les locuteurs avaient pour interlocuteur une personne connue pour obtenir de la parole plus familière ; un sujet neutre leur a été proposé préalablement (« Les dernières vacances »), ensuite, ils étaient libres du choix des thèmes de conversation. Les enregistrements ont été faits avec un équipement professionnel ; les locuteurs portaient un micro-casque individuel AKG HSC 200 et le mixeur Behringer 802A a été utilisé. À la base des enregistrements de parole spontanée, ont été sélectionnés des fragments monologiques sans disfluences (pauses d'hésitations, reprises, allongements non-structuraux) qui ont été convertis en un texte homogène. Ce texte a été proposé aux locuteurs en lecture lors de la deuxième phase d'enregistrement. Les deux phases d'enregistrement ont été espacées d'une semaine. Les enregistrements ont eu lieu entre janvier et juin 2002.

Une telle méthodologie a été adoptée dans l'objectif de disposer d'enregistrements de qualité permettant une analyse acoustique fine. Conjointement, le contenu segmental est le même dans les enregistrements des deux types de discours ; les caractéristiques acoustiques des différents phénomènes prosodiques et des segments peuvent donc être étudiées sans recours à une procédure de normalisation. Dans le développement du projet, seuls les passages choisis des enregistrements spontanés et leurs homologues en lecture ont été soumis à une segmentation phonémique et ont reçu une annotation et une analyse. Pour notre part, nous avons choisi de

travailler sur la totalité des propos, recueillis lors de la phase spontanée, d'une locutrice, N, âgée de 20 ans, étudiante à la faculté de philologie de l'université d'État de Saint-Pétersbourg.

La totalité du matériau spontané recueilli couvre 29,13 minutes de parole. Pourtant, nous avons restreint le matériau analysé dans le cadre de la présente étude sur la base des considérations suivantes : le corpus dont nous disposons est un corpus de parole spontanée contrainte. Certes, il s'agit d'une interaction entre locutrices et le discours est produit dans un objectif communicatif. Pourtant, les conditions d'enregistrement sont relativement artificielles, il n'est donc pas évident de parler sans à-coups ou même de trouver un sujet de conversation. Nous avons par conséquent exclu de l'analyse des passages que nous avons jugés artificiels. Également, de longs passages de silences, correspondant aux interventions prolongées de l'interlocutrice n'ont pas été pris en compte. Au final, nous disposons d'un corpus de 17 minutes (temps de parole, pauses comprises) sur lequel nous avons fondé la présente étude.

Nous avons choisi de travailler sur la totalité des enregistrements en espérant aboutir à une base de données enrichie comportant des annotations multiniveaux. Généralement, le terme d'« annotation linguistique » couvre le procédé d'une notation descriptive ou analytique des données linguistiques brutes : ainsi il s'agit d'un travail de dotation d'une description symbolique à un objet linguistique préexistant. Dans un contexte plus général, il peut s'agir de différents types de transcription, du marquage des parties du discours, de la structure syntaxique, annotation des cas de co-référence, *etc.* Dans la suite de notre travail analytique, le corpus annoté constitue la source de référence pour validation d'hypothèses.

4. Annotation du corpus

Nous considérons que l'annotation du corpus représente une valeur ajoutée importante, toute l'originalité de la démarche résidant dans les principes qui la sous-tendent. Pourtant, l'annotation, telle que nous la concevons, doit communiquer aux données brutes une structuration particulière, dont les fondements doivent être les plus neutres possibles vis-à-vis des théories linguistiques et/ou phonologiques. Simultanément, cette annotation initiale doit mettre en œuvre un cadre dynamique, permettant l'ajout d'informations nouvelles ou bien la formalisation des théories alternatives à celles qui sous-tendent les annotations de base. Pour exploiter alors toute la structure des données, nous avons besoin de disposer de la possibilité de formuler des requêtes ou de récupérer des données brutes observables pour des traitements statistiques postérieurs (par exemple, sous la forme des scripts de PRAAT).

Nous nous sommes limitée ici à l'annotation prosodique du corpus de travail. Nous avons procédé notamment à l'annotation de l'organisation tonale et du phrasé prosodique, l'alignement temporel

à chaque niveau retenu de la constituance prosodique rendant exploitable la dimension de l'organisation rythmique et temporelle.

4.1. Segmentation phonémique

Nous avons commencé le traitement de notre corpus en accomplissant une tâche d'alignement phonémique, considérant que le niveau phonémique était le pivot et la base pour l'alignement d'autres niveaux de l'analyse linguistique que nous proposons pour le corpus *(cf.* l'interprétation du niveau de squelette en phonologie autosegmentale).

Faute de disposer d'un aligneur automatique, nous avons procédé à une segmentation manuelle du corpus en phonèmes bien que cette tâche prenne beaucoup de temps. Ce choix méthodologique est fondé sur les considérations que les méthodes d'alignement automatique n'ont pas été exhaustivement testées sur de la parole spontanée.

Dans un discours spontané informel, le contenu prime sur la forme, et ceci entraîne un renforcement des processus de coarticulation : on constate alors des réalisations allophoniques non-conformes au modèle de la langue fondée sur la prononciation de mots isolés. Remarquons que les caractéristiques acoustiques sont modifiées pour les allophones, non seulement dans des positions faibles mais également dans des positions fortes.

Nous avons suivi les principes de segmentation décrits par Skrelin (1999), la démarche étant fondée sur une observation systématique de spectrogramme et d'oscillogramme, accompagnée d'une vérification auditive.

Les cas les plus litigieux sont :

- les occlusives sourdes à l'initiale : nous avons appliqué une contrainte temporelle stipulant la durée de l'occlusion supérieure à 30 ms, mais ne dépassant jamais 100 ms ;
- la séquence d'une voyelle suivie d'une consonne nasale ou liquide : nous nous sommes appuyée, dans un premier temps, sur des impressions auditives fondées toutefois sur l'examen du spectrogramme à bande large où nous avons cherché des micro-variations dans le comportement des formants et de l'intensité ;
- les désinences atones des adjectifs peuvent comporter une ou deux syllabes et subissent une réduction systématique. Très fréquemment l'identité du phonème vocalique n'est pas claire et nous avons choisi de transcrire toute la rime comme [ə] ([@] en SAMPA) ; de même, le /j/ final de la désinence du genre masculin n'était transcrit et segmenté que quand nous pouvions en observer des traces physiques sur les supports (sonagramme à bandes étroites, courbes d'intensité ou de f0, oscillogramme) ;
- les consonnes /v/ et /l/ dans les positions intervocaliques. Nous avons pourtant appliqué le critère temporel, considérant qu'un phonème ne pouvait avoir une durée inférieure à 25 ms.

- Nous avons pu également constater des cas d'insertion de voyelles atones dans des groupes consonantiques, ce phénomène impliquant une resyllabation.

En vue d'un traitement automatique des données ainsi que d'une transférabilité et d'échanges de données, nous avons utilisé dans les annotations l'alphabet SAMPA *(Speech Assessment Methods Phonetic Alphabet)*, l'adoption de l'API posant un problème technique, étant donné l'exigence de l'installation d'une police spéciale. De son côté, l'alphabet SAMPA est constitué uniquement de symboles ASCII et donc, compatible avec toutes les plate-formes sans installation préalable d'une police spécifique.

Dans les études menées à partir des corpus oraux, on retrouve plusieurs types de transcription de données : le choix d'un type particulier de transcription dépend directement du type d'étude que l'on envisage. Notamment, une distinction est faite entre une transcription phonémique large et une transcription phonémique étroite, cette dernière fixant avec exactitude les réalisations du locuteur. La transcription phonémique étroite étant un travail intéressant en soi mais très coûteux en temps, nous avons décidé de procéder plutôt à une transcription phonétique large, quasi allophonique : nous parlons d'une transcription quasi-allophonique car nous nous sommes fiée plus à notre perception du son, qu'aux principes phonologiques généraux. Nous adoptons donc 7 classes de voyelles (les variations contextuelles des consonnes étant moins problématiques à annoter, *cf.* tabl. 1).

Tableau 1

Liste des sept classes de symboles vocaliques utilisées lors de l'annotation du corpus au niveau phonémique

GROUPE	RÉALISATIONS
/a/	[ɑ, æ, a, ʌ, ɤ, ɐ], combinaisons /a/+j, /a/+V, /a/+ sonnante (en fonction de la réalisation)
/i/	[ɪ, i, j], combinaisons /i/+j, /i/+V, /i/+ sonnante (en fonction de la réalisation)
/o/	[o, ø, œ, œ, ɔ], combinaisons /o/+j, /o/+V, /o/+ sonnante (en fonction de la réalisation)
/u/	[ʊ, u, ɯ, ʏ, y], combinaisons /u/ + V
/e/	[ɛ, e], combinaisons /e/ + V
/ɨ/	[ɨ, ʉ], certaines réalisations de /a/, /u/
/ə/	[ɜ, ə, ɵ], certaines réalisations de /a/, /u/, /o/, /i/, /e/

Une fois l'alignement réalisé au niveau phonémique, nous avons procédé à un regroupement des phonèmes en unités supérieures, à savoir en syllabes, en mots phonétiques, en unités intonatives. Nous avons également appliqué l'algorithme MOMEL et le codage des points-cibles en sortie avec les étiquettes de l'alphabet INTSINT.

4.2. Regroupement en syllabes

Le niveau immédiatement supérieur au niveau phonémique est le niveau syllabique. Nous évoquons donc les principes qui ont guidé notre démarche dans le regroupement de phonèmes en syllabes.

Pour la langue russe, il existe deux approches et deux ensembles de principes de syllabation qui ont été proposés respectivement par Avanessov (1954) et Scherba (1963) : précisons que ces principes de syllabation font référence explicitement aux frontières de mots. En même temps, certaines études expérimentales attestent d'une forte tendance du russe à former des syllabes ouvertes (nous pouvons y voir l'actualisation du principe de l'attaque maximale) et une tendance à la restructuration des mots phonologiques à l'oral.

Nous avons décidé de suivre au maximum le principe de la formation des syllabes ouvertes lorsque nous avons procédé au regroupement des segments en syllabes. Selon les hypothèses phonologiques, les groupes de consonnes résultants doivent obéir au principe de sonorité croissante : ce principe n'a pas toujours été respecté dans l'annotation du corpus. Nous avons plutôt respecté le principe phonotactique selon lequel un groupe de consonnes est autorisé au début de syllabe s'il est autorisé au début de mot. Un deuxième critère a été de nature perceptive : la syllabe résultante devait sonner juste, sans rupture interne perceptible.

Nous avons appliqué un traitement spécifique aux cas de chute des voyelles atones, notamment dans des positions post-accentuelles et dans le contexte des consonnes sibilantes *(cf.* fig. 1). Pourtant, ce processus ne donne pas lieu à une resyllabation au niveau du mot, du moins en perception : la consonne devient syllabique et les paramètres temporels (durée de la syllabe) sont conformes aux modèles standard.

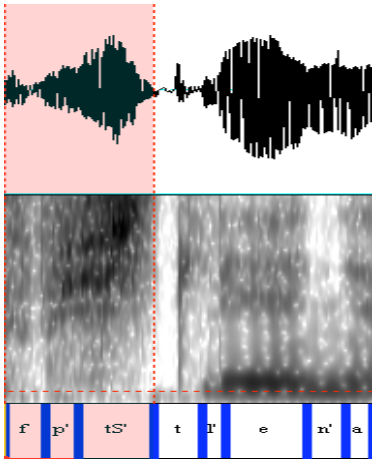

Figure 1

Omission des voyelles atones dans le mot /fpʲɪtʃʲətlʲenʲɪ/

4.3. Les mots

Avant de procéder à l'annotation du corpus au niveau du mot, nous avons dû décider quels critères guideraient notre démarche. Rappelons que la phonologie prosodique classique (Nespor et Vogel, 1986) établit un parallèle entre le constituant du mot dans la hiérarchie prosodique et le mot morphosyntaxique. À cette étape, aucune distinction n'est faite entre les mots grammaticaux, privés de l'accent lexical dans les positions syntagmatiques faibles, et les mots lexicaux : cette différence est traitée lors de la formation des groupes clitiques, niveau immédiatement supérieur de la hiérarchie. À cette structure bi-niveaux la pensée phonétique russe oppose la distinction systématique entre un mot phonologique et un mot phonétique (ce terme désignant la structure comportant un mot lexical et tous les clitiques qui lui sont associés au niveau phonétique).

Nous avons également souligné la forte soumission (révélée dans les travaux de Kassevitch, 2002) de l'organisation rythmique du mot à un contexte rythmique et intonatif plus large. Un troisième paramètre à ne pas négliger fait référence aux critères qui fondent la formation des syllabes, à savoir le principe des syllabes ouvertes. Cependant, persiste le risque que les frontières de mots ne coïncident pas avec les frontières de constituants syllabiques. Une solution possible serait l'utilisation du concept du mot phonologique restructuré, proposé par Peperkamp (1999).

En conséquence, au niveau du mot nous disposons d'une annotation en termes d'unités proches des groupes clitiques de la phonologie prosodique classique (Nespor et Vogel, 1986 ; Hayes,

1989) : de ce choix méthodologique, des cas de non-coïncidence des frontières des mots et de celles des syllabes sont attestés dans la représentation multi-niveaux.

4.4. Annotation du degré de proéminence

L'annotation du degré de proéminence nous permet de rendre compte de différents niveaux de l'organisation métrique de notre corpus. Nous avons adopté une annotation qui permet de reconstruire une grille métrique à trois niveaux (dans la lignée de Beckman et Edwards, 1990). Trois degrés de proéminence sont distingués et se résument en deux oppositions : premièrement les rapports d'opposition se créent entre les syllabes inaccentuées et les syllabes porteuses de l'accent lexical ; dans un deuxième temps, parmi les syllabes métriquement fortes au niveau du mot, nous différencions celles qui sont porteuses de la proéminence mélodique. Dans cette annotation, nous défendons le principe selon lequel tout mot lexical n'est pas forcément associé à une proéminence mélodique. Le concept de la proéminence mélodique, tel que le conçoit la tradition métrique-autosegmentale, n'a pas été entièrement assimilé dans les études sur l'intonation du russe. Lors de l'annotation, nous nous sommes principalement fiée à notre perception. Ainsi la figure 2 correspond à l'annotation du degré de proéminence pour un syntagme nominatif « очень красивый город » ([une]) *très belle ville*) : nous considérons que c'est seulement le dernier mot (« город », *ville*) qui porte la proéminence mélodique ; aucune autre proéminence mélodique (accent mélodique pré-nucléaire) n'est perçue.

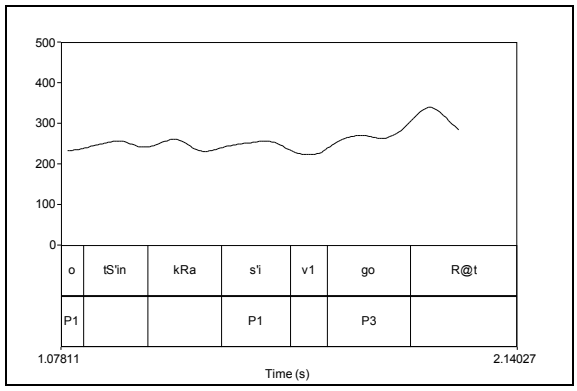

Figure 2
Exemple d'annotation du degré de proéminence

4.5. Unités intonatives

L'unité intonative paraît être une unité universellement reconnue dans l'analyse de l'organisation sonore du langage, le concept étant étroitement lié à la fonction structurelle de la prosodie. Il a été proposé une définition formelle par référence aux tons de frontières et à la présence d'une proéminence mélodique nucléaire (centre intonatif). On insiste également sur le rôle de la structure prosodique en tant qu'interface entre d'autres modules langagiers (syntaxique, informationnel, discursif). Dans le même esprit, on constate un lien étroit entre la structuration prosodique du discours et le domaine de la phonostylistique *(cf.* Zlatooustova et *al.*, 1986 : 79). Ainsi, selon les conditions de production, le même texte peut recevoir une structuration différente. Ces considérations nous font adopter une attitude de réticence vis-à-vis des approches qui présupposent l'identification de la structure prosodique sur la base du seul texte.

Une partie de notre corpus a reçu une annotation en syntagmes prosodiques dans le cadre du projet INTAS, mentionné ci-dessus. Nous disposons également d'une annotation du corpus par un expert francophone qui n'a aucune notion de russe : il s'agit donc d'une annotation fondée sur des propriétés formelles (mélodiques et temporelles) des unités résultantes. Nous avons pu constater des cas de divergence dans les représentations résultantes. La figure 3 reprend l'annotation d'une unité interpausale *« Ну у дедушки там своих забот хватает у него тоже там кошки всякие собака есть »* (« *Eh bien mon grand-père ne manque pas de soucis lui non plus, il a là-bas de nombreux chats, et il a un chien* »). L'expert français signale trois unités intonatives, il y en a quatre dans l'annotation des spécialistes russes. Une frontière est consensuelle, laquelle paraît séparer deux énoncés ; la dernière frontière paraît se décaler d'un mot dans les deux annotations : du point de vue de la grammaire l'adjectif « кошки » *(chats)* forme un syntagme nominal avec le nom « всякие » *(différents)* qui porte la proéminence mélodique ; du point de vue de l'organisation prosodique, aucun indice acoustique n'est associé avec la frontière syntaxique en question. L'annotation des experts russes est alors guidée par les connaissances linguistiques générales. Une annotation alternative aurait pu être proposée, établissant une distinction entre la Macro-unité Intonative, composée de l'unité intonative et le segment d'unité intonative (Di Cristo et Hirst, 1996). Enfin, nous observons une seule frontière discordante. La justification de l'annotation de cette frontière peut être linguistique : un syntagme prépositionnel au début de phrase induit la présence d'une frontière. Il est évident qu'une proéminence mélodique est associée avec le premier mot de cette unité de performance ; or, pour respecter la contrainte structurelle de la formation du syntagme intonatif, il est nécessaire qu'un des mots du syntagme résultant « своих забот хватает » *(il ne manque pas de soucis)* soit associé à un accent mélodique. Nous avons évoqué les controverses

sur l'annotation du degré de proéminence dans le paragraphe consacré à ce sujet : à notre avis, il est donc impossible, à l'étape actuelle, de prendre une décision sur la présence d'une frontière prosodique dans la position désignée.

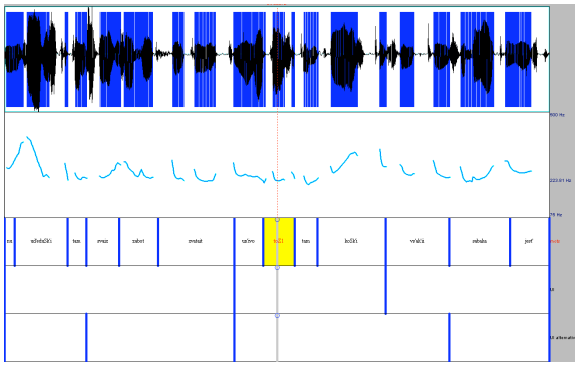

Figure 3
Exemple d'annotation d'une unité interpausale en unités intonatives
(palier 2 : expert français ; palier 3 : experts russes)

4.6. Unités interpausales

Au niveau le plus concret (et le plus bas dans la hiérarchie adoptée) nous disposons de la segmentation de notre corpus en unités de performance : unités interpausales *(inter pausal units*, IPU, Koiso et *al.,* 1997). Le terme IPU désigne un bloc de parole délimité du reste du discours par des pauses silencieuses d'une durée supérieure à 200 ms. Notons que cette unité est introduite dans une étude portant sur l'analyse de l'organisation discursive du message, et que le choix est argumenté par son caractère objectif. Ainsi, dans la transposition de la méthodologie du corpus, c'est l'unité interpausale qui peut servir de pivot du transfert des annotations pour tous les autres niveaux.

4.7. Annotation mélodique : MOMEL et INTSINT

Afin d'identifier l'organisation tonale de notre corpus, nous avons appliqué les traitements suivants : la courbe mélodique a reçu une modélisation par application de la procédure MOMEL et a été annotée par la suite à l'aide de l'alphabet INTSINT.

Cette annotation est semi-automatique. Nous nous sommes servie d'une implémentation de l'algorithme MOMEL sous la forme d'une fonction externe en C qui est appelée par un script de

l'intérieur du logiciel PRAAT (Boersma et Weenink, 2005). Lors de son application, ce script prévoit une possibilité de correction manuelle des points-cibles détectés. Comme nous l'avons précisé, lors de la correction manuelle, les points-cibles peuvent être supprimés, ajoutés ou déplacés.

Or, se pose la question de l'unité d'analyse pour le codage. Dans la version implémentée de l'algorithme, les points-cibles reçoivent leur codage sur la base de deux paramètres : clé *(key)* et étendue *(range)*. Ensemble, les paramètres de clé (niveau) et d'étendue définissent le registre mélodique, concept qui n'est pas négligeable dans les analyses prosodiques *(cf.* Laver, 1994, sur cinq lectures possibles de ce terme). Le registre tonal permet d'évaluer les caractéristiques globales de la f0 d'un locuteur ou d'un contour mélodique. La dimension linguistique du phénomène du registre tonal se retrouve associée à l'organisation informationnelle du message *(cf.* Ladd, 1992, travaux sur l'organisation mélodique au niveau du paragraphe vocal, paraton, domaine du développement d'un topique Yule, 1980 ; Brown et Yule, 1983). Pour des raisons de facilité de traitement, notre corpus est segmenté en fichiers d'une minute environ. Or, compte tenu des remarques faites ci-dessus, pour le traitement MOMEL-INTSINT nous avons choisi de travailler avec des unités plus petites correspondant aux interventions de la locutrice. De plus, Hirst (2004) propose un traitement unifié pour le bloc de cinq énoncés, formant un tout du point de vue discursif dans le corpus Eurom-1. Compte tenu du choix de l'unité de traitement pour l'annotation de la composante tonale, nous nous sommes intéressée aux valeurs fréquentielles pour toute catégorie d'étiquettes. En effet, surtout pour les catégories absolues de points cibles qui caractérisent l'étendue tonale, nous nous sommes posé la question de savoir si ces catégories étaient bien délimitées. Nous présentons également sur la figure 4 les histogrammes des distributions de la f0 selon la catégorie d'INTSINT. L'examen de ces histogrammes nous permet de voir que la variabilité la plus élevée est propre aux étiquettes attribuées aux sommets des pics mélodiques ; les points-cibles associés à des vallées ont un comportement plus stable dans le domaine fréquentiel *(cf.* les observations similaires chez Patterson, 2000 ; Portes et Di Cristo, 2003). Néanmoins, même pour des catégories absolues (T, B, M)[1] nous voyons que les bandes de fréquence qui leur sont propres se chevauchent.

1. L'alphabet INTSINT comporte les symboles suivants : T(*op*), B(*ottom*), M(*iddle*), H(*igh*), L(*ow*), U(*pstepped*), D(*ownstepped*), S(*ame*).

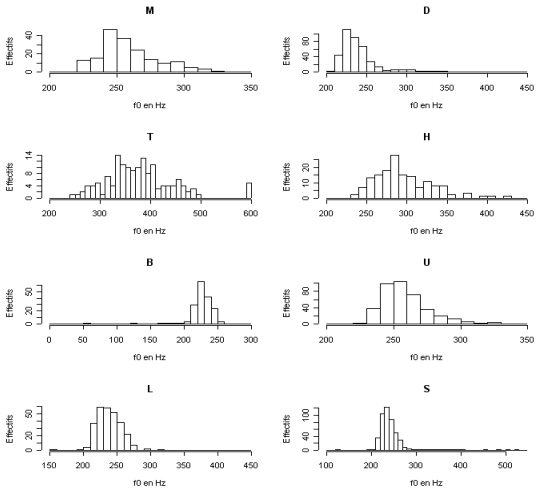

Figure 4
Histogrammes des fréquences de f0 selon la catégorie INTSINT

Nous avons également recouru à des mesures de dissymétrie et d'aplatissement pour caractériser les distributions résultantes. Le coefficient de dissymétrie permet de porter un jugement sur l'égalité de répartition autour de la moyenne. Nous prêtons attention surtout aux valeurs observées pour les symboles absolus (M, T, B) : pour les symboles relatifs, les valeurs éloignées de celles de la distribution normale paraissent cohérentes, vu leur caractère relatif et la dépendance des valeurs *(settings)* des paramètres initiaux. Nous constatons tout d'abord un coefficient de dissymétrie très élevé pour la catégorie B. En vérité, il est dû à quelques valeurs sporadiques (nombre d'observations égal à 6) dans la bande de fréquence de 100-150 Hz. Il s'agit des valeurs infra-basses pour notre locutrice, associées à des syntagmes à valeur pragmatique conclusive. Une fois ces valeurs exclues, le coefficient de dissymétrie affiche la valeur de 0,408, le coefficient d'aplatissement étant de -0,411. Prenons en compte également les coefficients de dissymétrie positifs pour les étiquettes M et T : une dissymétrie positive est révélatrice d'une longue queue à droite dans la distribution des valeurs de f0 (fig. 4). Ces tendances témoignent du fait que la locutrice « préfère » utiliser le registre élevé ; nous pouvons y voir l'effet des stratégies discursives adoptées (souci de ne pas être interrompue), mais également l'effet de l'organisation discursive en relation avec le développement de l'argumentation. Cependant, ces observations n'ont que le statut d'hypothèses et nécessitent une vérification dans une étude expérimentale spécifique.

Nous constatons également une dissymétrie positive importante pour la totalité des points-cibles (fig. 5) : la distribution a la forme d'une distribution de Weibul, plutôt que celle de la loi normale. Cette tendance peut être caractéristique de la locutrice de notre corpus. Or, elle fait penser également aux observations faites à propos de la dimension tonale de la parole spontanée avec une dominance des mouvements mélodiques ascendants. Il peut également s'agir d'une stratégie de la locutrice qui consiste à relever la ligne basse (ligne des vallées mélodiques) en raison de l'influence des contraintes discursives et des variations du débit de parole.

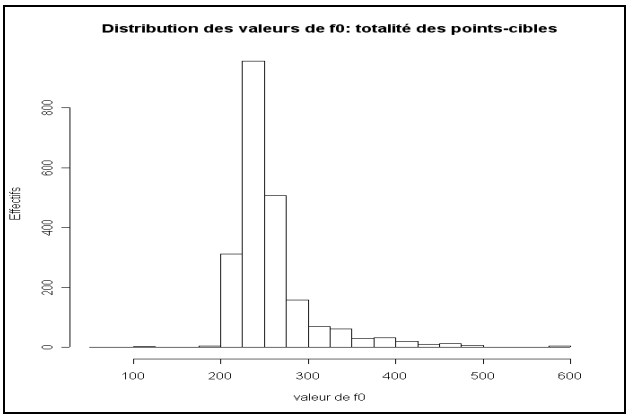

Figure 5
Histogramme des valeurs de f0 de tous les points-cibles

5. Représentation multiniveaux : bilan

Le format des fichiers d'annotation que nous avons développé correspond au modèle TextGrid pour le logiciel PRAAT. Ce format d'annotation permet d'isoler les différentes sources d'informations sur des paliers différents (« tiers »), en se rapprochant ainsi des représentations typiques de la phonologie multilinéaire ainsi que des représentations hiérarchiques de la phonologie prosodique. Simultanément, le logiciel PRAAT permet une formulation simple des requêtes, ce qui facilite et automatise la récupération des données phonétiques (nous évoquerons ces possibilités lors de la discussion dans les chapitres suivants des analyses effectuées à partir du corpus).

Nous disposons donc d'un corpus doté d'une annotation multi-niveaux, comportant les niveaux des phonèmes (1), des syllabes (2), des mots (3), des unités intonatives (4 et 5)[2], du degré de proéminence (6), des unités-interpausales (7), du codage INTSINT (9) avec les valeurs en Hz récupérées initialement du corpus (8) et prédites (10). Un extrait est présenté sur la figure 6.

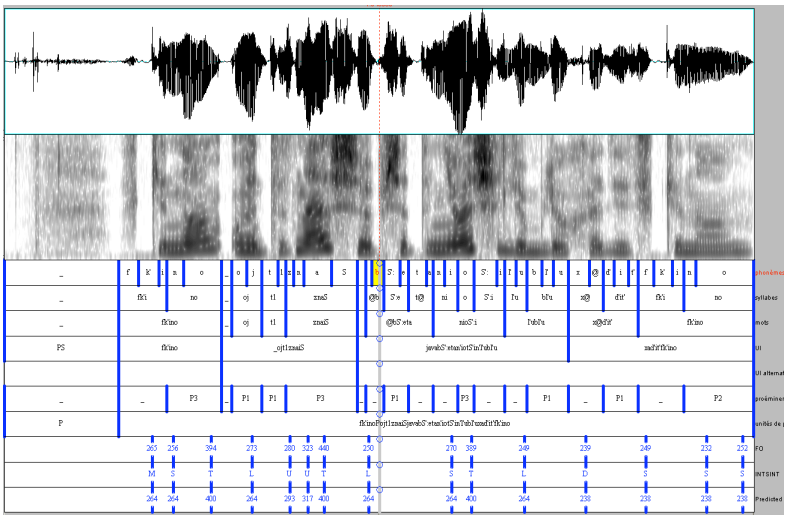

Figure 6
Exemple d'un TextGrid d'annotation du corpus

6. Observations préliminaires sur le phrasé prosodique

L'annotation multiniveaux s'avère être une ressource inestimable pour explorer les rapports entre les différents niveaux linguistiques, chacun représentant une partie des informations disponibles aux locuteurs en communication. Nous avons par conséquent décidé de proposer dans cette dernière partie de notre contribution quelques réflexions et observations portant sur la perception du phrasé prosodique en russe.

Rappelons que, dans l'annotation du phrasé prosodique, les experts avaient pour tâche de marquer les frontières des unités intonatives. Ce concept s'inscrit dans le paradigme de la phonologie prosodique, cadre de référence dans les travaux sur la structuration prosodique du flux de parole. Notamment, le phrasé prosodique reçoit dans le cadre de la phonologie prosodique une

2. Annotation de l'expert français pour la totalité du corpus et annotation des experts russes pour une partie du corpus.

modélisation formelle sous la forme d'une hiérarchie de constituants. Des hiérarchies multiples fondées sur des critères variés ont été proposées (Nespor et Vogel, 1986 ; Hayes, 1989 ; Selkirk, 1984 ; Beckman et Pierrehumbert, 1986). L'introduction de la hiérarchie des constituants prosodiques permet de rompre avec le parallélisme entre les structures syntaxique et phonologique. De même, la matérialité de la constituance prosodique a été démontrée dans de nombreuses études phonétiques, ainsi que dans les études neuro- et psycholinguistiques. Les chercheurs s'accordent sur l'interprétation de la constituance prosodique comme étant un cadre gérant l'implémentation des propriétés acoustiques du signal de parole.

Dans la hiérarchie classique des constituants, le syntagme intonatif est un des rares constituants définis par des propriétés acoustiques (donc, physiques) : entre autres, c'est le domaine d'un contour mélodique perceptivement cohérent et unifié. D'autre part, dans la phonologie prosodique de Nespor et Vogel (1986), l'algorithme de formation des unités intonatives s'appuie sur deux types d'informations : les informations morpho-syntaxiques et les informations sur la structure prosodique des niveaux inférieurs, cette structure étant supposée être connue, au moins partiellement, au moment de la dérivation de la structure intonative. Du point de vue de la morpho-syntaxe, il est nécessaire de disposer d'une partition de l'énoncé en termes de clauses et d'éléments appositionnels. Il est également considéré que seulement les frontières des clauses radicales *(root clauses)* [3] coïncident avec les frontières intonatives. Pour les clauses imbriquées, le marquage de la frontière gauche (et droite, selon la position dans la clause radicale) n'est pas obligatoire. Cependant, dans le travail de Halliday (1967), la structure clausale généralement détermine la distribution des frontières des unités intonatives.

Nous avons donc décidé, dans cette étude préliminaire, de procéder à une observation qualitative de l'articulation entre la structure clausale et la structure intonative perçue. Nos observations sont fondées sur l'annotation proposée par l'expert francophone ; cette segmentation a été réalisée sur la base de critères prosodiques formels, sans être influencée par les informations morpho-syntaxiques. L'unité initiale de notre analyse est l'unité inter-pausale *(cf. supra)*. Dans un premier temps, pour chaque unité inter-pausale, nous avons déterminé si elle comprend une ou plusieurs clauses. Dans le cas des unités comportant plusieurs clauses, les frontières prosodiques internes signalées ont été classées en fonction de leur correspondance ou non-correspondance avec la frontière inter-clauses. Nous avons également recueilli des informations sur l'éventuelle présence d'autres frontières prosodiques internes, les unités comportant des pro-phrases et des marques de communication ayant été exclues de l'analyse.

3. On définit la clause radicale comme une clause qui n'est pas contenue dans une autre clause.

Nous résumons les informations recueillies dans les tableaux 2 et 3 (voir *infra)*. Nous constatons que seulement un tiers des unités inter-pausales comportent plusieurs clauses. Ce résultat semble être la conséquence de la nature de notre corpus. En effet, en parole dialogique spontanée, les périodes d'échanges rapides de répliques alternent avec des périodes de développement étendu d'un sujet. En même temps, l'unité inter-pausale à la base de notre analyse étant une unité définie par des critères formels, elle ne correspond pas non plus à une unité de la structure discursive. Notamment, une réplique dans le dialogue peut comporter plusieurs IPUs, les pauses qui séparent les IPUs ayant une nature diverse : pauses structurelles *vs* pauses d'hésitation. Nous avons également pu constater que 38% des unités mono-clausales contiennent d'autres frontières prosodiques internes.

Tableau 2
Distribution des IPUs en fonction du nombre de clauses qu'elles contiennent

IPU comportant 1 clause	66%
IPU > 1 clause	34%

Si nous nous limitons aux IPUs comportant plusieurs clauses (tableau 3), nous constatons que dans 38% des cas les frontières morpho-syntaxiques entre deux clauses n'ont pas reçu de marquage intonatif particulier. Compte tenu du fait que d'autres frontières prosodiques internes ont été perçues par l'expert, nous avons pu constater que connaître la structure clausale du matériau verbal permet de prédire correctement environ 45% de frontières intonatives, et ceci dans des conditions où les informations sémantico-syntaxiques ne sont pas disponibles pour l'auditeur.

Tableau 3
Marquage prosodique de la structure clausale à l'intérieur des IPUs **≥** *2 clauses*

Marquage prosodique d'une frontière clausale	62%
Absence de marquage prosodique	38%

D'une part, le résultat obtenu corrobore l'hypothèse de la non-correspondance entre la structure prosodique et la structure syntaxique, pivot de la phonologie des domaines. Les études portant sur l'articulation prosodie ~ syntaxe constatent, que, entre autres, en présence d'indices grammaticaux, le marquage prosodique de la structure est optionnel. Des réflexions similaires sont menées dans le

cadre de l'approche de compétition de la génération du sens avec les informations apportées par différentes composantes linguistiques (Blache et Di Cristo, 2002), où les auteurs se penchent sur la question de savoir pourquoi la prosodie présente plus de variabilité dans certains contextes par rapport à d'autres. On rejoint également la problématique de la désambiguïsation des structures morpho-syntaxiques par des moyens prosodiques. Cependant, nous constatons que plus de 60% des unités non-ambiguës reçoivent un marquage prosodique. Nous observons donc un certain parallélisme entre les structures prosodique et syntaxique, assuré par le jeu d'indices acoustiques. Certes, le matériau sur lequel nous avons mené nos analyses n'est pas suffisamment large, par suite cette hypothèse mérite une exploration approfondie dans le futur.

Des recherches futures, fondées sur d'autres corpus et menées dans le cadre de l'approche de la phonologie de laboratoire, s'avèrent d'autant plus importantes que nous avons fondé tout d'abord ces observations préliminaires sur les données de l'expert froncophone, chez qui nous avons constaté une sensibilité plus marquée à la distribution de focus. Ainsi la figure 7 reprend l'annotation proposée pour l'IPU « Главное не куда поехать а главное с кем поехать » *(Le plus important ce n'est pas où aller, mais avec qui y aller).* La frontière entre deux clauses est absente dans cette annotation, cette frontière est même déplacée pour être alignée avec le mot porteur de l'accent nucléaire (en l'occurrence, l'adverbe куда, *« où »*). D'autre part, il semble difficile de recourir aux jugements des experts russophones, car dans les travaux expérimentaux, on dispose de peu de précisions sur la structuration du message en syntagmes : on s'appuie le plus souvent sur la structure syntaxique et on exige que les syntagmes soient des unités de sens, mais cette exigence n'est formulée qu'en termes très abstraits. La recherche d'un paradigme objectif d'annotation des frontières des unités intonatives s'impose. D. Hirst[4] propose une méthodologie objective, compatible avec l'organisation prosodique en lecture : comme en lecture, toute pause a une motivation structurelle et accompagne systématiquement une frontière prosodique, les positions ainsi marquées sont des sites inconditionnels de frontières prosodiques ; par conséquent, leurs réalisations acoustiques peuvent être analysées afin d'aboutir à un modèle acoustique. Cette approche est innovante dans la mesure où la méthodologie habituelle consiste à confronter l'annotation du phrasé par des linguistes selon les principes théoriques, souvent à la base des transcriptions de texte, et les productions des locuteurs. À l'étape suivante, l'algorithme d'annotation automatique pourrait être appliqué à l'annotation du corpus spontané et les

4. Projet de recherche 145 *:* Typologie prosodique : développement d'un protocole objectif d'analyse de la prosodie des langues.

annotations résultantes pourraient être confrontées aux structures morpho-syntaxique et discursive.

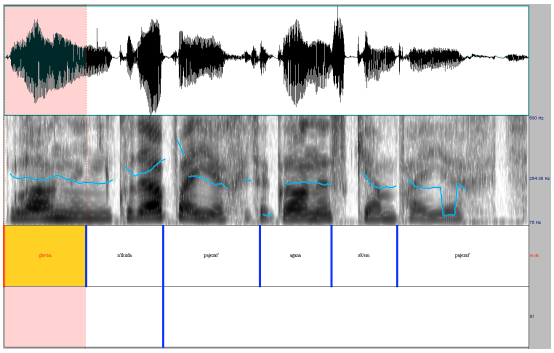

Figure 7
Phrasé prosodique perçu pour l'IPU « Главное не куда поехать а главное с кем поехать »
(Le plus important ce n'est pas où aller, mais avec qui y aller)

7. Résumé et conclusion

Dans notre contribution, nous avons présenté un corpus de parole spontanée en langue russe. Notre corpus étant enrichi d'une annotation miltiniveaux et aligné temporellement pour chaque niveau d'annotation, nous avons spécifié des principes qui ont guidé l'annotation à chaque niveau retenu. Enfin, nous avons procédé à une discussion critique de la performance de l'algorithme MOMEL-INTSINT dans l'annotation de ce corpus de parole spontanée en langue russe.

Par la suite, nous nous sommes penchée sur la question du phrasé prosodique d'autant plus actuelle pour le russe, les formalismes de la phonologie des domaines n'ayant pas encore été appliqués dans les analyses prosodiques pour cette langue. Nous avons notamment analysé l'articulation entre l'organisation clausale du corpus et le phrasé prosodique perçu par un expert francophone. Nous avons pu constater que la structure clausale (définie par référence au concept de prédicativité) ne détermine que 45% de frontières d'unités intonatives annotées. Dans nos analyses, nous avons cherché à démontrer tout l'intérêt qu'il y a à mener un travail de description formelle de l'organisation prosodique à partir d'un corpus enrichi d'annotations appropriées. Simultanément, les résultats obtenus nous poussent à nous interroger sur le statut de l'unité de *синтагма* (« syntagme »), concept de référence dans les analyses prosodiques du russe.

8. Références bibliographiques

AVANESSOV, R.I. (1954). Аванесов, Р.И., О словоразделе и строении слога в русском языке. *Вопросы языкознания*, 6.

BECKMAN, M.E. (1997). A typology of spontaneous speech, in Y. Sagisaka ; W. N. Campbell & N. Higuchi (eds) *Computing Prosody*, New York: Springer-Verlag, p. 7-26.

BECKMAN, M.E. & PIERREHUMBERT, J.B. (1986). Intonational Structure in Japanese and English, *Phonology Yearbook,* 3, p. 255-309.

BECKMAN, M.E. & EDWARDS, J. (1990). Lengthenings and Shortenings and the Nature of Prosodic Constituency, in Kingston, J. ; Beckman, M.E. (eds), *Papers In Laboratory Phonology I. Between the Grammar and Physics of Speech,* p. 152-178.

BLACHE, Ph. & DI CRISTO, A. (2002). Variabilité et dépendances des composants linguistiques, *Actes de la Conférence Traitement Automatique du Langage Naturel* (TALN-2002).

BOERSMA, P. & WEENINK, D. (2005). *Praat: a system for doing phonetics by computer,* <http://www.fon.hum.uva.nl/praat/>.

BROWN G. & YULE G. (1983). *Discourse analysis,* Cambridge: Cambridge University Press.

CEDERGREN, H. & SANKOFF, D. (1974). Variable rules: Performance as a statistical reflection of competence, *Language,* 50, p. 333-355.

D'IMPERIO, M. (2005). La Phonologie de Laboratoire : finalités et quelques applications, in Nguyen, N. ; Wauquier-Gravelines, S. ; Durand, J. (éds*), Phonologie et phonétique: Forme et substance.* Paris : Hermès, p. 241-264.

DI CRISTO, A. & HIRST, D. (1996). Vers une typologie des unités intonatives du français, *Actes des XXIes Journées d'Étude sur la Parole,* Avignon, p. 219-222.

HALLIDAY, (1967). *Intonation and grammar in British English,* The Hague: Mouton.

HAYES, B. (1989). The prosodic hierarchy in meter, in P. Kiparsky & G. Youmans (eds), *Rhythm and meter,* Orlando: Academic Press, p. 201-260.

HIRST, D. (2004). Lexical and Non-lexical Tone and Prosodic Typology, *Proceedings of International Symposium on Tonal Aspects of Language,* 28-30 mars 2004, Beijing, China, p. 81-88.

KASSEVITCH, V. (2002). Касевич, В.Б., О фонетическом слове. *Вопросы фонетики,* 4.

KOISO, H. ; HORUICHI, Y. ; TUTIYA, S. ; ICHIKAWA, A. & DEN, Y. (1998). An analysis of turntaking and backchannels based on prosodic and syntactic features in Japanese map task dialogs, *Language and Speech,* 41, p. 295-321.

LADD, D.R. (1992). An Introduction to Intonational Phonology (Prosody), in Docherty, G.J. ; Ladd, D. R. (eds), *Papers in Laboratory Phonology II. Gesture, Segment, Prosody,* p. 321-334.

LAVER, J. (1994). *Principles of Phonetics,* Coll. Cambridge Textbooks in Linguistics, Cambridge: Cambridge University Press.

LEVELT, W. (1989). *Speaking. From Intention to Articulation,* Cambridge: Cambridge University Press.

LICKLEY, R. ; SCHEPMAN, A. & D.R. LADD (2005). Alignment of "phrase accent" lows in Dutch falling-rising questions: Theoretical and methodological implications, *Language and Speech,* 48, p. 157-183.

NESPOR, M. & VOGEL, I. (1986). *Prosodic Phonology,* Dordrecht: Foris Publication.

PATTERSON, D. (2000). *A linguistic approach to pitch range modelling,* Thèse de doctorat, Edinburgh University.

PEPERKAMP, S. (1999). Prosodic words, *GLOT International,* 4(4), p. 15-16.

PIERREHUMBERT, J. ; BECKMAN, M. E & LADD, D. R. (2001). Conceptual Foundations of Phonology as a Laboratory Science, in Burton-Roberts, N. ; Carr, P. and Docherty, G. (eds) *Phonological Knowledge,* Oxford: Oxford University Press, p. 273-304.

PORTES, C. & DI CRISTO, A. (2003). Pitch range in spontaneous speech: semi-automatic approach *versus* subjective judgement*, Proceedings of the 15th ICPhS*, 3-9 août, Barcelone, Espagne, p. 583-586.

SCHERBA, L.V. (1963). Щерба, Л.В., *Фонетика французского языка.* 7-ое изд., Москва.

SELKIRK, E. (1984). *Phonology and Syntax : The Relation Between Sound and Structure,* Coll. Current Studies in Linguistics Series, 10, Cambridge: The MIT Press.

SKRELIN, P.A. (1999). Скрелин, П.А., *Сегментация и транскрипция*. Санкт-Петербург.

YULE, G. (1980). Speaker's topic and major paratones, *Lingua*, 52, p. 33-47.

ZLATOOUSTOVA, L. ; POTAPOVA, R. ; TROUNINE, V. (1986). *Phonétique générale et appliquée*, Moscou : Éditions de l'université de Moscou (en russe).