

Thèse de l'Institut National Polytechnique de Grenoble
Spécialité Signal, Image, Parole et Télécommunication
2 juin 2003

présentée par
Bleicke HOLM

SFC :

un modèle de superposition de contours multiparamétriques
pour la génération automatique de la prosodie –
Apprentissage automatique et application à l'énonciation de formules
mathématiques

Composition du jury :

Jean-Luc SCHWARTZ
Anne LACHERET-DUJOUR
Hansjörg MIXDORFF
Michel CONTINI
Gérard BAILLY



Institut de la Communication Parlée
46 av. Félix Viallet, 38031 Grenoble CEDEX 1 France

Résumé

Cette thèse traite de la génération automatique de la prosodie en vue de la synthèse de la parole à partir de texte.

Nous partons de l'hypothèse que la prosodie encode dans le continuum du signal acoustique les instances de fonctions linguistiques et para-linguistiques en nombre relativement restreint (hiérarchisation, segmentation, attitudes, focalisation etc.) par des contours multiparamétriques dont l'étendue correspond à la portée des fonctions. Ces objets (les contours) interagissent par superposition additive. Nous proposons un modèle, le SFC « Superposition of Functional Contours », qui réalise la factorisation de réalisations prosodiques en objets associés aux fonctions à l'aide d'une analyse par la synthèse. La décomposition obtenue est optimale et économique par rapport à un ensemble de données (corpus) au sens où la forme des contours relève uniquement des besoins statistiques sous les contraintes fonctionnelles que nous imposons.

Le SFC a été appliqué à des corpora d'oralisation de formules mathématiques et de phrases isolées en français. Il a été également appliqué à l'allemand et au galicien.

Mots clés : prosodie, synthèse de parole, génération automatique, superposition, contours

Abstract

The object of this thesis is the automatic generation of prosody in the framework of text-to-speech synthesis.

Our main hypothesis is that prosody encodes instances of a rather small number of linguistic and para-linguistic functions (signaling hierarchy, segmentation, attitudes, focus etc.) by means of multi-parametric contours that are coextensive with the range of the functions. These objects (contours) interact by additive superposition. We propose a model, the SFC "Superposition of Functional Contours", that uses an analysis-by-synthesis method in order to retrieve from the prosodic continuum the interacting objects associated with the functions. The factorization proposed is optimal with respect to a given set of observations (corpus) in the following sense: the shape of contours depends only on the statistical needs provided the imposed functional constraints.

The SFC has been applied to a corpus of spoken mathematical formulas and to several corpora of isolated French sentences. It has also been applied to German and Galician.

English title: SFC: Superposition of Functional Contours. A model for automatic prosody generation.

Keywords: prosody, speech synthesis, automatic generation, superposition, contours

Remerciements

Mon premier merci ne peut s'adresser qu'à Gérard Bailly, principal interlocuteur pour tout ce qui se trouve sur ces pages. Sans lui, cette thèse n'existerait simplement pas. Je le remercie également de m'avoir laissé faire à ma tête (dure) – qualité importante d'un directeur de thèse – je parle de « laisser faire » non pas de la « tête dure » . . .

Pour l'accueil au sein de l'ICP, je remercie Pierre Escudier, son ancien directeur, et Jean-Luc Schwartz, le nouveau. Par ailleurs, c'est ce dernier qui, il y a six ans, a répondu d'une manière ouverte et encourageante à mon premier contact avec ce laboratoire.

Le travail de cette thèse a nécessité un recours permanent aux moyens informatiques du laboratoire. Je remercie l'équipe technique d'avoir assumé la tâche peu gratifiante de garantir son fonctionnement au quotidien.

Un grand merci à Stephan Raidt : travailler avec lui sur l'application du modèle à l'allemand et l'évaluation comparative a été efficace et enrichissant. Ce projet n'aurait pas été possible sans l'« encadrement allemand » de la part de Hansjörg Mixdorff – merci !

Pour leur accueil chaleureux à la Dublin City University pendant une « courte mission » dans le cadre de COST 258, je remercie Alex Monaghan et Donal Fitzpatrick. Les nombreuses discussions avec eux ont été une source de motivation importante.

Je tiens à remercier Eduardo Banga de l'université de Vigo : son séjour à Grenoble, également dans le cadre de COST, a permis à notre modèle de faire ses premiers pas en galicien.

Puis, je dois un grand merci à Hélène Løevenbruck et Guillaume Rolland qui ont conçu et enregistré un grand corpus de phrases déclaratives en français. Ce corpus a été un objet d'étude idéal pour notre modèle. Ce corpus mérite encore beaucoup d'analyses dans l'avenir.

Je remercie Jean-Pierre Orliaguet du LPNC à Grenoble pour son aide à la mise en place de la tentative d'une évaluation perceptive du modèle dans une tâche de dictée.

Je tiens à remercier énormément beaucoup Matthias Odisio qui a relu et corrigé la majeure partie de ce manuscrit. Si toutes fois vous trouvez une seule faute d'orthographe, une phrase maldite, charabiaque ou tout simplement incompréhensible, je vous prie de vous adresser directement à lui – il en est le seul responsable !

Table des matières

Introduction	17
1 État de l'Art	19
1.1 Modèles phonétiques de f_0	22
1.1.1 Stylisation par points-clés	24
1.1.2 Modèles de commande	25
1.1.3 Stylisation ancrée sur la structure segmentale	28
1.1.4 Pertinence perceptive de la description phonétique	29
1.1.5 Modèles basés sur des événements	31
1.1.6 Résumé	33
1.2 Modèles phonologiques de f_0	34
1.2.1 Classification fonctionnelle	34
1.2.2 Systèmes de transcription tonale	36
1.2.3 Descriptions phonologiques « ad hoc »	37
1.2.4 Résumé	39
1.3 Entre phonétique et phonologie	40
1.3.1 Le grand saut	40
1.3.2 Contraintes phonétiques fortes	44
1.3.3 Une grammaire de mouvements élémentaires	44
1.3.4 Enchaînement de contours stockés	45
1.3.5 ToBI vers f_0	46
1.3.6 ToBI – mais pas trop	47
1.3.7 Prédiction d'une description phonologique	48
1.3.8 Résumé	49
1.4 Sélection d'unités	50
1.5 Modélisation des durées	53
1.5.1 Prédiction directe des durées segmentales	53
1.5.2 Prédiction par l'intermédiaire des durées syllabiques	54

2	Le modèle	57
2.1	Préliminaires	58
2.1.1	La composante rythmique	58
2.1.2	La composante mélodique	70
2.1.3	Contours multiparamétriques	72
2.2	Motivation	74
2.2.1	Un modèle issu de l'ICP	74
2.2.2	Arbres de performance	77
2.3	Description	83
2.3.1	Fonctions et leurs portées	84
2.3.2	Générateurs et familles de contours	85
2.3.3	Superposition	86
2.4	Apprentissage	88
2.4.1	Hypothèses et définitions	90
2.4.2	La distribution du résidu	92
2.4.3	L'initialisation	99
2.4.4	Les contours de référence	101
2.4.5	Les générateurs de contours	102
2.4.6	Résumé	107
2.5	Résumé	108
3	Résultats et applications	109
3.1	Présentation des corpora	109
3.1.1	Formules mathématiques lues	109
3.1.2	Six attitudes	117
3.1.3	1001 phrases	117
3.1.4	Le corpus de Vigo	117
3.2	Résultats quantitatifs	120
3.2.1	Considérations générales	120
3.2.2	Prédiction et caractéristiques des générateurs	122
3.3	Évolution	131
3.3.1	Un début très simple : la version dc	132
3.3.2	Ajouter une relation de dépendance : la version d	133
3.3.3	Diversifier les dépendances : la version dd	133
3.3.4	Une fonction de segmentation : la version x	134
3.3.5	Identifier une fonction spécialisée : la version dv	136
3.3.6	Réinterprétation unifiante : la version dx	137
3.3.7	Les erreurs de prédiction	141
3.3.8	Résumé	142
3.4	Expansion du mouvement	143
3.5	Qualification et nature du qualifiant	147

3.5.1	Résumé	158
3.6	Formules mathématiques – français	160
3.7	Formules mathématiques – allemand	166
3.8	Corpus de Vigo	170
4	Évaluation perceptive	173
4.1	Une tâche de dictée	174
4.1.1	Mise en place du test	174
4.1.2	Les erreurs	175
4.1.3	Les mesures temporelles	178
4.1.4	Le questionnaire	179
4.1.5	Conclusions	180
4.2	Évaluation comparative	181
4.2.1	Adaptation de l'IGM	181
4.2.2	Méthodologie	182
4.2.3	Les résultats	183
4.2.4	Conclusions	192
	Conclusions et perspectives	195
A	Corpora	199
A.1	Les formules mathématiques – français	199
A.2	Les formules mathématiques – allemand	202
A.3	Les 1001 phrases	211
B	Erreurs et corrélations	221
C	Evaluation	227
C.1	Test de préférence	227
C.1.1	Contours des stimuli de test	227
C.1.2	Tableaux des réponses – français	229
C.1.3	Tableaux des réponses – allemand	230
	Bibliographie	231

Table des figures

1	Synthèse de parole à partir de texte	17
1.1	Le modèle prosodique entre texte et signal	20
1.2	Paramétrisation de f_0 par rapport à la syllabe	23
1.3	Schéma du modèle de Fujisaki	26
1.4	Modèle phonétique de Traber	29
1.5	Schéma des modèles de prédiction prosodique	41
1.6	Sélection d'unités	51
2.1	Optimisation des paramètres de répartition	61
2.2	Distribution des z-scores	62
2.3	Génération des pauses : déterminer un seuil	64
2.4	z-score réel vs. z-score virtuel	66
2.5	Microprosodie	71
2.6	Contour multiparamétrique	72
2.7	Du contour vers l'arbre de performance	78
2.8	Structure syntaxique vs. structure de performance	80
2.9	L'attachement d'un opérateur mathématique	81
2.10	Attachement et balance syllabique	82
2.11	Exemple de superposition	87
2.12	Cycle d'analyse par la synthèse	89
2.13	Distribution du résidu	93
2.14	Distribution du résidu : lissage	96
2.15	Distribution du résidu : renforcement	99
2.16	Attribution des contours aux niveaux hiérarchiques	101
2.17	Implémentation des générateurs de contours	105
3.1	Formule 117 analysée par la grammaire	112
3.2	Caractéristiques du corpus français	113
3.3	Distribution des empans du corpus de formules (français)	113
3.4	Caractéristiques du corpus de formules (allemand)	116
3.5	Distribution des empans du corpus allemand	116

TABLE DES FIGURES

3.6	Caractéristiques du corpus « 1001 phrases »	118
3.7	Distribution des empanns du corpus « 1001 phrases »	118
3.8	Caractéristiques du corpus de Vigo	119
3.9	Distribution des empanns du corpus de Vigo	119
3.10	Erreurs de prédiction vs. types de rampe d'entrée	125
3.11	Exemple synthétisé avec des générateurs à entrée réduite	126
3.12	Erreurs de prédiction vs. types de rampe d'entrée (2 entrées)	127
3.13	Erreurs de prédiction vs. types de rampe d'entrée	128
3.14	Erreurs de prédiction vs. nombre d'unités cachées	130
3.15	Exemple avec la version dc	132
3.16	Exemple avec la version d	133
3.17	Exemple avec la version dd	134
3.18	Exemple avec la version x	135
3.19	Exemple avec la version x (une fonction négligée ?)	136
3.20	Exemple avec la version dv	137
3.21	Exemple avec la version dx	138
3.22	Familles de contours	139
3.23	Familles de contours X	140
3.24	Evolution du modèle : les erreurs de prédiction	141
3.25	Expansion de mouvements porteurs, f_0	144
3.26	Expansion de mouvements porteurs, C	145
3.27	Extrapolation de mouvements porteurs, f_0	146
3.28	Qualification, 2+3 syllabes, f_0	149
3.29	Qualification, 2+3 syllabes, C	150
3.30	Qualification, 4+2 syllabes, f_0	151
3.31	Qualification, 4+2 syllabes, C	152
3.32	Qualification, 4+3 syllabes, f_0	153
3.33	Qualification, 4+3 syllabes, C	154
3.34	Qualification, 3+4 syllabes, f_0	155
3.35	Qualification, 3+4 syllabes, C	156
3.36	Qualification, 2+5 syllabes, f_0	157
3.37	Qualification, 2+5 syllabes, C	157
3.38	Qualification – résumé	158
3.39	Qualification – interprétations	159
3.40	Structure de performance : observée vs. prédite	160
3.41	L'attachement : observé vs. prédit	161
3.42	Attachement vs. balance syllabique : observations et prédictions	162
3.43	Attachement : observations vs. prédictions	162
3.44	Prédiction de f_0 : deux exemples	163
3.45	Prédiction de f_0 : dernier exemple	164
3.46	Famille de contours : DG	165

3.47	Structure de performance : observée vs. prédite (allemand)	166
3.48	Attachement vs. balance syllabique : observation et prédiction (allemand)	167
3.49	Attachement : observations vs. prédictions (allemand)	167
3.50	Prédiction de f_0 : formules en allemand	169
3.51	Deux exemples pour le galicien	171
4.1	Tâche de dictée : acquisition des données	176
4.2	Tâche de dictée : résultats du MOS	180
4.3	Test de préférence – capture d’écran	183
4.4	Préférences totales en comparaison directe.	184
4.5	Fouille des données – histogrammes	185
4.6	Fouille des données – triangles	188
4.7	Fouille des données – hexagones	189
4.8	Erreurs objectives vs. préférences subjectives	193
C.1	Contours du test de préférence – français	228
C.2	Contours du test de préférence – allemand	228

Liste des tableaux

2.1	Erreurs de répartition – GIPCs sans pauses	60
2.2	Taux d’erreurs de la prédiction des pauses	65
2.3	Erreurs de prédiction : pauses silencieuses	67
2.4	Erreurs de répartition – GIPCs avec et sans pauses	67
2.5	Répartition – Evaluation	69
2.6	Les attributs du lexique posodique chez Aubergé	75
B.1	Erreurs & corrélations : formules mathématiques	223
B.2	Erreurs & corrélations : attitudes (DC, QS, DI)	224
B.3	Erreurs & corrélations : attitudes (EV, EX, SC)	225
B.4	Erreurs & corrélations : 1001 Phrases & Vigo	226
C.1	Test de préférence – français. SFC–igm	229
C.2	Test de préférence – français. Org–SFC	229
C.3	Test de préférence – français. Org–igm	229
C.4	Test de préférence – allemand. SFC–igm	230
C.5	Test de préférence – allemand. Org–SFC	230
C.6	Test de préférence – allemand. Org–igm	230

Introduction

Cadre de travail. Ce mémoire présente un modèle de génération automatique de la prosodie. La figure 1 schématise d'une manière très simplifiée la chaîne de traitement dans le cadre de la synthèse de parole à partir du texte par concaténation de diphtones : une analyse linguistique du texte (qui peut être enrichi d'informations contextuelles) fournit d'une part la séquence de segments à enchaîner, et d'autre part l'information requise par le modèle prosodique pour la prédiction de paramètres tels que la fréquence fondamentale, les durées segmentales etc. Ces derniers sont utilisés pour la mise en forme de l'enchaînement « brut » des segments à l'aide de techniques de traitement du signal afin de produire le signal acoustique de parole.

La présente étude ne se préoccupe que d'une petite partie de cette chaîne qui est cependant primordiale pour la qualité de la parole synthétique : le modèle prosodique.

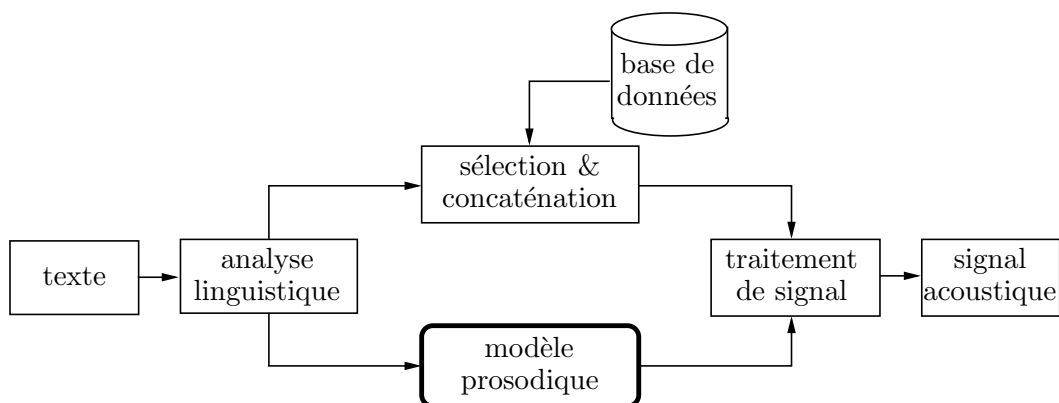


FIG. 1: **Synthèse de parole à partir de texte.** Le modèle de génération de la prosodie s'intègre dans la chaîne de traitement du texte vers la parole en spécifiant les paramètres nécessaires à la mise en forme de l'enchaînement des segments stockés dans une base de données.

Le modèle SFC (Superposition of Functional Contours) que nous proposons est caractérisé par un encodage direct de fonctions linguistiques ou paralinguistiques par des contours multiparamétriques, ancrés sur la structure syllabique¹. L'étendue de ces contours correspond à la portée des fonctions et leur morphologie ne dépend que de la fonction et des propriétés phonotactiques des parties de l'énoncé concernées. L'hypothèse centrale est qu'une même fonction peut être appliquée à tout niveau hiérarchique. Leurs différentes contributions se superposent et fournissent ainsi le contour prosodique global.

Cette étude se concentre particulièrement sur les fonctions de segmentation et de hiérarchisation. L'analyse d'un corpus contenant des oralisations de formules mathématiques qui exigent, avec leur structure profondément emboîtée, un fort recours aux capacités hiérarchisantes de la prosodie, montre que seulement deux fonctions qui expriment une dépendance gauche ou droite entre deux groupes de mots suffisent à la génération d'une réalisation prosodique riche et proche de la référence humaine naturelle.

Le modèle sera également appliqué à un large corpus de phrases déclaratives isolées, ainsi qu'à deux autres langues : le galicien et l'allemand.

Plan du manuscrit. Ce manuscrit est organisé en quatre parties : la première propose un bref état de l'art qui tente de décortiquer les présupposés phonétiques et phonologiques de quelques modèles de génération de la prosodie afin de pouvoir situer notre approche.

La deuxième est consacrée à la présentation du modèle. Son inversion, c.-à.-d. l'apprentissage des générateurs de contour qu'il met en oeuvre, y prend une place importante. En effet, le modèle est étroitement lié au processus d'apprentissage, car ce n'est que par la confrontation aux contours prosodiques observés dans un corpus soigneusement construit que la morphologie des contours prend forme et sens : dans un cycle d'analyse par la synthèse, les contraintes de haut niveau permettront la décomposition en contours attachés aux fonctions (para-)linguistiques préalablement identifiées.

La troisième partie contient la description des différents corpora utilisés ainsi que les résultats de synthèse réalisée par le SFC. De plus, on verra comment utiliser le modèle comme un outil d'analyse permettant l'identification des manifestations prosodiques de nouvelles fonctions.

La dernière partie propose d'abord l'esquisse d'une évaluation perceptive visant à accéder à la charge cognitive (plus ou moins lourde selon l'adéquation des contours prosodiques générés) d'un auditeur en observant son écriture dans une tâche de dictée de formules mathématiques. Puis, l'application du modèle à l'allemand permet une évaluation plus « classique » en le comparant au modèle de Mixdorff (grâce à son adaptation pour l'oralisation de formules mathématiques par Stephan Raidt).

¹Plus précisément : la séquence des GIPCs (« group inter perception center »).

État de l'Art

Introduction

Cette étude a comme objet la génération automatique de la prosodie à partir du texte. Le nombre et la variété de modèles proposés à ce sujet dans la littérature sont tels qu'il me semble impossible de fournir ici un véritable « état de l'art » comme le titre de cette partie le promet. Ce que cette partie propose est donc une sélection relativement restreinte de différentes approches. Les exemples détaillés nous serviront de prétexte pour tisser la toile de fond des problèmes auxquels elles se heurtent et des solutions proposées. Le lecteur à la recherche d'une véritable revue des modèles prosodiques du français peut se référer à l'excellent livre de Lacheret-Dujour et Beaugendre [1999] qui ouvre l'horizon à des théories telles que celle de Rossi [1999] qui dépassent notre cadre de la synthèse de parole.

La figure 1.1 reprend de la figure 1 la partie qui nous préoccupera tout au long de ce manuscrit : le modèle prosodique. Il me semble bénéfique de le décomposer afin de pouvoir comparer les différentes propositions étape par étape et de mettre en valeur les différents choix adoptés. Une telle comparaison n'est pas sans risque car les modèles sont très hétérogènes. Chaque grille d'étapes que l'on peut proposer sera plus ou moins adaptée à un modèle donné. Nous allons prendre ce risque en proposant une analyse selon la figure 1.1.

Plan de la partie. Les deux extrêmes de la chaîne de génération sont le *texte* et le *signal acoustique*. Le texte peut être enrichi d'informations linguistiques et paralinguistiques (p.e. dans le cadre de génération à partir de concepts).

Le *modèle phonétique* extrait du signal ce qui est jugé pertinent pour la prosodie. Il s'appuie sur une paramétrisation de la forme d'onde fournie par le traitement du signal. Il délivre des indices caractéristiques de ces paramètres et doit permettre la resynthèse du continuum prosodique. En général, le modèle phonétique traite de la fréquence fondamentale et de la segmentation en phonèmes. On peut y ajouter l'énergie,

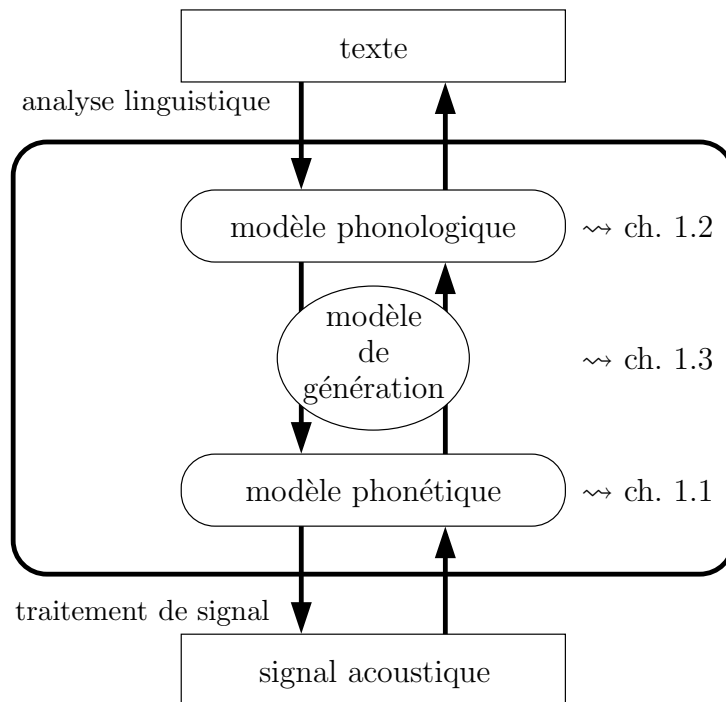


FIG. 1.1: Le modèle prosodique entre texte et signal.

des caractéristiques spectrales etc. Dans le chapitre 1.1, nous détaillerons quelques propositions faites dans la littérature.

Côté texte, nous identifions la *description phonologique*. Souvent on distingue la phonologie profonde et la phonologie de surface. La première permet une organisation de la structure fonctionnelle selon les besoins de l'oralisation. La seconde décrit la réalisation de la phonologie profonde en terme d'éléments jugés perceptivement distinctifs. Il en découle une relation importante entre phonétique et phonologie : le modèle phonétique a la charge d'« implémenter » dans le signal ces éléments distinctifs. Le chapitre 1.2 se concentrera surtout sur des aspects de la phonologie de surface.

Finalement, dans le chapitre 1.3, nous tenterons de comprendre comment les différents auteurs mettent ces deux descriptions en correspondance.

Remarques. Il convient d'insister sur le fait que les étapes distinguées ici – bien qu'elles soient largement reconnues (cf. par exemple Hirst et al. [2000]) – ne sont pas obligatoires. Les modèles qui les emploient de manière distincte et explicite sont plutôt rares. Il ne s'agit pas non plus d'étapes indépendantes : les choix faits à un endroit influent souvent sur les autres choix. Ainsi, la distinction entre modèle phonologique (de surface) et le modèle phonétique n'est pas toujours aisée. C'est à la frontière entre

phonologie et phonétique que s'opère la transition du monde symbolique et discret vers la réalisation physique continue. Les différents modèles tentent donc dans les étapes avoisinant cette frontière de réduire l'écart, et l'attribution des opérations employées à l'un ou à l'autre module peut quelquefois sembler artificielle. La grille proposée nous permettra néanmoins de mettre en parallèle *dans la mesure du possible* les différents choix des modèles.

Dans la figure 1.1, les étapes sont liées par des flèches descendantes dont la signification est évidente dans le cadre de la génération de la prosodie à partir du texte. Les flèches montantes nécessitent en revanche une explication : aucun modèle valable ne peut être mis au point sans référence à des observations. D'une manière ou d'une autre, il contient des paramètres qui doivent être ajustés, ce qui implique qu'il doit être possible d'inverser au moins une partie des transitions entre étapes. Nous allons voir pour les différents modèles où se situe sur notre échelle le point de rencontre entre les *a priori* des modèles et l'observation des données. Remarquons déjà ici que l'inversion peut prendre des formes très différentes : elle peut consister en des méthodes d'apprentissage qui servent uniquement à l'ajustement des paramètres du modèle, ou elle peut être très étroitement liée au modèle en procédant à une analyse des observations par la synthèse – comme c'est le cas du modèle que nous allons proposer dans la deuxième partie de ce manuscrit.

1.1 Modèles phonétiques de f_0

Selon le Petit Robert la phonétique est la « branche de la linguistique qui étudie les sons des langues naturelles ». Cette définition nécessite quelques précisions, notamment dans le cadre de la présentation des modèles prosodiques que nous nous proposons dans ce chapitre. La phonétique identifie et décrit les sons des langues. Elle ne procède pas directement à leur classement selon les fonctions qu'ils remplissent (ce qui est la tâche de la phonologie). La description phonétique d'un énoncé le caractérise dans son unicité en extrayant d'une paramétrisation (obtenue grâce aux moyens de traitement de signal) des indices qui permettent la resynthèse de l'énoncé. Le modèle phonétique sert à préciser l'objet de l'étude : voilà, de quoi on traite et comment l'on le paramétrise.

Nous allons d'abord nous restreindre au traitement de la fréquence fondamentale (f_0). Nous présumons la connaissance de son évolution au cours d'un énoncé. Cette courbe peut – là où elle est définie, donc dans les parties voisées du signal – être obtenue de manière automatique et avec une assez bonne fiabilité grâce à des méthodes d'analyse de signal.

On peut penser que l'axe « naturel » de l'évolution de f_0 est le temps : c'est ce que l'on obtient d'un point de vue « traitement de signal ». Or, la parole offre un autre axe (tout aussi naturel) : la structure segmentale. La première grande distinction entre les modèles phonétiques proposée ici est donc de déterminer s'ils font explicitement référence à la structure segmentale ou non. Ce critère peut sembler surprenant car évidemment, tant que les durées des phonèmes¹ sont connues (ou correctement prédites) le changement de repère secondes vers syllabes est toujours réversible – tout comme l'est la transformation de f_0 en échelle logarithmique. Le choix du repère conditionne néanmoins ce que les modules supérieurs voient comme contour mélodique. Les deux contours à gauche de la figure 1.2 ont des allures très différentes si leur évolution est exprimée en fonction du temps, mais ils ont la même forme par rapport au repère syllabique. L'inverse, c.-à.-d. des contours ayant une forme identique dans le domaine temporel et différent par rapport aux syllabes, est évidemment tout aussi possible. Ceci peut paraître banal, mais illustre bien le rôle du modèle phonétique : mettre les données acoustiques en forme – une forme qui doit faciliter la mise en relation avec le modèle phonologique, et qui est en même temps un filtre à travers lequel nous regardons les données.

Nous avons vu que le choix du repère a une influence sur la dimension descriptive du modèle phonétique. La distinction devient encore plus importante si l'on regarde sous l'angle de la synthèse : le but d'un modèle de génération de la prosodie est de prédire la réalisation phonétique d'un énoncé. Or, traiter la fréquence fondamentale comme une

¹Tout au long de ce manuscrit, j'utiliserai le terme « phonème » d'une manière simpliste : soit en tant que segment auquel on peut attribuer, par exemple, une durée ; soit en tant que classe de ces segments qui est, entre autre, caractérisée par une durée moyenne.

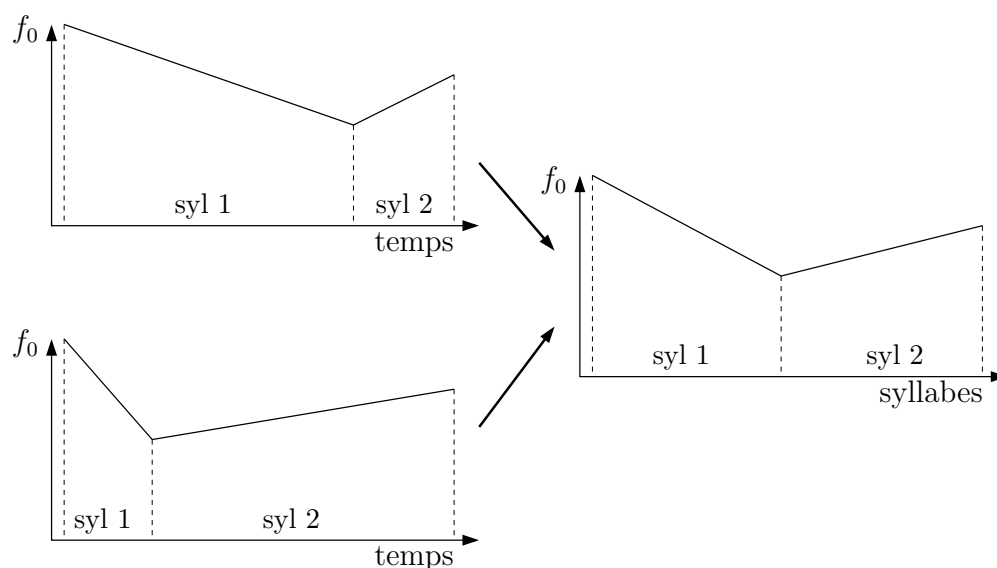


FIG. 1.2: **Changement de repère : temps \longleftrightarrow syllabes.** À gauche : paramétrisation de la courbe de f_0 par rapport au temps ; à droite : par rapport aux syllabes. Un simple changement de repère dans la description des courbes modifie ainsi la notion de ce que l'on considère comme semblable.

fonction du temps revient soit à dire qu'elle est indépendante de la structure temporelle des segments, soit à inclure dans le modèle de génération un mécanisme d'alignement entre les deux axes. Ou, vu de l'autre côté : un modèle faisant *a priori* que le distinctif de l'évolution de f_0 s'exprime par rapport à la structure segmentale tente ainsi de factoriser dès la description phonétique la prédiction rythmique et mélodique.

Le deuxième aspect sous lequel nous allons regarder les modèles phonétiques est, s'ils isolent dans le signal acoustique des parties significatives. Nous verrons d'un côté des modèles faisant référence à des points cibles, des mouvements linéaires ou des formes d'accents – et de l'autre côté des modèles qui attachent une importance plus homogène à tout l'énoncé. Les premiers ont l'avantage d'accrocher précocement la description phonétique aux modèles phonologiques car ils identifient des objets qui semblent plus faciles à mettre en relation avec une description symbolique. Ce rapprochement est alors payé au prix d'un *a priori* plus fort au niveau phonétique.

Le troisième point qui me semble important est étroitement lié au précédent : il s'agit de déceler si les différents modèles attachent une importance à la forme globale des courbes de f_0 (ce qui amène à des notions de contour ou de mouvement) ou s'ils préconisent une vue locale.

Plan du chapitre. La sélection des modèles présentés ici n'est bien évidemment pas exhaustive. Le but est d'*illustrer* le large éventail de choix possibles avec leur *a priori* plus ou moins fort. Nous allons commencer l'exposé avec les modèles qui ne font pas de référence à la structure segmentale des énoncés. Nous verrons ainsi des stylisations par points cible (cf. 1.1.1) et des modèles de commande (cf. 1.1.2) qui mettent en oeuvre la superposition de contours et qui se distinguent ainsi très nettement des précédents de par leur approche globale. Dans le paragraphe 1.1.3, nous rencontrerons des modèles relativement simples qui fournissent une discrétisation de f_0 par rapport à la structure segmentale de l'énoncé. Ensuite, nous discuterons dans le paragraphe 1.1.4 l'approche IPO qui met en avant la qualité perceptive de la stylisation phonétique, avant de venir dans 1.1.5 à des modèles qui introduisent davantage d'*a priori* sur les portions du signal qui sont dignes d'intérêt : les événements.

1.1.1 Stylisation par points-clés

L'idée de base des modèles présentés dans ce paragraphe est qu'un locuteur – en produisant une courbe de f_0 – vise à atteindre certains points cibles, les transitions entre ceux-ci n'étant que des « maux nécessaires » qui n'ont pas de signification et qui, par conséquence, peuvent être omises de la description phonétique et remplies par interpolation.

Un exemple classique est celui décrit par **Pierrehumbert [1980]**. Il a la particularité d'attacher une grande importance aux maxima de la fréquence fondamentale ; les creux de f_0 ne sont pas forcément des cibles motivées, mais ils peuvent être assimilés à la transition entre deux cibles hautes.

Un autre modèle de ce type est MOMEL, proposé par **Hirst et Espesser [1993]**; **Hirst et al. [2000]**. Il se présente sous forme d'un algorithme dont le but est de paramétriser la composante macroprosodique d'une courbe de f_0 mesurée en une suite de points cibles interpolés par des « splines » de degré deux. À la différence du modèle précédent, les points cibles ne sont pas uniquement les maxima de f_0 , mais plus généralement des points à tangente horizontale (maxima, minima et éventuellement points d'inflexion).

Nous allons par la suite détailler et commenter MOMEL. Ces points cible sont déterminés en trois étapes :

1. Les auteurs considèrent la courbe de la fréquence fondamentale comme étant une superposition (linéaire ou logarithmique) d'une composante macroprosodique et d'une composante microprosodique. Cette dernière comprend les influences du contenu segmental et est essentiellement négative (baisse de f_0 pour les occlusives voisées par exemple). C'est seulement à la frontière entre voisement et non-voisement que l'on observe une composante microprosodique positive. La première étape de MOMEL vise donc à éliminer ces parties de la courbe de f_0 .

2. La seconde étape tente de fournir des points cibles potentiels : dans une fenêtre

d'analyse centrée sur la valeur courante, la courbe est approximée par une fonction quadratique ; des valeurs de f_0 déviant trop vers le bas de l'approximation sont éliminées (influence présumée de la microprosodie). La régression quadratique est renouvelée tant qu'il y a des valeurs à éliminer. L'extremum de la parabole d'approximation est le point cible pour la trame en question s'il se situe à l'intérieur de la fenêtre d'analyse.

3. La suite des points cibles est, dans une troisième étape, segmentée en sorte que les moyennes des cibles dans des segments² adjacents soient le plus espacées possible. La dernière étape consiste à éliminer les cibles déviant plus que d'un écart-type de la moyenne pour le segment, et à retenir comme point cible représentatif du segment la moyenne des cibles y restant.

Selon les chiffres donnés pour plusieurs corpora, l'algorithme MOMEL fournit une stylisation avec en moyenne environ 3 points cibles par seconde.

Les auteurs affichent une distance moyenne ($|1 - hz'/hz|$) entre les courbes reconstituées et originales comprise entre 4,7% et 6,3%. Malheureusement, ceci n'est pas comparable aux erreurs citées des autres modèles phonétiques.

1.1.2 Modèles de commande

Après les modèles « points cibles » qui adoptent une vue très locale sur les objets d'intérêt de l'évolution de f_0 dans le temps, nous abordons maintenant les modèles de commande qui s'intéressent à des contours *étendus* qui se *superposent*.

Fujisaki et Hirose [1984] proposent – initialement pour le japonais – un modèle inspiré par Öhman [1967] en décrivant les contours prosodiques comme étant la superposition additive de deux composantes : groupes³ et accents mélodiques (cf. figure 1.3). Sa formulation mathématique permet de l'utiliser de manière transparente en analyse *et* en génération. Elle se lit comme suit :

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^I Ap_i Gp(t - T_{oi}) + \sum_{j=1}^J Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})]$$

avec

$$Gp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{pour } t \geq 0 \\ 0, & \text{pour } t < 0 \end{cases}$$

$$Ga(t) = \begin{cases} \min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{pour } t \geq 0 \\ 0, & \text{pour } t < 0 \end{cases}$$

F_b sert à adapter le contour au registre du locuteur. Les termes avec Ap_i correspondent aux groupes intonatifs. Les courbes atteignent rapidement un maximum

²à ne pas confondre avec les segments d'ordre phonémique.

³anglais : phrase

après T_{oi} et descendent ensuite lentement. Elles peuvent être assimilées à des lignes de déclinaison. L'insertion d'une composante de groupe revient à une réinitialisation de la ligne de déclinaison. Les termes avec Aa_j décrivent les accents : une montée, un plateau éventuel (selon la différence entre T_{1j} et T_{2j}) et une chute rapide. Leur étendue est beaucoup plus courte que celui des composantes de groupe.

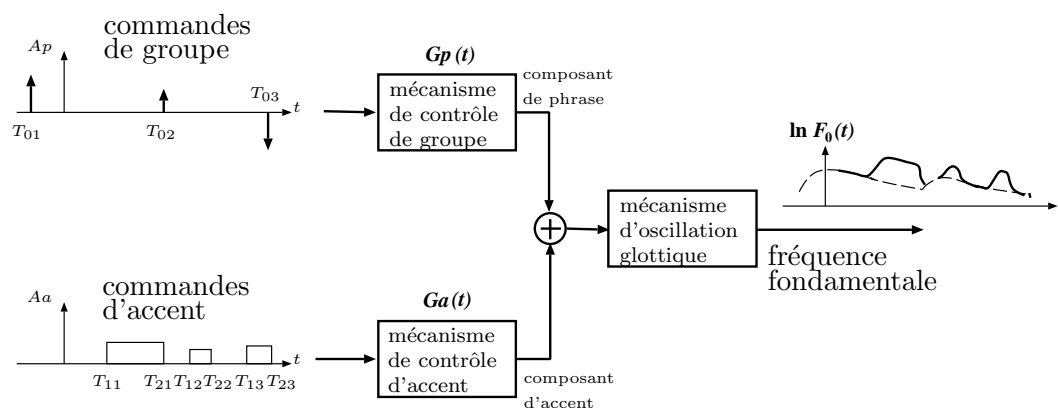


FIG. 1.3: Schéma du modèle de Fujisaki.

Ancrage syllabique. Bien que les paramètres T_i peuvent être exprimés par rapport à la structure segmentale (p.e. au début de la syllabe), la description des contours reste fondamentalement ancrée sur un axe temporel : ce sera à la charge du modèle de génération des paramètres d'assurer l'alignement des commandes par rapport aux segments.

Interprétation physiologique. On peut interpréter les deux composantes comme la réponse de systèmes linéaires de second degré à des entrées en forme d'impulsion (groupes) ou de fonction rectangulaire (accents). Fujisaki associe ces systèmes aux systèmes de contrôle du larynx (pression subglottique et tension des cordes vocales).

1.1.2.1 Inversion

À ce stade, le modèle peut être considéré comme purement phonétique : en principe, chaque courbe de f_0 peut être décomposée avec une précision arbitraire et l'ensemble des paramètres équivaut à la courbe de la fréquence fondamentale. Or, la décomposition n'est pas unique mais dépend de la méthode d'inversion. Les adeptes du modèle utilisent des méthodes itératives afin de déterminer les paramètres optimaux. Selon les contraintes imposées pendant ce processus, les paramétrisations d'un même énoncé seront différentes. On peut distinguer deux approches différentes :

- par le haut : la procédure d'inversion est guidée *phonologiquement* ;
- par le bas : elle l'est *phonétiquement*.

Avec a priori phonologique. La méthode de Mixdorff [1998] pour l'allemand, le modèle MFGI⁴, consiste à lier les commandes d'accents avec des intonèmes d'un modèle phonologique inspiré par Isačenco et Schädlich [1964]. Ainsi, il évite que la stylisation suive le contour observé à outrance et au détriment de l'interprétabilité des commandes. Une analyse de contours de ce genre est en général supervisée par un expert afin de corriger des problèmes au cas où l'approximation d'un énoncé donné nécessite plus ou moins de commandes que prévu par le modèle phonologique.

Le pari du modèle de Fujisaki est que ses paramètres peuvent être prédits à partir d'une description phonologique, il me semble donc justifiable que l'on guide le modèle phonétique par le modèle phonologique afin d'assurer une plus grande cohérence entre les deux. Or, le souhait de pouvoir analyser des grands corpora d'une manière automatique favorise des approches strictement phonétiques afin de réduire l'intervention de l'expert.

Sans a priori phonologique. La proposition de Fujisaki et Narusawa [2002] rentre dans ce cadre. La difficulté principale est de trouver une première estimation des paramètres qui peut être affinée itérativement par la suite. Les auteurs proposent une série de prétraitements de la courbe mesurée de f_0 afin de pouvoir estimer nombre, position et amplitude des composants nécessaires à sa paramétrisation :

- détection et élimination des erreurs grossières dans la mesure de f_0 ;
- correction des effets de microprosodie ;
- interpolation dans les parties non-voisées du signal ;
- l'approximation par des « splines » cubiques.

Ensuite, les débuts et fins des composants d'accent (T_{1j} , T_{2j}) sont simplement identifiés avec les points d'inflexion de la courbe prétraitée. La première approximation de A_{aj} est fixée par l'amplitude absolue moyenne entre T_{1j} et T_{2j} . La contribution des accents peut ainsi être estimée et soustraite de la courbe de f_0 afin d'en extraire les composants de groupe intonatif en procédant de gauche à droite. Puis, le nombre de commandes étant alors fixé, il devient possible d'affiner leurs paramètres par optimisation.

Mixdorff [2000] se tourne vers une procédure très semblable. Lissage et correction de microprosodie sont confiés à MOMEL (cf. page 24). Le contour est ensuite filtré en passe haut ($> 0,5$ Hz) afin d'en extraire la partie due aux accents (CHF). Le résidu (CBF) correspond à la composante « groupes ». Puis, CHF et CBF sont d'abord séparément approximés avec des paramètres Fujisaki avant d'être superposés et optimisés conjointement. Dans une dernière étape, les paramètres sont optimisés en tenant compte de la fiabilité des mesures de f_0 dans les différentes parties de l'énoncé

⁴Mixdorff-Fujisaki German Intonation model

(priorité aux segments voisés et de forte énergie).

1.1.3 Stylisation ancrée sur la structure segmentale

L'évolution quasi-continue de f_0 en fonction du temps étant connue, beaucoup d'auteurs considèrent qu'il est plus pertinent de l'exprimer en fonction de la structure (supra-)segmentale de l'énoncé. Il reste alors des choix à prendre sur les unités d'ancrage (phonèmes, voyelles, syllabes, mots) et sur le nombre de paramètres par unité.

La méthode utilisée par **Black et Hunt [1996]** illustre bien cette approche. Ils procèdent en deux étapes : d'abord, est calculée pour chaque segment voisé la valeur moyenne de f_0 en Hz. Dans les parties non-voisées, les contours sont interpolés. Cette approximation comporte par rapport aux contours « bruts » d'une voix féminine une erreur de 9,9 Hz (RMSE) avec une corrélation de 0,9. Ensuite, les auteurs extraient trois valeurs par syllabe : au début, au milieu de la voyelle et à la fin de la syllabe. C'est l'ensemble de ces valeurs qui constitue leur modèle phonétique et qu'ils tentent de prédire à partir de leur modèle phonologique (cf. page 46). On peut regretter que les auteurs n'indiquent pas l'erreur des contours reconstitués par interpolation à partir des trois valeurs par syllabe.

Ainsi le choix de trois valeurs par syllabe⁵ est assez courant (cf. p.e. [Emerard et al., 1992] : trois valeurs extraites de la voyelle). En s'appuyant sur des tests de perception de préférence et de type MOS⁶, **de Tournemire [1994]** propose à ce sujet une comparaison systématique de stylisations avec un nombre différent de valeurs par syllabe/mot qui peut se résumer comme suit : l'évolution de f_0 à l'intérieur des consonnes ne nécessite pas de modélisation – une interpolation linéaire entre les voyelles étant suffisante. L'utilisation de trois points par voyelle est jugée légèrement avantageuse par rapport à un seul point. Si l'on veut réduire le nombre de paramètres, la valeur de f_0 au milieu des voyelles semble la plus pertinente. Une stylisation avec trois points par *mot* serait selon l'auteur acceptable si l'on fait abstraction des phénomènes d'emphase.

Bien qu'il soit légèrement plus complexe, le modèle phonétique de **Traber [1992]** rentre dans la même catégorie. Nous le présentons rapidement ici, surtout à cause de l'intérêt que nous portons à son modèle de génération de la prosodie (cf. page 40). Les courbes de f_0 sont représentées par 8 valeurs de f_0 par syllabe. Comme esquissé dans la figure 1.4, chaque syllabe est d'abord divisée en deux parties par le point d'énergie maximale de la voyelle. Ces deux parties sont ensuite divisées en deux intervalles de même durée à l'intérieur desquelles la courbe de f_0 (apparemment exprimée en Hz) est approximée par une droite. Les intersections des quatre droites avec les frontières de « quarts » de syllabe fournissent les 8 valeurs.

⁵Selon la vitesse d'élocution (typiquement 4 à 6 syllabes par seconde), ceci correspond à environ 15 valeurs par seconde – à comparer avec les 3 s^{-1} extraites en moyenne par MOMEL.

⁶Mean Opinion Square

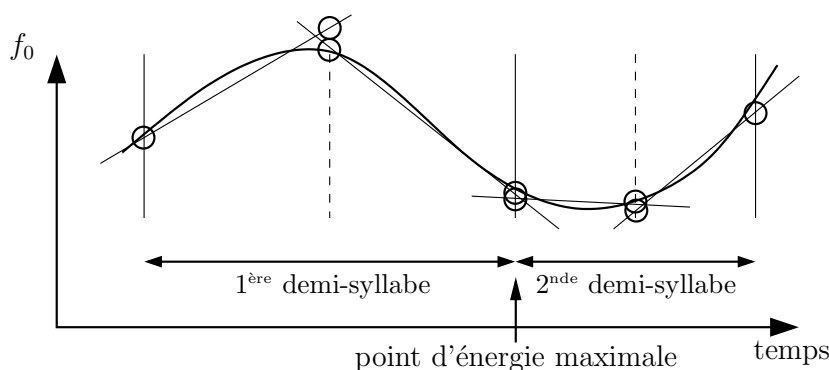


FIG. 1.4: **Modèle phonétique de Traber.** Afin d'illustrer l'extraction selon Traber de 8 valeurs de f_0 par syllabe, nous avons représenté ici l'évolution de f_0 au cours d'une seule syllabe. Les valeurs extraites sont les intersections des droites de régression avec les frontières des « quarts » de la syllabe.

Notons finalement la proposition intéressante de **Monaghan [1992]** qui vise à corriger les erreurs de stylisation attribuables à des propriétés intrinsèques des phonèmes en réintroduisant au moment de la synthèse par concaténation une composante micro-prosodique propre aux diphones. Nous y reviendrons plus loin (cf. page 71).

1.1.4 Pertinence perceptive de la description phonétique

Les modèles discutés jusqu'ici peuvent être qualifiés comme acoustiques : les approximations de f_0 proposées sont guidées par des critères numériques.⁷ Nous avons vu comment l'on peut utiliser des connaissances *a priori* afin d'éliminer l'influence de la microprosodie des courbes de f_0 dans l'espoir d'obtenir une paramétrisation plus pertinente, c.-à.-d. dans ce cas, indépendante du contenu segmental de l'énoncé. L'approche qui est présentée dans ce paragraphe recherche la pertinence sous un autre angle d'attaque : l'auditeur – et ajoute donc une dimension *psycho*-acoustique. Elle a été développée au sein de l'IPO⁸ : [de Pijper, 1983; Hart et al., 1990] etc. Nous allons suivre ce modèle à travers son adaptation pour le français par **Beaugendre [1994]**.

L'approche IPO – partie phonétique. La méthode IPO consiste à remplacer, dans un premier temps, les mesures quasi-continues de f_0 par une suite de segments de droite, dénommés *mouvements mélodiques*. Le but de cette étape est de simplifier les courbes tout en préservant une *égalité perceptive*, ce qui justifie l'appellation *copies-exactes* des

⁷À l'exception des modèles du type Fujisaki, lorsque l'emplacement des commandes est contraint phonologiquement.

⁸Instituut voor Perceptie Onderzoek (Institut pour la recherche perceptive)

stylisations. Il s'en suit une étape de classification fonctionnelle de ces mouvements que nous détaillerons dans le prochain chapitre (cf. page 35).

Restons, pour le moment, à l'aspect phonétique du modèle : qu'est-ce qu'il faut entendre par « égalité perceptive » ? Dans un premier temps, c'est l'expérimentateur qui en décide en comparant la resynthèse du stimulus avec les valeurs de f_0 de l'original et de la stylisation. Le jugement doit ensuite être validé par un test de perception : Beau-gendre choisit 28 phrases d'un corpus de 59 phrases stylisées. Ces phrases sont « courtes de manière à faciliter la mémorisation de l'intégralité de l'intonation par les sujets » [p.64]. Le critère de sélection montre un problème lié à la notion d'« égalité perceptive » : il est potentiellement très difficile d'en juger. Néanmoins, les résultats de Beau-gendre montrent que ses copies-exactes sont perçues comme presque aussi « égales » à l'original que l'original même. Il utilise quatre types de paires de stimuli : (A) deux fois l'original resynthétisé, (B) original resynthétisé et la copie exacte, (C) la copie exacte et une 1^{ère} version alternative et (D) la copie exacte et une 2^{ème} version alternative, où les versions alternatives sont des modifications plus ou moins importantes des copies exactes. Les stimuli du type A – donc strictement identiques – sont jugés égaux à 89,9% par les 17 sujets. Ceux de type B à 85,1% et les autres à 48,9% et 16,3% respectivement. L'auteur conclut donc que les copies-exactes sont perceptivement aussi « égales » aux originaux que les originaux même. . .

Discussion. Cette approche a le mérite d'intégrer dans le processus de stylisation ce que tous les modèles phonétiques tentent de faire : rendre la courbe de f_0 maniable tout en préservant l'information qu'elle contient. La question qui se pose est si l'on peut se contenter d'appliquer une méthode globale en utilisant par exemple trois valeurs de f_0 par syllabe ou s'il vaut mieux utiliser une méthode qui s'adapte à chaque énoncé mais qui nécessite un difficile jugement humain⁹ et qui n'est donc pas déterministe (la stylisation dépend de l'expérimentateur).

La partie phonétique du modèle IPO n'est pas forcément ancrée sur la structure segmentale : les mouvements mélodiques peuvent être spécifiés par rapport à l'axe temporel. Cependant, la classification des mouvements qui suivra dans la partie phonologique (cf. page 35) aura recours au repère syllabique.

Notons pour conclure la description du modèle phonétique « IPO », que les éléments de base de sa stylisation sont des segments de droite et non pas des points. Au niveau de la description phonétique la différence importe peu, puisque une droite équivaut évidemment à deux points, mais elle conditionne le modèle *phonologique* comme on va le voir dans le chapitre 1.2.

⁹Il est certainement envisageable d'automatiser l'obtention des copies-exactes, mais sans test perceptif, on pourrait tout au plus prétendre fournir une stylisation qui soit en général perceptivement très proche des paramètres originaux, et qui ne mériterait plus le titre de « copie-exacte ».

1.1.5 Modèles basés sur des événements

Dans le paragraphe 1.1.1 nous avons vu avec MOMEL un modèle qui réduit la description phonétique de f_0 à une suite de points cibles – les transitions entre eux étant jugées secondaires. Cependant, l'emplacement des points cibles était sans *a priori* et guidé uniquement par le souci d'une bonne approximation économique de f_0 qui lisse les effets de la microprosodie. Les modèles dont nous traitons dans ce paragraphe vont plus loin dans leurs présupposés : les parties du signal dignes d'intérêt sont identifiées *avant* leur paramétrisation.

Taylor [1998, 2000] propose le « Tilt Intonation Model » (TIM) : Ce modèle décrit la réalisation phonétique d'un énoncé comme une séquence d'événements : les *accents* et *tons de frontière*. Seule l'évolution de la fréquence fondamentale au cours des événements est modélisée, le reste est interpolé. Le modèle suppose que tous les événements sont réalisés par une montée de f_0 suivie d'une descente (l'étendue de chacune de ces parties pouvant être nulle). Chaque événement est caractérisé par des paramètres continus : l'amplitude, la durée, la position temporelle¹⁰, la valeur de f_0 au milieu de l'événement et le « tilt ». Ce dernier décrit la proportion entre la partie montante et la partie descendante de l'événement.

Le pari du TIM est :

- que la stylisation du signal obtenue comprenne tout ce qui est prosodiquement important ;
- qu'il soit – en vue de la synthèse à partir de texte – plus facile de prédire les paramètres des événements qu'un contour non-structuré ; et
- que cette prédiction soit plus aisée dans l'espace « tilt ».

La principale différence par rapport aux modèles discutés plus haut est donc la notion d'événement. Par exemple, chez Black et Hunt [1996], l'interpolation ne porte que sur des parties non-voisées du signal où f_0 n'est pas définie, et n'est donc qu'un moyen d'abstraire du contenu segmental afin d'obtenir des contours de f_0 complets tout au long de l'énoncé.

Qu'est-ce qu'un événement ? Sa définition dépasse largement une caractérisation en termes de f_0 :

« Pitch accents [...] are f_0 excursions associated with syllables which are used by the speaker to give some degree of emphasis to a particular word or syllable. [...] boundary tones are rising f_0 excursions which occur at the edges of intonational phrases and as well as giving the hearer a cue as to the end of the phrase can also signal effects such as continuation and questioning. » [Taylor, 2000, p. 4]

Cette définition relève clairement de la phonologie (de surface) et peut faire hésiter à classer le TIM comme étant un modèle principalement *phonétique*, bien qu'il « can be regarded as a *phonetic* model » selon Taylor [2000, p. 4]. Evidemment, la classification

¹⁰par rapport au début de l'énoncé ou par rapport à la syllabe associée

des différentes parties d'un modèle n'est pas une fin en soi, mais il faut se rendre compte qu'obtenir la paramétrisation selon le TIM à partir du signal, demande plus d'efforts que les modèles phonétiques indépendants de la notion d'événements, parce que les accents et les tons de frontière sont définis par rapport à leur fonction (« emphasis », « continuation », « questioning ») – et ceci nécessite le recours au jugement d'experts.

Taylor propose de modéliser l'expertise de l'identification des événements à l'aide de HMM¹¹ s'appuyant sur des entrées de f_0 et de l'énergie du signal. Il obtient des taux d'identification entre 47% et 70% (à comparer avec 74% d'accord entre experts). Taylor remarque que

« ... events which are not distinct acoustically are detected with much lower accuracy than those which have prominent acoustic features. » [Taylor, 2000, p. 11]

Ceci reflète assez explicitement que le TIM va au-delà de la phonétique. Notons que Taylor inclut l'analyse des événements sans prééminence acoustique à des fins de test seulement – car selon lui, ils pourraient être *a posteriori* ignorés. Il se pose alors la question si le recours aux fonctions des accents est nécessaire et/ou bénéfique.

Les erreurs de stylisation données par Taylor sont obtenues par analyse d'un corpus multi-locuteur : 15,25 Hz (erreur RMS¹²) avec une corrélation de 0,644 par rapport aux contours bruts et 7,51 Hz (0,833) par rapport aux contours lissés. Des erreurs pour d'autres corpora sont données par Dusterhoff et al. [1999]. Elles varient entre 3,9 Hz (voix d'homme : $f_0 = 126,9 \pm 27,9$ Hz) et 14,5 Hz (voix de femme : $f_0 = 163,5 \pm 42,2$ Hz) par rapport aux contours lissés. Si je cite ces chiffres ici, c'est surtout pour montrer à quel point il est difficile de quantifier le succès d'une méthode. Les auteurs donnent les explications suivantes à l'égard des difficultés de stylisation rencontrées pour un des corpora :

« This suggests that perhaps the labels are not of as high a quality, or perhaps that there is more movement in the non-event (connection) portions of the original f_0 ... » [Dusterhoff et al., 1999]

Ceci souligne les deux problèmes principaux du TIM :

- sa dépendance à la détection cohérente des événements ; et
- la distinction entre événement et non-événement, car on peut se poser la question si le mouvement de f_0 dans le non-événement est vraiment sémantiquement non-significatif.

Ne perdons pas de vue que le rôle du modèle phonétique est d'aboutir à une description pertinente et *maniable* par les modules supérieurs. Le modèle de **Möhler et Conkie [1998]** constitue un bon exemple de comment « digérer » la substance prosodique déjà au niveau phonétique. Tout comme Taylor [2000], ce modèle présuppose que l'on connaît l'emplacement des accents. Toute la stylisation du contour observé se fait donc par rapport aux accents : dans une fenêtre de trois syllabes centrée sur la

¹¹Hidden Markov Model

¹²Root Mean Square

syllabe portant l'accent, la courbe de f_0 mesurée en Hz est approximée par une somme de deux « sigmoïdes » (l'une montante et l'autre descendante). Avec ses six paramètres libres, cette fonction d'approximation représente une courbe en forme de marche avec une transition plus ou moins douce et un pic éventuel au seuil. Tout ce qui se passe en dehors de ces fenêtres est ignoré et rempli par interpolation. Les auteurs obtiennent ainsi des erreurs RMS d'approximation entre 10.7 Hz pour un corpus d'article de journal et 13,6 Hz pour un corpus contenant des messages courts. Apparemment, la voix enregistrée est celle d'une femme, mais les auteurs ne donnent aucune indication sur le registre de la voix, ce qui rend les données difficilement interprétables. Il semble cependant raisonnable que l'erreur soit plus élevée que celle obtenue par Black et Hunt [1996] (cf. page 28) car ces derniers n'imposent aucune contrainte aux contours (mis à part leur discrétisation avec trois valeurs par syllabe).

Bien que l'espace des courbes de f_0 soit ainsi fortement restreint, il reste néanmoins continu et la différence entre ce modèle et le TIM se résume à une paramétrisation différente. Si j'ai décidé de présenter brièvement ce modèle, c'est parce que Möhler et Conkie entreprennent dans une seconde étape de discrétiser l'espace des courbes possibles en utilisant une quantification vectorielle. Ainsi, chaque réalisation d'accent peut être associée à un prototype. Le nombre de prototypes varie entre 4 et 32 – donnant selon les auteurs les meilleurs résultats pour 8 ou 16. Evidemment, l'erreur de stylisation augmente, p.e. pour une quantification avec 8 prototypes à 27,9 Hz et 35,6 Hz pour les deux corpora respectivement.

Nous nous trouvons donc ici en face d'un modèle phonétique qui s'étend dans le domaine du symbolique. Les mesures de f_0 sont transformées en une série de prototypes d'accent qui émergent d'une analyse statistique d'un corpus. Le modèle reste néanmoins *phonétique* puisqu'il ne fait aucune référence directe au rôle que joue tel ou tel prototype. La seule entrée *phonologique* est le positionnement des accents qui pourrait probablement être remplacée par une détection de pics dans la courbe de f_0 mesurée.

1.1.6 Résumé

Ce chapitre devrait avoir donné un aperçu du large spectre des modèles phonétiques : allant de la simple discrétisation par un nombre fixe de points par syllabe jusqu'à l'encodage complexe par contours superposés. En dépit de toutes les différences, ils partagent tous un but commun : paramétriser la substance prosodique et ainsi la rendre accessible au traitement de plus haut niveau. Il n'a pas été question de juger lequel de ces modèles y parvient le mieux, mais plutôt de montrer les différents angles d'attaque possibles.

1.2 Modèles phonologiques de f_0

Selon le Petit Robert la phonologie est la « science qui étudie la fonction des sons dans les langues naturelles et dégage ainsi les phonèmes ». Le mot clé ici est « fonction ». Il ne s'agit plus de décrire un énoncé dans son unicité, mais d'identifier ce qu'il le caractérise par rapport à sa fonction communicative. Les éléments dégagés ne sont – si l'on parle de prosodie – pas les phonèmes, mais des objets tels que des tons, des intonèmes ou des clichés.

Plan du chapitre. Habituellement, on distingue phonologie de surface et phonologie profonde. La première exprime l'instanciation de la dernière dans la substance phonétique. Il s'en suit que la phonologie de surface est plus proche de la phonétique : un moyen de dégager une telle description est de procéder à une classification des observations décrites en terme d'un modèle phonétique. Nous allons voir un exemple de cette approche dans le paragraphe 1.2.1 avec la deuxième partie du modèle IPO. Le système de transcription prosodique INTSINT peut également être vu sous cet angle, mais nous le classons ici, de même que ToBI, sous les systèmes de transcription tonale (cf. paragraphe 1.2.2).

La phonologie profonde se présente comme couche d'abstraction qui peut être établie à partir de la structure syntaxique et d'autres sources d'information de haut niveau, et qui tient compte des contraintes d'oralisation (p.e. l'eurythmie). Elle identifie souvent plusieurs niveaux hiérarchiques (pieds, groupes d'accent, groupes intonatifs etc.) et doit permettre de déduire une représentation de surface linéaire. L'appréciation des différentes théories proposées dans la littérature nécessiterait un examen détaillé des structures et règles mises en oeuvre, ce qui dépasse largement le cadre de ce rapide tour d'horizon des modèles de génération prosodique, lesquels se satisfont généralement de descriptions plus « *ad hoc* ».

Revenons donc au rôle plus profane que nous attribuons ici aux modèles phonologiques : être l'interface entre les informations issues du domaine symbolique (lexique, syntaxe etc.) et la description phonétique. Plus haut, j'avais pris la perspective ascendante en parlant de « classification des observations ». Cependant, nous allons rassembler sous « phonologie » également les descriptions descendantes qui ne sont guère plus qu'un ré-encodage des informations disponibles à partir du texte en vue de leur mise en correspondance avec les réalisations phonétiques. Ainsi, nous verrons dans le paragraphe 1.2.3 avec les modèles de Traber et de Emerard et al. deux exemples très simples (mais typiques) de ce type de modèle phonologique.

1.2.1 Classification fonctionnelle

Nous commençons ici la petite sélection de modèles phonologiques avec ceux qui offrent une prolongation classifiante des modèles phonétiques avec lesquels ils vont

de pair. La différence avec la classification *phonétique* des formes d'accent opérée par Möhler et Conkie (cf. page 32) est mince. Elle consiste dans le statut des classes : soit elles sont descriptives, soit elles sont censées être porteuses de signification.

L'approche IPO – partie phonologique. La seconde étape de l'approche IPO est la *standardisation*. Il s'agit là de représenter tous les mouvements possibles par un nombre restreint de classes. Le contour obtenu pour un énoncé donné à partir des représentants des classes est appelé *copie-conforme*. L'idée est qu'il existerait une *équivalence perceptive* entre la copie-conforme et l'original. Nombre et caractérisation des classes dépendent de la langue étudiée et – ce qui peut être problématique – du corpus utilisé ainsi que de celui qui procède à la classification. On a déjà rencontré l'idée d'identifier dans la description phonétique des classes d'objets avec le modèle de Möhler et Conkie [1998] (cf. page 32). Cependant, il y a deux différences importantes :

- les objets identifiés par Möhler et Conkie sont localisés aux « événements » tandis que les segments de droite de l'approche IPO concernent l'énoncé dans sa totalité ;
- la quantification vectorielle que Möhler et Conkie proposent est purement phonétique, tandis que le modèle IPO fait explicitement référence à la fonction des classes identifiées.

C'est ce deuxième point qui justifie que la standardisation de l'approche IPO prend sa place dans les modèles *phonologiques*.

Les distributions des paramètres qui caractérisent les 500 mouvements mélodiques recensés dans le corpus analysé par **Beaugendre [1994]** ne permettent pas la mise en évidence de classes « naturelles ». Mais selon Beaugendre « la mise en correspondance des courbes stylisées et de la segmentation phonémique montre clairement qu'il existe une corrélation entre le timing des mouvements et la structure segmentale. » [p.81] La nature de cette « corrélation » reste cependant floue et – mis à part le fait que les critères de classification sont contraints par des considérations de production et de perception de la parole – la définition des classes est issue d'un jugement d'expert. L'analyse du corpus aboutit ainsi à l'identification de neuf classes de mouvements mélodiques.

Tout comme les copies-exactes, les copies-conformes sont soumises à l'évaluation perceptive. Neuf sujets jugent sur une échelle de 1 à 5 le naturel de 15 phrases en quatre versions différentes : original resynthétisé, copie-conforme et de deux versions alternatives. Le meilleur résultat est obtenu – comme attendu – par l'original, suivi d'assez près par la copie-conforme. Les autres versions sont nettement moins bien notées. Ceci permet certes de constater que la stylisation à l'aide des 9 classes fournit une prosodie assez naturelle, mais de là à conclure qu'il s'agit de *copies-conformes* avec une *équivalence* perceptive ne me semble pas évident.

L'approche IPO a une particularité intéressante : les objets de base sont des segments de droite – donc des mouvements. J'avais remarqué qu'au niveau phonétique ceci n'est pas particulièrement important, puisque une droite n'est rien d'autre que deux points. Or, dès que l'on procède à la classification phonologique, ce choix d'objet

élémentaire prend de l'importance car on y attache une interprétation/signification. Le choix du modèle IPO peut être paraphrasé par « la signification prosodique est portée par des mouvements caractérisés par leurs pentes et leurs ambitus ».

1.2.2 Systèmes de transcription tonale

Nous allons nous tourner maintenant vers deux modèles qui attachent, de manière plus prononcée que le modèle précédent, la signification aux points cibles : INTSINT¹³ et ToBI¹⁴. Ces modèles sont, en effet, sous une forme ou une autre très utilisés dans les chaînes de génération prosodique.

INTSINT. Proposé par **Hirst et Di Cristo [1998]**, il s'agit d'un système de transcription de la structure phonologique de surface. INTSINT repose sur une représentation phonétique d'une courbe de f_0 sous forme de points cibles comme elle peut être obtenue par exemple avec MOMEL (cf. page 24). Ces points cibles sont traduits en suite de tons. Les auteurs distinguent deux types de tons différents :

- les tons relatifs : H(igher), L(ower), S(ame), D(ownstep), U(pstep)
- les tons absolus : T(op), B(ottom), M(id)

Le deuxième élément de description fourni par INTSINT sont les marqueurs de frontière. Ils peuvent être utilisés pour délimiter les unités intonatives¹⁵. Les doubles crochets sont prévus pour des unités de plus haut niveau auxquelles s'associent des phénomènes comme la réinitialisation de la ligne de déclinaison ou des tons extrêmes.

A l'opposé de ToBI, INTSINT est conçu pour être applicable à toute langue sans qu'il soit nécessaire d'établir au préalable un inventaire de mouvements prosodiques spécifiques à la langue en question. Ceci se traduit par une forte cohérence avec le modèle *phonétique* sous-jacent : l'algorithme de codage automatique proposé par Hirst et al. [2000] consiste en une simple catégorisation des points cible obtenus par MOMEL en tons de INTSINT. Le pari du modèle est que la suite des tons associée à un énoncé encode l'information véhiculée par la prosodie à la fois simplement et complètement.

ToBI. Ce système de transcription de l'intonation a été proposé par **Silverman et al. [1992]**. Suivant les travaux de **Pierrehumbert [1980, 1981]** on distingue entre accents et tons de frontière, les deux pouvant être de type H(igh) ou L(ow). Les accents possibles sont : H*, !H*, L*, L+H*, L*+H ou H+!H* où « * » indique l'alignement avec la syllabe portant l'accent (stress) et « ! » signale une baisse de f_0 par rapport à l'accent précédent (downstep).

L'analyse suppose une structure hiérarchique : chaque groupe intonatif¹⁶ contient

¹³International Transcription System for INTonation

¹⁴Tone and Break Indices

¹⁵anglais: "intonation units"

¹⁶anglais: "intonational phrase"

au moins un groupe intermédiaire. Ces derniers sont caractérisés par un accent de phrase (noté H-, !H- ou L-) après le dernier accent. Les groupes intonatifs se terminent avec un ton de frontière H% ou L%.

Le deuxième élément de ToBI sont les indices de rupture¹⁷ : ils classifient la cohésion entre mots sur une échelle de 0 (cliticalisation) et 4 (entre groupes intonatifs).

Les deux éléments – tons et indices de rupture – ne sont cependant pas indépendants : l'indice 4 est obligatoirement lié à la présence d'un ton de frontière et l'indice 3 est lié à un accent de phrase le précédant.

De part sa construction, ToBI dépend de la langue à laquelle il est appliqué – une différence majeure avec un système tel qu'INTSINT (cf. page 36). Mais s'il a été initialement conçu pour l'anglais américain, il a été cependant adapté depuis à plusieurs autres langues¹⁸.

La mise en place d'un système de type ToBI présuppose la connaissance des schémas intonatifs présents dans la langue. La version esquissée plus haut, par exemple, implique que les énoncés de l'anglais américain sont analysables en groupes intonatifs et que tous ces groupes se terminent par un ton de frontière.

Discussion. Selon Wightman [2002] le rôle de ToBI est double :

« (1) It needed to capture the “meaning” of intonational events such as prominent, continuation, final *etc.* and (2) it had to describe the shape of the pitch-track. »

Le premier rôle est *phonologique* : décrire ce qui est distinctif ; le second est *phonétique*. L'utilisation de ToBI et des variations qui en ont été proposées montrent une divergence selon ces deux axes : premièrement, affiner la richesse de la description phonétique comme le font Grice et al. [à paraître] pour l'allemand ce qui

« suggests that the system is rather designed for offering a close phonetic transcription of observed f_0 contours than a phonological description as claimed. »
[Mixdorff, 2002b]

Le deuxième axe consiste à grouper plusieurs types d'accents en une seule classe. Ce sont surtout les auteurs qui s'intéressent à la synthèse à partir de texte qui favorisent de telles simplifications, car les étiquettes complètes sont très difficilement prédictibles et c'est l'étiquetage des distinctions fines qui est le moins fiable (cf. [Pitrelli et al., 1994]).

1.2.3 Descriptions phonologiques « ad hoc »

Nous venons maintenant à deux exemples d'approches alternatives de l'interface phonologique. Ils ont été choisis d'une part parce qu'ils me semblent typiques et d'autre

¹⁷anglais: “break indices”

¹⁸cf. <http://www.ling.ohio-state.edu/~tobi/#languages>

part afin de compléter les bases nécessaires à la compréhension des modèles correspondants, qui feront le lien entre phonologie et phonétique et que nous rencontrerons dans le chapitre suivant.

Indices de frontières et d'accents. Traber [1992] présuppose une description phonologique de surface : les phrases sont découpées en groupes intonatifs ; les frontières entre groupes peuvent être de trois types différents : entre phrases, avec ou sans pause. Les syllabes sont annotées par le type d'accent qu'elles portent (0 : pas d'accent, 1 : accent principal, 2 à 4 : accents secondaires). Traber considère que

« Accentuation and phrasing are two of the weak points in our system (as they are in many other systems), partly because syntactic information simply is not enough to derive a reasonable sentence stress pattern and phrasing, and partly because the the accentuation and phrasing rules are not elaborate enough. »
[Traber, 1992, p. 291]

La description phonologique de Traber pourrait être assimilée à un étiquetage de type ToBI réduit. La raison pour laquelle j'ai préféré l'exposer ici, est que ses indices semblent moins dépendants des réalisations phonétiques et plus guidées « par le haut ».

Marquage en fonction des catégories grammaticales. Emerard et al. [1992] constatent que pour le français

« a prosodic boundary can usually be determined from the pattern formed by the grammatical category of the word itself and that of its 2 or 3 left and right hand neighbors. » [p. 229]

Ils n'utilisent donc pas moins de 153 règles afin d'attribuer à chaque mot d'une phrase un marqueur mélodique entre m1 et m10. Les marqueurs m9 et m10 correspondent par exemple à des mots (fonctionnels resp. lexicaux) considérés comme prosodiquement irrelevantes – au moins tant que l'on ignore emphase et focus. Les autres marqueurs tiennent compte du type de frontière prosodique qui suit le mot. Cette interface phonologique semble un peu plus riche que celle de Traber. Nous allons voir (cf. page 45) que ceci est nécessaire car la prédiction de f_0 à partir de ces marqueurs se fait localement sans s'appuyer sur des « fenêtres contextuelles ».

Les auteurs présentent dans la suite de l'article une méthodologie qui permet d'affiner semi-automatiquement le marquage. Le but est de générer un ensemble de règles qui permette de choisir pour chaque mot un cliché de f_0 . Le nombre de clichés augmente considérablement par rapport aux 10 marqueurs mélodiques, et il ne convient plus de les considérer comme interface phonologique, mais plutôt phonétique. On se retrouve donc ici aussi dans le champ de tension entre phonologie et phonétique que nous avons déjà rencontré à plusieurs reprises.

1.2.4 Résumé

La présentation des modèles phonologiques dans ce chapitre était certes un peu réductrice – surtout en ce qui concerne l’omission des modèles de phonologie profonde qui s’associent souvent aux modèles de surface. Nous avons vu que ce qui constitue l’interface phonologique chez les différents auteurs est plus ou moins proche du monde phonétique. Il est évident que ceci aura des reflets importants dans les modèles qui opèrent la transition entre phonologie et phonétique – une affaire à suivre dans le prochain chapitre...

1.3 Entre phonétique et phonologie

Les deux chapitres précédents auront – espérons-le – permis de clarifier les deux pôles de la synthèse à partir de texte : les descriptions phonétique et phonologique. Il s'agit maintenant de compléter la chaîne : comment lier ces deux mondes souvent hostiles ? La conciliation sera plus ou moins facile selon l'effort investi dans les modèles phonétique et phonologique qui tentent de les rapprocher. Néanmoins, il ne faut pas perdre de vue que ces efforts de rapprochement ont été payés au prix d'introduire des *a priori* plus forts.

Plan du chapitre. L'ordre de la présentation suit la complexité des entrées phonologiques utilisées. Le schéma de la figure 1.5 permettra de suivre le fil (ou plutôt : filet) des exemples. Il sera parcouru du haut à droite vers le bas gauche. Nous allons donc commencer dans le paragraphe 1.3.1 avec le modèle de Traber : il emploie une description phonétique très proche du signal acoustique (8 valeurs de f_0 par syllabe) et un modèle phonologique simple qui n'incorpore que peu d'indices sur les réalisations phonétiques présumées. Il s'en suit la proposition de Jokisch et al. qui se base sur une entrée phonologique semblable, mais qui utilise un modèle phonétique de type Fujisaki dont les paramètres sont censés être plus faciles à mettre en correspondance avec des objets phonologiques. Puis, dans le paragraphe 1.3.4, nous rencontrerons le modèle d'Emerard et al. qui tente une association simple et compréhensive des deux mondes. Ensuite, nous venons à des modèles qui s'appuient d'une manière ou d'une autre sur ToBI. Ils se distinguent par les modèles phonétiques employés : la simple discrétisation de f_0 avec le modèle Black et Hunt (1.3.5), les prototypes phonétiques de Möhler et Conkie et la prédiction des paramètres TIM selon Dusterhoff et Black. Nous verrons d'un côté que la richesse de ToBI ne s'avère pas toujours nécessaire et que d'un autre côté, s'appuyer sur la connaissance de l'emplacement (et du type) des accents fait sentir le besoin de prédire ces informations. Ceci va finalement nous amener dans le paragraphe 1.3.7 à la proposition de Véronis et al. qui fait confiance à un modèle simpliste de correspondance entre les descriptions phonologique (INTSINT) et phonétique (MOMEL) et qui se concentre sur une prédiction de la première à partir d'une entrée très simple (catégories grammaticales des mots). Ainsi, ce modèle sort du cadre que nous avons posé au début de ce petit état de l'art (cf. la figure 1.1, p. 20), car son modèle de génération se place entre le texte et le modèle phonologique.

1.3.1 Le grand saut

Le premier modèle que nous allons examiner est celui de **Traber [1992]**. Par rapport aux modèles suivants il s'attaque à la tâche la plus ardue : prédire la réalisation phonétique directement à partir d'une description phonologique relativement pauvre. Il s'inscrit dans la ligne de plusieurs propositions qui tentent d'établir la correspondance

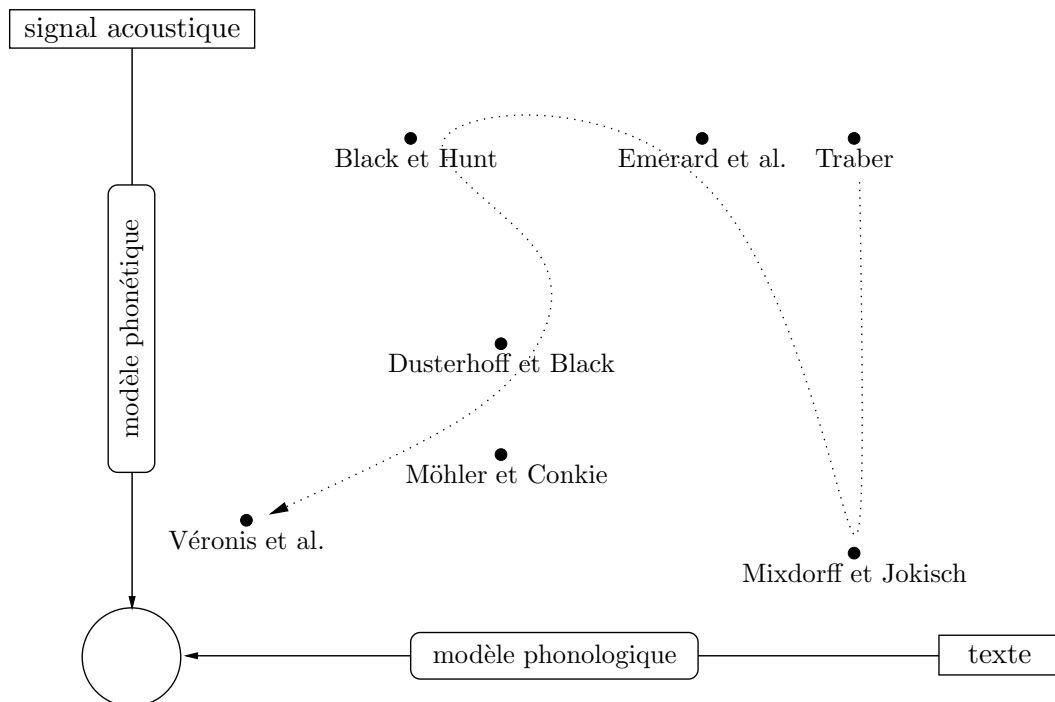


FIG. 1.5: **Modèles de génération prosodique présentés.** Plus un modèle est placé à gauche, plus la description phonologique qu'il utilise s'approche du monde phonétique. Plus il est placé en bas, plus les indices fournis par la description phonétique sont liés aux objets du modèle phonologique. La flèche pointillée indique simplement l'ordre de présentation dans ce chapitre.

entre phonologie et phonétique par un dispositif plus ou moins opaque d'apprentissage automatique (p.e. Ljolje et Fallside [1986] : prédiction de mouvements de f_0 par HMM ; Scordilis et Gowdy [1989] : utilisation de deux réseaux de neurones (RN) pour la prédiction de f_0 au niveau mot ; Sagisaka [1990] : prédiction au niveau syntagme mineur par RN). L'apport original de Traber est d'étendre l'utilisation d'un réseau de neurones à la prédiction de phrases complètes. Cette approche renonce à l'idée que l'on puisse décrire la transition entre description phonologique et phonétique de manière compréhensive. La transition est vue comme un « simple » problème de mise en correspondance de l'information : codée d'un côté phonologiquement et de l'autre côté par les paramètres phonétiques codant la courbe de f_0 . L'avantage qui réside dans l'utilisation d'un RN est qu'il peut gérer la nature très différente des espaces d'entrée et de sortie et qu'il peut apprendre des relations non-linéaires entre les deux. Traber montre que le RN est capable d'apprendre la génération de contours complexes.

Le facteur déterminant pour le succès du RN est le codage¹⁹ de l'information phonologique à son entrée. Traber utilise deux types d'information : l'une peut être qualifiée de suprasegmentale – et correspond donc à son modèle phonologique que nous avons déjà discuté dans le chapitre précédent :

- des marqueurs d'accents entre 0 et 4 ;
- des indices de frontière entre 0 et 2 ;
- un marqueur binaire reflétant si la syllabe appartient à une phrase en position finale ou non ;
- un marqueur binaire indiquant si la syllabe est avant ou après l'accent principal du groupe.

L'autre partie regroupe des informations phonotactiques et segmentales qui sont censées avoir une influence sur la microprosodie, celle-ci n'étant pas éliminée par le modèle phonétique :

- voyelle longue ou courte,
- voyelle avec f_0 intrinsèque haute ou basse et,
- deux marqueurs indiquant le (non-)voisement du contexte consonantique à gauche et à droite.

Traber utilise une fenêtre glissante de trois syllabes pour les données locales et de 13 syllabes pour les informations suprasegmentales autour de la syllabe en cours de traitement. Les marqueurs à l'intérieur de ces fenêtres sont fournis en entrée d'un RN à deux couches cachées. A ces entrées s'ajoutent 5 valeurs de feedback provenant de la deuxième couche cachée émises lors du traitement de la syllabe précédente. Le RN renvoie à sa sortie 8 valeurs de f_0 par syllabe. Le réseau est donc composé de 133 noeuds (95 pour l'entrée codée, 20+10 dans les couches cachées et 8 en sortie) et de 2180 connexions. Pendant la phase de l'apprentissage il s'agit d'ajuster les poids qui sont associés à ces connexions et éventuellement des biais associés aux noeuds – soit

¹⁹ « codage » ici dans un sens purement technique.

un total d'environ 2200 paramètres.

Il est important de noter l'utilisation des indices d'accent par Traber. Ces indices (de 0 à 4) sont déterminés par des règles et reposent sur une segmentation préalable de chaque énoncé en groupes d'accent. L'attribution des indices est liée à la réalisation prosodique présumée et facilite la tâche du RN, mais augmente le traitement symbolique en amont. Ces indices sont néanmoins une entrée phonologique plus pauvre que des étiquettes de type ToBI.

Métaphore spatiale du temps. Le relativement grand nombre d'entrées du RN est intimement lié à l'utilisation de fenêtres glissantes sur la chaîne des symboles d'entrée. Ce procédé sert à capturer le contexte de la syllabe en cours de traitement. La tâche du RN est d'associer à chaque contexte les paramètres prosodiques adéquats. Cette technique a l'avantage de ne nécessiter aucune décision *a priori* sur ce qui importe à la caractérisation pertinente des contextes. Le choix des indices et surtout de la taille des fenêtres est guidé par le succès du RN à fournir des contours satisfaisants.

Boîte noire. Non seulement les contextes pertinents sont modélisés implicitement, mais aussi côté phonétique, aucune contrainte n'est imposée aux contours – le RN est donc une véritable boîte noire réalisant la mise en correspondance de la description symbolique vers la réalisation prosodique. Traber lui-même regrette que l'on n'acquiert aucune connaissance explicite en utilisant des techniques d'apprentissage automatique et il souligne l'importance de l'approche « compréhensive » en parallèle :

« The greatest disadvantage of using concatenative or automatic learning approaches to transform symbolic data into physical parameters is that they do not yield (much) scientific knowledge about the problem at hand and usually lack generality. » [Traber, 1992, p. 303]

Corpus. On peut s'attendre à ce qu'une augmentation du nombre d'entrées améliore la capacité du RN à s'adapter aux paires entrées/sorties du corpus d'apprentissage. Cependant, il faut agrandir le corpus d'apprentissage si on veut éviter un sur-apprentissage : l'ensemble des sorties ne doit pas devenir une collection de cas uniques. Le corpus d'apprentissage contient des extraits de bulletins météorologiques, d'infos²⁰ et d'informations touristiques. Il comprend 100 phrases, ce qui devrait correspondre à environ²¹ 3500 syllabes. Compte tenu du fait que le RN doit prédire 8 valeurs de f_0 par syllabe, ce corpus semble tout juste suffisant pour l'apprentissage des 2200 paramètres.

²⁰anglais: "news"

²¹Traber ne donnent que le nombre moyen de syllabes par phrase (35,4) sur le corpus entier dont 86 phrases sont réservées à l'évaluation.

Résultats. Traber obtient des erreurs RMS de prédiction de 9,2 Hz pour un locuteur avec un registre de 75–150 Hz. Cette valeur est calculée sur un corpus de test de 86 phrases qui n'a pas été utilisé pour l'apprentissage du RN. Le modèle améliore la prédiction de f_0 par rapport à une autre approche proposée dans le même article qui opère par concaténation de contours stockés dans une base de données et qui fournit une prédiction avec une erreur de 9,6 Hz. Reste à savoir si cette différence est significative d'un point de vue perceptif.

1.3.2 Contraintes phonétiques fortes

Le modèle IGM²² proposé par Mixdorff et Jokisch [2001] ressemble en deux points à celui de Traber : le type d'entrées et l'utilisation d'un RN. La grande différence est que les auteurs n'essaient pas de prédire directement f_0 , mais les paramètres de contrôle du modèle MFGI (cf. page 27). Si Traber a besoin de larges fenêtres glissantes pour coder les contextes et permettre ainsi la génération de contours complexes, le nombre d'entrées se réduit ici à 20 – la forme des contours étant contrainte, il suffit de prédire amplitude, début et fin²³ des commandes d'accent et de groupe pour toutes les syllabes associées à un accent ou un début de groupe intonatif. Un problème essentiel est alors de savoir si une syllabe porte ou ne porte pas un accent. Une solution serait de s'appuyer sur un système de règles, une autre pourrait porter sur les amplitudes de commandes prédites en introduisant un seuil en dessous duquel la syllabe serait considérée comme inaccentuée.

Corpus et Résultats. L'entraînement du RN et les tests sont effectués avec le corpus prosodique de l'université de Stuttgart. Il contient 48 minutes de texte d'information radio lu par un locuteur masculin. L'entraînement se base sur environ un tiers de ce corpus, c.-à.-d. 5000 syllabes dont 1165 accentuées et 522 associées à des débuts de groupes intonatifs. Les erreurs RMS du modèle sur un autre tiers du corpus sont autour de 18Hz ce qui se compare plutôt favorablement à l'erreur de stylisation d'environ 14Hz induite par le modèle phonétique.

1.3.3 Une grammaire de mouvements élémentaires

Au chapitre 1.2.1, nous avons vu comment l'approche IPO selon Beaugendre [1994] aboutit à l'identification de 9 mouvements linéaires qui permettent la stylisation des contours observés sous la contrainte d'une équivalence perceptive. Afin de pouvoir mettre cet inventaire en correspondance avec une analyse linguistique des énoncés, l'auteur identifie par un système de règles d'abord des frontières de groupes à partir

²²Integrated Model of German prosody

²³Les contours de groupe s'étendent toujours jusqu'à la fin de la phrase. Ce paramètre est donc à prédire uniquement pour les commandes d'accents.

d'un découpage syntaxique. Puis, sont attribués des accents de groupe qui sont, enfin, mis en relation avec les mouvements mélodiques.

On peut considérer ce système de règles comme le concentré de l'expertise de l'auteur par rapport au corpus étudié. Cependant, la généralisation de cette méthode rigoureuse à des corpora de taille plus conséquente semble relativement lourde.

1.3.4 Enchaînement de contours stockés

Nous avons constaté (cf. page 38) qu'**Emerard et al. [1992]** utilisent une description phonologique plus riche que Traber. Comment est-elle exploitée pour la génération des contours de f_0 ? L'idée est très simple : les valeurs de f_0 à attribuer à un mot donné sont stockées dans une table indexée par le marqueur mélodique (m1–m10) et le nombre de syllabes du mot. La table comprend des contours entre une et cinq syllabes. Pour les mots de plus de cinq syllabes, les valeurs pour la syllabe centrale sont dupliquées.

Un point crucial du modèle est évidemment la génération de la table de contours. Ils sont extraits d'un corpus de 150 phrases déclaratives, isolées. Les critères de sélection ne sont malheureusement pas très explicites : on choisit un contour « représentatif » qui ne montre pas de « discontinuité significative » entre la fin d'une voyelle et le début de la voyelle suivante.

Une caractéristique importante à souligner dans cette approche, est sa nature locale : les contours de mots sont enchaînés tels quels – aucun mécanisme ne permet de les mettre dans un contexte plus global. Ainsi par exemple, tous les mots étiquetés avec m9, donc les mots fonctionnels, et ayant le même nombre de syllabes, auront exactement les mêmes courbes de f_0 . Ces mots sont, par ailleurs, considérés par les auteurs comme « prosodically irrelevant in text-to-speech systems where neither emphasis nor focus are processed » [p.230]. Il me semble néanmoins étonnant de fixer leur fréquence fondamentale d'une manière absolue au lieu de l'intégrer dans un mouvement plus global.

Règles vs. modélisation implicite. La piste suivie par les auteurs semble être d'enrichir l'ensemble de clichés possibles et d'établir à partir d'un corpus un système de règles qui permettrait un choix optimal en fonction des informations phonémique et syntaxique. Les auteurs semblent même vouloir abandonner les contours de *mots* en attachant les règles directement aux syllabes (c.-à.-d. voyelles). Les règles devraient alors se complexifier, et elles remplissent finalement le même rôle que le réseau de neurones de Traber. L'avantage serait la lisibilité des règles par rapport à la modélisation implicite et opaque. Or, disposer des règles lisibles ne garantit pas en soi l'acquisition de connaissance pertinente. Une observation telle que : « Oral and nasal vowels follow a similar pitch pattern quite different from that of schwa whose attack is lower » [p.240] semble identifier la réduction d'une voyelle comme *cause* de sa fréquence fondamentale basse bien qu'il s'agisse probablement d'une corrélation entre deux *effets* dus à une

cause commune (position dans les mots p.e.). La modélisation implicite par un RN exploite également ce genre de corrélation, sans pour autant suggérer une relation de cause à effet. Elle n'a donc à mon avis rien à envier à la formulation sous forme de règles.

Enchaînement dynamique. Malfrère et al. [1998] proposent un enchaînement dynamique de contours d'extension plus grande : un corpus est segmenté en groupe d'accents et tous les contours de f_0 correspondants sont stockés dans un dictionnaire. Les unités sont indicées par les types d'accents et le nombre de syllabes qu'elles contiennent, par la moyenne de f_0 et par la position de l'unité dans la phrase. Prédire un contour revient ainsi à enchaîner et lisser les contours du dictionnaire qui correspondent à la suite d'accents souhaités et qui impliquent le moins de distorsion aux points de jonction. Cette approche se focalise sur une création de contours ayant l'air naturel, sans pour autant modéliser de manière compréhensive la prosodie. Il convient de noter la similitude avec une des méthodes proposées par Traber [1992] où des contours stockés dans un dictionnaire sont enchaînés. La différence réside dans le fait que Traber choisit de moyenniser des contours partiels si il y en a plusieurs de disponibles et que Malfrère et al. adoptent une stratégie plus globale en choisissant celui qui s'intègre le mieux dans la suite des contours. L'enchaînement dynamique de contours stockés se trouve aussi dans la proposition de López-Gonzalo et Rodríguez-García [1996]. Mais les contours ont l'empan d'une seule syllabe ce qui nécessite évidemment plus de calculs pour trouver un enchaînement sans « gêne » mais rend le corpus de contours stockés plus maniable.

1.3.5 ToBI vers f_0

Nous venons maintenant aux modèles qui s'appuient sur ToBI. Nous commençons par celui de **Black et Hunt [1996]**. Les étiquettes ToBI contiennent de l'information sur f_0 . Elles devraient alors fournir des prédicteurs puissants, ce qui réduit la complexité du modèle tentant de déterminer la réalisation phonétique d'un énoncé. Le modèle phonétique étant simple (3 valeurs de f_0 par syllabe), les auteurs ont recours aux fenêtres glissantes (cinq syllabes autour de la syllabe traitée) pour coder le contexte. Pour chaque syllabe ils aboutissent ainsi à 75 indices²⁴ binaires. A ces indices issus de la description phonologique s'ajoute l'information phonotactique par rapport à la phrase courante (8 indices). La transition vers la réalisation phonétique est opérée par trois modèles linéaires : un pour les valeurs de f_0 en début de syllabe, un autre pour celles de milieu et le dernier pour les valeurs de fin de syllabe.

Corpus et résultats. Les auteurs utilisent un corpus d'apprentissage de 4000 syllabes extrait du Boston University FM Radio corpus [Ostendorf et al., 1995] du locuteur f2b (une femme avec un registre de $f_0 = 163,5 \pm 42,2$ Hz selon [Dusterhoff et al., 1999]) et

²⁴anglais: "features"

un corpus de test de 926 syllabes. L'erreur de prédiction rapportée (34,8 Hz) montre une amélioration par rapport à une génération faite à base de règles [Anderson et al., 1984] : 44,7 Hz. Ces erreurs peuvent paraître très grandes par rapport aux valeurs citées par Traber (9,2 Hz). D'une part ceci s'explique probablement par le registre différent des locuteurs et d'autre part ceci montre bien la difficulté de comparer les résultats de différentes études.

Discussion. Cette approche a comme principal atout la simplicité de sa mise en oeuvre qui permet une application aisée à d'autres langues. Sa valeur en synthèse à partir de texte repose sur la bonne prédiction des étiquettes ToBI. Notons que Black et Hunt n'exploitent ici pas toute la richesse de la description phonologique car certains types d'accent sont trop rares dans le corpus d'apprentissage.

1.3.6 ToBI – mais pas trop

Nous avons vu l'effort que Möhler et Conkie [1998] ont porté sur l'extraction de prototypes phonétiques des accents (cf. page 32). Les auteurs utilisent des arbres de classification et régression (CART) afin de les prédire. Les critères sont une quarantaine d'indices comprenant des informations phonotactiques, la présence d'une frontière de phrase, les types de syllabe et de voyelle et, les durées de la syllabe et de la pause silencieuse suivante (s'il y en a une) et optionnellement le type d'accent (ToBI).

Les erreurs RMS de prédiction du système complet sont entre 33 et 48Hz. Elles sont plus petites pour le corpus d'articles de journal que pour le corpus de messages courts²⁵, ce qui est au moins en partie dû au modèle phonétique (cf. page 32).

Discussion. Ce qui est remarquable est que l'utilisation de ToBI dans les prédicteurs n'améliore les performances que légèrement ; ce qui indique que l'information contenue dans ces étiquettes est redondante par rapport aux autres entrées. Il serait cependant intéressant d'analyser la relation entre les prototypes phonétiques et les accents (H, L, etc.) sous-jacents et d'éclaircir ainsi la transition effective opérée du modèle phonologique vers la réalisation phonétique.

Un autre point à relever est l'utilisation des durées syllabiques ainsi que des pauses : il s'agit là certainement de bons prédicteurs, mais on doit se demander s'ils ne sont pas justement trop bons, car dans le cadre de la synthèse à partir de texte, ils ne sont évidemment pas connus d'avance. La génération de la fréquence fondamentale dépend ainsi de la qualité de la prédiction rythmique.

Un autre exemple de modèle tenté par la richesse des étiquettes ToBI en tant que prédicteurs est celui de Dusterhoff et Black [1997]. Ils utilisent 12 arbres de régression afin de prédire les paramètres du TIM (cf. page 31) à partir de 40 indices comparables à ceux employés par Black et Hunt [1996] (cf. page 46). Les auteurs montrent

²⁵anglais: "prompts"

que l'utilisation des *types* d'accents (selon ToBI) comme prédicteurs n'améliore pas la qualité des contours de f_0 générés. Dusterhoff et Black présupposent dans cet article la connaissance des événements (accents et tons de frontière), mais ils affirment qu'il serait beaucoup plus facile de prédire leurs emplacements que de prédire des étiquettes plus riches comme ceux de ToBI.

Une méthode de prédiction des événements est proposée par **Dusterhoff et al. [1999]** : 83% des événements-accents sont prédits correctement. Les erreurs des contours de f_0 données dans cet article sont basées sur des événements étiquetés à la main, et ne permettent donc pas de juger de l'influence d'événements mal prédits. Elles sont de 34,3 Hz avec une corrélation de 0,6 pour le corpus f2b (le même corpus qu'utilisent Black et Hunt [1996] en obtenant une erreur de 34,8 Hz – cf. page 46) et de 9,1 Hz (corrélation : 0,74) pour un corpus de 450 phrases isolées d'un locuteur masculin avec un registre de $f_0 = 126,9 \pm 27,9$ Hz. Ici aussi, on se retrouve donc face à des grandes différences d'erreurs selon les corpora. Les auteurs attribuent ceci au fait qu'il soit « *easier to predict the intonation of a database when it is for mostly declarative, isolated sentences* ».

1.3.7 Prédiction d'une description phonologique

Bien que les modèles précédents semblent mettre en question l'utilité d'une description phonologique forte, une autre perspective est possible : **Véronis et al. [1998]** font confiance à un modèle simpliste (et à améliorer ultérieurement) de transition entre la description phonologique en terme d'INTSINT (cf. page 36) et la réalisation phonétique (MOMEL). Le problème se déplace alors vers la prédiction des étiquettes phonologiques.

L'originalité du modèle est de se contenter d'une information syntaxique simple. Le présupposé est que le découpage d'un énoncé qu'une analyse syntaxique peut fournir constitue la base nécessaire et suffisante à la prédiction des étiquettes phonologiques. À la différence d'Emerard et al. (cf. page 45), l'analyse syntaxique est ici implicite : les auteurs n'utilisent que la classe grammaticale C_i des mots qui, selon eux, peut être obtenue d'une manière automatique à partir du texte avec une fiabilité de 95%. Véronis et al. formulent deux hypothèses de base qui permettront la mise en place d'un modèle stochastique : (1) pour prédire les étiquettes INTSINT P_i , on peut faire abstraction des mots en ne considérant que leurs classes grammaticales C_i et (2) on peut réduire l'information prosodique à exactement une étiquette INTSINT P_i par mot.

En génération de prosodie, il s'agit de trouver la suite P d'étiquettes qui est la plus probable sachant que l'on est en présence d'une suite C de classes grammaticales. On a :

$$Pr(P|C) = \frac{Pr(P) \cdot Pr(C|P)}{Pr(C)}$$

Étant donné que le dénominateur ne dépend pas de P , il suffit d'examiner le numé-

rateur. Afin de pouvoir estimer les distributions de probabilité à partir d'un corpus, il est indispensable d'introduire des contraintes/approximations de la forme des distributions. Ainsi, les auteurs proposent les factorisations suivantes :

$$\begin{aligned} Pr(P) &= \prod Pr(P_i|P_{i-1}P_{i-2}) \\ Pr(C|P) &= \prod Pr(C_i|C_{i-1}P_i) \end{aligned}$$

Ceci revient à dire que la probabilité de trouver une étiquette P_i donnée ne dépend que des deux étiquettes précédentes ce qui permet de calculer la probabilité pour la suite P comme le produit de ces probabilités individuelles. Il suffit maintenant d'estimer les probabilités par les fréquences d'occurrence dans un corpus. Le corpus utilisé consiste de 500 phrases du corpus EUROM 1 avec un total de dix locuteurs. La prédiction des étiquettes INTSINT se fait en choisissant pour la suite C d'une phrase donnée la suite P ayant la plus grande probabilité qui est maintenant devenue calculable.

Résultats. Pour évaluer le modèle, Véronis et al. mènent un test perceptif de type MOS²⁶ portant sur quatre versions différentes :

1. f_0 d'origine (stylisé selon MOMEL) ;
2. f_0 à partir des étiquettes INTSINT originales ;
3. f_0 à partir des étiquettes INTSINT prédites ;
4. f_0 générée à partir de points cibles aléatoires.

Les durées segmentales pour toutes les versions sont égales aux durées moyennes des phonèmes. L'ordre des résultats est – comme attendu – décroissant : la version 1 est meilleure que la version 2 et la version 3 est supérieure à la 4. Entre les deux versions basées sur les étiquettes INTSINT la différence n'est pas significative. Le modèle semble donc bien capable de prédire ces étiquettes.

1.3.8 Résumé

L'intention de ce chapitre n'était pas de faire un tour complet des modèles de génération prosodique. A l'aide de quelques exemples concrets, j'espère avoir montré l'interaction complexe entre les présupposés phonétiques et phonologiques adoptés. Le classement selon les deux axes a été une aide d'orientation – il n'est pas une fin en soi. Le chapitre suivant montrera qu'il existe des alternatives au découpage du processus texte vers signal que j'avais proposé jusqu'ici.

²⁶Mean Opinion Score

1.4 Sélection d'unités

Après la présentation des modèles phonétiques et phonologiques, faisons un pas en arrière et resituons leur place dans la chaîne de génération automatique de la prosodie à partir du texte (cf. figure 1.6 qui est une version enrichie de la figure 1 sur la page 17). À partir d'une analyse linguistique et à travers le modèle phonologique, le modèle phonétique fournit une description paramétrique de la prosodie (durées segmentales, f_0 , énergie etc.). Celle-ci pilote la mise en forme par des techniques de traitement du signal de l'enchaînement des unités de base (diphones p.e.) disponible dans une base de données. Ces unités sont choisies selon les besoins phonémiques que l'analyse linguistique a établis.

Or, si la synthèse à l'aide d'un dictionnaire de diphones permet de se passer d'une modélisation détaillée de la coarticulation entre phonèmes, un nombre croissant de chercheurs tente depuis **Sagisaka [1988]** d'étendre cette idée de modélisation implicite à des propriétés suprasegmentales. On n'utilise plus un dictionnaire avec des entrées de taille fixe (p.e., diphones), mais on enchaîne au moment de la synthèse les meilleures unités disponibles sélectionnées dans un grand corpus.

Cette approche est très hétérogène :

1. Quelques auteurs [e.g. Sagisaka, 1988; Donovan et al., 2001] l'utilisent simplement afin d'optimiser le choix d'unités de base, qui seront adaptées par traitement du signal aux besoins phonétiques spécifiés par d'autres modules de plus haut niveau (tout comme c'est le cas pour l'utilisation de diphones), et de réduire ainsi la distorsion induite par le traitement du signal. Cette approche est représentée dans la figure 1.6 par la flèche ①.
2. Ou l'on essaie de choisir les meilleures unités non pas seulement par rapport à leur durée et fréquence fondamentale d'origine, mais aussi par rapport à d'autres critères comme leur appartenance à un accent d'un type donné [e.g. Campbell et Black, 1997; Wightman et al., 2000] (cf. flèche ②).
3. D'autres auteurs vont beaucoup plus loin : la base de données est étiquetée avec les caractéristiques phonotactiques et prosodiques des unités et on essaie de se passer de toute modification des unités enchaînées dans la mesure du possible. Pour une revue récente on peut se référer à Möbius [2000].

La question évidente à résoudre est : Quelles unités du corpus sont les *meilleures* pour synthétiser un énoncé donné ? En général, on distingue deux types de critères : est-ce que l'unité choisie a des propriétés proche de la cible, et est-ce que la suite des unités évite des distorsions aux points d'enchaînement ? À chacun de ces critères, on peut attribuer une fonction de coût que l'on essaie de minimiser. Ceci n'étant qu'une reformulation de la question posée, il s'agit de quantifier ces fonctions et de pondérer soigneusement les différentes contributions.

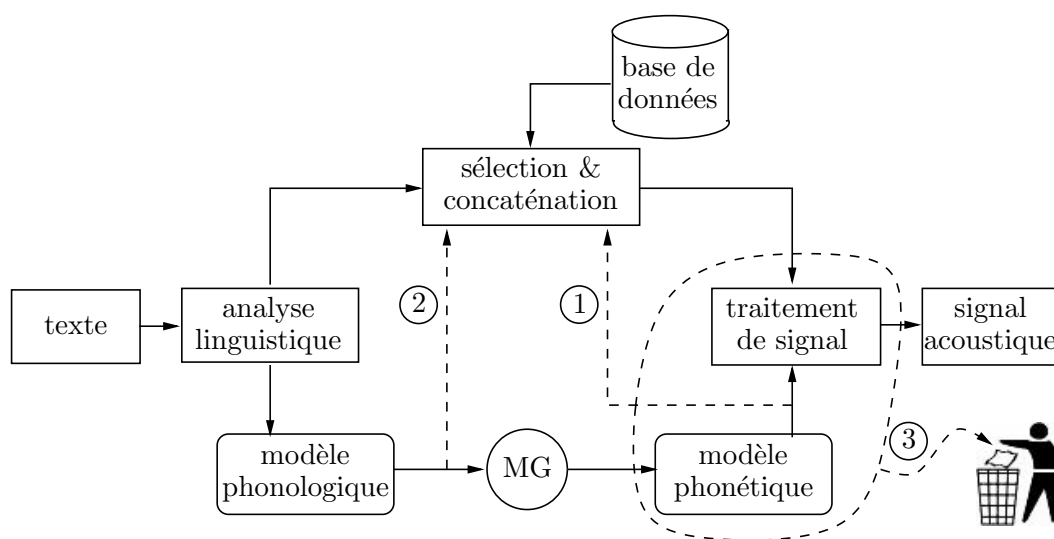


FIG. 1.6: **Sélection d'unités.** Les modèles phonologique et phonétique ainsi que le modèle de génération (MG) ont été l'objet des trois chapitres précédents. Les modèles présentés ici identifient le traitement de signal comme maillon faible de la chaîne de génération du texte vers le signal acoustique, et essaient de réduire son influence : en sélectionnant des unités proches des caractéristiques phonétiques spécifiées par le modèle (flèche ①) et/ou en incluant des critères phonologiques dans le processus de sélection (flèche ②) dans le but d'éliminer tout traitement de signal et le modèle phonétique (cf. ③).

Afin de marquer la différence avec les modèles discutés plus haut, nous ne nous intéressons ici qu'aux approches qui tentent de déplacer au moins une partie de la génération de la prosodie dans le processus de la sélection d'unités. L'idée est donc de choisir des unités qui sont phonémiquement correctes – condition *sine-qua-non* pour l'intelligibilité – et qui sont de plus dans le même contexte phonologique que les cibles, ce qui garantit implicitement leur bonne forme prosodique. Ces modèles n'ont donc théoriquement pas besoin de modèle phonétique.

Taylor et Black [1999] emploient cette voie. La sélection se fait à partir des structures phonologiques. La phrase à synthétiser est analysée en tant qu'arbre (les feuilles étant les phonèmes). Commenant par son sommet les nœuds sont comparés aux nœuds présents dans le corpus. Si aucun nœud avec la même structure *et* le même contenu phonémique n'est trouvé, la recherche se poursuit pour les nœuds « fils » jusqu'à ce que la recherche aboutisse. Comme il serait plus facile pour les modules de « haut niveau » de générer la structure phonologique d'un énoncé que de prédire sa réalisation phonétique, les auteurs espèrent ainsi pouvoir copier la prosodie naturelle du corpus vers l'énoncé à synthétiser. Apparemment, cette idée tentante ne peut pas être poursuivie jusqu'au bout, mais doit être secondée par un modèle de génération prosodique « classique » qui fixe des cibles phonétiques (f_0 , durée etc.). Si les unités

choisies sont trop éloignées des cibles, elles sont transformées par des moyens du traitement du signal. Les auteurs clament d'obtenir de cette manière « the best of both worlds ». Or, la pondération de ces deux mondes me semble extrêmement difficile : comment savoir à quel moment le modèle classique a vu plus juste que la prosodie implicite ? Si on fait confiance au modèle « classique », l'amélioration apportée par la sélection d'unités sera du même type que dans [Campbell et Black, 1997; Wightman et al., 2000] où l'on vise plus simplement une réduction des distorsions en ayant recours à un choix plus large d'unités, et la question qui se pose est si l'on peut envisager de faire ① que de temps en temps.

En général, les plus grands bénéfices d'une sélection dynamique d'unités sont attendus pour des applications en domaine restreint. **Bulyko et Ostendorf [2001]** proposent de lier la prédiction prosodique et la sélection d'unités. Le corpus est étiqueté selon ToBI²⁷. L'apport original est l'utilisation d'un même formalisme pour décrire les structures prosodiques présentes dans le corpus et pour prédire celles qui doivent l'être : les « weighted finite state transducers ». Ceci permet de pondérer plus flexible-ment l'apport de la prosodie « prédite ».

Citons finalement la proposition récente de **Prudon et d'Alessandro [2001]** pour le français qui est d'une étonnante simplicité : non seulement les auteurs se passent de modèle phonétique (comme attendu dans le cadre des approches non-modifiantes) mais ils n'utilisent pas de modèle phonologique non plus. Les phonèmes du corpus qui fournit les unités à enchaîner sont étiquetés par deux indices : leur position dans la syllabe et leur position dans le mot. La sélection obéit à quatre critères :

- préférer les unités avec le bon indice de mot
- sinon, préférer les unités avec le bon indice syllabique
- préférer des unités adjacentes
- sinon, préférer des unités avec une f_0 semblable

Une évaluation perceptive est menée afin de comparer la méthode décrite ci-dessus avec un système classique (concaténation de diphones plus modification selon un modèle prosodique) : dix sujets jugent 18 stimuli (entre 8 et 22 secondes) en six voix différentes dont une naturelle, deux à base de diphones et trois basées sur la sélection d'unités. Les résultats semblent favorables à la nouvelle proposition surtout en ce qui concerne l'impression globale.

Ce qui caractérise la proposition de Prudon et d'Alessandro est avant tout qu'elle ne fait aucune allusion aux fonctions que la prosodie peut assumer. Les unités sont choisies en fonction de leur emplacement dans les mots mais pas par rapport au rôle que les mots remplissent dans la phrase. On peut supposer qu'il s'agit là d'une première implémentation. Mais la voie est ouverte de générer la prosodie par un lien direct entre fonctions et leurs réalisations sans passer par un interface paramétrique entre phonologie et phonétique.

²⁷Certains types d'accents étant trop rares, seulement trois catégories d'accents sont retenus.

1.5 Modélisation des durées

Après avoir donné un aperçu de quelques modèles qui visent la génération de l'évolution de la fréquence fondamentale, nous allons nous tourner maintenant vers la prédiction de la structure temporelle d'un énoncé. Dans le cadre d'une synthèse par concaténation, le problème posé est simple : il faut spécifier la durée des phonèmes à enchaîner.

La modélisation rythmique étant un domaine complexe (prédiction de pauses, débit etc.), la présentation que nous proposons ici est fortement restreinte. Elle oppose deux approches possibles : la première répond directement au problème posé ci-dessus, la seconde suppose qu'une factorisation entre micro- et macro-rythme est possible, ce qui permet de porter l'effort de prédiction au niveau syllabique.

1.5.1 Prédiction directe des durées segmentales

Bartkova et Sorin [1987] proposent un modèle de prédiction des durées segmentales pour le français dans la tradition de Klatt [1979, 1987] et de O'Shaughnessy [1981, 1984]. Il s'agit d'un modèle multiplicatif qui s'appuie d'une part sur des caractéristiques du phonème à traiter et d'autre part sur des coefficients qui dépendent du contexte syntaxique/prosodique :

$$\begin{aligned} \text{durée de voyelle} &= DI \cdot V_i \cdot m_c \\ \text{durée de consonne} &= DI \cdot C_{ij} \end{aligned}$$

où DI est la durée intrinsèque du phonème et m_c un coefficient co-intrinsèque reflétant l'influence de la consonne²⁸ suivante sur la voyelle en question. Ces paramètres sont obtenus grâce à un corpus de 308 logatomes (quatre locuteurs). Les durées intrinsèques sont considérées comme indépendantes du locuteur et les m_c ont été déterminés pour les deux locuteurs extrêmes (en termes de vitesse d'élocution).

Les facteurs V_i et C_{ij} sont censés capturer l'influence du contexte sur les durées et dépendent, en principe, du locuteur. Ils sont calculés à l'aide de deux corpora (2 paragraphes et 70 phrases construites pour l'analyse des groupes consonantiques). Les contextes considérés par Bartkova et Sorin sont : mot lexical/fonctionnel, pause (longue/courte/finale), frontière syntaxique, position de syllabe à l'intérieur du mot et groupe consonantique. Les coefficients sont associés aux feuilles d'un arbre de décision binaire portant sur ces contextes. Les V_i sont indépendants de la voyelle, mais les C_{ij} dépendent de l'appartenance de la consonne à une des sept classes j .

La formule générale de durée est secondée par quelques règles spéciales qui gèrent des exceptions dans l'attribution des coefficients C_{ij} (p.e. pour des liaisons ou des voyelles nasales suivies d'une consonne nasale).

²⁸plus précisément : l'appartenance de la consonne à une classe parmi six

Les attraits principaux de ce modèle sont sa simplicité au niveau calcul et la séparation lisible d'une part entre les influences phonémiques (DI, m_c) vs. positionnelles (V_i, C_{ij}) et d'autre part entre les paramètres dépendant vs. indépendant du locuteur. Suivant cette logique, il peut sembler étonnant que l'appartenance d'une consonne à un groupe consonantique soit gérée par le choix des C_{ij} . Apparemment, le modèle multiplicatif ne permet pas de séparer cet effet du contexte suprasegmental.

L'inconvénient du modèle réside dans la difficulté d'établir l'arbre de décision : quels critères prendre en compte et dans quel ordre ? Une extension des règles nécessiterait probablement le recours aux techniques d'apprentissage automatique des arbres, ce qui comprend le risque de perdre la lisibilité des règles en tant que connaissances. Les « règles spécifiques » laissent également pressentir la nécessité d'en ajouter dès que des corpora plus conséquents sont inclus dans l'analyse.

Les écart-types entre durées mesurées et prédites se situent entre 10 et 13ms pour la relecture par les deux locuteurs modélisés du texte qui a servi à l'établissement des coefficients V_i et C_{ij} . L'erreur augmente à 14–17ms pour la lecture d'un nouveau texte et reste donc relativement petite.²⁹

Van Santen [1992, 1994] propose une analyse systématique des différents facteurs supposés avoir une influence sur les durées segmentales. Cette approche nécessite le regroupement des phonèmes en classes (voyelles, consonnes et consonnes à l'intérieur de groupes consonantiques) et aboutit ainsi à une généralisation (sous forme de sommes de produits) des modèles de Klatt qui s'appuie sur un grand corpus contenant plus de 40000 réalisations phonémiques.

1.5.2 Prédiction par l'intermédiaire des durées syllabiques

Nous avons vu qu'il est possible de prédire les durées segmentales en s'appuyant sur les relations statistiques entre des indices qui peuvent être extraits du texte et des durées observées. Or, il semble qu'un auditeur soit plus sensible à l'organisation temporelle *supra*-segmentale qu'aux « détails » des durées phonémiques (cf. par exemple Barbosa [1994]). Il est donc souhaitable de ne pas accumuler les erreurs de prédiction sur ces dernières, mais de prédire directement les durées *supra*-segmentales. Sous cette perspective, le calcul des durées de phonèmes peut être confié à un modèle de « répartition ».

Campbell [1992] choisit cette voie en proposant un algorithme de répartition qui est basé sur l'idée que l'élasticité d'un phonème à l'intérieur d'une syllabe correspond à l'écart-type de la distribution de ses durées mesurées dans un corpus de logatomes : plus sa durée varie dans ce corpus « neutre », plus il sera allongé/comprimé afin de remplir la durée de la syllabe.

À quel point certains auteurs considèrent les durées segmentales comme secondaires par rapport aux durées syllabiques, se reflète dans des remarques brèves telles que

²⁹Les auteurs proposent de la comparer à une « just-noticeable difference » de 20ms.

« Phone duration is calculated from the superordinate syllable's duration taking into account the phone properties found in the database » [Mixdorff et Jokisch, 2001] ou : « a relatively simple algorithm spreads the syllabic duration over its segments according to the segment durational class and the type of syllabic structure » [Zellner, à paraître].

Notons finalement que van Santen [1997] s'oppose farouchement à la possibilité de prédire les durées segmentales par l'intermédiaire des durées syllabiques et ceci pour deux raisons : « Accurate prediction of syllabic duration requires taking into account the full segmental makeup of a syllable. » et « These interactions make it obviously impossible to accurately predict segmental duration merely from knowledge of total syllable duration » (p. 239). En général, les modèles tiennent compte de la première observation en faisant référence au contenu phonémique des syllabes pour la prédiction de leurs durées. La deuxième objection touche plus sévèrement aux fondements de l'approche suprasegmentale – nous y reviendrons dans le paragraphe 2.1.1.4 (cf. page 70) après avoir exposé la partie rythmique de notre modèle.

Le modèle

Cette partie centrale du manuscrit est consacrée à la motivation, la description et l'apprentissage du modèle qui fait l'objet de la présente étude : le SFC « Superposition of Functional Contours ». Il génère la réalisation prosodique d'un énoncé en superposant des contours (potentiellement chevauchants) qui ont à charge d'encoder directement des fonction communicatives.

Plan de la partie. Nous allons commencer par détailler notre description phonétique (de bas niveau) de la prosodie en tant que phénomène multiparamétrique. Elle est caractérisée par des contours à multiples dimensions, ancrés sur la structure segmentale. Une partie relativement importante de ce chapitre sera occupée par la séparation entre micro- et macro-rythme ; seul ce dernier sera traité par le SFC.

Le chapitre suivant donnera les motivations qui nous ont amené à proposer un modèle qui opère en superposant des contours associés directement aux fonctions (para-)linguistiques. Le SFC, inspiré des propositions théoriques d'Aubergé, hérite d'un nombre important de propriétés du modèle de Morlec [1997] que nous exposerons avant de venir aux limitations de ce dernier qui sont surtout liées à sa notion de hiérarchie stricte de contours. Une rapide excursion dans la forêt des arbres de performance nous aidera à motiver pourquoi nous attachons beaucoup d'importance aux poids syllabiques des groupes de mots et pourquoi nous croyons qu'il s'agit là du facteur principal qui détermine la morphologie des contours.

Ensuite, nous serons prêts à décrire le modèle proprement dit : après l'identification des fonctions (para-)linguistiques et de leurs portées, nous introduirons les générateurs de contours et nous verrons comment les contours complets sont obtenus par superposition.

Le principe du SFC étant très simple, la difficulté pratique réside dans l'entraînement des générateurs : ce n'est qu'en observant en détail l'inversion du modèle – c.-à.-d. l'apprentissage dans un cycle d'analyse par la synthèse – que le rôle des générateurs peut être clarifié. L'algorithme d'apprentissage contient également la clé pour l'interprétation des résultats que le SFC peut fournir.

2.1 Préliminaires

Ce chapitre présente une étape préparatoire à la description du modèle de génération de la prosodie. Il s'agit de la factorisation entre micro- et macro-prosodie afin que le SFC proprement dit puisse se concentrer sur la prédiction de cette dernière. J'entends ici par macro-prosodie les phénomènes *inter*-syllabes, et par micro-prosodie tout ce qui relève du niveau *intra*-syllabe.

Ce qui suit peut être abordé suivant deux points de vue : d'une part, comme extraction ascendante de paramètres macro-prosodiques à partir du signal acoustique, d'autre part, nous ne perdrons pas de vue que ce processus doit être réversible, c.-à.-d. les paramètres macro-prosodiques doivent être utilisables en tant qu'entrée d'un modèle micro-prosodique qui prédit, par exemple, les durées segmentales. En effet, la présentation de l'émergence de pauses silencieuses est très fortement guidée par cette deuxième perspective descendante bien que l'ajustement de ses paramètres fait forcément appel aux observations et donc à une vue ascendante.

Plan du chapitre. Après avoir détaillé séparément les composantes rythmique et mélodique de notre description phonétique de la prosodie, nous allons clarifier ce que nous entendons par « contour prosodique ». Il ne s'agit ni plus ni moins que d'un rassemblement des deux composantes, aligné avec la structure syllabique d'une phrase. Son intérêt est surtout d'ordre pratique car il permet au SFC de gérer les deux composantes simultanément et par le même formalisme.

2.1.1 La composante rythmique

Le modèle phonétique de la durée est inspiré par le modèle de Campbell [1992] (cf. page 54) qui a été adapté au français par Barbosa [1994]. La présentation ci-dessous se base sur une définition du GIPC et d'un coefficient d'allongement qui a la vocation de représenter la composante macro-rythmique. La suite de ce chapitre sera consacrée aux détails micro-rythmiques. Le fil rouge sera implicitement la perspective de synthèse : connaissant le coefficient d'allongement et en s'en tenant à la définition du GIPC, comment obtenir les durées de phonèmes¹ et comment décrire l'émergence de pauses ?

Définition du GIPC. L'unité rythmique que nous utilisons est le « groupe inter-P-center » (GIPC), c.-à.-d. l'intervalle entre deux centres perceptifs (PC). En première approximation, nous définissons un PC comme la fin d'une pause silencieuse ou l'attaque d'une voyelle (s'il n'y a pas de pause depuis la voyelle précédente). Cette définition pourrait être affinée en tenant compte du fait que le centre perceptif semble, selon Scott [1993], caractérisé phonétiquement par une montée d'énergie dans une bande spectrale

¹voir remarque 1 sur p. 22

centrée autour de 1000Hz : cela conduirait à avancer légèrement le PC dans le cas où la voyelle est précédée par une consonne voisée.

Le coefficient d'allongement. Le coefficient d'allongement C a pour vocation d'exprimer la déviation de la durée d'un GIPC par rapport à une durée neutre (ou attendue) Δ' . Nous définissons :

$$\Delta' = (1 - r) \cdot \sum_i \bar{d}_i + r \cdot D \quad (2.1)$$

où i indexe les phonèmes contenus dans le GIPC, \bar{d}_i est la durée moyenne (déterminée si possible à partir d'un corpus indépendant) du phonème i , D est la durée moyenne des GIPCs et r un facteur de pondération actuellement fixé² à 0,6. Le premier terme tente d'allonger les GIPCs qui contiennent beaucoup de phonèmes tandis que le second assure une tendance vers l'isochronie (cf. [Fant et Kruckenberg, 1996]).

Pour un GIPC avec une durée mesurée Δ nous définissons le coefficient d'allongement :

$$C = \ln(\Delta/\Delta') \quad (2.2)$$

Dans une perspective de synthèse, la connaissance de C pour un GIPC donné permet de calculer sa durée. Notons ici, qu'il serait certainement intéressant d'affiner la définition de C en s'appuyant sur des résultats en perception de la vitesse locale d'élocution comme proposé par Pfitzinger [1999].

2.1.1.1 Répartition

Principe de base. Comment obtenir les durées segmentales à partir de la durée du GIPC? Nous faisons l'hypothèse que l'élasticité des phonèmes est proportionnelle à l'écart-type de leur durées (mesurées – ici aussi – dans un corpus indépendant) et qu'à l'intérieur d'un GIPC tous les phonèmes sont comprimés (ou étirés) dans le même sens. Une manière simple de réaliser ceci est de déterminer le z -score k de sorte que

$$\Delta = \sum_i e^{\mu_i + k\sigma_i}, \quad (2.3)$$

où μ_i et σ_i sont respectivement la moyenne et l'écart-type de la distribution des logarithmes des durées du phonème i . On attribue alors au phonème i la durée $e^{\mu_i + k\sigma_i}$. L'inversion de l'équation (2.3) ne peut pas être effectuée analytiquement, mais déterminer k à partir des μ_i , σ_i et Δ avec des méthodes d'inversion numérique, est simple et rapide.

²En variant r , nous avons synthétisé quelques phrases de sorte que toutes les durées des GIPCs égalent leur Δ' . Nous avons retenu la valeur de r qui procure la perception de la plus grande monotonie rythmique. Une meilleure assise expérimentale de ce paramètre serait certainement souhaitable.

Optimisation de la répartition. La répartition peut être améliorée en optimisant les paramètres μ_i et σ_i pour un corpus donné : pour tous les GIPCs qui ne contiennent aucune pause, et connaissant la durée totale des GIPCs, nous minimisons au sens des moindres carrés l’erreur de prédiction des durées segmentales. Les μ_i et σ_i perdent alors leur signification de moyenne et écart-type d’un phonème i , et deviennent paramètres libres de la répartition. Il aurait peut-être été plus judicieux de les renommer $\hat{\mu}_i$ et $\hat{\sigma}_i$ afin de signaler ce changement. Nous ne l’avons pas fait, parce que l’optimisation est une étape optionnelle : les paramètres optimisés prenant simplement la place des paramètres standard (sauf dans la définition de la durée attendue Δ' où nous continuons d’utiliser la « vraie » moyenne μ_i). La figure 2.1 montre l’effet de l’optimisation sur les paramètres.

Afin de donner une idée de l’effet de cette optimisation, il semble utile de considérer les erreurs sur un corpus de référence. Ici, et par la suite, nous utilisons celui des formules mathématiques (cf. la description de ce corpus à partir de la page 109). Il consiste en 157 phrases – chacune répétée deux fois. L’optimisation est effectuée avec les données de la première répétition (appelée ici corpus A). Les chiffres cités correspondent aux erreurs RMS en ms. Nous excluons systématiquement le dernier GIPC d’une phrase et nous ne considérons évidemment que les GIPCs avec le même contenu phonémique – des erreurs dans la prédiction des pauses (voir ci-dessous) pouvant changer ce contenu par rapport à l’original. Le tableau 2.1 montre que l’erreur baisse substantiellement. Un test de rang selon Wilcoxon appliqué aux erreurs absolues révèle que cette différence est fortement significative – et ceci également pour le corpus B qui n’a pas été utilisé pour l’optimisation des paramètres.

En guise de référence, nous avons ajouté dans la table l’erreur d’une répartition « naïve » qui consiste à attribuer à tous les phonèmes à l’intérieur d’un même GIPC la même durée (c.-à.-d. μ_i, σ_i indépendants de i). Un test – certes conservateur – de Wilcoxon indique que l’amélioration de la répartition « standard » par rapport à la répartition « naïve » n’est pas statistiquement significative. Un test-t appliqué aux erreurs absolues logarithmées³ a confirmé cette observation.

	naïf		standard		optimisé	
	#GIPCs	RMSE	#GIPCs	RMSE	#GIPCs	RMSE
corpus A	2347	19,7	2283	17,9	2293	13,4
corpus B	2345	18,3	2303	17,4	2292	13,4

TAB. 2.1: **Erreurs de répartition (RMS en ms) – GIPCs sans pauses.** Les colonnes « #GIPCs » indiquent le nombre de GIPCs pris en compte.

³Afin de rendre les distributions fortement asymétriques « plus normales ». Les erreurs égales à zéro ont été écartées ; elles proviennent de toutes façons des GIPCs avec un seul phonème, qui sont sans intérêt pour la répartition.

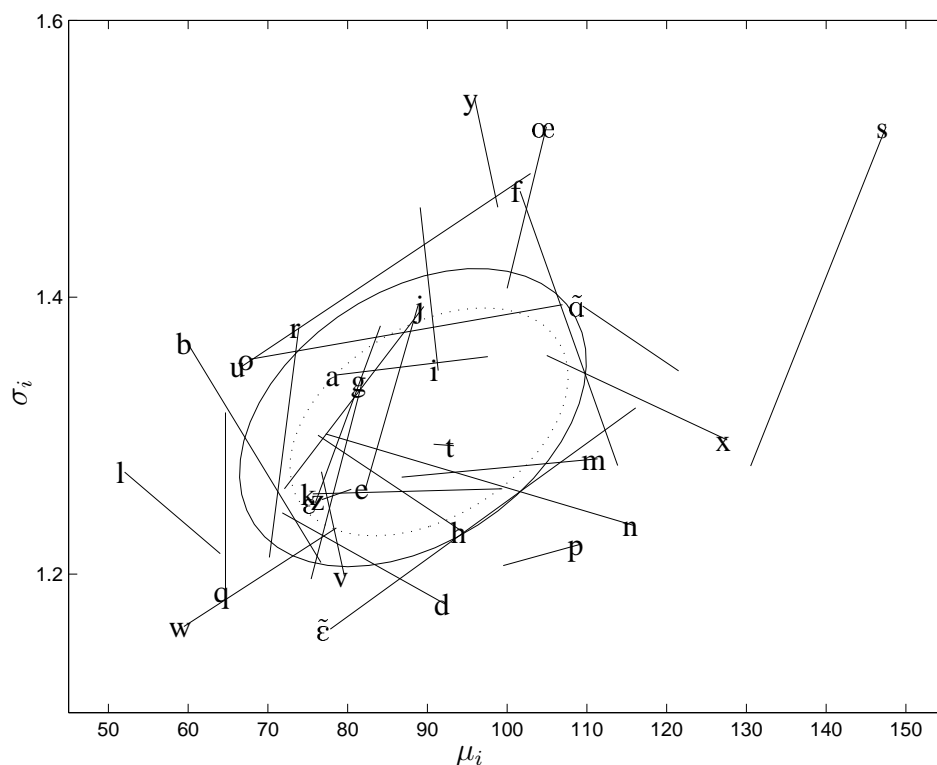


FIG. 2.1: **Déplacement des paramètres par l'optimisation.** μ_i et σ_i sont (avant l'optimisation) respectivement la moyenne et l'écart-type de la distribution des logarithmes des durées du phonème i . Afin d'améliorer la lisibilité μ_i et σ_i sont exprimés en ms.

Les étiquettes des phonèmes sont placées à l'endroit correspondant aux valeurs *optimisées*. Les ellipses de dispersion (pointillé : avant ; trait plein : après l'optimisation) montrent que la moyenne des μ_i diminue légèrement et que leur écart-type augmente. La moyenne des σ_i reste constante – il s'agit là d'une contrainte à laquelle nous avons soumis l'optimisation, car la multiplication des σ_i avec un facteur commun et constant n'influe pas sur la répartition puisqu'il peut être absorbé par le z-score k (cf. équation (2.3)). Globalement, la dispersion des paramètres devient plus grande après l'optimisation.

En détail, on s'aperçoit, par exemple, que les phonèmes [f] et [s] deviennent plus élastiques.

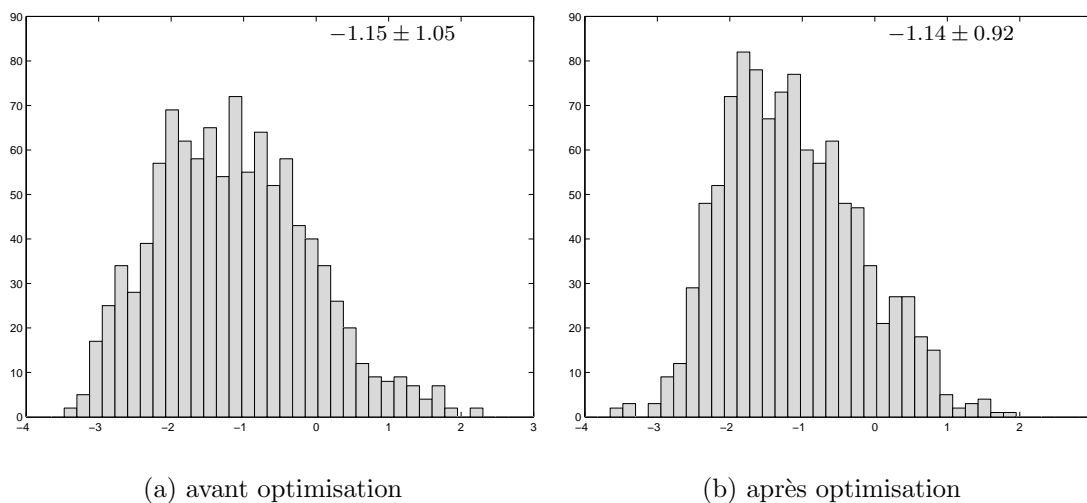


FIG. 2.2: **Distribution des z-scores.** Après l'optimisation, la distribution des z-scores pour les GIPCs ne change que légèrement, son écart-type devient plus petit et la moyenne reste plus ou moins inchangée.

2.1.1.2 Génération des pauses

Les pauses silencieuses font naturellement partie des GIPCs : la pause est considérée comme un phénomène émergent – au sens où elle résulte d'un fort allongement de la syllabe avant la pause. En analyse, ceci ne pose aucun problème : allongement final et/ou pause se traduisent par des coefficients d'allongement grands – reflétant ainsi une cohésion moindre entre les mots en question. Or, l'inversion n'est pas unique, c.-à.-d. un GIPC avec des phonèmes courts et une pause (pause silencieuse) peuvent avoir le même coefficient d'allongement que le même GIPC sans pause et des durées phonémiques plus élevées (pause subjective). Par la suite, nous allons donc détailler comment cette inversion est réalisée.

Critères d'insertion de pauses. Nous suivons l'idée de Barbosa [1994] qu'allonger progressivement un GIPC consiste à allonger les phonèmes comme décrits plus haut jusqu'à ce qu'il devienne plus « économique » d'introduire une pause. Afin de déterminer un critère de génération de pauses silencieuses à partir de notre corpus, nous partons de l'information disponible en synthèse : la chaîne des phonèmes, les emplacements possibles des pauses (c.-à.-d. entre deux mots) et les coefficients d'allongement (CA). Nous ne pouvons pas utiliser directement la durée des GIPCs observés, car leur contenu dépend de la présence des pauses.

Le critère le plus simple serait un seuil sur le CA. Or, ceci s'avère trop simple

puisque nous observons des insertions systématiques de pauses en fonction du contenu phonémique du GIPC concerné : le corpus A contient, dans des contextes comparables et avec des CA comparables, 17 fois la séquence « égal un » et 25 fois la séquence « égal deux ». La première présente toujours une pause silencieuse, la seconde seulement dans 12% des cas. Une explication possible est le fait que le « d » au début de « deux » contribue à l’extensibilité du GIPC « frontalier » et rend ainsi l’insertion d’une pause superflue.

Le critère que nous retenons porte donc sur le z-score des GIPCs « frontaliers » sous l’hypothèse qu’il n’y a pas de pause. S’il dépasse un seuil, nous insérons une pause. Le seuil est déterminé sur la base de tous ces GIPCs dans le corpus A. Nous devons donc donner la définition du GIPC de référence pour lequel le z-score est calculé : compte tenu du fait que la génération des pauses doit procéder de gauche à droite à l’intérieur de la phrase, nous utilisons comme début le début *réel* (qui peut être influencé par une pause *observée* dans le GIPC précédent). Sa fin est déterminée sous l’hypothèse qu’il ne contient lui même aucune pause. Prenons l’exemple du deuxième GIPC « frontalier » dans la séquence « racine de deux » avec sa transcription $[\text{rasin}(-)\text{d}\emptyset(-)\text{d}\emptyset]^4$ où « (-) » indique une pause éventuelle. Le début du GIPC dépend du fait que la pause après « racine » soit réalisée dans le corpus ou non. La fin étant indépendante de la réalisation de la pause après « de », nous avons donc deux possibilités : $[\text{d}\emptyset]$ si nous observons une pause après « racine », ou $[\emptyset\text{d}]$ sinon. Ensuite, nous calculons grâce au CA mesuré⁵ la durée hypothétique Δ_h du GIPC de référence selon les équations 2.1 et 2.2. Rappelons que cette durée peut être différente de celle observée – ce que nous retenons de l’observation est uniquement le CA et le *début* du GIPC. Puis, nous calculons un z-score « hypothétique » k_h de sorte que :

$$\Delta_h = \sum_i e^{\mu_i + k_h \sigma_i} \quad (2.4)$$

Un seuil κ_h est déterminé afin que la règle

$$\begin{aligned} k_h \leq \kappa_h &\implies \text{pas de pause} \\ k_h > \kappa_h &\implies \text{insérer une pause} \end{aligned}$$

minimise l’erreur⁶ de la prédiction des pauses. La figure 2.3 montre le taux de prédiction correcte en fonction de κ_h . Le tableau 2.2 donne les taux de prédiction obtenue avec la

⁴Le lecteur pourrait être surpris de voir « de » transcrit par $[\text{d}\emptyset]$ et non $[\text{d}\emptyset]$. La raison à cela est qu’il semble ici (au moins pour notre locuteur) préférable de le traiter en tant que syllable à part entière.

⁵Le CA mesuré est basé sur le contenu *observé* du GIPC qui peut être différent de celui que nous utilisons comme référence.

⁶Nous ne distinguons, pour l’instant, pas entre erreurs de première ou de seconde espèce.

valeur optimale de κ_h . Optimiser ou non les paramètres de répartition ne semble pas avoir d'influence sur la fiabilité de la méthode de génération de pause proposée ici : bien que le seuil doive évidemment être adapté ($\kappa_h = 0,95$ pour les paramètres standard et $\kappa_h = 0,79$ pour les paramètres optimisés) les taux d'erreurs restent comparables et en dessous de 3%. En guise de référence, notons que le critère rejeté plus haut, basé sur le CA, produit entre 4 et 5% d'erreurs avec notamment une augmentation des omissions de pauses.

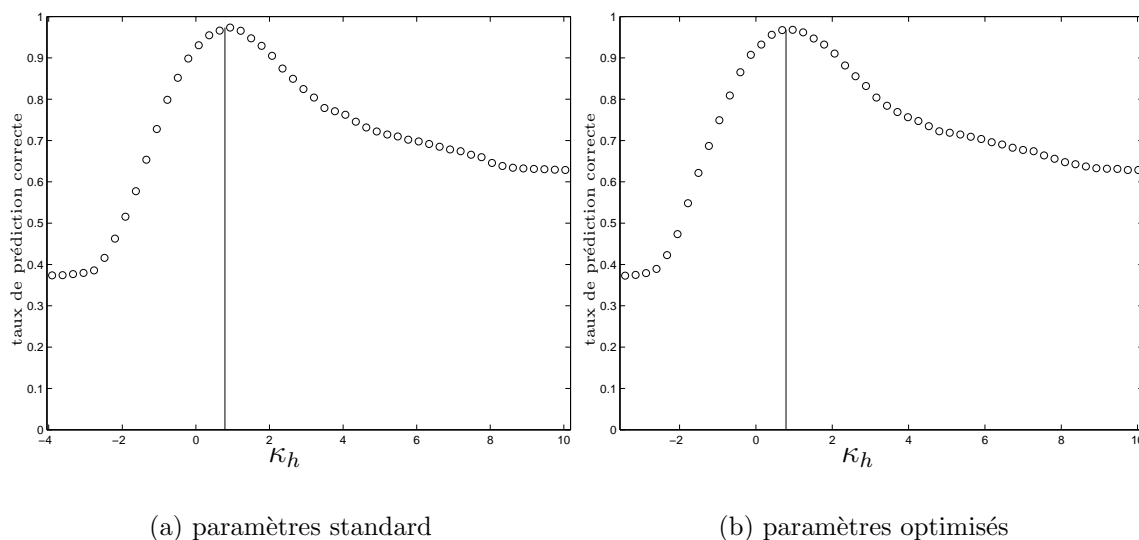


FIG. 2.3: **Taux de prédiction correcte de la génération de pauses.** Ce taux est représenté en fonction du seuil sur le « z-score hypothétique ». À gauche : calcul des z-score avec les paramètres standards ; à droite : avec les paramètres optimisés pour la répartition.

La durée des pauses générées. Emplacement et durée des pauses sont liés car un allongement peut être perçu comme une pause subjective bien qu'il n'y ait pas de pause silencieuse, et une pause silencieuse de faible durée n'est pas toujours facile à identifier (surtout avant des plosives sourdes). Néanmoins, pour la synthèse nous sommes obligés de décider s'il convient d'insérer une pause et de calculer sa durée le cas échéant. Dans le paragraphe précédent, nous avons développé un critère permettant de décider *où* insérer des pauses. Nous allons maintenant voir comment *prédire leurs durées*.

Sachant l'emplacement des pauses, nous connaissons le contenu phonémique de chaque GIPC « frontalier » ce qui permet de calculer sa *durée totale* à l'aide du CA. Il reste à déterminer la répartition de cette durée entre phonèmes et pause silencieuse. En suivant encore les travaux de Barbosa [1994], nous supposons que l'étirement

paramètres standard					
		correct		erreur	
prédiction	pause	non-pause	pause	non-pause	taux d'erreurs
corpus A	601	1003	36	14	3,0%
corpus B	603	1012	26	13	2,4%
paramètres optimisés					
		correct		erreur	
prédiction	pause	non-pause	pause	non-pause	taux d'erreurs
corpus A	591	1016	23	24	2,8%
corpus B	594	1015	23	22	2,7%

TAB. 2.2: **Prédiction des pauses silencieuses.** Les seuils retenus sont $\kappa_h = 0,95$ pour les paramètres standard et $\kappa_h = 0,79$ pour les paramètres optimisés.

des phonèmes (donc leur z-score réel) peut être prédit d'après l'étirement qui serait nécessaire pour remplir la durée totale du GIPC (donc avec une durée de pause égale à zéro). Nous déterminons donc pour tous les GIPCs « frontaliers » de notre corpus un z-score réel k_r et un z-score virtuel k_v de sorte que :

$$\Delta = \sum_i e^{\mu_i + k_v \sigma_i} \quad (2.5)$$

$$\Delta - d_p = \sum_i e^{\mu_i + k_r \sigma_i} \quad (2.6)$$

où Δ est la durée totale du GIPC et d_p la durée de la pause silencieuse. On a $k_r \leq k_v$ avec $k_r = k_v$ si et seulement si $d_p = 0$. À la différence des calculs concernant le z-score hypothétique plus haut, nous pouvons maintenant utiliser les durées observées puisque la question du contenu du GIPC ne se pose plus. La figure 2.4 présente les mesures des z-scores réels versus les z-scores virtuels. Les points qui nous intéressent ici, sont ceux en dessous de la diagonale. Afin de pouvoir prédire k_r nous approximations l'ensemble de ces points par

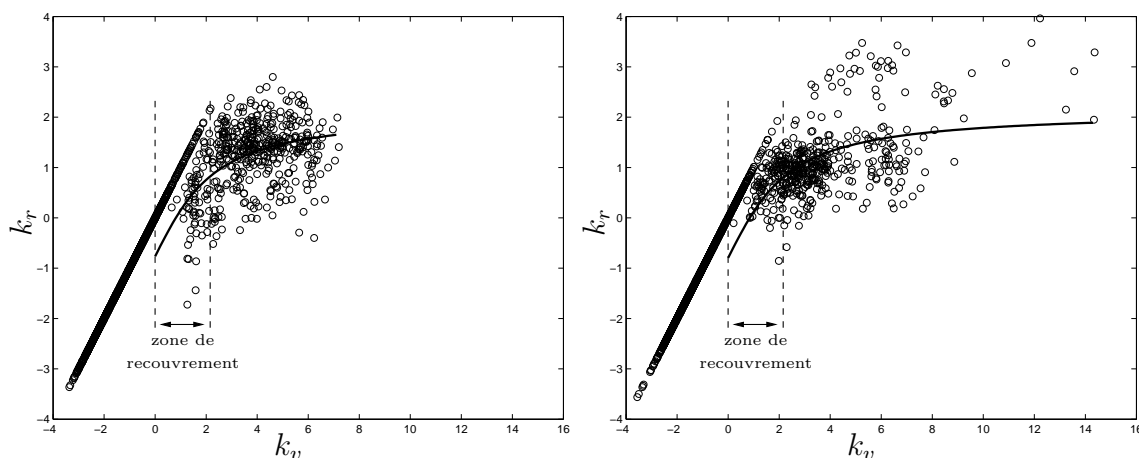
$$k_r(k_v) = \alpha \cdot \arctan(k_v/\alpha) + \beta \quad (2.7)$$

avec

$$(\alpha; \beta) = (1, 83; -0, 76) \quad \text{pour les paramètres standards, et}$$

$$(\alpha; \beta) = (1, 86; -0, 80) \quad \text{pour les paramètres optimisés.}$$

Cette fonction a été choisie pour deux propriétés : la saturation de k_r pour les valeurs grandes de k_v et sa tangente parallèle à la diagonale pour $k_v = 0$. Notons que β doit être inférieur à zéro.



(a) paramètres standard

(b) paramètres optimisés

FIG. 2.4: **z-score réel vs. z-score virtuel.** La zone de recouvrement est celle où pour un même z-score virtuel k_v le corpus présente des GIPCs avec pause et d'autres sans. C'est dans cette zone que se situe le seuil κ_h bien qu'il porte sur le z-score « hypothétique » k_h . C'est également dans cette zone où le rejet d'une pause avec une trop petite durée peut intervenir.

Le z-score réel prédit d'après le z-score virtuel permet de calculer les durées des phonèmes. La durée de pause est alors selon l'équation (2.6) :

$$d_p = \Delta - \sum_i e^{\mu_i + k_r(k_v)\sigma_i} \quad (2.8)$$

Durée minimale de pause. En fonction des paramètres α et β , il est possible que certaines rares pauses ainsi générées aient des durées très faibles. Il peut donc être souhaitable d'introduire un seuil sur d_p . En général, nous le fixons à 30 ms. Si d_p est inférieur à ce seuil, l'insertion d'une pause silencieuse est annulée avant de procéder au GIPC suivant.

Les erreurs de prédiction des durées de pause (dont l'insertion a été correctement prédite) sont rassemblées dans le tableau 2.3. Il s'avère que l'optimisation nuit légèrement aux résultats. Le lecteur attentif pourrait s'étonner du fait que le nombre de pauses cité dans le tableau 2.3 n'égale pas le nombre de pauses correctes du tableau 2.2. Il y a deux raisons à cela. D'abord, nous avons calculé les erreurs de durées uniquement pour les pauses appartenant à des GIPCs de contenu identique dans l'original et dans la prédiction. Il y a un petit nombre de cas où une fausse prédiction de l'existence d'une pause dans le GIPC précédent invalide cette condition pour une pause correctement

prédite. Puis, nous avons exclu toutes les pauses après le mot introductif « résous » qui précède toutes les phrases du corpus. Ce mot est toujours suivi d'une pause relativement longue. Malheureusement, il s'agit là de l'unique occurrence du phonème [u] dans le corpus. Les paramètres de répartition ne peuvent donc pas être optimisés et la relation entre k_r et k_v sort du cadre de l'approximation donnée en équation (2.7). En effet, le [u] forme à lui seul le dernier GIPC de « résous ». La répartition est donc ici toujours celle entre pause silencieuse et durée de phonème (équation (2.8)), mais jamais celle entre phonèmes gérée par l'équation (2.3). Pour ce corpus, nous pouvons donc fixer manuellement ses paramètres de sorte qu'ils aillent au mieux avec l'équation (2.7) : $\mu_u = 4.2$ et $\sigma_u = 0,3$. Il me semblait donc préférable d'exclure ce cas bien particulier du calcul d'erreur des durées de pauses.

	standard		optimisé	
	#pauses	RMSE	#pauses	RMSE
corpus A	428	43,8	422	46,5
corpus B	428	43,3	424	48,0

TAB. 2.3: **Erreurs de prédiction : pauses silencieuses.**

Dans le tableau 2.4, nous donnons les erreurs de répartition complète. Tous les GIPCs sont pris en compte, à part ceux de « résous », le dernier d'une phrase et ceux qui n'ont pas le même contenu à cause d'une erreur dans la prédiction de pauses. Je rappelle que l'optimisation des paramètres et l'approximation de la relation entre les scores réels et virtuels sont réalisées à partir du corpus A seulement. Les résultats obtenus pour le corpus B montrent qu'il ne s'agit pas d'une procédure qui optimise les paramètres à outrance et au détriment de la généralisation.

Les erreurs de répartition constituent une référence pour les erreurs de prédiction complète (c.-à.-d. basée sur des *CA prédits*). Ces erreurs sont, par ailleurs, pour la version optimisée, inférieures au seuil de la JND⁷ de 20 ms cité par Bartkova et Sorin [1987].

	#GIPCs	standard	#GIPCs	optimisé
corpus A	2554	21,8	2558	17,4
corpus B	2574	20,7	2559	17,3

TAB. 2.4: **Erreurs de répartition – GIPCs avec et sans pauses.**

⁷Just Noticeable Difference

2.1.1.3 Un petit test de préférence

Nous avons entrepris un petit test de préférence pour savoir si les erreurs objectives des différents modèles de répartition se reflètent dans la perception de stimuli synthétiques. À cette fin, nous avons sélectionné 7 formules que nous avons synthétisées, à partir de diphtongues, en quatre versions différentes :

- répartition naïve ;
- répartition standard ;
- répartition optimisée ;
- avec les durées segmentales de l'original.

Aux GIPCs avant les pauses nous avons toujours attribué les durées segmentales de l'original, car ils sont perceptivement très proéminents et nous voulions éviter que les jugements se basent principalement sur ces GIPCs. 7 sujets ont écouté des paires de stimuli et ont exprimé leur préférence à travers un interface graphique simple. Chaque sujet a ainsi jugé toutes les paires de versions possibles dans les deux ordres pour toutes les 7 formules, soit un total de $7 \times 6 \times 2 = 84$ jugements.

La quasi-unanimité des sujets a rapporté de très grandes difficultés à distinguer les stimuli. Le tableau 2.5 montre les résultats classés selon les paires de versions, les nombres dans les cases étant les votes en préférence de la version en haut en face de la version à gauche. Pour référence : sous l'hypothèse de réponses au hasard, le score attendu dans chaque case serait 49 (7 phrases \times 7 sujets \times 2 répétitions/2). Un test de χ^2 sur les scores totaux par versions (la ligne des sommes) donne $\chi^2 = 18,0$, ce qui serait, avec trois degrés de liberté, largement significatif, mais ce qui est à interpréter avec beaucoup de prudence, puisque les votes ne sont évidemment pas tous indépendants. Retenons cependant que la tendance des résultats va exactement dans le même sens que les erreurs objectives.

Notons finalement un indice de la grande difficulté qu'ont éprouvée les sujets : le second stimulus des paires a été choisi 359 fois contre seulement 229 pour le premier malgré le parfait équilibre de la présentation, ce qui semble indiquer une stratégie « dans le doute pour le second » chez plusieurs sujets.

Ma conclusion de ces résultats est que l'ajustement fin des durées segmentales à l'intérieur d'une durée de GIPC fixée est un problème secondaire (en vue de grand nombre de « confusion » même entre original et version naïve). Il est probablement en partie masqué par la qualité de voix synthétique qui peut même dans certains cas pénaliser la version « originale » si celle-ci doit avoir recours à des diphtongues plus fortement déformés qu'une autre version. Il va cependant de soi, que la tendance de préférence qui se manifeste dans ce test n'incite pas à écarter l'utilisation de la répartition optimisée – bien au contraire.

	naïf	standard	optimisé	original
naïf	–	48	61	63
standard	50	–	56	64
optimisé	37	42	–	57
original	35	34	41	–
Σ	122	124	158	184

TAB. 2.5: **Répartition – Evaluation.** Le tableau rassemble le nombre de votes en faveur de la version indiquée en haut en comparaison avec celle notée à gauche. La dernière ligne contient les scores totaux pour chaque version.

2.1.1.4 Résumé

Ce chapitre a surtout été consacré à la composante micro-rythmique de notre modèle. Un certain nombre de détails ont dû être présentés et il est certainement nécessaire de résumer les étapes parcourues :

1. GIPC : de voyelle à voyelle, une pause silencieuse étant attribuée au GIPC précédent.
2. Coefficient d’allongement : il mesure la vitesse locale d’élocution et permet le calcul de la durée totale d’un GIPC (pause incluse).
3. Répartition : la durée du GIPC qui peut être calculée grâce au CA, est distribuée aux phonèmes selon le modèle z-score.
4. Optimisation : elle est optionnelle mais permet de diminuer l’erreur de répartition de manière significative.
5. Insertion de pause : partant du CA nous calculons la durée d’un GIPC sous l’hypothèse qu’il ne contient pas de pause. Si le z-score associé dépasse un seuil, nous insérons une pause.
6. Durée de pause : nous déterminons un z-score virtuel comme si la durée de la pause était nulle.⁸ Nous avons ensuite établi une relation empirique entre celui-ci et le z-score réel qui permet le calcul des durées phonémiques ainsi que de la pause.
7. Seuil : l’insertion d’une pause trop courte est annulée.

Dans la suite de ce manuscrit, nous nous concentrons uniquement sur la composante *macro-rythmique*, matérialisée par le coefficient d’allongement.

⁸Ceci n’est pas la même chose qu’une pause absente.

Une réponse à van Santen. Dans le brève chapitre de l'état de l'art qui a été consacré à la prédiction des durées segmentales (cf. page 53), j'ai promis de revenir aux critiques de van Santen [1997] qui consistent – rappelons-le – d'une part en refutant l'indépendance des durées syllabiques de leur structures segmentales et d'autre part en clamant l'impossibilité à déterminer les durées segmentales uniquement à partir des durées syllabiques. Nous tenons compte de la première objection en ne prédisant pas directement les durées syllabiques mais leur coefficient d'allongement. Celui-ci est la tentative – certes à parfaire – d'abstraire du contenu phonémique.

La deuxième objection est plus importante : elle met en question la base de l'approche que nous avons adoptée. Je suis convaincu que les résultats de van Santen relatifs à l'anglais américain et au mandarin s'appliquent également au français : la répartition des durées segmentales ne dépend certainement pas uniquement de la durée de la syllabe et des propriétés intrinsèques des segments. Dans ce sens, notre modèle (et ceux de Campbell [1992], Barbosa [1994] etc.) est faux. Il propose une factorisation qui n'est pas parfaite. Elle pourrait être améliorée en incluant dans la description macro-rythmique d'autres paramètres que le coefficient d'allongement, qui généreraient, par exemple, les comportements différents de l'attaque ou de la coda. L'idée serait alors que ces paramètres supplémentaires seraient des indices d'ordre macro-prosodique : un allongement particulièrement fort d'une partie de la syllabe pourrait ainsi être liée à la structure phonologique.

Soulignons finalement l'avantage de la factorisation : elle tente d'optimiser la prédiction des unités macro-rythmiques, donc les durées entre deux centres de perception. Ceci est payé au prix d'une perte de précision à l'intérieur des GIPCs. En vue des difficultés rapportées par les sujets du test ci-dessus à distinguer même les stimuli en « répartition naïve » de ceux qui ont été resynthétisés, il me semble justifiable d'accepter pour le moment l'erreur de factorisation.

2.1.2 La composante mélodique

Suivant la même idée que dans le chapitre précédent pour le rythme, nous allons maintenant préparer le terrain mélodique, afin de pouvoir nous concentrer par la suite sur ses propriétés supra-segmentales.

Discrétisation de la fréquence fondamentale. Nous approximations f_0 par trois valeurs par voyelle ce qui est, comme nous l'avons vu dans l'état de l'art, une approche assez classique (cf. page 28). Plus précisément, le cours de f_0 en échelle logarithmique à l'intérieur d'une voyelle est approximé par un polynôme de degré deux, et nous retenons les valeurs de cette approximation à 10%, 50% et 90% de l'intervalle occupé par la voyelle. La fréquence fondamentale « porteuse » des consonnes voisées sera restituée par interpolation linéaire.

Traitement implicite de la micro-mélodie. L'approximation parabolique induit évidemment des erreurs. La figure 2.5 montre ceci pour une phrase typique : la courbe continue est celle du f_0 observé ; les petits cercles représentent celle que l'on obtient à partir de l'approximation quadratique. L'erreur dans la partie encadrée est particulièrement évidente. Il s'agit là de la baisse de f_0 caractéristique des occlusives voisées (ici en occurrence le [b]). Or, dans le modèle de génération des paramètres prosodiques, nous souhaitons pouvoir abstraire du contenu segmental. C'est exactement ce que nous avons fait pour la composante rythmique : établir le coefficient d'allongement afin de représenter le macro-rythme et laisser au modèle de répartition les détails micro-rythmiques (durées segmentales et pauses).

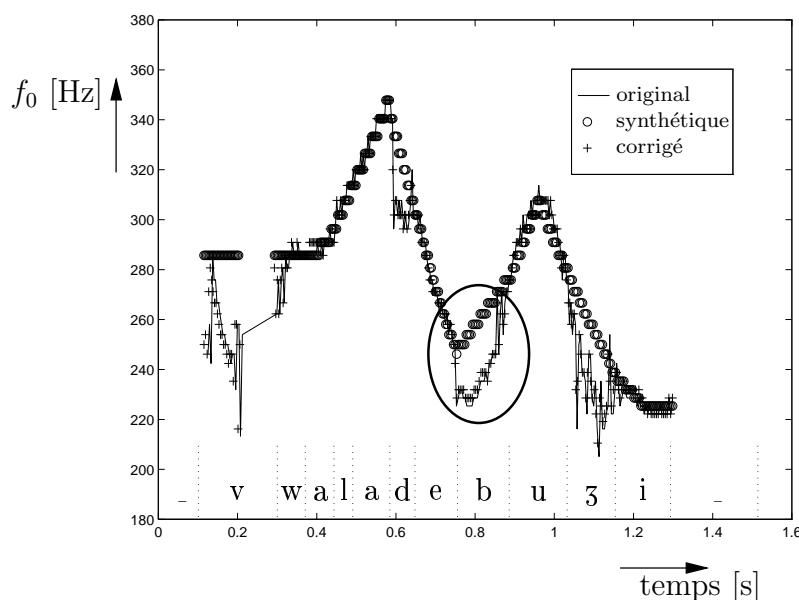


FIG. 2.5: **Microprosodie.** La courbe observée de f_0 (trait continu) dévie d'une manière typique de l'approximation quadratique (cercles) basée sur les noyaux vocaliques : au cours des fricatives/occlusives voisées (p.e. le [b] encerclé), on observe des creux. On peut tenir compte de cette composante microprosodique en la sauvegardant ensemble avec les diphones.

Pour la fréquence fondamentale, nos trois valeurs extraites de la voyelle joueront le rôle du coefficient d'allongement. Les effets de micro-mélodie sont laissés à la charge de la synthèse par concaténation de diphones. En effet, au moment de la création du dictionnaire de diphones, la fréquence fondamentale est approximée de la même manière que décrit ci-dessus. Pour chaque diphone nous sauvegardons donc les paramètres de l'approximation ainsi que son erreur détaillée par rapport au f_0 mesuré. Au moment de la synthèse, quand le diphone le plus adapté a été sélectionné, nous corrigeons, comme proposé initialement par Monaghan [1992], la courbe parabolique par la courbe

de l'erreur d'approximation. Nous considérons cette erreur comme composante micro-mélodique propre au diphone. Le résultat de cette correction est symbolisé par les croix de la figure 2.5. Cette figure ne fait office que d'illustration, car les diphones utilisés pour la resynthèse sont extraits de la phrase même ce qui explique la parfaite égalité entre le f_0 corrigé et l'observation.

2.1.3 Contours multiparamétriques

Nous voilà prêts à rassembler les paramètres prosodiques avec lesquels le modèle de génération, le SFC, travaille. Il s'agit simplement des trois valeurs de f_0 par GIPC et du coefficient d'allongement. L'ensemble de ces paramètres pour tous les GIPCs d'une phrase constitue ce que nous appelons son contour prosodique et exprime la représentation du SFC de la prosodie.

Graphiquement, nous séparons la partie f_0 de la partie temporelle comme dans la figure 2.6. Les valeurs du contour y sont indiquées par les points. Notons que les trois valeurs de f_0 sont simplement alignées à $-1/3$, 0 et $+1/3$. Pour la représentation des contours, nous nous contenterons par la suite, des lignes liant les points. Il ne s'agit pas là de l'approximation quadratique qui est la base de la synthèse, mais d'une visualisation intuitive des paramètres prosodiques tel qu'ils sont vus par le SFC.

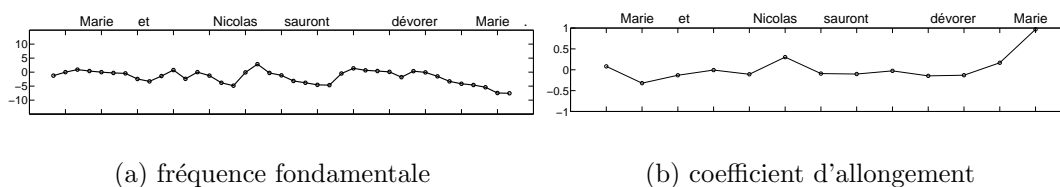


FIG. 2.6: **Contour multiparamétrique.** L'ensemble des deux courbes représente le contour multiparamétrique de la phrase. Les traits sur l'abscisse indiquent les GIPCs. Les mots de la transcription orthographique de la phrase sont alignés avec la fin de leur dernier GIPC.

Si nous sommes amenés par la suite, à calculer des erreurs de prédiction, c.-à.-d. la différence entre un contour observé et un contour prédit, ce sera toujours fait par rapport aux points discrets des contours et non pas par rapport aux contours quasi-continus mesurés ou restitués par interpolation.

Les valeurs de f_0 seront toujours données en demi-tons par rapport à une fréquence de référence proche du registre moyen du locuteur. Le coefficient d'allongement est « interne » multiplié avec un facteur 10 afin d'aboutir à des valeurs de même ordre de grandeur que la fréquence fondamentale. Cette mise à échelle est cependant inversée pour les représentations graphiques ainsi que pour les erreurs de prédiction et devrait donc rester invisible dans le reste de ce manuscrit.

Ancrage segmental. Une caractéristique très importante des contours prosodiques est qu'ils sont par définition ancrés sur la structure segmentale : leur abscisse n'est pas le temps, mais la succession des GIPCs. Nous évitons ainsi les difficultés d'alignement entre prédictions temporelle et mélodique.

Au-delà du rythme et de la mélodie. Les contours multiparamétriques offrent un cadre facilement extensible à d'autres paramètres prosodiques, comme par exemple l'énergie. On appliquerait la même idée : un paramètre par GIPC à la charge du modèle de génération, et un modèle « micro-énergétique » gérant l'ajustement au niveau segmental, par exemple en fonction des énergies intrinsèques des phonèmes.

Dans un cadre plus large, on peut également inclure dans les contours des paramètres prosodiques multimodaux comme les mouvements de la tête (cf. p.e. Yehia et al. [2000]) ou les gestes synchronisés avec l'acte de parole (cf. p.e. Kendon [1994]).

2.2 Motivation

2.2.1 Un modèle issu de l'ICP

Le SFC n'a bien évidemment pas émergé du néant. Il doit un héritage très important aux travaux de Morlec et à travers eux aux propositions d'Aubergé.

2.2.1.1 Un lexique structuré de formes prosodiques globales

Aubergé [1991, 1992] établit à partir d'un corpus soigneusement construit en appliquant la méthode de paires minimales, introduite par Delattre [1969], un lexique de contours prosodiques. Elle propose la notion de rendez-vous structurel entre intonation et syntaxe qui matérialise l'idée que ces deux agents linguistiques collaborent afin de véhiculer le message sous la contrainte d'un alignement souple : unités intonatives et constituants syntaxiques partagent des points d'ancrage.

Aubergé identifie quatre niveaux syntaxiques (phrase, proposition, groupe et sous-groupe) avec lesquels sont associés des contours prosodiques indépendants. Suivant la même intuition que Thorsen [1980]/Grønnum [1992], les contours interagissent par simple superposition afin de produire le contour global de l'énoncé. En analyse, les contours sont obtenus d'une manière hiérarchique. Commencant par le niveau de phrase, des contours moyens sont stockés dans le « lexique » en fonction des attributs qui les caractérisent. L'analyse des niveaux subordonnés est faite de la même manière, après avoir soustrait la contribution des niveaux supérieurs du contour observé.

La table 2.6 rassemble les oppositions mentionnées dans [Aubergé, 1992] qui ont servi à la construction du corpus et qui fournissent ainsi les attributs selon lesquels les contours du lexique sont classifiés. Il s'ajoute à tous les niveaux la longueur des unités en terme de nombre de syllabes. Plusieurs classes de contours peuvent être réunies dans une seule, si leur représentants sont très semblables. Ainsi, cette méthode permet de déterminer quelles caractéristiques linguistiques ont un impact sur la structuration prosodique des énoncés. Ou, autrement dit : quand et comment syntaxe et prosodie respectent leurs rendez-vous.

Précisons que dans la première implémentation de ce modèle, les contours des niveaux non-terminaux sont des simples lignes de déclinaison et que seules les unités en bas de l'échelle hiérarchique accèdent à des contours "détaillés".

Résumons comment ce modèle opère la transition entre les informations linguistiques vers une réalisation prosodiques : à chaque sous-groupe (ou groupe s'il est terminal) est associé un contour du lexique selon les attributs qui le caractérisent. Ces contours sont enchaînés, et ajustés par les lignes de déclinaisons des niveaux supérieurs. Les contours manipulés sont donc :

- associés au niveau hiérarchique ;
- coextensifs aux unités identifiées préalablement ;

- caractérisés par des attributs relativement précis.

L'apport original du modèle est qu'il associe des formes prosodiques globaux directement aux besoins linguistiques et que ces formes sont extraites des observations par le biais de leur impact statistique [Aubergé et Bailly, 1995].

phrase :	proposition :	groupe :	sous-groupe :
– déclarative	– position absolue	– nature	– groupe simple
– positive	– initiale	– nominale	– nature
– négative	– finale	– verbale	– nom
– interrogative	_____	_____	– adjectif
– directe	– liée	– fonction	– position relative
– introduite	– au verbe	– sujet	– pré-
– inversée	– dominant	– objet	– nominale
– impérative	– dominé	_____	– post-
	– au nom	– position absolue	– nominale
	– dominante	– initiale	– groupe complexe
	– dominée	– interne	– position
	– relative	– finale	– d'énumération
	– insérée	_____	– initiale
	– non-liée	– position relative	– interne
	– isolée	– préverbale	– dépendance
	– juxta-posée	– post-verbale	

TAB. 2.6: Exemples d'attributs pris en considération selon Aubergé [1992] pour la construction du lexique de contours prosodiques.

2.2.1.2 Contours dynamiques et coarticulés

Morlec [1997] apporte à ce modèle une implémentation d'une plus grande souplesse qui comporte des aspects dynamiques. Les contours stockés dans un lexique sont remplacés par leur génération coarticulée en fonction des emplacements des points d'ancrage par des réseaux de neurones récurrents associés aux quatre niveaux syntaxiques.

La génération dynamique des contours permet de s'affranchir de la contrainte de les devoir stocker pour toutes les longueurs possibles. De plus, elle rend possible l'extraction de contours de haut niveau qui sont plus complexes que des lignes de déclinaison. Ainsi, les travaux de Morlec sur les attitudes prosodiques ont mis en évidence l'existence de clichés mélodiques (cf. Fónagy et al. [1984]) décrits par un modèle d'expansion du mouvement.

À la différence du modèle d'Aubergé, les contours manipulés ne sont plus coextensifs aux unités : chaque niveau hiérarchique génère un contour qui englobe la phrase entière. Il est censé contenir toutes les manifestations intonatives associées à ce ni-

veau. La caractérisation détaillées des unités chez Aubergé est remplacée par un marquage relativement simple des relations de dépendance grammaticale. Ce marquage est l'héritage de Bailly [1983] qui, à son tour, s'est inspiré de Hays [1964]. La disposition des marqueurs sur un niveau donné permet la génération du contour associé à ce niveau. Le contour global est obtenu – comme chez Aubergé – par superposition des quatre niveaux.

Il est à noter que la souplesse gagnée par la génération coarticulée dynamique des contours est payée par une plus grande opacité de la relation entre les représentations linguistique et prosodique : mis à part le niveau phrase, les contours sont le résultat complexe de l'interaction de toutes les relations de dépendance du niveau. On est donc loin de l'interprétabilité attrayante d'un lexique prosodique.

L'approche de Morlec a été poursuivie par Brichet et Aubergé [2002] pour l'analyse de l'accent d'emphase portant sur le mot ou la syllabe. Retenons également que des tests de « gating » (cf. [Grépillat, 1996]) ont démontré que la forme globale de ces contours est exploitée perceptivement, permettant ainsi une identification précoce, c.-à.-d. avant la réalisation des « événements » les plus saillants des différentes attitudes.

2.2.1.3 Limitation de l'approche hiérarchique

Le point de départ de la présente étude a été la tentative d'appliquer le modèle de Morlec à un type d'énoncé très profondément hiérarchisé : l'oralisation de formules mathématiques. La solution évidente semblait l'extension à six niveaux syntaxiques du modèle de génération (pour traiter le cas de formules les plus complexes). Or, il nous est rapidement apparu insatisfaisant de charger cinq⁹ réseaux de neurones d'apprendre à faire la même chose : indiquer des relations de dépendances (bien qu'à des niveaux hiérarchiques différents).

Le SFC est le fruit de notre volonté de mettre les fonctions communicatives au premier plan et de postuler que leurs reflets prosodiques devraient être semblables indépendamment du niveau hiérarchique sur lequel elles opèrent.

⁹Six niveaux moins le niveau porteur.

2.2.2 Les arbres de performance

La structuration syntaxiquement profonde des formules algébriques nous a amené à examiner de plus près la structure rythmique de leurs oralisations. L'outil principal que nous avons employé sont les arbres de performance.

2.2.2.1 Introduction et définition

Les travaux de Grosjean et al. [1979] et Gee et Grosjean [1983] introduisent les *structures de performances*. Il s'agit d'une représentation graphique de la structure temporelle d'un énoncé : des phrases isolées sont lues à cinq vitesses d'élocution différentes. Un indice de séparation entre chaque paire de mots adjacents est ensuite calculé comme la durée moyenne des pauses silencieuses entre ces mots. Plus l'indice entre deux mots est petit, plus ces mots sont attachés l'un à l'autre. Ces indices permettent de représenter sous forme d'un arbre les groupements rythmiques des énoncés : en commençant par l'indice le plus petit, les mots adjacents sont liés par des nœuds dont la hauteur égale l'indice. Si un mot fait déjà partie d'un nœud, c'est ce nœud qui est lié au mot (ou nœud) adjacent. C'est-à-dire, les mots séparés par les pauses les plus petites forment les nœuds les plus bas dans la hiérarchie. On appelle cet arbre *structure de performance* ou *arbre de performance* (cf. figure 2.7).

Par définition, ces arbres représentent une structure rythmique *moyenne* de plusieurs énoncés d'une même phrase. L'indice de séparation ne peut être calculé que s'il y a une pause silencieuse entre *tous* les mots – d'où la contrainte d'utiliser (aussi) des énoncés en vitesse d'élocution lente. Se basant sur les travaux de Duez [1987], Monnin et Grosjean [1993] étendent l'approche à l'analyse des énoncés uniques : l'indice de séparation n'est plus la pause silencieuse seule (qui peut être zéro), mais on y ajoute la durée de la dernière voyelle avant la frontière de mots. Ceci permet d'obtenir les arbres de performance complets.

Dans le cadre de notre description de la structure temporelle d'un énoncé, il semble logique d'adapter légèrement cette dernière proposition : nous utilisons comme mesure pour la distance entre les mots, le coefficient d'allongement du GIPC « frontalier ». Ce GIPC contient évidemment la dernière voyelle avant la jonction ainsi qu'une éventuelle pause silencieuse. L'utilisation du coefficient d'allongement nous permet de tirer profit de la normalisation qu'il opère par rapport aux durées intrinsèques des phonèmes. La figure 2.7 montre la transition entre un contour de coefficient d'allongement vers l'arbre de performance associé. Les structures de performance contiennent moins d'informations que les contours mais ils permettent un autre regard sur les groupements rythmiques.

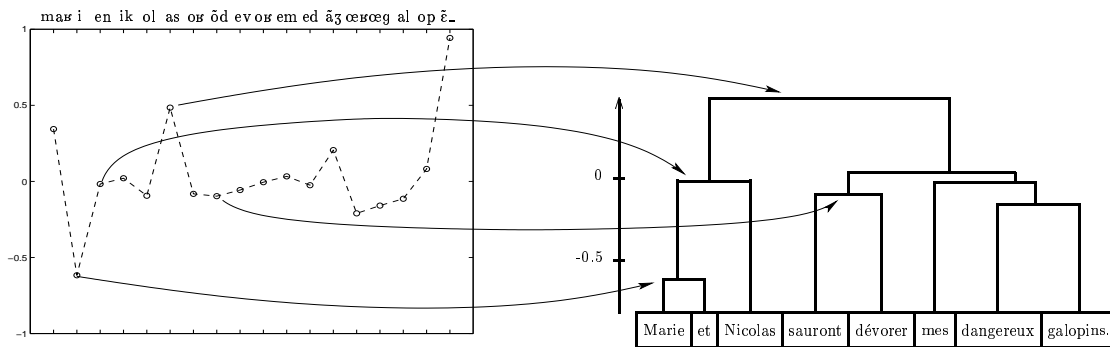


FIG. 2.7: **Contours et arbres de performance.** L'arbre de performance représente la cohésion rythmique des mots : la hauteur d'un nœud liant deux mots correspond au coefficient d'allongement du dernier GIPC du premier mot.

2.2.2.2 Structure syntaxique vs. Structure de performance

Les travaux sur les structures de performance s'inscrivent dans une recherche visant à mettre en évidence la réalité psychologique des structures de surface des phrases, telles que les définit la grammaire générative transformationnelle. Des tâches de natures très différentes ont été proposées. Ainsi, Johnson [1965] mesure la fréquence de non-rappel d'un mot par rapport à la fréquence de rappel du mot précédent et montre une corrélation de cette « erreur de transition » avec l'importance de la coupure syntaxique. C'est-à-dire, plus la coupure est importante, plus les sujets ont tendance à se tromper sur le mot suivant la coupure. Martin [1970] demande à des sujets d'organiser les mots d'une phrase en « groupes naturels » ce qui permet d'établir des structures hiérarchiques qui peuvent être confrontées aux structures syntaxiques. Grosjean et al. [1979] utilisent une méthode très similaire sur les mêmes phrases qui ont servi à établir des structures de performance basées sur les durées de pauses. L'observation essentielle de ces travaux est que les structures obtenues sont largement indépendantes de la tâche expérimentale¹⁰ – peu importe si elle cible la production ou la compréhension langagière. Plus étonnant encore – et renforçant l'idée qu'il s'agit d'une représentation pertinente de la réalité psychologique de la structure des phrases – même pour la langue des signes américaine, on obtient selon Grosjean et al. [1981] des résultats similaires.

En général, les propriétés des structures de performance sont résumées comme suit :

1. Les unités de base sont de longueurs plus ou moins égales en nombre de syllabe.¹¹
2. Les arbres sont hiérarchiques.

¹⁰Grosjean et al. [1979] mesurent un coefficient de corrélation de $r = 0,92$ entre les indices à partir des pauses et à partir de la segmentation.

¹¹Par unités de base on entend les groupes de mots formant des nœuds de petite hauteur (en dessous d'un seuil).

3. Ils sont relativement symétriques.

Vu la méthode de construction des arbres, la propriété 2. ne peut guère surprendre. Puis, 1. et 3. sont les deux revers d'une même médaille : si le regroupement en nœuds a tendance à être équilibré en nombre de syllabe *sur tous les niveaux*, il en découle que les unités de base ont des tailles comparables (1.) et que la coupure principale se situe vers le milieu de la phrase (3.). Cette propriété de symétrie ou d'équilibrage syllabique des arbres de performance contrasterait, selon certains auteurs¹², avec les structures syntaxiques souvent très asymétriques, et montrerait la nécessité d'introduire une hiérarchie d'unités prosodiques (p.e. syntagme phonologique, syntagme intonational).

De nombreux auteurs (cf. Zellner [1994]) ont essayé de trouver des algorithmes qui prédisent la structure de performance. Grosjean et al. [1979] utilisent la complexité syntaxique (hauteur dans l'arbre syntaxique) de la frontière entre deux mots, combinée de manière *ad hoc* avec un principe de bisection qui simule la tendance à générer des pauses plutôt au milieu de (sous-)phrases. Gee et Grosjean [1983] comparent cet algorithme avec celui de Cooper et Paccia-Cooper [1980] et proposent par la suite un nouvel algorithme qui emploie comme unité d'analyse les « syntagmes phonologiques », qui forment, par regroupement en « syntagmes intentionnels », des arbres à partir desquels se fait la prédiction de la structure de performance.

Notre but premier n'est pas la prédiction des arbres de performance, mais la prédiction de la structure temporelle de l'énoncé entier à partir de l'information qu'il véhicule. Les arbres de performance ne seront que le reflet de cette structure plus fine. De notre point de vue, il convient de ne pas porter *trop* d'importance à ces niveaux hiérarchiques et aux unités qui les composent. Au lieu d'essayer de déterminer par un jeu de règles ces unités dites prosodiques, nous préférons qu'elles émergent de la prédiction du modèle. Néanmoins, les observations sur les structures de performances nous sont utiles afin de déterminer quelle information semble importante à l'entrée du modèle :

1. les groupements syntaxiques, puisque c'est aux frontières de ces groupes que l'on trouve les coupures prosodiques, et
2. la taille de ces groupes en nombre de syllabes, car la plupart des « non-congruences » entre syntaxe et prosodie semble liées à des groupes syntaxiques de tailles asymétriques.

La contradiction n'est qu'apparente entre le premier point et l'affirmation citée plus haut que les structures de performances diffèrent de la structure syntaxique : ce que l'on observe est un décalage de hauteur pour *certaines* nœuds. L'étroite relation entre les structures n'est jamais mise en question (pour un exemple particulièrement convaincant cf. figure 2.8) – sans une telle relation, les structures de performance n'auraient pas sollicité autant d'intérêt. Par ailleurs, les syntagmes prosodiques proposés ne se définissent pas sans rapport à la syntaxe.

¹²cf. par exemple de Tournemire [1998].

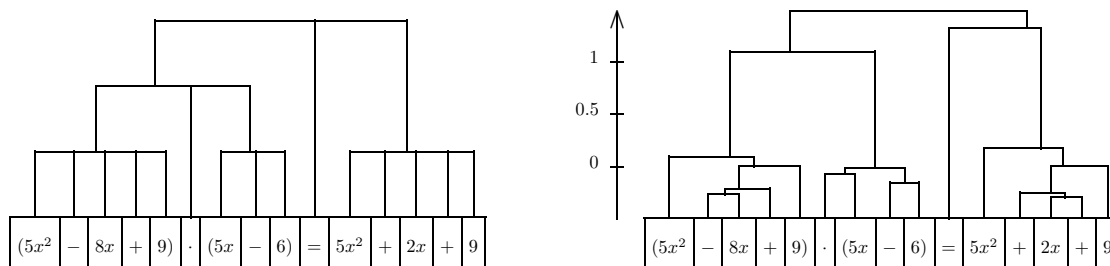


FIG. 2.8: Structure syntaxique vs. structure de performance.

2.2.2.3 Attachement

Jusqu'ici, nous avons regardé les structures de performance sous l'angle des groupements hiérarchiques. Dans ce paragraphe, nous allons examiner l'attachement de certains mots. Nous allons voir l'importance des poids syllabiques des groupes impliqués.

La figure 2.9 montre les structures de performance de deux formules mathématiques. Les deux contiennent un opérateur qui lie deux groupes de taille fortement asymétrique (traits épais dans les structures syntaxiques à gauche). Rythmiquement, l'opérateur est dans les deux cas attaché au groupe plus petit (à gauche dans l'exemple en haut et à droite dans l'exemple en dessous). Ce comportement suit une tendance générale que l'on peut observer dans le corpus de formules mathématiques. Il convient de souligner qu'il s'agit ici d'une caractéristique de notre locuteur pour ce type d'énoncés. Un autre locuteur pourrait *a priori* employer d'autres stratégies.

Afin de préciser le rapport entre attachement et poids syllabique, nous définissons comme l'attachement A d'un mot la différence entre le coefficient d'allongement à sa droite et celui à sa gauche. Un attachement à gauche correspond donc à $A > 0$ et un attachement à droite à $A < 0$. Les opérateurs mathématiques binaires ($=, +, -, \dots$) lient deux groupes : « G op D ». Nous appelons la différence de nombre de syllabes entre D et G « balance syllabique » B . Dans le premier exemple de la figure 2.9, nous avons pour le premier « $+$ » comme opérande gauche (G) : « neuf », et comme opérande droite (D) : « cinq X moins neuf sur sept X moins six » et donc une balance syllabique de $B = 9 - 1 = 8$. L'attachement mesuré est de $A = 1,49 - 0,55 = 0,94$. Pour le second exemple, on obtient $B = -6$ et $A = -1,25$. La figure 2.10 montre les attachements de tous les opérateurs du corpus avec au moins une opérande de plus de deux syllabes en fonction de la balance syllabique. En guise de repère, les points qui correspondent aux exemples de la figure 2.9 sont indiqués.

On peut regretter le manque de données pour $B > 10$ – il est dû aux limitations du corpus (cf. page 109). Néanmoins, la droite de régression $A = 0,063 \cdot B - 0,348$ indique une légère préférence de l'attachement à droite : $A = -0,348$ pour une balance

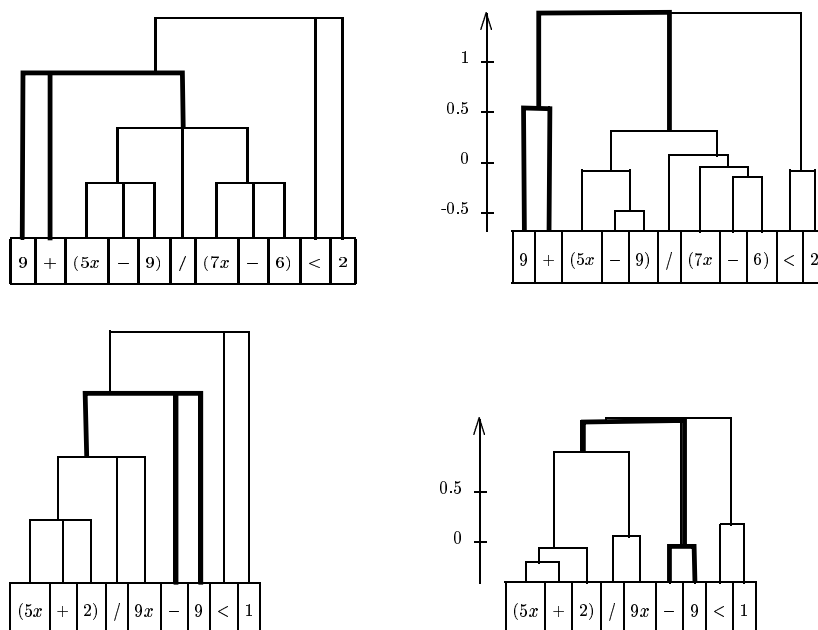


FIG. 2.9: L'attachement d'un opérateur mathématique. À gauche : structures syntaxiques ; à droite : arbres de performance. Les opérateurs mathématiques ont tendance à être attachés aux opérandes plus courts.

syllabique égale à zéro. Le coefficient de corrélation entre A et B est de 0,64 et montre donc que déjà une mesure aussi simple que la différence de syllabes explique 41% de la variance dans l'attachement des opérateurs en question.

Résumé. Outre les groupements de mots opérés par la prosodie à l'intérieur d'un énoncé, les arbres de performance incitent à considérer le sens de l'attachement d'une feuille avec ses voisins. Nous avons montré à quel point cet attachement est déterminé par le poids syllabique des constituants de la phrase – une propriété qu'un modèle doit être capable de reproduire. Le nombre de syllabes des différentes parties mises en relation par la syntaxe (que nous appellerons l'*empan* de la relation) joue donc un rôle important dans l'entrée du modèle que nous proposons.

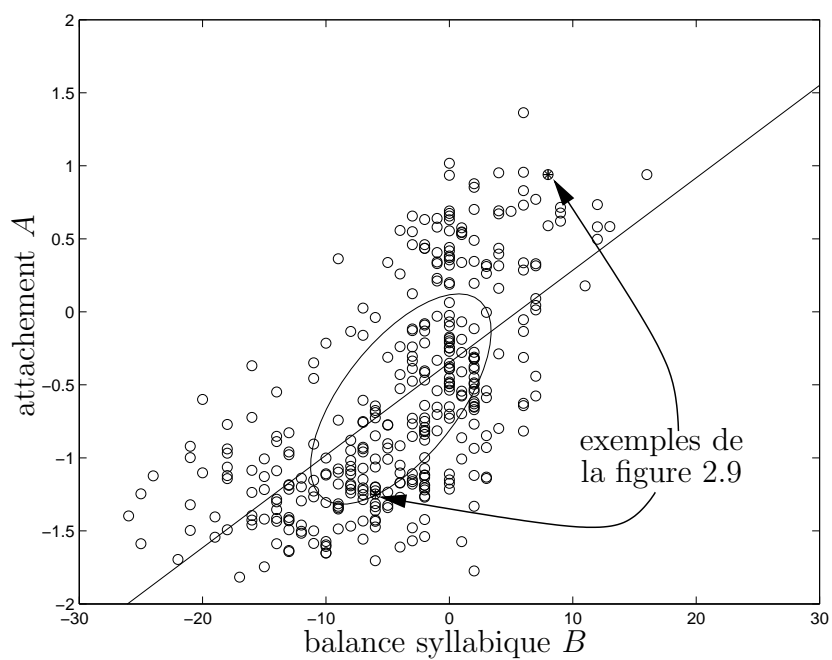


FIG. 2.10: **Attachement des opérateurs vs. balance syllabique.** Nous avons ajouté à la présentation l'ellipse de dispersion et la droite de régression.

2.3 Description du modèle

Au début du chapitre précédent nous avons rappelé les points de départ du SFC : la proposition d'Aubergé [1991] pour les fondements théoriques, et pour l'implémentation, le travail de Morlec [1997]. Il est temps de présenter le SFC.

Vue d'ensemble. Le SFC pose au premier plan des fonctions de haut niveau. Nous allons présenter bientôt, celles qui ont fait l'objet de cette étude. Pour le moment, l'important à retenir est que nous supposons que la prosodie d'un énoncé est le reflet plus ou moins direct de **fonctions communicatives** que le locuteur applique afin de véhiculer son message. Nous supposons que ces fonctions sont en nombre relativement restreint, mais qu'une même fonction peut être appliquée à plusieurs reprises au sein d'un même énoncé. Nous appellerons donc une telle application une **instance** de la fonction. Une instance de fonction n'est pas uniquement caractérisée par la fonction à laquelle elle est associée mais également par sa **portée**, c.-à.-d. la partie de l'énoncé à laquelle elle se réfère ou les parties qu'elle met en relation.

L'hypothèse forte du SFC est que la manifestation prosodique d'une instance de fonction est un contour multiparamétrique coextensif à sa portée. Nous appelons ces contours associés à l'action d'une fonction **contours fonctionnels**¹³ et nous parlons de l'**empan** du contour pour référer à la suite de syllabes pour lesquelles le contour est défini. En bref : l'empan d'un contour fonctionnel correspond à la portée d'une instance de fonction.

Nous supposons que les contours fonctionnels associés à une *même* fonction forment une **famille de contours** qui est morphologiquement cohérente. Le SFC implémente cette hypothèse en associant à chaque fonction un **générateur de contour** qui a la charge d'engendrer l'ensemble des contours associés à une fonction. Il doit pouvoir faire ceci en tenant compte des besoins d'une instance de fonction, c.-à.-d. notamment sa portée. L'expérience avec le modèle de Morlec ainsi que les observations décrites dans le chapitre précédent par rapport aux arbres de performance nous ont amené à restreindre l'information disponible aux générateurs à celle qui est uniquement d'ordre phonotactique (nombres de syllabes relatifs à la portée de l'instance).

La dernière hypothèse centrale du SFC est que les instances de fonction interagissent par **superposition** additive des contours fonctionnels afin de produire le **contour global** tel qu'il a été décrit dans le chapitre 2.1.3.

L'ensemble de ces hypothèses se résume par le tableau suivant qui met en évidence l'association directe entre intentions communicatives et implémentation prosodique que le SFC propose.

¹³Ou quelquefois : contours *localisés* si nous souhaitons souligner la distinction par rapport au contour global de l'énoncé.

fonction	\implies	générateur
instance	\implies	contour fonctionnel
portée	\implies	empan
interaction	\implies	superposition

Une dernière remarque avant de détailler les différents éléments du modèle : nous avons exposé ci-dessus les « hypothèses » du SFC. Le but de la présente étude n'est pas de prétendre qu'elles décrivent la réalité cognitive de la génération de prosodie. Il s'agit plutôt de prendre ces hypothèses, qui sont – à mon avis – simples et séduisantes, au sérieux et de les confronter aux observations empiriques.

2.3.1 Fonctions et leurs portées

Dans ce paragraphe nous présentons brièvement les fonctions communicatives qui ont fait l'objet de cette étude. Le SFC nécessite évidemment que nous spécifions également leur portée.

Fonctions attitudinales. Ces fonctions ont une portée qui comprend la phrase entière. Au sein du SFC, elles jouent le même rôle que dans le modèle de Morlec [1997]. Les contours fonctionnels associés peuvent être considérés comme contours « porteurs ». Nous avons appliqué le SFC au corpus enregistré par Morlec afin de vérifier la compatibilité des résultats (cf. page 143). Les attitudes représentées dans le corpus sont : déclaration (DC), question (QS), doute-incrédulité (DI), évidence (EV), exclamation (EX) et ironie de soupçon (SC).

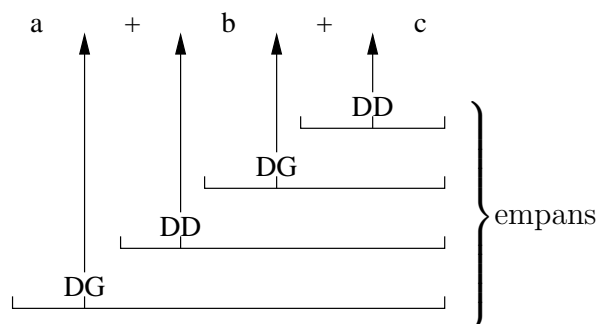
Fonctions hiérarchisantes. Un des objectifs principaux de cette thèse est d'explorer les limites de la capacité de la prosodie à véhiculer une organisation hiérarchique très profonde d'un énoncé. Le corpus d'oralisation de formules mathématiques a été enregistré dans ce but. Nous viendrons à sa construction en détail dans la troisième partie de ce manuscrit (cf. page 109).

Nous distinguons deux fonctions exprimant la relation de dépendance entre deux parties adjacentes d'un énoncé :

1. La dépendance gauche (DG) signale que, de deux groupes de mots celui de gauche est dominé par celui de droite. L'exemple le plus courant est la relation entre le groupe verbal et le sujet. Nous analysons donc : (La belle fille) DG (lit un livre rouge). Mais également : (belle) DG (fille).
2. La dépendance droite (DD) réfère au cas inverse où le groupe droit est dominé par celui de gauche, comme c'est le cas du complément d'objet direct par rapport au verbe : (lit) DD (un livre rouge), et – sans surprise : (livre) DD (rouge).

Pour le corpus des formules mathématiques nous nous sommes inspiré de l'analyse des phrases SVO en laissant la place du verbe aux opérateurs mathématiques binaires

et en postulant un branchement à droite en présence de plusieurs opérateurs du même niveau hiérarchique. Cela donne p.e. :



D'un point de vue technique, une différence importante entre les fonctions attitudinales présentées avant et les fonction hiérarchisantes est que ces dernières lient *deux* groupes de mots. L'empan associé est ancré aux frontières de ces groupes.

Fonctions de segmentation. La morphologie caractéristique des contours prosodiques aux alentours des articles et pronoms nous a amené à postuler l'existence d'une fonction de segmentation, appelée X, dont la portée serait uniquement l'article (ou pronom) et le mot suivant : (1e) X (livre). Elle servirait à mettre en relief ce dernier par rapport au premier. Dans le chapitre 3.3 nous verrons qu'il est possible de postuler la manifestation de cette même fonction pour la segmentation entre auxiliaire de mode et infinitif.

Fonctions d'emphase et de focalisation. L'analyse des réalisations prosodiques de la mise en valeur particulière de certaines parties d'un énoncé n'est pas l'objet du présent travail. Cependant, la façon dont elle pourrait être incluse semble relativement claire : la portée d'une telle fonction serait simplement la syllabe, le mot ou le groupe de mots mis en valeur (cf. Bricet et Aubergé [2002]). Il est également possible d'étendre sa portée aux parties voisines.

2.3.2 Générateurs et familles de contours

Répetons que le SFC associe à chaque instance d'une fonction (para-)linguistique un contour multiparamétrique dont l'empan correspond à la portée de la fonction. L'hypothèse est que tous ces contours fonctionnels appartenant à une même fonction forment une famille cohérente. À l'intérieur de ces familles, la morphologie des contours serait uniquement déterminée par des contraintes phonotactiques.

Afin de donner vie au SFC, il faut donc des dispositifs capables d'engendrer ces familles. Nous les appelons *générateurs*. Il y en a un par fonction communicative et ils doivent effectuer le « mapping » entre configurations phonotactiques et contours prosodiques.

Actuellement, ces générateurs sont implémentés en tant que réseaux de neurones. On reviendra sur les détails à la fin du chapitre sur l'apprentissage du modèle (cf. page 102).

2.3.3 Superposition

Le contour prosodique global prédit est obtenu en tant que superposition additive des contours fonctionnels. Il s'agit donc d'une interaction très simple des manifestations prosodiques des différentes fonctions de haut niveau. Elle permet d'appliquer une même fonction à des niveaux hiérarchiques différents et plusieurs fois sur un même niveau. Les portées des instances de fonctions et avec eux les emfans des contours fonctionnels peuvent se chevaucher et s'emboîter.

Par ailleurs, la notion de niveaux hiérarchiques dans l'organisation syntagmatique de l'énoncé devient complètement secondaire. Leurs reflets restent perceptibles dans les contours globaux, mais ils sont le produit de la morphologie des contours en fonction des propriétés phonotactiques des parties associées de l'énoncé (les niveaux plus *hauts* contiennent simplement des emfans plus *longs*).

La figure 2.11 montre un exemple de génération du contour global par superposition des contours fonctionnels. On retrouve le contour prosodique observé de la figure 2.6 en haut des graphiques en tirets. Sur les quatre niveaux de la partie inférieure, sont représentés les contours fonctionnels avec leur point d'ancrage et le nom de la fonction associée. Ils sont générés – pour l'instant : comme par miracle – par quatre générateurs différents (un par fonction). Les deux contours DD sont issus du même générateur – ils appartiennent à la même famille. En superposant additivement les contours fonctionnels on obtient le contour global prédit (traits pleins, en haut des figures).

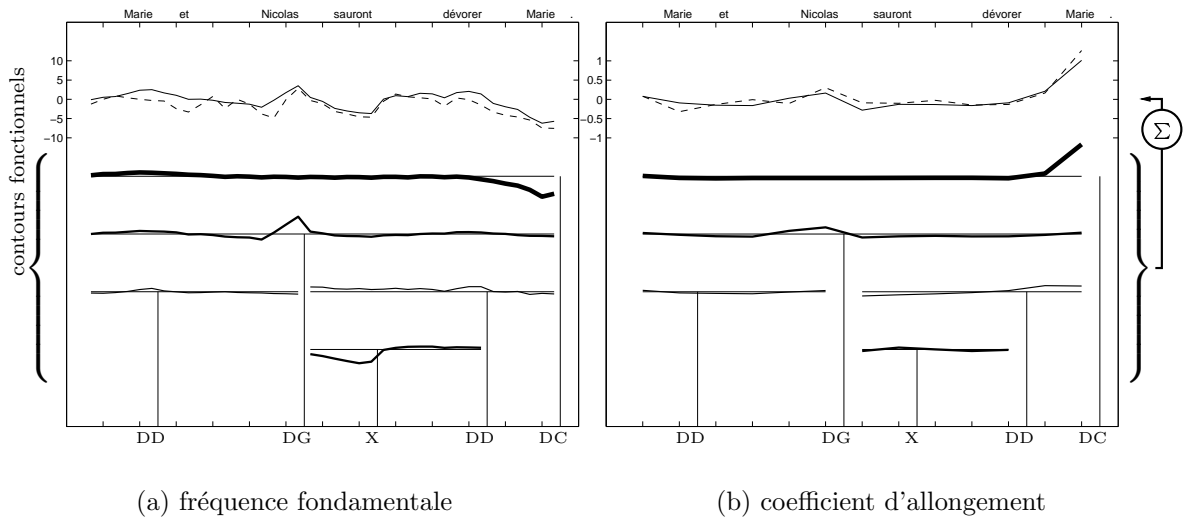


FIG. 2.11: **Exemple de superposition.** Les résultats des prédictions du SFC seront, par la suite, souvent représentés sous cette forme. On y trouve tout en haut la transcription orthographique (les mots étant alignés avec la fin de leur dernier GIPC). La courbe en tiret représente le contour global observé ; celle en trait plein le contour prédit par superposition des contours fonctionnels qui, à leur tour, se trouve dans la partie inférieure des figures.

Les étiquettes des fonctions associées aux contours fonctionnels sont notées tout en bas à l'endroit du point d'ancrage central du contour. Quelques fois DG et DD seront abrégés par D resp. G. L'épaisseur des contours tente – manque de couleurs – d'identifier leurs appartenances aux différentes familles, bien que ce ne soit pas très visible. L'échelle de f_0 sont les demi-tons.

2.4 L'inversion du modèle : l'apprentissage

L'inversion du modèle SFC – c'est-à-dire, l'apprentissage de contours fonctionnels à partir d'un corpus – est une tâche difficile. Traber [1992] a montré que la prédiction de contours complexes à partir d'indices phonologiques et phonotactiques simples est possible avec un réseau de neurones (RN). Pour l'entraînement de ce réseau, il dispose des contours cibles : les observations. Notre modèle nécessite une décomposition des contours observés avant de pouvoir entraîner des générateurs de contours du type RN (ou autre). Or, cette décomposition des contours observés en contours se superposant n'est pas unique. Elle doit être contrainte. Une possibilité est d'imposer une forme particulière aux contours – comme on le fait dans les modèles du type Fujisaki. Une autre est de définir une procédure d'inversion. Morlec [1997] choisit cette voie en proposant un apprentissage hiérarchique partant des contours porteurs et ajoutant successivement les contours portés.

Or, un but de notre modèle est de s'affranchir de la notion de hiérarchie des contours : les contours du même type peuvent être porteurs ou portés. Cependant, l'interprétation que nous attachons aux contours nous fournit la contrainte dont nous avons besoin : chaque contour est associé à une fonction (para-)linguistique et nous supposons que les contours associés à une *même* fonction sont cohérents. Nous garantissons cette cohérence par la méthode de génération : chaque fonction se voit attribuer *un* générateur responsable de la création de tous les contours associés. L'entrée des générateurs est uniquement de nature phonotactique et décrivant l'empan de manière très simple.

Analyse par la synthèse. La figure 2.12 schématise l'algorithme d'inversion du modèle que nous proposons. Au début

- nous avons un corpus conçu en vue de l'étude d'une ou plusieurs fonctions communicatives ;
- nous connaissons (ou avons des hypothèses sur) ses instances et leurs portées ;
- nous disposons des contours prosodiques globaux *observés* et
- nous avons un modèle de synthèse : le SFC.

Ce qui nous fait défaut sont les générateurs de contours. Ou : on pourrait supposer qu'après une initialisation plus ou moins fiable, nous sommes en possession de générateurs, mais les contours fonctionnels qu'ils génèrent ne sont pas entièrement dignes de confiance. Il faut alors entraîner les générateurs afin qu'ils accomplissent leur tâche en accord avec les observations. Or, pour qu'un générateur puisse effectuer la mise en correspondance entre configurations phonotactiques et contours fonctionnels multiparamétriques pour une fonction communicative donnée, il lui faut un ensemble de paires entrée-sortie pour son apprentissage. Toute la difficulté réside – comme nous avons déjà remarqué – dans le fait que les observations ne délivrent pas directement les contours cibles des générateurs, mais uniquement leur superposition. Le cycle d'ana-

lyse par la synthèse résout ce problème en utilisant la différence entre les prédictions préliminaires (obtenues grâce à la superposition des contours fonctionnels préliminaires que nous appelons aussi contours « prototypiques ») et les contours globaux observés afin de corriger les contours fonctionnels. Les contours prototypiques ainsi corrigés sont triés selon leur appartenance aux différentes fonctions et nous fournissons alors les ensembles de sorties cibles nécessaires à l'entraînement des générateurs. L'apprentissage des générateurs constitue un filtre de mise en cohérence des familles de contours cibles et nous pouvons espérer une amélioration de l'état des générateurs. L'erreur de prédiction du SFC est ainsi réduite au prochain cycle. Nous interrompons le cycle dès que cette erreur cesse de diminuer.

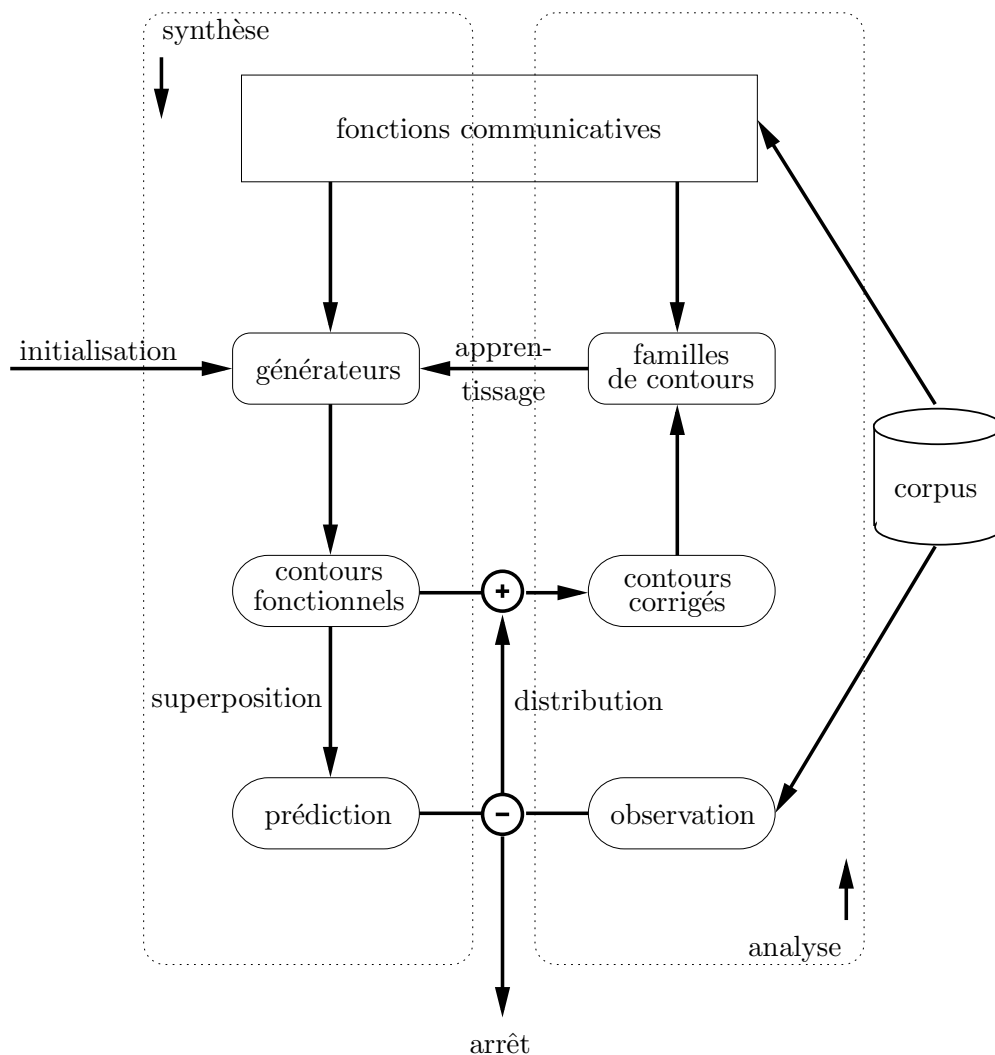


FIG. 2.12: Cycle d'analyse par la synthèse.

2.4.1 Hypothèses et définitions

L'hypothèse de base de notre modèle est que chaque contour encode une *information*. L'empan d'un contour doit être en relation avec la portée de l'information encodée par ce contour. Pour cette raison, nous utilisons des contours de petit empan pour encoder des informations locales et des contours plus longs si l'information relie des parties plus grandes ou plus éloignées.

Si nous disons que les contours sont porteurs d'information, il faut se poser la question de déterminer *comment* cette information est *encodée*. Nous supposons que l'information est contenue dans la *déviaton* d'un contour par rapport à un contour *porteur*. Cette hypothèse est à l'origine de toute approche superpositionnelle. Son rôle est double : d'abord, en décrivant la prosodie par des contours superposés, on suppose que l'être humain peut (dé)coder ces contours. Si ceci semble assez facile en production, il est moins évident en décodage : on doit supposer que l'auditeur a une attente sur le contour porteur qui lui permet de distinguer le contour porté – c'est-à-dire que l'auditeur est capable de percevoir et d'interpréter la déviation du contour *entendu* par rapport au contour *attendu*.

Son deuxième rôle est lié à la notion de la perception des contours en tant que formes. Qu'est-ce que le (proto-)typique d'un contour en tant qu'évolution de paramètres physiques dans le temps ? Quand est-ce que deux contours ont la même forme ou une forme similaire ? Nous rencontrons donc encore la notion de différence entre contours ! Dans cet esprit, on peut définir comme le *typique* d'un contour sa déviation par rapport à un contour plat.¹⁴ Nous considérons donc que ce qui identifie un contour et ce qu'il porte comme information s'expriment par cette différence.

Afin de pouvoir décrire l'algorithme d'inversion, nous avons besoin d'une notation précise. Le prochain paragraphe présente la plupart des variables utilisées afin de ne pas interrompre le déroulement de l'exposé de l'algorithme.

Les variables.

p : index de phrase

k : index de GIPC par rapport à la phrase

l : index de GIPC par rapport au contour localisé

k^p : nombre de GIPC de la phrase p (longueur de la phrase)

l_c^p : nombre de GIPC appartenant au contour c de la phrase p (longueur du contour)

k_c^p : index du premier GIPC appartenant au contour c de la phrase p

$\mathcal{K}^p = \{0; 1; \dots; k^p - 1\}$: ensemble des indices de GIPC de la phrase p

$\mathcal{L}_c^p = \{0; 1; \dots; l_c^p - 1\}$: ensemble des indices de GIPC du contour c de la phrase p

¹⁴On pourrait peut-être remplacer « plat » par « standard ». Mais nous voulons pour l'instant éviter une définition plus compliquée de « standard ».

\mathcal{C}^p : ensemble de contours contribuant à la phrase p

$\mathcal{C}_k^p = \{c \in \mathcal{C}^p : k_c^p \leq k \leq k_c^p + l_c^p\}$: ensemble de contours contribuant au GIPC k de la phrase p

$n^p(k) = \#\mathcal{C}_k^p$: nombre de contours contribuant au GIPC k de la phrase p

$i \in \mathbb{N}$: index d'itération ($i = 0$: initialisation)

Ψ^p : contour observé (phrase p)

$\Phi^{p;i}$: contour prédit (phrase p) après i cycle d'apprentissage

$\psi_c^{p;i}$: cible c (phrase p) pour le cycle i d'apprentissage

$\varphi_c^{p;i}$: contour prototypique c (phrase p) après i cycle d'apprentissage

$F_1^p(k), F_2^p(k), F_3^p(k), C^p(k)$: valeurs de f_0 et du coefficient d'allongement du GIPC k de la phrase p

Notations. Le contour prédit

$$\begin{aligned} \Phi^p : \mathcal{K}^p &\rightarrow \mathbb{R}^4 \\ k &\mapsto \Phi^p(k) = (F_1^p(k), F_2^p(k), F_3^p(k), C^p(k)) \end{aligned}$$

est obtenu à partir des contours localisés

$$\begin{aligned} \varphi_c^p : \mathcal{L}_c^p &\rightarrow \mathbb{R}^4 \\ l &\mapsto \varphi_c^p(l) \end{aligned}$$

par superposition :

$$\Phi^p(k) = \sum_{c \in \mathcal{C}_k^p} \varphi_c^p(k - k_c^p)$$

Nous utilisons la notation : $\Phi = \sum \varphi_c$ s'il n'y a pas de confusion possible.

Remarques. La formulation mathématique de l'algorithme d'inversion a certes l'avantage de permettre au lecteur intéressé une compréhension détaillée. Néanmoins, elle comporte le risque de rebuter ceux qui se contenteront des grandes lignes. Je m'efforcerai à ce que les formules ne deviennent pas une fin en soi mais une illustration des explications. Il n'est donc pas nécessaire pour la compréhension de ce chapitre de suivre en détail les (nombreux) indices.

Le choix des noms de variables a été guidé par le souci de rendre une lecture rapide possible :

- le numéro de GIPC est k ou l , le dernier renvoie aux contours localisés,
- tous les ensembles ont des noms en lettres calligraphiés,
- des lettres grecques sont utilisées pour des contours dans un sens large, c.-à.-d. aussi pour des pondérations etc.,
- ces lettres sont en minuscules si elles réfèrent à des contours localisés – leurs superpositions portent le même nom en majuscules.

2.4.2 La distribution du résidu

L'idée générale. Supposons qu'après i cycles d'itération, nous disposons d'une estimation des contours prototypes $\varphi_c^{p;i}$. En superposant ceux-ci on obtient la prédiction $\Phi^{p;i} = \sum \varphi_c^{p;i}$. Nous définissons l'erreur – ou le *résidu* – de cette prédiction comme :

$$\Delta^{p;i} = \Psi^p - \Phi^{p;i}$$

Nous considérons le résidu comme information contenue dans le contour observé qui n'est pas (encore) représenté par les contours prototypes. La question suivante se pose donc : *Comment distribuer cette information* aux contours participant à la superposition. Il faut obtenir à partir des $\varphi_c^{p;i}$ des nouvelles cibles $\psi_c^{p;i+1}$ qui pourront être utilisées afin d'améliorer l'état d'apprentissage des générateurs. La contrainte principale que nous imposons est que la superposition des contours cibles soit égale au contour observé :

$$\Psi^p = \sum \psi_c^{p;i+1} \quad (2.9)$$

Le but de ce chapitre est de proposer une méthode $\varphi^i, \Delta^i \longrightarrow \psi^{i+1}$ paramétrable et interprétable. Afin d'illustrer l'idée de la *distribution du résidu*, regardons d'abord la solution la plus simple :

$$\psi_c^{p;i+1}(k - k_c^p) = \varphi_c^{p;i}(k - k_c^p) + \frac{\Delta^{p;i}(k)}{n^p(k)} \quad (2.10)$$

Le résidu au GIPC k est réparti de manière homogène sur tous les contours qui contribuent à ce GIPC (cf. figure 2.13). On vérifie facilement que l'équation (2.9) est satisfaite :

$$\begin{aligned} \sum \psi_c^{p;i+1} &= \sum \left(\varphi_c^{p;i} + \frac{\Delta^{p;i}}{n^p} \right) \\ &= \Phi^{p;i} + \Delta^{p;i} \sum \frac{1}{n^p} \\ &= \Phi^{p;i} + \Delta^{p;i} \\ &= \Psi^p \end{aligned}$$

Cette distribution selon l'équation (2.10) est une solution tout à fait valable à notre problème et le lecteur moins friand de détails mathématiques pourra poursuivre la lecture à la page 99...

Cependant, j'espère garder l'un ou l'autre avec moi pour regarder comment améliorer cette prescription. La caractéristique de l'équation (2.10) est de traiter tous les contours de la même manière sans *a priori*. Or, on peut souhaiter exploiter au moins les points suivants.

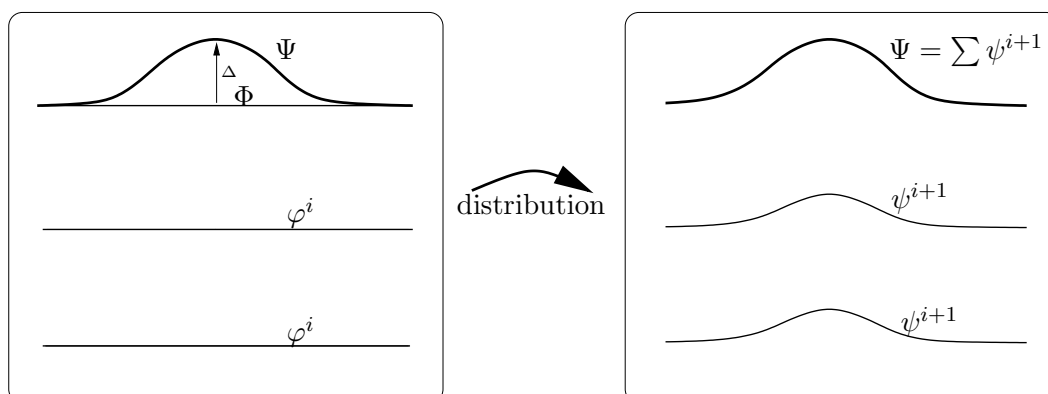


FIG. 2.13: **Distribution du résidu.** L'illustration montre le cas simple où l'on a deux contours contributeurs plats ($\varphi = 0$) à gauche en bas. Leur superposition Φ est alors également égale à zéro (trait fin en haut). Le résidu est partagé équitablement entre ces deux contours de sorte que leur superposition à droite soit égale au contour observé (trait épais à gauche et à droite).

1. Certains endroits des contours peuvent être privilégiés pour accueillir l'information associée au contour. L'expérience montre que ceci est, par exemple, le cas pour le dernier GIPC immédiatement avant la frontière entre deux groupes – lié à l'accentuation démarcative finale du français.
2. Ce qui a déjà été capturé dans les contours prototypes devrait être renforcé s'il s'agit de quelque chose de pertinent, et lissé dans le cas contraire.

2.4.2.1 La distribution en quatre étapes

Au lieu de répartir naïvement le résidu entier entre tous les contours, nous allons procéder en plusieurs étapes et nous réservons à chaque étape une partie paramétrable du résidu.

Afin de préparer le terrain, il nous faut encore quelques notations qui permettront de préciser les connaissances sur les contours que nous souhaitons exploiter. D'abord : comment pondérer les GIPCs des différents contours ? Pour chaque contour c , nous définissons un contour de pondération :

$$\begin{aligned} \omega_c^p : \mathcal{L}_c^p &\rightarrow [0; 1] \\ l &\mapsto \omega_c^p(l) \end{aligned}$$

Donc, si l'on veut privilégier, pour les GIPCs avant des frontières de groupes, uniquement les contours associés avec ces frontières, on définira $\omega_c(l) = 1$ là où l renvoie au GIPC en question pour le contour c , et $\omega_c(l) = 0$ partout ailleurs. Pour les calculs,

nous aurons besoin du cumul de ces pondérations :

$$\Omega^p(k) = \sum_{c \in \mathcal{C}_k^p} \omega_c^p(k - k_c^p)$$

Le deuxième point est un peu plus délicat : comment décider si un contour exprime une information pertinente qui sera alors à renforcer ? L'idée que nous poursuivrons est de renforcer les contours qui contiennent une information allant dans le sens du résidu $\Delta^{p;i}$ et d'affaiblir ceux qui vont à son encontre. Alors, nous devons définir l'information contenue dans un contour. Comme esquissé plus haut, je considère qu'il s'agit de sa déviation par rapport à un contour *attendu*. Nous voilà bien avancés : qu'est-ce le contour attendu ? Supposons pour l'instant que pour chaque contour φ_c^p , nous le connaissons et que nous l'appelons ϕ_c^p .¹⁵ On y reviendra un peu plus loin (cf. page 101). Ainsi, nous pouvons définir la modulation d'un contour :

$$\gamma_c^{p;i}(l) = \varphi_c^{p;i}(l) - \phi_c^{p;i}(l) \quad (2.11)$$

Et nous pouvons pour chaque GIPC k de la phrase p distinguer trois ensembles de contours. D'abord, celui¹⁶ des contours qui ont une modulation dans le sens du résidu :

$$\mathcal{C}_{k;+}^{p;i} = \{c \in \mathcal{C}_k^p : \gamma_c^{p;i}(k - k_c^p) \cdot \text{sgn}(\Delta^{p;i}(k)) > \varepsilon\} \quad (2.12)$$

Puis, celui des contours avec une modulation dans le sens opposé :

$$\mathcal{C}_{k;-}^{p;i} = \{c \in \mathcal{C}_k^p : \gamma_c^{p;i}(k - k_c^p) \cdot \text{sgn}(\Delta^{p;i}(k)) < -\varepsilon\} \quad (2.13)$$

Et, nous rassemblons les contours n'ayant aucune modulation significative dans

$$\mathcal{C}_{k;0}^{p;i} = \mathcal{C}_k^p \setminus \{\mathcal{C}_{k;+}^{p;i} \cup \mathcal{C}_{k;-}^{p;i}\}.$$

Comme pour les pondérations, nous aurons pour les calculs besoin des cumuls des modulations – ici triés selon leur signe $s \in \{+; -; 0\}$:

$$\Gamma_s^{p;i}(k) = \sum_{c \in \mathcal{C}_{k;s}^{p;i}} \gamma_c^{p;i}(k - k_c^p) \text{ pour } s \in \{+; -; 0\}$$

Nous sommes enfin prêts à décrire les quatre étapes de distribution du résidu. Pour chaque étape n nous donnerons une prescription $\varphi_{n-1} \longrightarrow \varphi_n$ en commençant évidemment avec $\varphi_0 = \varphi$. Après chaque étape nous avons la superposition $\Phi_n = \sum \varphi_n$

¹⁵J'ai choisi la lettre ϕ qui a donc le même nom que φ mais une forme plus simple – ce qui correspond bien à l'idée que l'on doit avoir du contour attendu.

¹⁶Puisque modulation et résidu ont pour chaque GIPC quatre composants (F_1, \dots, C) il s'agit en réalité de quatre ensembles.

qui permettra de déterminer le résidu restant : $\Delta_n = \Psi - \Phi_n$. Notre but est que ce résidu décroisse : $|\Delta_n| \leq |\Delta_{n-1}|$, et que la partie maximale du résidu ainsi absorbée par Φ_n durant une étape soit paramétrable (avec ι, κ et λ). À l'issue de la dernière étape, nous aurons trouvé les nouvelles cibles :

$$\psi_c^p = \varphi_{4;c}^p$$

Voici donc les quatre étapes :

1. Ajustement proportionnel aux poids *a priori* des contours :
 $\varphi_{1;c} = \varphi_{0;c} + a_1 \omega_c$
2. Lisser les contours ayant une modulation dans le sens opposé du résidu :
 $\varphi_{2;c} = \varphi_{1;c} + a_2 \gamma_c$ pour $c \in \mathcal{C}_-$
3. Renforcer les contours portant une information pertinente :
 $\varphi_{3;c} = \varphi_{2;c} + a_3 \gamma_c$ pour $c \in \mathcal{C}_+$
4. Distribuer le reste :
 $\varphi_{3;c} = \varphi_{2;c} + a_4$

Les prochains paragraphes seront uniquement consacrés à l'établissement des coefficients a_n . Ils dépendent – précisons-le puisque je ne voudrais pas surcharger les écritures – du GIPC k .

1. Pondération a priori. Nous présumons que $\Delta_0 \neq 0$ (autrement nous n'avons aucune correction à apporter au contour) et que $\omega_c \neq 0$ car dans le cas contraire, le choix de a_1 est sans importance. Cette étape est contrainte par deux conditions :

- (a) $a_1 \omega_c \cdot \Delta_0 \geq 0$
 (b) $|\sum_c a_1 \omega_c| \leq \iota |\Delta_0|$ avec $\iota \in [0; 1]$

La première condition exprime simplement que la correction apportée va dans le même sens que le résidu et la seconde spécifie que la fraction du résidu que nous voulons consacrer à cette étape n'est pas supérieure à ι .¹⁷ Puisque a_1 n'est pas censé dépendre de c , nous avons $|\sum a_1 \omega_c| = |a_1| \Omega$, et on voit que le choix suivant remplit au mieux nos deux conditions :

$$\boxed{\begin{array}{l} \varphi_1 = \varphi_0 + a_1 \omega \text{ avec :} \\ a_1 = \begin{cases} 0 & \text{si } \Omega = 0, \\ \frac{\iota \Delta_0}{\Omega} & \text{sinon.} \end{cases} \end{array}} \quad (\text{pondération})$$

¹⁷On aurait aimé remplacer le « \leq » par un « $=$ », mais ceci n'est pas possible au cas où $\omega_c(k) = 0$ pour tous les $c \in \mathcal{C}_k$

2. Lissage. Nous cherchons ici à utiliser le résidu pour lisser les contours qui ont une modulation dans le sens opposé. Ceci permettra d'éviter les situations où, au cours des cycles d'apprentissage, on aboutirait à des contours qui s'annulent mutuellement – les uns ayant des pics, les autres des creux. En principe, rien n'interdit ce cas de figure au niveau du modèle : il est tout à fait concevable qu'un mouvement montant se superpose à un mouvement descendant. Ou, en terme de durées syllabiques : qu'une tendance à un ralentissement du débit d'élocution induit par une fonction (para-)linguistique soit compensée par une accélération due à une autre fonction. Néanmoins, il est souhaitable que ceci soit réduit aux cas où l'analyse¹⁸ des contours le rend indispensable.

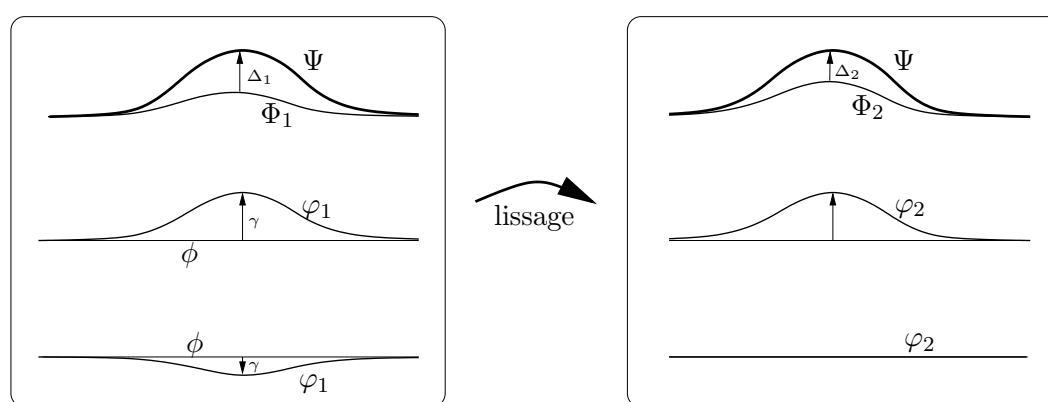


FIG. 2.14: **Distribution du résidu : lissage.** Le schéma montre à gauche l'état des contours avant l'étape de lissage et à droite après. Le contour en trait épais en haut est le contour observé Ψ . Il reste évidemment inchangé. Les contours en trait fin en haut correspondent à la superposition des contours contributeurs qui, quant à eux, sont représentés sur les deux lignes inférieures. La flèche Δ_1 indique que nous sommes en présence d'un résidu positif. Le contour qui sera affecté par le lissage est alors celui du bas ayant une modulation négative. Après le lissage, sa modulation a complètement disparu permettant ainsi la diminution du résidu Δ_2 .

Supposons pour un instant que le résidu soit positif pour un GIPC k donné. Ceci veut dire que la superposition des contours tels qu'ils sont à ce stade de l'apprentissage, ne parvient pas à atteindre la hauteur du contour observé (cf. figure 2.14 à gauche). Les contours qui nous intéressent dans ce paragraphe sont uniquement ceux de l'ensemble $\mathcal{C}_{k;-}$, c.-à.-d. ceux qui ont – pour le GIPC en question – un creux : ils « aggravent » donc la situation. Une interprétation que l'on peut en faire est la suivante : au cours des cycles d'apprentissage précédents, ces contours se sont vus attribués une information qui n'est pas de leur ressort, c.-à.-d. qui n'est pas véritablement en relation avec les fonctions (para-)linguistiques associées à ces contours.

¹⁸J'entends par « analyse » le processus de l'inversion du modèle.

Ouvrons ici une parenthèse : on pourrait être tenté de voir en cela une lacune de la procédure d'inversion du modèle. Or, il s'agit là au contraire de toute sa force : au début de l'apprentissage nous ne pouvons pas savoir comment décomposer les contours observés – l'intérêt de la répartition du résidu est justement de pouvoir déceler les erreurs de décomposition, commises nécessairement aux cycles d'apprentissage précédents. C'est cette capacité de l'algorithme qui nous permet de démarrer avec une décomposition naïve et sans *a priori* (on y reviendra un peu plus loin). Il y a là aussi la raison pour laquelle je préfère parler du *résidu* et non pas de l'*erreur* de la prédiction : on y tient une source d'information – le tout est de la mettre au profit de la meilleure décomposition possible.

Mais revenons aux contours trop creux : en vue de l'information que le résidu nous apporte pour le GIPC en question, il est alors souhaitable d'atténuer ces creux. Nous allons le faire proportionnellement à l'étendue des creux, c.-à.-d. à leurs modulations (cf. figure 2.14 à droite). Ainsi, nous traitons tous les contours de cette catégorie de la même manière : les creux seront remplis à un pourcentage égal. Les contraintes que nous devons respecter se résument à :

(a) $a_2\gamma_c \cdot \Delta_1 \geq 0$

(b) $|\sum_c a_2\gamma_c| \leq \kappa|\Delta_1|$ avec $\kappa \in [0; 1]$

(c) $a_2 \leq 0$

(d) $a_2 \geq -1$

Les premières sont les mêmes que pour la répartition pondérée du paragraphe précédent : (a) la correction va dans le sens du résidu, et (b) la fraction du résidu que nous réservons à la tâche du lissage est paramétrée par κ . Les contraintes (c) et (d) spécifient ce que nous entendons par « lisser » : la correction apportée au contour φ_c l'approche du contour attendu ϕ_c (cf. équation (2.11)) et elle n'inverse pas le signe de la modulation – c.-à.-d. elle ne crée pas de pic où il y avait un creux auparavant.

Nous avons :

$$|\sum_{c \in \mathcal{C}_-} a_2\gamma_c| = |a_2||\Gamma_-| = -a_2|\Gamma_-|$$

Et, puisque Γ_- et Δ_1 ont par définition (cf. équation (2.13)) des signes opposés, on voit que $a_2 = \kappa\Delta_1/\Gamma_1$ satisfait l'égalité de la contrainte (b)¹⁹ ainsi que (c). La contrainte (a) est par ailleurs également satisfaite (grâce à (c) et $\gamma\Delta_1 \leq 0$).

En évitant l'inversion du signe de la modulation (d), nous aboutissons à :

¹⁹Nous pouvons ici tranquillement assumer $\Gamma_- \neq 0$, car autrement nous aurions aussi $\gamma = 0$ et la définition de a_2 serait sans importance.

$$\varphi_2 = \varphi_1 + a_2\gamma \text{ avec :}$$

$$a_2 = \begin{cases} -1 & \text{si } \kappa|\Delta_1| \geq |\Gamma_-|, \\ \frac{\kappa\Delta_1}{\Gamma_-} & \text{sinon.} \end{cases} \quad (\text{lissage})$$

La condition $\kappa|\Delta_1| \geq |\Gamma_-|$ se prête par ailleurs à une interprétation toute simple : si la fraction du résidu que nous voulons distribuer dans cette étape dépasse la somme de tous les « creux », nous nous contentons d'une correction qui efface ces creux et qui nous laisse davantage de résidu pour l'étape suivante.

3. Renforcement. Cette étape est complémentaire au lissage : si nous avons, au cours de l'étape précédente, enlevé l'« information » des contours aux endroits où elle semblait en contradiction avec l'observation, nous allons maintenant renforcer les modulations qui vont dans le bon sens (cf. figure 2.15). Pour un GIPC k donné, les corrections apportées portent donc sur d'autres contours – ceux de l'ensemble $\mathcal{C}_{k,+}$. Les calculs sont cependant identiques : partant des contraintes maintenant bien familières

(a) $a_3\gamma_c \cdot \Delta_2 \geq 0$

(b) $|\sum_c a_3\gamma_c| \leq \lambda|\Delta_2|$ avec $\lambda \in [0; 1]$,

et sans avoir à nous soucier d'une inversion de la modulation, nous obtenons :

$$\varphi_3 = \varphi_2 + a_3\gamma \text{ avec :}$$

$$a_3 = \begin{cases} 0 & \text{si } \Gamma_+ = 0, \\ \frac{\lambda\Delta_2}{\Gamma_+} & \text{sinon.} \end{cases} \quad (\text{renforcement})$$

4. Distribution du reste. Et si, après tous ces efforts, il reste toujours du résidu ? Ceci est tout à fait possible : il suffit que nous tombions nez à nez avec un GIPC pour lequel aucun contour n'est particulièrement privilégié, que les modulations à lisser ne suffisent pas à absorber le résidu et qu'aucun contour ne dévie de sa forme attendue dans le sens du résidu. Dans ce cas, nous avons recours à la répartition naïve que l'on a déjà rencontrée au tout début de ce chapitre. La seule contrainte est de faire disparaître le résidu restant :

(a) $\sum a_4 = \Delta_3$

Et nous retrouvons l'analogie de l'équation (2.10) :

$$\varphi_4 = \varphi_3 + a_4$$

$$a_4 = \frac{\Delta_3}{n} \quad (\text{reste})$$

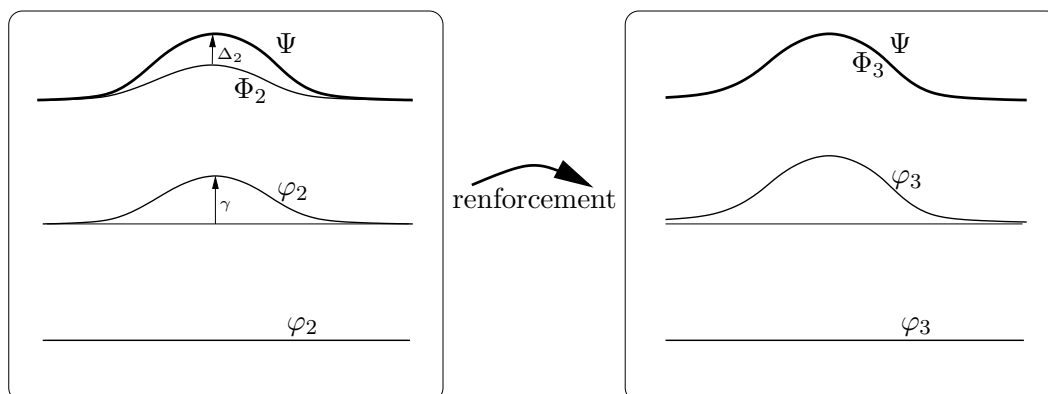


FIG. 2.15: **Distribution du résidu : renforcement.** Le contour contributeur au milieu à gauche ayant une modulation dans le sens du résidu, l'étape de renforcement permet de lui faire absorber la totalité du résidu (si l'on choisit $\lambda = 1$).

Evidemment, nous courrons ici le risque d'inverser des modulations apprises pendant les cycles d'apprentissage précédents. Mais nous n'avons pas le choix : la confrontation de la superposition des contours avec le contour observé doit avoir le dernier mot.

Notons ici que la répartition naïve de l'équation (2.10) n'est finalement rien d'autre qu'un cas spécial de la procédure en quatre étapes : on l'obtient pour $\iota = \kappa = \lambda = 0$. Les exemples et résultats présentés dans la suite de ce manuscrit ont été obtenus avec $\iota = 0,5$ et $\kappa = \lambda = 1$, donc une pondération modérée et des lissages et renforcements maximaux.

Résumé. J'espère avoir montré dans ce chapitre comment mettre l'erreur de prédiction au profit d'une meilleure décomposition des contours observés. Les détails de la répartition – bien qu'ils ne soient ni indispensables à la compréhension de l'algorithme ni à sa convergence – auront été nécessaires à l'explication de ce que j'entends par *meilleure* décomposition : chaque contour porte l'information *minimale* requise pour expliquer les formes des contours observés. Dans ce sens, il s'agit d'une *décomposition économique*.

2.4.3 L'initialisation

Jusqu'ici nous avons supposé disposer d'une estimation des contours prototypiques. Nous allons maintenant décrire comment on peut obtenir une *première* estimation. *A priori* tous les moyens devraient être bons : si le corpus est bien construit et si nos hypothèses ci-dessus sont justes, l'algorithme de répartition du résidu permet de rattraper une mauvaise initialisation. Toutefois, une bonne initialisation peut éviter un transfert lent de l'information d'un type de contour à un autre au cours des cycles

d'apprentissage et, elle peut atténuer quelques lacunes d'un corpus qui contient des régularités accidentelles.

Initialisation standard. Il existe une initialisation particulièrement simple qui s'appuie sur le pas d'itération : partant des contours prototypiques $\varphi_c^{p;0} = 0$, on peut directement utiliser les formules de répartition du résidu, données ci-dessus afin d'obtenir des premières cibles. Les étapes de lissage et de renforcement ne contribuent pas, puisque la modulation des contours est encore nulle ($\Gamma = 0$) :

$$\psi_c^1 = \varphi_c^0 + a_1\omega_c + a_4$$

Initialisation hiérarchique. Une autre façon de procéder est basée sur l'organisation hiérarchique des contours ; elle est inspirée par Morlec [1997]. Pour une phrase p , j'appelle un contour c' dominant c si son empan est plus grand que celui de c et si il l'englobe complètement – ou en écriture formelle : si $k_{c'}^p \leq k_c^p$, $k_c^p + l_c^p \leq k_{c'}^p + l_{c'}^p$ et $l_c^p < l_{c'}^p$. J'appelle alors $\mathcal{D}_c^p = \{c' \in \mathcal{C}^p : c' \text{ domine } c\}$ l'ensemble des contours dominant c .

Nous pouvons donc définir récursivement l'appartenance d'un contour à un niveau hiérarchique h :

$$\mathcal{H}_1^p = \{c \in \mathcal{C}^p : \mathcal{D}_c^p = \emptyset\}$$

$$\mathcal{H}_h^p = \{c \in \mathcal{C}^p \setminus \bigcup_{i < h} \mathcal{H}_i : \mathcal{D}_c^p \subset \bigcup_{i < h} \mathcal{H}_i\}$$

Cette définition peut sembler difficilement lisible, l'idée est pourtant fort simple : nous introduisons un nouveau niveau hiérarchique (plus bas) pour tous les contours *directement* dominés par les contours déjà attribués à un niveau supérieur (cf. figure 2.16).

L'initialisation hiérarchique est réalisée comme suit :

1. Une première « synthèse » est effectuée uniquement avec les contours du niveau \mathcal{H}_1 dans leur état « prénatal » : $\varphi = 0$.
2. La répartition du résidu, appliquée à ces contours, permet d'obtenir un premier ensemble de contours cibles qui seront confrontés au « filtre d'apprentissage » des générateurs de contours.
3. La seconde synthèse exploite les contours appris du premier niveau et ajoute des contours $\varphi = 0$ pour le deuxième niveau.
4. L'ensemble des contours peut être initialisé en répétant la répartition (2.) et en ajoutant niveau par niveau les contours manquants (3.)

Cette méthode repose sur l'hypothèse que le processus de la mise en cohérence des cibles pendant l'apprentissage des générateurs, parvient à éliminer des contours porteurs ce qui est porté. L'intérêt de cette méthode est surtout d'établir un lien avec la méthode d'apprentissage proposée par Morlec [1997]. Il est cependant important de noter que la notion de hiérarchie que nous avons introduite ici est *a posteriori* tandis qu'elle prend une place importante et *a priori* dans le modèle de Morlec.

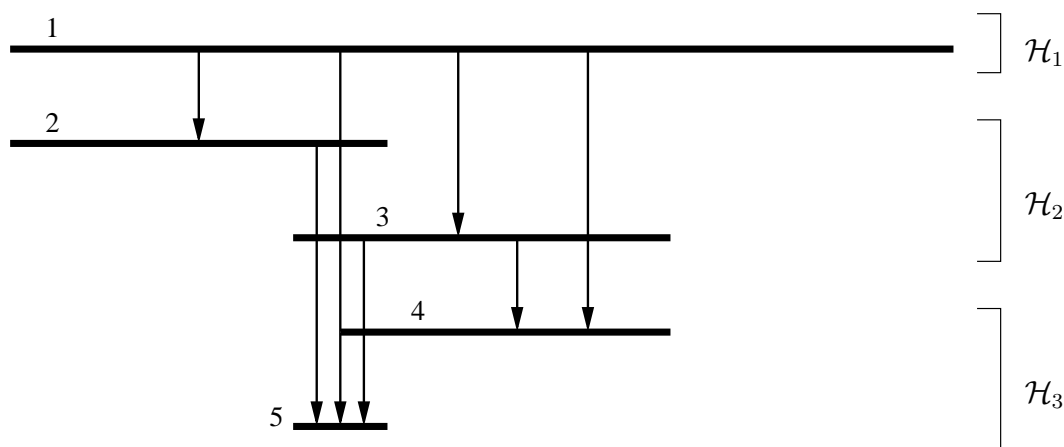


FIG. 2.16: **Attribution des contours aux niveaux hiérarchiques.** Les lignes horizontales schématisent les empan de cinq contours. Les flèches verticales expriment la dominance d'un contour sur un autre. Seul le contour 1 n'est dominé par aucun autre, on a donc $\mathcal{H}_1 = \{1\}$. Les contours 2 et 3 ne sont dominés que par le contour 1 – ils appartiennent donc au deuxième niveau : $\mathcal{H}_2 = \{2; 3\}$. Les contours 4 et 5 sont également attribués à un même niveau car aucun des deux ne domine l'autre : $\mathcal{H}_3 = \{4; 5\}$.

Initialisation avec connaissances préalables. Finalement, il existe une troisième façon d'obtenir un premier ensemble de contours cibles. Elle est apparentée à la méthode précédente bien qu'elle soit sans rapport avec la hiérarchie des contours.

Si nous disposons de générateurs de contours entraînés pour un sous-ensemble des contours présents dans le corpus, ils peuvent être utilisés dès la première synthèse du cycle d'apprentissage ce qui permet un test rapide soit d'un ajout d'un nouveau type de contour, soit d'une extension du domaine des contours déjà identifiés.

2.4.4 Les contours de référence

Au cours des étapes de lissage et de renforcement nous avons eu besoin de la notion de « contour attendu » ϕ afin de pouvoir définir ce que nous voulons entendre par bonne ou mauvaise information (par rapport au résidu) contenue dans un contour. Il est temps d'y apporter quelques précisions.

La première solution qui vient à l'esprit est d'utiliser un contour plat $\phi = 0$. Ce serait un contour qui n'apporte aucune contribution à la superposition. Or, au moins pour la composante f_0 des contours, le point 0 correspond à la fréquence de référence utilisée pour la transformation en échelle logarithmique. Il n'est pas souhaitable que celle-ci ait une influence sur la forme des contours – elle devrait plutôt être absorbée par les contours porteurs du niveau phrase.

Il est donc certainement préférable d'utiliser la moyenne des contours contributeurs par rapport à leur évolution de GIPC à GIPC : $\phi = \bar{\varphi}$. Puisque les contours porteurs absorbent la valeur moyenne des contours, les contours portés ont en général une moyenne proche de zéro, et la nouvelle définition revient pour eux pratiquement à $\phi = 0$.

Or, l'expérience montre que les formes des contours émergents de l'analyse de différents corpora comportent souvent des parties qui peuvent être apparentées à des lignes de déclinaison. Il me semble peu judicieux de « lisser » ces parties en les ramenant vers zéro. On pourrait donc remplacer l'utilisation de la moyenne d'un contour comme référence par la droite de régression (ce qui évoque les patrons mélodiques que l'on trouve chez Aubergé [1991]) ; ou bien, puisque l'on observe souvent un phénomène de « reset » à la frontière entre deux groupes liés par un même contour, par deux droites de régression.

Cependant, cette méthode étant très invasive et incluant beaucoup d'*a priori* sur les formes de contours, je préfère utiliser une prescription plus locale : pour un GIPC donné le contour de référence est obtenu comme la moyenne des valeurs voisinant ce GIPC. Cette définition a l'avantage de permettre d'identifier les maxima et minima des contours en vue de leur renforcement ou lissage. De plus, les parties linéaires échappent à ces étapes de la répartition du résidu.

2.4.5 Les générateurs de contours

La description du modèle SFC touche à sa fin. Mais, il manque toujours un élément qui peut sembler très important : comment générer concrètement des contours multiparamétriques associés à une fonction pour un empan donné ? Jusqu'à maintenant, nous avons parlé – d'une manière volontairement floue – de « générateurs de contours » capables d'engendrer des familles de contours « homogènes ». Si nous avons pu nous passer de préciser l'implémentation de ces générateurs, c'est dû au fait qu'elle n'est pas tellement importante pour le modèle. Une boîte noire est suffisante tant qu'elle remplit les tâches suivantes :

- accepter en entrée une caractérisation de l'empan,
- fournir en sortie un contour multiparamétrique correspondant, et
- être capable d'apprendre ce « mapping » à partir d'un (grand) ensemble de paires entrée-sortie.

Le point crucial est évidemment l'apprentissage. La description du cycle d'analyse par la synthèse a laissé entrevoir les attentes que nous avons vis à vis de cet apprentissage : elle doit mettre les paires entrée-sortie « en cohérence » – c.-à.-d. capturer les propriétés communes aux contours, et filtrer (ou lisser) ce qui semble être des cas uniques.

Caractérisation des empan. L'information sur les empan, fournie en entrée des générateurs, spécifie quels contours seront *a priori* identiques. Nous utilisons la contrainte forte qu'elle peut être spécifiée en terme de nombre de GIPCs uniquement. Ceci correspond à l'hypothèse de base que la morphologie des contours d'un même type ne dépend que des propriétés phonotactiques des groupes de mots concernés.²⁰ Concrètement, cela revient à fournir simplement le nombre de GIPCs si la fonction associée s'applique à un seul groupe (comme les contours de phrase), ou les nombres de GIPCs des deux groupes mis en relation (de dépendance p.e.) par la fonction. Nous appelons cette simple caractérisation des empan leur *configuration phonotactique*.

Implémentation des générateurs. Après avoir caractérisé les empan de la sorte, l'ensemble des paires entrée-sortie que le générateur doit apprendre, est naturellement séparé en plusieurs classes – une pour chaque configuration unique. Le moyen le plus simple est de déterminer pour chaque classe le contour moyen – une méthode qui est quelque peu apparentée au lexique structuré implémenté par Morlec et al. [1995]. Cet « apprentissage » remplit certainement les critères de capture des régularités et de lissage des « accidents ».

Cependant, un tel générateur de contour moyen ne peut générer que des contours pour les empan vus pendant l'apprentissage. Ceci ne pose aucun problème au cours du cycle d'analyse par la synthèse, mais afin de pouvoir les utiliser en synthèse d'une phrase nouvelle, il faudrait soit un corpus d'apprentissage très complet, soit une méthode d'interpolation (et d'extrapolation) pour des éventuels contours non-appris. Un autre désavantage de cette méthode est qu'elle n'exploite aucune métrique de l'espace des entrées : des contours associés à des groupes de 7 ou de 8 GIPCs peuvent potentiellement être complètement différents.

Pour ces deux raisons, nous préférons utiliser un simple réseau de neurones (RN) pour la génération des contours. La méthode plus simple décrite ci-dessus reste néanmoins une option de « test rapide » par exemple pour l'introduction d'une nouvelle fonction.

L'utilisation d'un RN est fortement inspirée par Morlec [1997]. Il y a cependant une différence importante : les RN de Morlec ont à charge la génération de contours dont l'empan est toujours coextensif à la phrase entière. Ils sont attachés à un niveau *hiérarchique* et expriment la morphologie de *toutes* les fonctions (para-)linguistiques opérant à ce niveau. Les RN doivent donc prévoir un mécanisme de coarticulation entre les fonctions – raison pour laquelle ils contiennent une boucle qui leur permet d'implémenter une notion d'« histoire » au cours d'un contour. Une coarticulation est également rendue possible chez Morlec, en fournissant aux RN une information sur la nature de la fonction suivante (toujours sur le même niveau hiérarchique).

Dans le cadre du SFC, chaque RN est consacré à la génération des contours associés

²⁰Nous reviendrons à cette hypothèse au cours de la présentation des résultats (cf. page 147).

à une seule fonction avec un contour coextensif uniquement à la portée de cette fonction. Toute interaction entre fonctions, même d'un même niveau hiérarchique, passe – du moins pour l'instant – par la superposition et le chevauchement des contours associés.

Néanmoins, l'implémentation des RN reste très voisine de celle proposée par Morlec : la prédiction des contours se fait GIPCs par GIPCs. Il faut à ce point distinguer entre l'entrée globale fournie au générateur en tant que boîte noire, qui est simplement la configuration phonotactique des empans détaillée plus haut, et l'entrée locale fournie au RN pour un GIPC donné. La transformation de l'entrée globale vers l'entrée locale fait partie intégrante de l'implémentation du générateur. Elle doit contenir de l'information positionnelle sur le GIPC afin de permettre au RN de produire au fil des GIPCs une suite de valeurs de sortie qui forment ainsi un contour cohérent.

Morlec a montré que la forme des contours peut être comprise en tant que mouvement de *préparation* suivi d'un geste de *capture* à l'approche des points d'ancrage²¹. En conséquence, l'entrée locale est obtenue en tant que valeurs sur des rampes de *deux* types différents : relatif (permettant la génération du mouvement préparatoire) et absolu (capture). Comme illustré dans la figure 2.17, les rampes absolues comptent la distance (en termes de nombre de GIPCs) aux points d'ancrage et les autres rampes décrivent la position relative du GIPC entre deux points d'ancrage.

Nous utilisons des rampes ascendantes et descendantes ce qui est redondant en combinaison avec la rampe relative, mais facilite l'apprentissage du RN. Ce codage des configurations phonotactiques des empans par plusieurs rampes en vue de la génération séquentielle des contours reste donc très simple. Il est néanmoins assez riche (en identifiant la position de chaque GIPC et la taille des groupes concernés) pour permettre la génération de contours complexes.

Notons que les RN prennent la forme de simples réseaux « feed-forward » avec une couche cachée à 17 nœuds. La fonction d'activation des unités cachées est la fonction logistique ; les unités d'entrée ayant une simple identité comme fonction d'activation, et celles de la sortie l'identité plus un biais. L'implémentation et l'apprentissage (à l'aide de l'algorithme de « resilient propagation ») des réseaux ont été effectués avec SNNS²². Nous reviendrons un peu plus loin (cf. page 122) sur quelques détails de l'implémentation et leurs retombées en terme de performance du SFC.

²¹Les points d'ancrage sont le début et la fin de l'empan et, le cas échéant, la frontière entre groupes.

²²cf. <http://www-ra.informatik.uni-tuebingen.de/SNNS/>

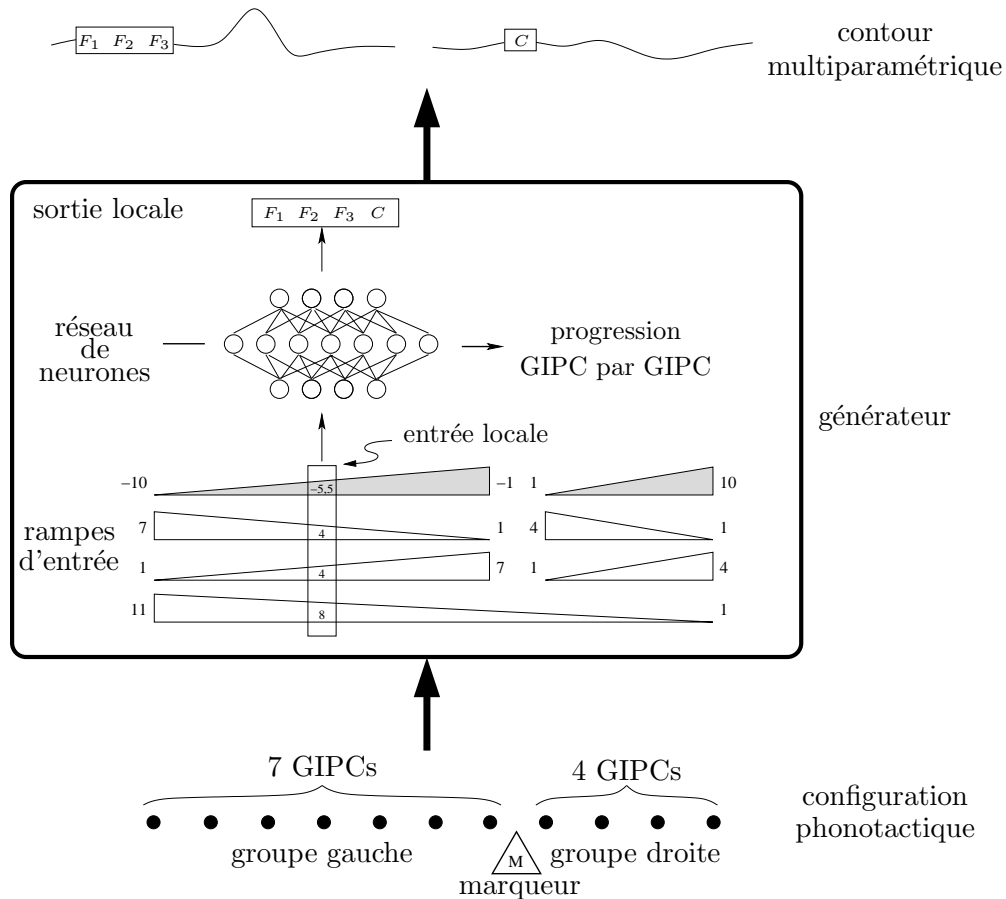


FIG. 2.17: **Implémentation des générateurs de contours.** Les caractéristiques de l'empan (en bas) sont transformées en rampes d'entrée qui permettent au RN de prédire les quatre valeurs de sortie par GIPC. La suite de ces sorties est le contour cherché. Il est représenté en haut : sa composante de f_0 à gauche et celle du coefficient d'allongement à droite.

L'exemple porte sur une fonction qui est appliquée à deux groupes adjacents – la frontière étant indiquée par le marqueur. Les rampes relatives (en gris) sont normalisées entre 1 et 10 avec des valeurs négatives pour le premier groupe.

Remarques sur l'utilisation de réseaux de neurones. J'insiste sur le fait que l'utilisation de RN pour la génération de contours prosodiques n'a aucune connotation cognitive ou neuro-mimétique. La seule raison pour laquelle nous les emploierons est leur capacité d'apprendre et de réaliser le « mapping » non-linéaire entre deux espaces très différents : les configurations phonotactiques et les contours prosodiques.

Rappelons brièvement les principales différences des RN que le SFC mobilise en tant que générateurs de contours fonctionnels et ceux employés par Morlec ou Traber : le RN de Morlec a à sa charge la génération coarticulée des manifestations prosodiques associées à des niveaux d'une hiérarchie prosodique présupposée. Il en découle la nécessité d'implémenter à l'intérieur du RN l'interaction entre plusieurs fonctions (cf. la boucle de rétroaction et le renseignement sur des points d'ancrage distants en entrée du réseau). Le RN de Traber a une tâche encore beaucoup plus complexe puisqu'il ne s'appuie pas sur un modèle phonétique de superposition. C'est le RN qui doit accomplir toute la négociation entre les différents facteurs qui influent sur la réalisation prosodique. La complexité accrue de la tâche trouve son reflet direct dans le nombre beaucoup plus élevé de paramètres à ajuster pendant l'apprentissage (plus d'entrée, plus d'unités cachées etc.) et la plus grande richesse informative des entrées (pas question de se contenter d'une information phonotactique).

Terminons ces remarques sur l'utilisation de RN en tant que générateurs de contours en évoquant une de leurs faiblesses majeure : si les RN fournissent, d'une manière générale, une interpolation raisonnable, leurs capacités d'*extrapolation* sont très limitées. Les conséquences pour le SFC sont immédiates : dès que nous demandons au modèle de générer le contour prosodique d'un énoncé qui met en œuvre des instances de fonction ayant une portée plus large que celles vues pendant l'apprentissage, nous devons nous attendre à ce que les contours aient des allures malsaines. De mon point de vue, ce problème est étroitement lié à une caractéristique des contours fonctionnels de grand empan : le nombre de points décrivant le contour multiparamétrique augmente de manière linéaire avec la taille de l'empan. Il est donc de plus en plus difficile pour les générateurs de fournir des contours raisonnablement lisses, c.-à.-d. en adéquation avec l'information qu'ils sont censés véhiculer et qui n'augmente certainement pas de la même manière que la longueur des empan. En effet, grâce à l'encodage de la configuration phonotactique en entrée, les RN en sont capables de manière suffisante pour les corpora que nous avons analysés. Cependant, à moyen terme, il me semble indispensable de revoir l'implémentation des générateurs par rapport à ce problème. Des pré- et post-traitements en entrée et sortie des RN pourraient être une solution. Borner les valeurs absolues des entrées locales serait une possibilité simple ; contraindre la forme des sorties par une paramétrisation, p.e. du type Fujisaki ou par DCT, serait une méthode complémentaire.

2.4.6 Résumé de l'inversion du SFC

Dans ce chapitre, nous avons exposé la solution que nous apportons au problème de l'inversion de notre modèle. Étant donné que les observations ne fournissent pas directement des contours exploitables pour l'entraînement des générateurs, la solution proposée passe par un cycle d'analyse par la synthèse.

Guidé par l'idée que la différence entre observation et synthèse préliminaire contient l'information utile à l'amélioration de l'état des générateurs, nous avons consacré beaucoup d'effort à détailler comment cette information peut être diffusée aux contours fonctionnels.

L'étape d'apprentissage des générateurs au cours de chaque cycle remplit le rôle d'un filtre : de l'information diffusée aux contours, il laisse passer uniquement celle qui est pertinente pour chaque famille de contours. Ainsi, la cohérence postulée des familles devient la contrainte de haut niveau permettant l'inversion de la superposition des contours fonctionnels.

2.5 Résumé

Retraçons les étapes par lesquelles nous sommes passés au cours de cette partie du manuscrit, consacrée à la présentation du SFC. Dans le premier chapitre, nous avons pris soin d'établir la description phonétique sur laquelle le modèle de génération proprement dit se base. Le but principal était de fournir une factorisation entre micro-prosodie (à l'échelle intra-syllabique) et macro-prosodie (inter-syllabique). Au passage, nous avons pu détailler l'émergence de pauses silencieuses.

Après la brève présentation des propositions et des implémentations d'Aubergé et de Morlec auxquelles le SFC doit un héritage important, et quelques observations relatives aux arbres de performance qui nous ont convaincu de l'importance des poids syllabiques des constituants syntaxiques pour la morphologie de la structure rythmique, nous avons pu enfin exposer le SFC. Il postule une relation étroite entre fonctions communicatives et leurs implémentations prosodiques par des contours dits fonctionnels. Il prévoit l'interaction des fonctions par la superposition de ces contours. Leur morphologie est déterminée par les configurations phonotactiques de la partie de l'énoncé à laquelle ils réfèrent.

Avec le grand chapitre consacré à l'apprentissage du modèle, ou – comme je préfère le dire – son inversion, j'espère avoir pu clarifier ce qui caractérise le plus le SFC : il est un moyen efficace de projeter des hypothèses sur des observations et d'en extraire les conséquences. Si dans la description du modèle, les familles « cohérentes » de contours fonctionnels ont dû paraître quelque peu floues, l'inversion du modèle leur donne un sens : il s'agit de la manifestation prosodique d'une fonction communicative sous la contrainte d'une décomposition économique. Cette interprétation est bien entendu uniquement valable dans le cadre du SFC, mais elle permet – comme nous allons voir dans la partie suivante – l'analyse de phénomènes prosodiques.

Terminons cette partie avec un petit tableau qui résume ce que le SFC partage et ce qui le différencie des propositions d'Aubergé et de Morlec :

	Aubergé [1992]	Morlec [1997]	SFC
interaction	superposition	superposition	superposition
empans	constituants syntaxiques	phrase	portées des fonctions
caractérisation	attributs & relations morphosyntaxiques, sémantiques	suite des relations de dépendance par niveau	fonctions communicatives (méta-)linguistiques
hiérarchie	explicite	explicite	implicite
génération	basée-lexique	dynamique	dynamique
apprentissage	hiérarchique	hiérarchique	global

Résultats et applications

Plan de la partie. Nous allons commencer cette partie avec la description des corpora auxquels nous avons appliqué le SFC. Il s'en suit le chapitre 3.2 qui prend les résultats quantitatifs du modèle comme prétexte pour discuter quelques paramètres techniques de l'implémentation des générateurs. Les chapitres 3.3 à 3.5 montrent le SFC « en action » : d'abord par rapport aux fonctions communicatives que l'on peut identifier dans un corpus de phrases déclaratives, puis, en vérifiant l'expansion du mouvement mise en évidence par Morlec [1997] et finalement, en analysant de plus près les réalisations prosodique d'une fonction communicative. Nous terminerons cette partie avec quelques résultats descriptifs.

3.1 Présentation des corpora

Cinq corpora ont été soumis à l'analyse par le SFC au cours de cette étude : le point de départ a été un corpus de formules algébriques lues. Il a été ensuite étendu et « traduit » en allemand. La présentation du corpus mettant en œuvre six attitudes, conçu par Morlec [1997], n'est rien de plus qu'un rappel de ce que l'on trouve en détail dans sa thèse. Je tiens ici à remercier Hélène Lœvenbruck et Guillaume Rolland pour la mise à disposition du grand corpus de phrases déclaratives isolées, ainsi qu'Eduardo Banga pour le corpus en galicien enregistré à l'université de Vigo.

3.1.1 Formules mathématiques lues

Le corpus de formules mathématiques a été conçu afin de pouvoir étudier la manière selon laquelle la prosodie peut encoder des relations de dépendance à l'intérieur d'une phrase. Nous avons choisi des formules mathématiques lues parce qu'on y trouve une structure syntaxique profonde qui est souvent ambiguë à l'oral et qui nécessite un recours renforcé à la structuration de l'énoncé par la prosodie. Nous avons restreint le domaine des formules aux équations algébriques issues de l'enseignement de niveau 3^e.

Le corpus est lu par un locuteur français à qui nous avons donné la consigne de ne pas utiliser d'indicateurs lexicaux de structure comme « ouvrez les parenthèses ». Chaque formule est énoncée deux fois de suite. Le locuteur peut annuler et répéter l'enregistrement d'une formule s'il n'est pas satisfait (hésitations, erreurs etc.).

3.1.1.1 Grammaire de génération

Le corpus a été généré automatiquement en variant systématiquement la profondeur syntaxique et la longueur syllabique des constituants pour couvrir statistiquement l'influence de ces paramètres sur la stratégie prosodique – tout en gardant une taille de corpus raisonnable. Nous donnons par la suite la grammaire de génération. Elle engendre 157 formules.

Le vocabulaire terminal est :

$$V_T = \{0; 1; 2; 5; 6; 7; 8; 9; x; =; <; >; +; -; \times; /; \sqrt{\quad}; \backslash; (;); ^2\}^1$$

Puis, on a :

$$V_N = \{G; D; P_1; P_2; P_3; P_4; S; S'; S''; R\}$$

Et un vocabulaire « quasi-terminal » :

$$V_{TN} = \{a; e; p; m; r\}$$

L'axiome est F.

Ensuite, il y a les règles de production. Pour les éléments de V_{TN} :²

$$\begin{aligned} e &\longrightarrow = \mid < \mid > \\ p &\longrightarrow + \mid - \\ m &\longrightarrow \times \mid / \\ rR &\longrightarrow \sqrt{R} \mid \backslash \mid R \backslash \mid \\ [=, <, >]a\sqcup &\longrightarrow [=, <, >]\{0; 1; 2; 6; 9\} \sqcup \\ \sqcup a[=, <, >] &\longrightarrow \sqcup\{0; 1; 2; 6; 9\}[=, <, >] \\ ax &\longrightarrow \{2; 5; 6; 7; 8; 9\}x \\ a[(, \times] &\longrightarrow \{2; 6; 9\}[(, \times] \\ [), \times, /]a &\longrightarrow [), \times, /]\{2; 6; 9\} \\ a &\longrightarrow \{1; 2; 6; 9\} \end{aligned}$$

Toute formule est générée par :

$$F \longrightarrow G e D$$

¹On utilise le symbole \backslash pour banaliser le symbole suivant. Il s'agit d'éviter des confusions entre « valeur absolue » et le « \mid » utilisé pour décrire la grammaire.

²La réécriture de « a » se fait selon le contexte. On utilise seulement la première règle applicable pour « a ». \sqcup signifie le début ou la fin d'une formule

Pour les formules (1)–(15)³ les deux côtés de l'équation sont des polynômes :

$$\begin{aligned}
 D, G &\longrightarrow P_i & i \in \{1; 2; 3; 4\} & (3.1) \\
 P_1 &\longrightarrow a \\
 P_2 &\longrightarrow ax \\
 P_3 &\longrightarrow ax \text{ p } a \\
 P_4 &\longrightarrow ax^2 \text{ p } ax \text{ p } a
 \end{aligned}$$

Les formules (16)–(75) sont générées par :

$$G \longrightarrow (P_i) \text{ m } (P_j) \mid (P_3)^2 \quad i, j \in \{1; 2; 3; 4\} \quad (3.2)$$

$$D \longrightarrow P_i \quad i \in \{1; 2; 3; 4\} \quad (3.3)$$

Puis, (77)–(122) sont issues de :

$$G \longrightarrow S' \text{ p } S' \mid P_1 \text{ p } S \mid S \text{ p } P_1 \quad (3.4)$$

$$D \longrightarrow P_1 \quad (3.5)$$

$$S \longrightarrow (P_i) \text{ m } (P_j) \mid (P_3)^2 \quad i \in \{1; 2; 3; 4\}$$

$$S' \longrightarrow (P_i) \text{ m } (P_j) \mid (P_3)^2 \quad i \in \{2; 3\}$$

Les formules (123)–(153) suivent le schéma suivant :

$$G \longrightarrow rR \quad (3.6)$$

$$D \longrightarrow P_1 \quad (3.7)$$

$$R \longrightarrow x \mid P_i \mid (P_i) \text{ m } (P_j) \mid P_1 \text{ p } S'' \mid S'' \text{ p } P_1 \mid S'' \text{ p } S'' \quad i \in \{1; 2; 3; 4\}$$

$$S'' \longrightarrow (P_i) \text{ m } (P_j) \quad i, j \in \{2; 3\}$$

Et finalement, pour (154)–(157) :

$$D, G \longrightarrow rR$$

$$R \longrightarrow P_i \quad i \in \{2; 3\}$$

Pour satisfaire les conventions usuelles de notation, on applique un post-traitement :

$$(P_i)/(P_j) \longrightarrow P_i/P_j$$

$$(a) \longrightarrow a$$

$$(ax) \longrightarrow ax$$

$$a \times (P_i) \longrightarrow a(P_i)$$

$$ax \times (P_i) \longrightarrow ax(P_i)$$

³Les numéros correspondent à ceux que l'on trouve dans l'annexe A.1.

Cette grammaire est capable de générer les formules que nous voulons produire (elle est suffisante). On en voit un exemple dans la figure 3.1. Outre des restrictions pour les indices, il y a un certain nombre de formules générées par la grammaire qui ne se trouvent pas dans le corpus. Il s'agit de toutes les formules qui contiennent P_1 m P_1 ou P_2 m P_2 et de la formule P_1 e P_1 . Par ailleurs, on a ajouté à notre corpus la formule (76). Avec ces limitations, la grammaire génère exactement notre corpus.

On aurait pu éviter le vocabulaire « quasi-terminal ». On ne l'a introduit que pour indiquer que l'on considère quelques opérateurs et les chiffres comme de simples variations lexicales. Après l'application de toutes les règles sauf celles pour les éléments de V_{TN} , on substitue ces derniers en utilisant une seule alternative. Ainsi, chaque structure syntaxique est représentée exactement une fois dans le corpus.

Les règles pour la substitution d'« a » facilitent l'analyse phonétique (p.e. pas de consonnes plosives à la fin d'un mot ni au début. Ceci afin de pouvoir segmenter avec précision les réalisations de l'énoncé.).

Caractéristiques du corpus. La figure 3.2 rassemble quelques informations sur le corpus ainsi généré : le nombre de syllabes (ou : GIPCs) par phrase varie entre 7 et 36 avec une moyenne de 20. La durée moyenne des GIPCs est de 210 ms. La fréquence fondamentale du locuteur français est de 145 ± 37.9 Hz.

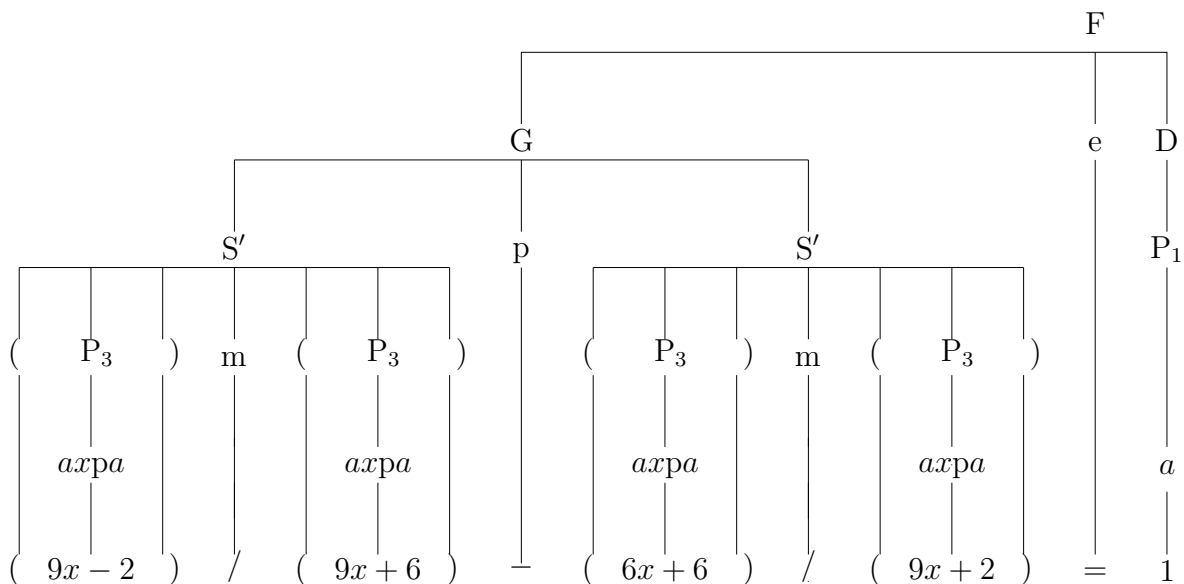


FIG. 3.1: Formule 117 analysée par la grammaire.

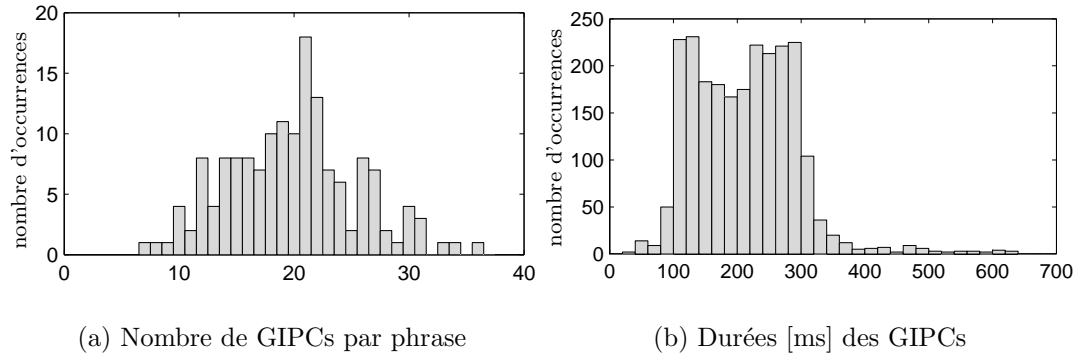


FIG. 3.2: Caractéristiques du corpus français.

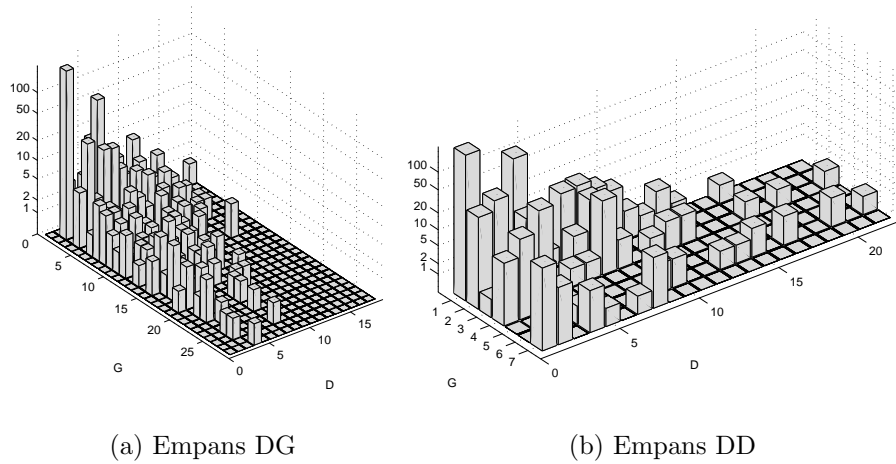


FIG. 3.3: **Distributions des empans du corpus de formules (français).** L'axe « G » réfère au nombre de GIPCs de la partie gauche d'un empan ; l'axe « D » à celui de la partie droite. La hauteur des « tours » correspond au nombre d'occurrences des configurations phonotactiques spécifiée par G et D.

3.1.1.2 Formules mathématiques – allemand

L’oralisation de formules mathématiques offre la perspective très intéressante d’effectuer des comparaisons *inter*-langues : il est effectivement possible d’utiliser un même corpus (sous sa forme écrite) prononcé en différentes langues. Le processus d’oralisation est plus complexe qu’une simple lecture car le locuteur dispose d’une certaine liberté concernant les choix lexicaux et la linéarisation de la structure d’une formule : pour une même langue, il faut donc s’attendre à des différences inter-locuteurs. Les différences auxquelles on peut s’attendre dans une autre langue sont alors du même type. Les principes de construction du corpus restent donc valables. En collaboration avec Stephan Raidt, nous avons exploré cette possibilité en enregistrant un corpus de formules algébriques en allemand.

Le corpus sous sa forme écrite a été changé en trois points :

- Il a été élargi afin de pallier aux asymétries présentes dans le corpus français – les outils de traitements des enregistrements (notamment la segmentation et la vérification de f_0) étant plus performants qu’au début de nos travaux, il a été possible d’étendre la taille « gérable » du corpus.
- Afin de pouvoir examiner ultérieurement l’influence de la portée d’un opérateur préfixé nous avons ajouté des formules du type $\sqrt{2x} + 3 = 4$.
- Les choix des nombres utilisés dans le corpus a été adapté à l’allemand – guidé par le même souci : une analyse phonétique simple.

Dans le détail ces changements reviennent à :

$$D, G \longrightarrow P_i \mid (P_i) \text{ m } (P_j) \mid (P_3)^2 \qquad i, j \in \{1; 2; 3; 4\}$$

Ceci englobe et complète les formules générées par les règles 3.1 à 3.3 – on les retrouve dans l’annexe A.2 avec les numéros 1 à 287.

Les formules générées par les équations (3.4) et (3.5) ont les numéros (288)–(329) et se trouvent complétées ((330)–(371)) par :

$$\begin{aligned} G &\longrightarrow P_1 \\ D &\longrightarrow S' \text{ p } S' \mid P_1 \text{ p } S \mid S \text{ p } P_1 \end{aligned}$$

Ainsi que les formules issues des règles (3.6) et (3.7) par :

$$\begin{aligned} G &\longrightarrow P_1 \\ D &\longrightarrow \text{r}R \end{aligned}$$

Les formules contenant des opérateurs préfixés ($\sqrt{\dots}$ ou $\mid \dots \mid$) avec des portées plus courtes que précédemment sont obtenues grâce à (cf.(438)–(449)) :

$$\begin{aligned} G &\longrightarrow \text{r}R \text{ p } T \\ D &\longrightarrow P_1 \\ R &\longrightarrow (P_i) \text{ m } (P_j) & i, j \in \{2; 3\} \\ T &\longrightarrow P_1 \mid (P_i) \text{ m } (P_j) & i, j \in \{2; 3\} \end{aligned}$$

Et symétriquement pour les formules (450)–(461) :

$$\begin{array}{l} G \longrightarrow P_1 \\ D \longrightarrow rR p T \end{array}$$

Finalement, nous ajoutons (462)–(475) :

$$\begin{array}{l} G \longrightarrow rP_i m P_i \\ D \longrightarrow P_1 \end{array} \quad i, j \in \{1; 2; 3; 4\}$$

Et, (476)–(489) :

$$\begin{array}{l} G \longrightarrow P_1 \\ D \longrightarrow rP_i m P_i \end{array} \quad i, j \in \{1; 2; 3; 4\}$$

Les règles pour la substitution des nombres sont adaptés en fonction de leur réalisation phonétique en allemand (et simplifiées) :

$$\begin{array}{l} ax \longrightarrow \{3; 4; 5; 6; 7; 8; 9\}x \\ a \longrightarrow \{3; 4; 5; 6; 7; 9\} \end{array}$$

Caractéristiques du corpus. Ce corpus a été enregistré par un locuteur masculin de langue maternelle allemande. Sa fréquence fondamentale moyenne est d'environ 106 ± 19 Hz. Les formules ont entre 5 et 44 syllabes avec une moyenne de 20. La durée des GIPCs qui ne contiennent pas de pause varie entre 95 ms et 870 ms⁴ (moyenne : 330 ms, cf. figure 3.4).

Au niveau de la représentation des empanns de différentes tailles dans le corpus, la figure 3.5 témoigne de l'effort fait pour symétriser le corpus : la partie droite des empanns DG peut devenir plus longue que dans le corpus enregistré en français (cf. figure 3.3). Si le côté gauche des empanns DD est plus limité ici (entre un et trois syllabes), ceci est dû à la coïncidence systématique de cette partie avec les opérateurs mathématiques qui sont plus courts en allemand – notamment il n'y a pas de « strictement supérieur à » avec ses 7 syllabes.

⁴Cette durée de GIPC exceptionnellement longue est due à une réduction de voyelle : [zibən] → [zibn] et donc une fusion de deux GIPCs en un seul.

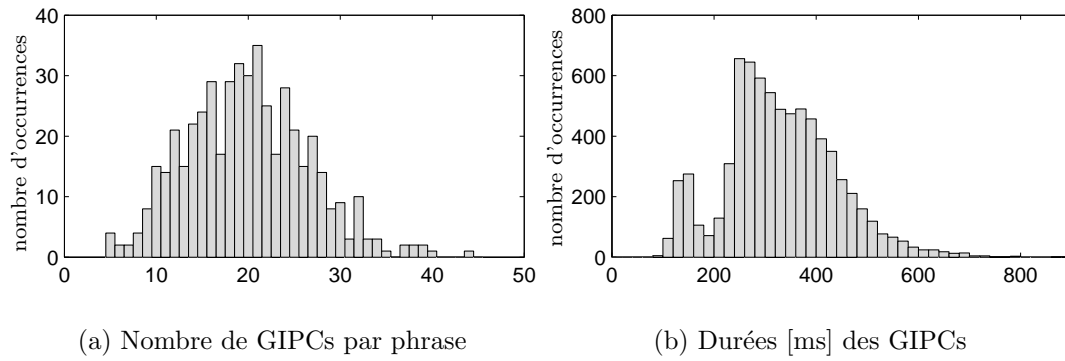


FIG. 3.4: **Caractéristiques du corpus de formules (allemand).** Le pic de GIPCs ayant une durée autour de 150 ms est dû aux GIPCs [in] de l'opérateur fréquent « minus ».

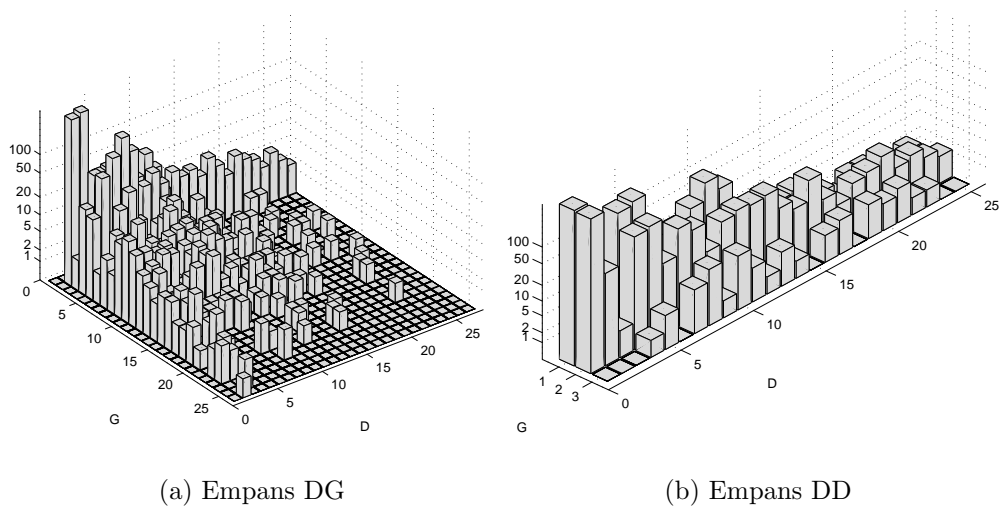


FIG. 3.5: **Distributions des empans du corpus allemand.**

3.1.2 Six attitudes

Il s'agit ici du corpus mis en place par Morlec [1997]. Il avait permis de mettre en évidence l'expansion de mouvements porteurs en fonction de six attitudes différentes (déclaration : DC ; doute-incrédulité : DI ; évidence : EV ; exclamation : EX ; question : QS ; ironie de soupçon : SC). Les phrases de ce corpus sont relativement courtes (entre 1 et 14 syllabes pour les déclarations et 1 et 8 syllabes pour les autres attitudes). Pour plus de précision sur la construction du corpus et ses caractéristiques le lecteur pourra se référer au manuscrit de Morlec [1997]. Le modèle SFC a été appliqué à ce corpus afin de vérifier si son analyse est compatible au modèle d'expansion de mouvements.

3.1.3 1001 phrases

Ce corpus, conçu et enregistré par Guillaume Rolland et Hélène Lœvenbruck de l'ICP, contient des phrases entre 4 et 20 syllabes. Il est enregistré par une locutrice française avec une fréquence fondamentale de 217 ± 32 Hz. La durée moyenne des GIPCs est de 170 ms (cf. figure 3.6). Un des buts de ce corpus est de pouvoir observer les réalisations prosodiques d'une relation de caractérisation par des moyens différents : par exemple une qualification par un adjectif, un complément avec préposition ou une subordonnée relative. Un autre aspect intéressant est qu'il comporte les mêmes syntagmes nominaux alternativement en position de sujet et d'objet direct.

3.1.4 Le corpus de Vigo

Ce corpus en galicien a été développé et enregistré à l'université de Vigo (Espagne). Les 200 premières phrases du corpus ont été étiquetées selon les besoin du SFC au cours d'une mission de courte durée d'Eduardo Banga à l'ICP dans le cadre du projet COST 258 « The Naturalness of Synthetic Speech ». Les phrases ont une longueur entre 16 et 36 syllabes. La fréquence fondamentale du locuteur est de 110 ± 19 Hz et la durée moyenne des GIPCs de 144 ms (cf. figure 3.8).

Le corpus présente un certain nombre de problèmes pas encore résolus :

- La représentation d'empans de tailles différentes est moins bonne que pour les autres corpora (cf. figure 3.9) – ce qui n'est pas étonnant puisqu'il n'a pas été conçu pour l'apprentissage d'un modèle du type SFC.
- La définition des GIPCs devrait être révisée (statut des diphtongues etc.)
- La qualité de l'enregistrement rend la détection de la fréquence fondamentale un peu plus difficile – notamment en fin d'énoncé.

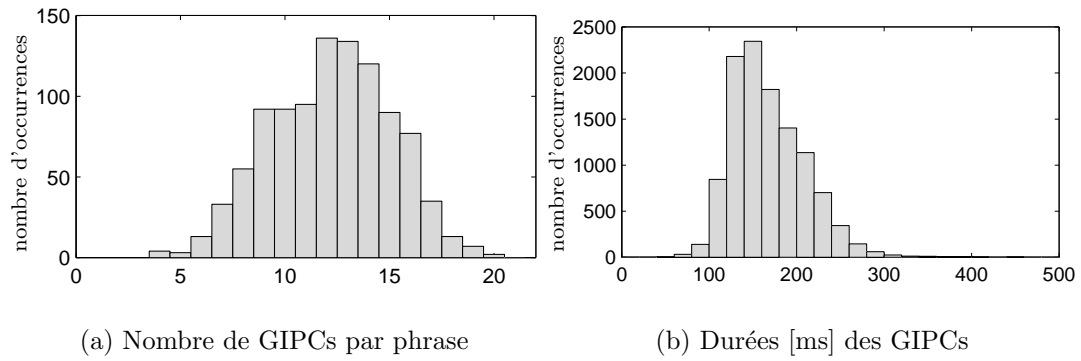


FIG. 3.6: Caractéristiques du corpus « 1001 phrases ».

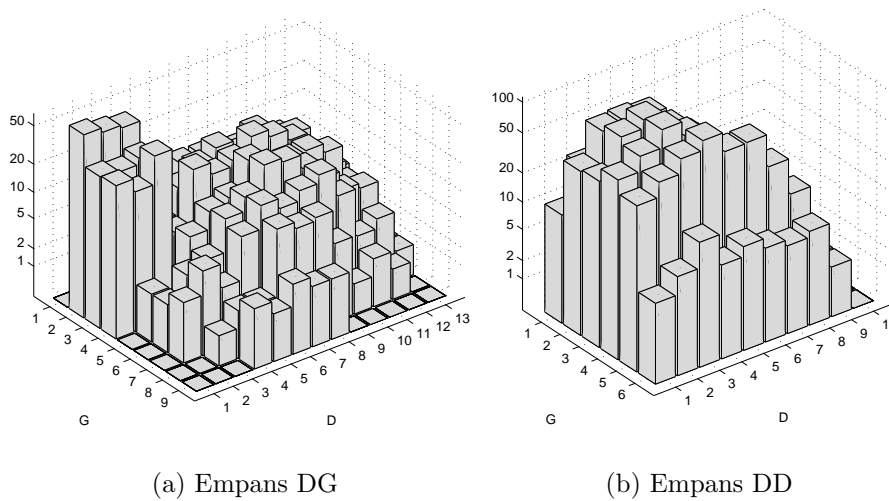


FIG. 3.7: Distributions des empans du corpus « 1001 phrases ».

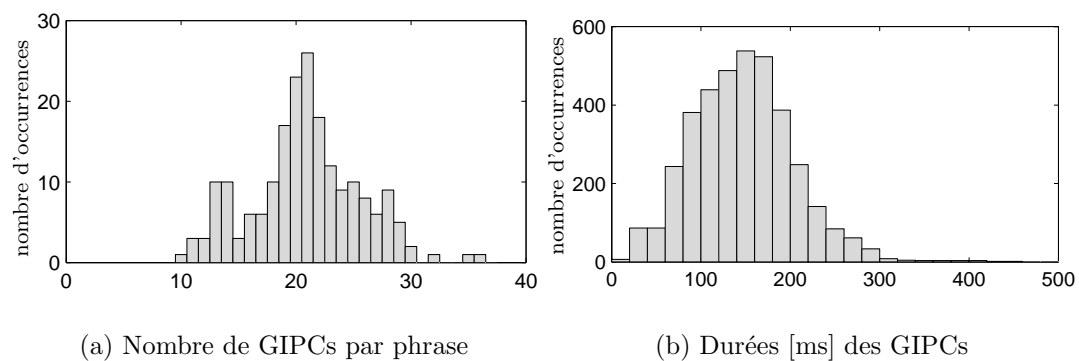


FIG. 3.8: Caractéristiques du corpus de Vigo.

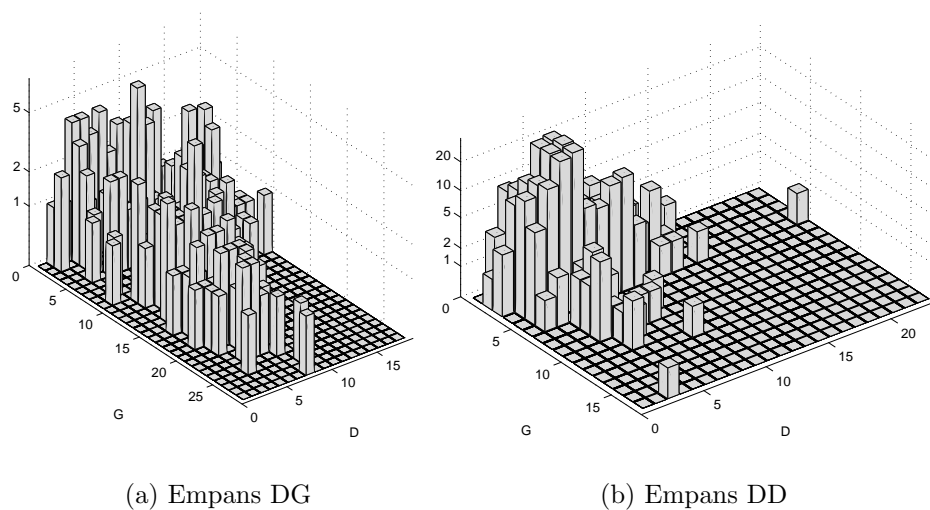


FIG. 3.9: Distributions des empans du corpus de Vigo.

3.2 Résultats quantitatifs

3.2.1 Considérations générales

L'évaluation objective de la performance d'un modèle de la génération de la prosodie est une tâche difficile. Nous avons vu dans l'état de l'art au début de ce manuscrit que les erreurs de prédictions données sur des corpora différents varient d'une manière importante. L'interprétation de ces chiffres en tant que comparaison entre modèles est donc – au moins – très délicate.

Néanmoins, les mesures objectives fournissent un repère relativement facile à obtenir et indispensable quant aux évolutions que l'on voudrait apporter au sein d'un même modèle. Dans l'annexe B, on trouvera donc un certain nombre de mesures sur les différents corpora auxquels nous avons appliqué le SFC. Elles constituent en quelque sorte un benchmark de l'état actuel de notre modèle par rapport à ces corpora.

Naturellement, les mesures portent sur les deux paramètres modélisés par le SFC : la fréquence fondamentale et la structure temporelle.

La fréquence fondamentale. Toutes les valeurs sont exprimées en demi-tons et non en Hz, car nous pensons qu'elles deviennent ainsi – au moins un peu plus – comparables entre différents corpora puisqu'elles dépendent moins du registre du locuteur. De plus, on peut soutenir qu'elles décrivent mieux la perception de la mélodie, et équilibrent ainsi mieux la pondération entre les erreurs dans le haut et dans le bas du registre – les premières ayant une influence plus importante que les dernières en échelle non-logarithmique par rapport à l'échelle en demi-tons.

Nous suivons notre modèle phonétique pour calculer les erreurs. Plus précisément, nos mesures se basent uniquement sur les trois valeurs de f_0 extraites/prédites par voyelle. Nous négligeons alors les erreurs dans les parties consonantiques qui sont induites par l'interpolation entre les noyaux vocaliques car elles sont perceptivement beaucoup moins importantes. Les chiffres ne tiennent pas non plus compte de l'erreur de stylisation par une fonction quadratique à l'intérieur des voyelles ; sachant que cette erreur est réduite grâce à la restauration de la composante microprosodique au moment de la synthèse (cf. chapitre 2.1.2 sur la page 71).

La structure temporelle. Ici, nous avons deux types de paramètres différents : le coefficient d'allongement et les durées (supra-)segmentales. Le premier est intimement lié à notre modèle phonétique – il n'offre donc aucun point de repère pour une comparaison inter-modèle. Cependant, outre sa facilité d'accès (il est une des sorties du SFC), il comporte un grand avantage : il permet la comparaison directe de toutes les syllabes (ou plus précisément : GIPCs). Notamment, il fait abstraction de la présence des pauses et rend ainsi comparable une syllabe fortement allongée et une syllabe suivie d'une pause. Ceci présuppose évidemment que sa définition permet de décrire de manière pertinente le macro-rythme.

Les mesures portant sur les durées (supra-)segmentales sont à la fois plus simples à interpréter (indépendamment du modèle phonétique) et plus difficiles à obtenir : puisque la prédiction des pauses est complètement intégrée dans la transition du coefficient d’allongement vers les durées phonémiques, les erreurs inévitables d’insertion ou d’omission d’une pause rendent la comparaison des durées des phonèmes⁵ avoisinants problématique. De plus, comme les frontières des GIPCs dépendent de la présence des pauses, leurs durées deviennent simplement incomparables là où les pauses observées et prédites ne concordent pas. Nous excluons donc de l’analyse, dans un premier temps, tous les GIPCs concernés ; les pourcentages sur les erreurs de prédiction de pauses confirmeront que ceci n’invalide pas les résultats.

Bien que nous considérions les pauses comme appartenant aux GIPCs précédents, nous comptabilisons leurs durées à part afin d’éviter la sur-pondération de leurs erreurs par rapport aux parties sonores. Nous séparons également les erreurs de durées des phonèmes/GIPCs selon la présence/absence d’une pause dans le GIPC et ce pour deux raisons :

- ceci nous permet d’évaluer la performance de l’algorithme de répartition en cas d’émergence de pause, et
- avant une pause, on peut considérer des erreurs plus grandes comme tolérables puisque les durées observées sont déjà moins fiables (la baisse d’énergie rend parfois l’étiquetage difficile) et on peut supposer dans une certaine mesure une équivalence entre allongement et pause silencieuse.

Le dernier GIPC de chaque phrase est écarté systématiquement des analyses de la prédiction de la structure temporelle : la durée de la dernière pause n’ayant aucune signification pour des phrases isolées, le SFC n’a aucune possibilité d’apprendre l’allongement du dernier GIPC d’une manière « motivée ». Il apprendra en général un allongement fort parce que les enregistrements se terminent typiquement avec une pause longue. Ce comportement est correct pour la synthèse de phrases isolées, mais il n’est pas raisonnable de calculer une erreur de « prédiction » à cet endroit.

Les mesures choisies. Nous avons retenu deux mesures : l’erreur RMS et la corrélation. La première, très courante dans la littérature, a le désavantage (ou l’avantage – selon la perspective) de donner plus de poids aux grands écarts entre observation et prédiction si on la compare aux moyennes des différences absolues. C’est une des raisons pourquoi nous découperons parfois les données (p.e. durées des GIPCs suivis ou non d’une pause) afin d’assurer que l’erreur n’est pas principalement dominée par des écarts particulièrement grands dans un sous-ensemble des données.

Bien qu’en général la corrélation augmente parallèlement à une diminution de l’erreur RMS (p.e. si l’on compare les résultats sur un corpus d’apprentissage et un corpus de test), l’information contenue dans ces deux mesures est complémentaire : une petite

⁵voir remarque 1 sur p. 22

erreur mais avec une faible corrélation peut indiquer que le paramètre en question est passablement prédit mais que le modèle n'arrive pas à établir une véritable relation entre prédicteurs et observations. De même, une grande erreur peut être relativisée par une forte corrélation qui indique que le modèle réussit néanmoins à suivre les tendances.

Effets aléatoires de l'apprentissage. L'erreur de prédiction du modèle est influencée par deux biais inévitables de l'apprentissage :

- l'entraînement des générateurs de contours implémentés en tant que réseaux de neurones n'est pas déterministe – il dépend de l'ordre de présentation des couples entrée-sortie qui est (et doit l'être) pseudo-aléatoire ;
- le choix du sous-corpus utilisé pour l'apprentissage qui peut être plus ou moins « facile » à apprendre et/ou plus ou moins apte à produire une bonne capacité de généralisation.

Afin de pouvoir juger des capacités du modèle en lui même, il est impératif de limiter dans la mesure du possible ces deux effets. En même temps, les variations « aléatoires » donnent une idée de la stabilité du modèle.

Le premier effet est simple à observer : il suffit de répéter plusieurs fois l'entraînement avec le même corpus d'apprentissage. Nous nous contentons de trois répétitions donnant lieu aux erreurs en versions « a », « b » et « c ».

Le deuxième point pourrait faire l'objet d'une analyse détaillée mais ici aussi nous nous contentons d'une procédure relativement simple : la taille du sous-corpus utilisé pour l'apprentissage sera toujours la moitié du corpus total. Nous partageons donc le corpus à trois reprises de manière aléatoire en paires de sous-corpora : A-B, C-D et E-F. L'union des sous-corpora A et B est le corpus entier – et il est de même pour les couples C-D et E-F. Si le corpus A est utilisé comme corpus d'apprentissage le corpus B fait office de corpus de test – et nous prenons soin d'inverser les rôles des sous-corpora. Nous aboutissons ainsi à $6 \times 3 = 18$ apprentissages (Aa, Ab, Ac, ..., Fc) pour lesquels nous déterminons les mesures d'erreurs sur le corpora d'apprentissage et de test et sur le corpus entier. Les moyennes de ces mesures nous donnent une idée de la performance du modèle, et leurs écart-types permettent de juger de la stabilité par rapport aux effets aléatoires. On retrouve ces chiffres sous la forme **moyenne ± écart-type** dans les tableaux de l'annexe B.

3.2.2 Prédiction et caractéristiques des générateurs

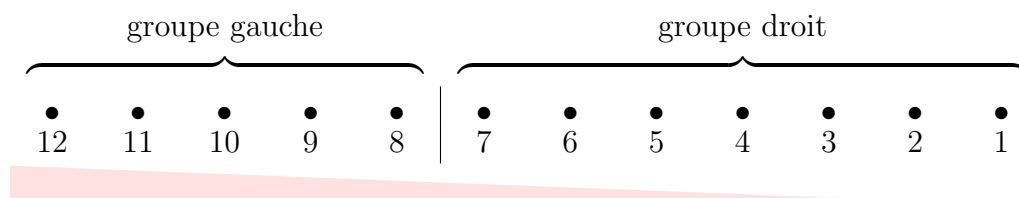
Introduction. Au lieu de discuter les erreurs mesurées d'une manière abstraite, il me semble plus intéressant de les voir « en œuvre » dans le domaine qui leur est le plus propice : l'évaluation intra-modèle. Nous allons en profiter pour approfondir quelques détails d'implémentation du SFC, concernant notamment celle des générateurs de contours. Le corpus utilisé pour cette discussion est celui des formules mathématiques en français.

Codage des entrées. Le choix de réseaux de neurones (RN) comme générateurs de contours a été guidé par leur capacité d'apprendre et de générer des contours prosodiques complexes à partir d'une entrée simple (cf. [Traber, 1992; Morlec, 1997]). Leurs sorties (trois valeurs de f_0 et le coefficient d'allongement par GIPC) sont fixées par les besoins du SFC.

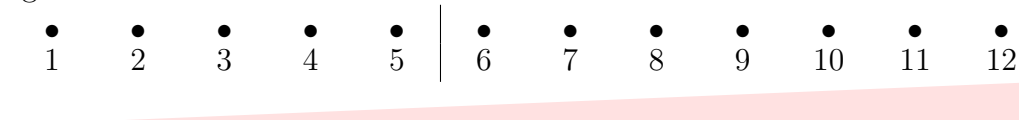
Le codage des informations phonotactiques fournies en entrée n'est en revanche pas déterminé *a priori*. Le but du codage est de permettre aux RN d'identifier chaque GIPC à l'intérieur d'un empan de configuration phonotactique⁶ donnée. Implicitement, le codage définit une métrique dans l'espace des entrées exprimant la similitude des GIPCs par rapport aux contours et détermine ainsi l'interpolation des contours.

L'idée générale, héritée du travail de Morlec [1997], est de synchroniser des rampes d'entrée avec les empan. Nous ne considérons ici que les empan liant deux groupes (p.e. associés aux fonctions « dépendance gauche » ou « dépendance droite » comme décrites sur la page 84). Nous distinguons des rampes « globales » synchronisées avec l'empan entier, des autres qui sont alignées avec les deux groupes et qui forment ainsi une double-rampe. Considérons par exemple un empan liant un groupe de 5 GIPCs à un groupe de 7 GIPCs, c.-à.-d. de configuration phonotactique « 5 + 7 ». Dans les schémas qui suivent chaque point représente un GIPC ; les nombres correspondent aux entrées du RN. Nous avons trois rampes globales :

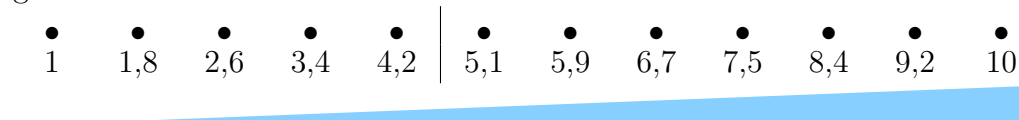
gdec : global-descendant



gasc : global-ascendant



grel : global-relatif

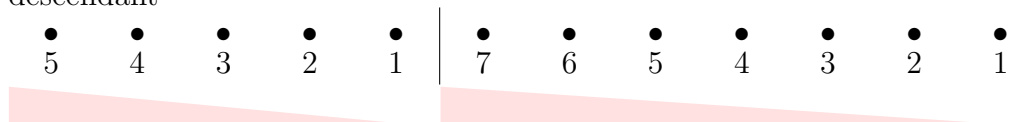


La rampe relative est normalisée de telle sorte que le premier GIPC a toujours la valeur 1 et le dernier la valeur 10 et ce afin d'assurer que toutes les entrées sont du même ordre de grandeur.

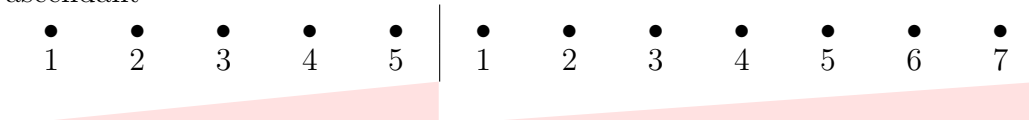
⁶Pour rappel : le nombre de GIPCs des deux groupes liés par le contour fonctionnel.

À ceci, nous ajoutons trois double-rampes :

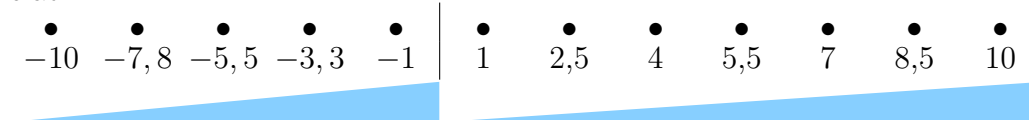
dec : descendant



asc : ascendant



rel : relatif



La normalisation de la dernière rampe, en particulier entre -1 et -10 pour la première partie, peut surprendre : le RN est ainsi informé que le GIPC en cours de traitement fait partie du premier ou du second groupe si l'entrée est positive. On aurait pu y parvenir également en ajoutant une entrée binaire.

Afin d'avoir une idée de l'efficacité du codage, nous allons évaluer des RN utilisant seulement une des rampes d'entrée. Il est évident qu'aucune des rampes ne peut seule fournir un codage satisfaisant : les rampes globales ne contiennent aucune information sur l'emplacement de la frontière entre les groupes, et les autres ne distinguent certainement pas non plus assez les empan différents (p.e. les débuts de tous les contours basés uniquement sur l'entrée **asc** seront strictement identiques – ce qui semble une contrainte trop forte).

Dans la figure 3.10, sont représentées, pour ces six RN avec une seule entrée, les erreurs et les corrélations ainsi que des barres d'erreurs qui correspondent aux écart-types discutés ci-dessus. Le trait plein relie les points relatifs aux corpus de test et le trait interrompu ceux du corpus d'apprentissage. Les renseignements que l'on peut tirer de ces « benchmarks » sont les suivants :

- La rampe **dec** fournit les meilleures prédictions. En codant l'approche vers une frontière de manière caractéristique, elle permet l'apprentissage d'un mouvement de capture.
- La rampe **asc** semble à elle seule très inefficace. Sa spécificité est d'identifier toutes les parties de contours aux débuts des groupes et ceci indifféremment du fait qu'il s'agisse du groupe gauche ou du groupe droit. Néanmoins, l'échec ici n'implique pas forcément qu'il faut abandonner ce type de rampe, car elle pourra avoir un impact positif en combinaison avec des entrées portant une information complémentaire.
- La rampe **rel**, qui réalise par sa normalisation une relation de compression horizontale entre contours longs et contours courts, ne parvient pas à exploiter l'in-

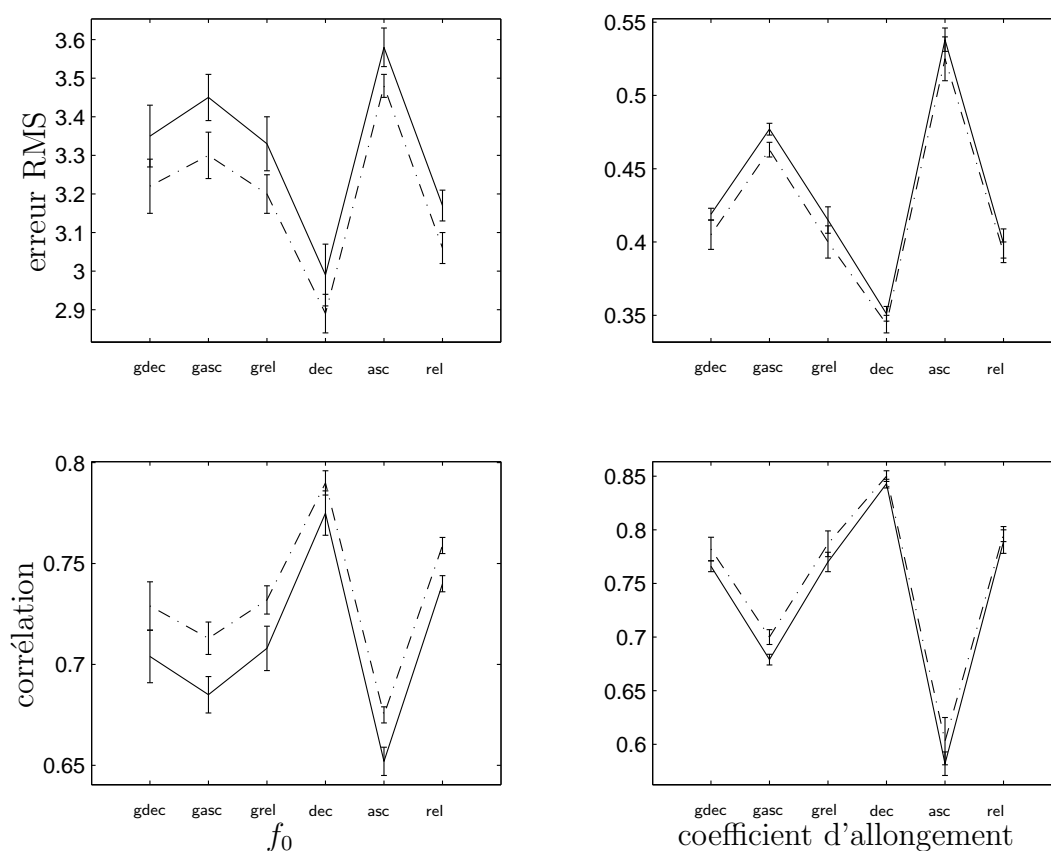


FIG. 3.10: **Erreurs de prédiction vs. types de rampe d'entrée.** À gauche sont représentées les mesures pour la fréquence fondamentale et à droite pour le coefficient d'allongement. En haut : les erreurs RMS ; en bas : les corrélations. Les mesures sur le corpus d'apprentissage sont reliées par une ligne interrompue, celles du corpus de test par une ligne continue. Ces lignes servent uniquement à guider l'œil : l'ordre de présentation des rampes est sans signification.

formation sur le groupe gauche/droit pour une prédiction comparable à celle de **dec**. Or, l'inspection des contours générés montre que ceux issus de **dec** s'annulent mutuellement entre contours DG et DD pour les groupes droits (cf. figure 3.11). Ce comportement n'est pas souhaitable, mais explique comment l'ensemble des RN **dec** contourne le manque d'information dans l'entrée.

- Des rampes globales, la rampe ascendante semble aussi la moins efficace.

Nous poursuivons notre exploration de l'efficacité du codage de l'entrée en évaluant des RNs qui disposent de deux entrées. La figure 3.12 montre les résultats pour toutes les combinaisons contenant toujours la rampe **dec** dont on vient de voir l'importance. Il est intéressant de noter les comportements différents pour la fréquence fondamentale et pour le coefficient d'allongement : l'entrée composée des deux rampes **dec** et **gdec**

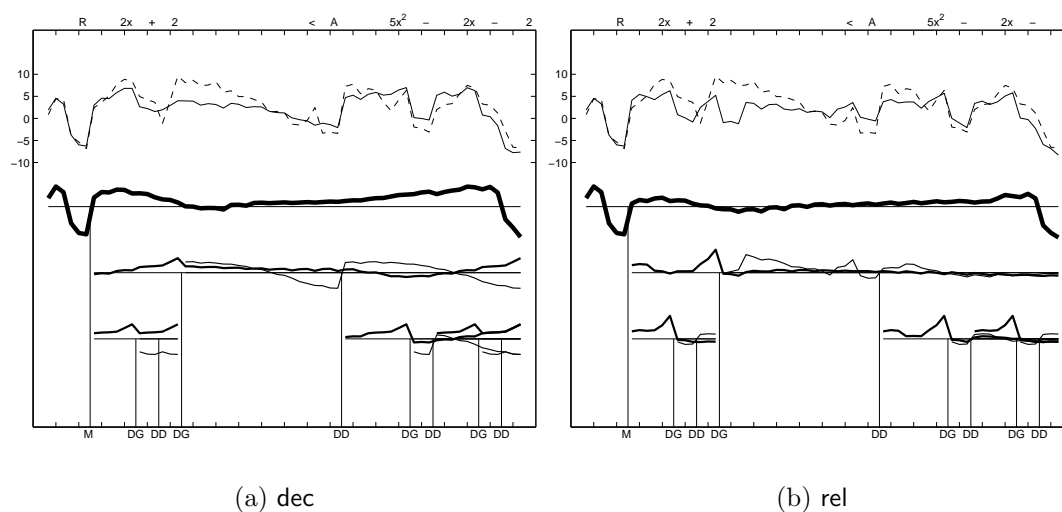


FIG. 3.11: **Exemple synthétisé avec des générateurs à entrée réduite.** Les contours fonctionnels à gauche sont générés par des RN ne disposant que d'une seule entrée : *dec*; ceux de droite sont basés sur l'information contenue uniquement dans la rampe *rel*. On s'aperçoit de la forte influence du codage de la configuration phonotactique. Ce phénomène est lié à la métrique des empans définie implicitement par le codage. Son impact est largement diminué dès que le codage est assez fin pour désambiguïser les configuration phonotactiques – ce qui n'est pas le cas avec une seule entrée.

est la plus efficace pour l'allongement mais le moins pour f_0 .

L'importance de l'entrée *gdec* pour la prédiction efficace de la structure temporelle des formules mathématiques est soulignée par les résultats de la figure 3.13 : en utilisant cinq des six entrées proposées, les erreurs et corrélations sont stables *sauf* si l'on omet ladite rampe.

Notons finalement que l'écart entre les résultats sur le corpus d'apprentissage et ceux du corpus de test se creuse avec l'augmentation du nombre d'entrées ce qui devient surtout visible en le comparant aux écart-types. Dans la pratique, il s'avère que les quatre entrées *gdec*, *asc*, *dec*, *rel* constituent un bon compromis. Tous les autres résultats présentés dans ce manuscrit ont été obtenus avec ce codage.

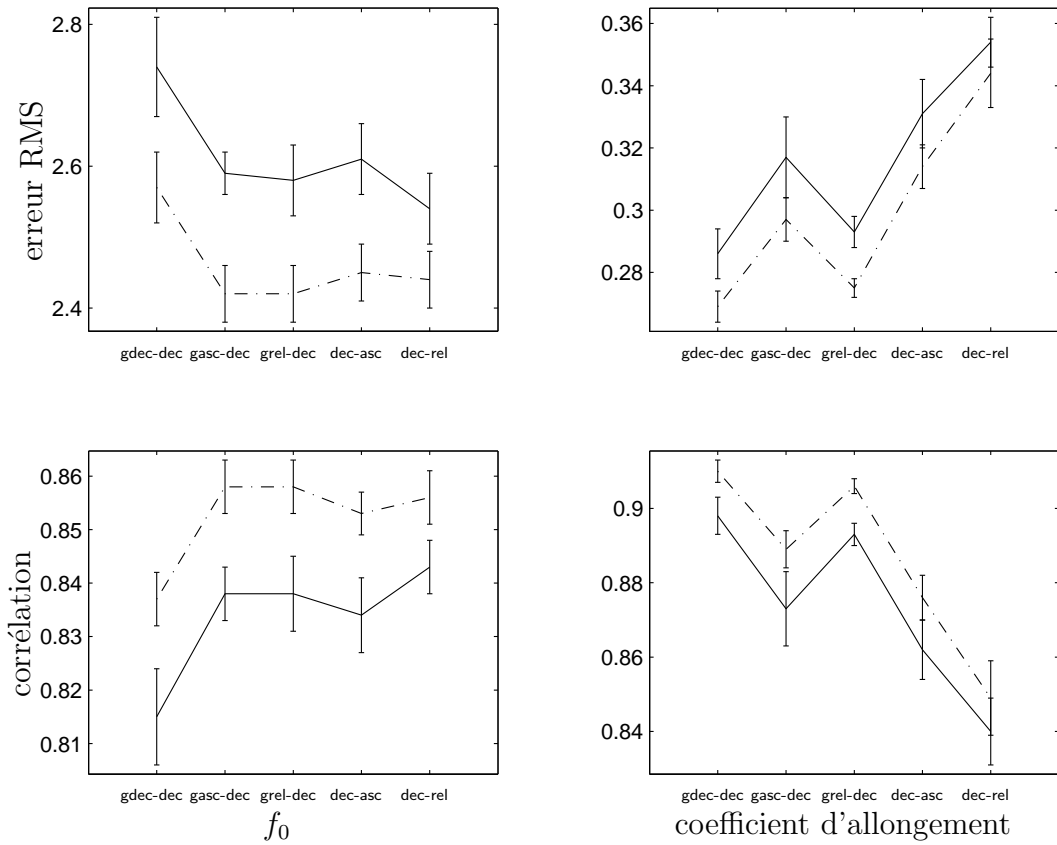


FIG. 3.12: **Erreurs de prédiction vs. types de rampe d'entrée (2 entrées).** L'importance des entrées n'est pas identique en ce qui concerne f_0 et le coefficient d'allongement : la combinaison gdec-dec, par exemple, qui donnent les meilleurs résultats pour le coefficient, semble plutôt inefficace pour la prédiction de f_0 .

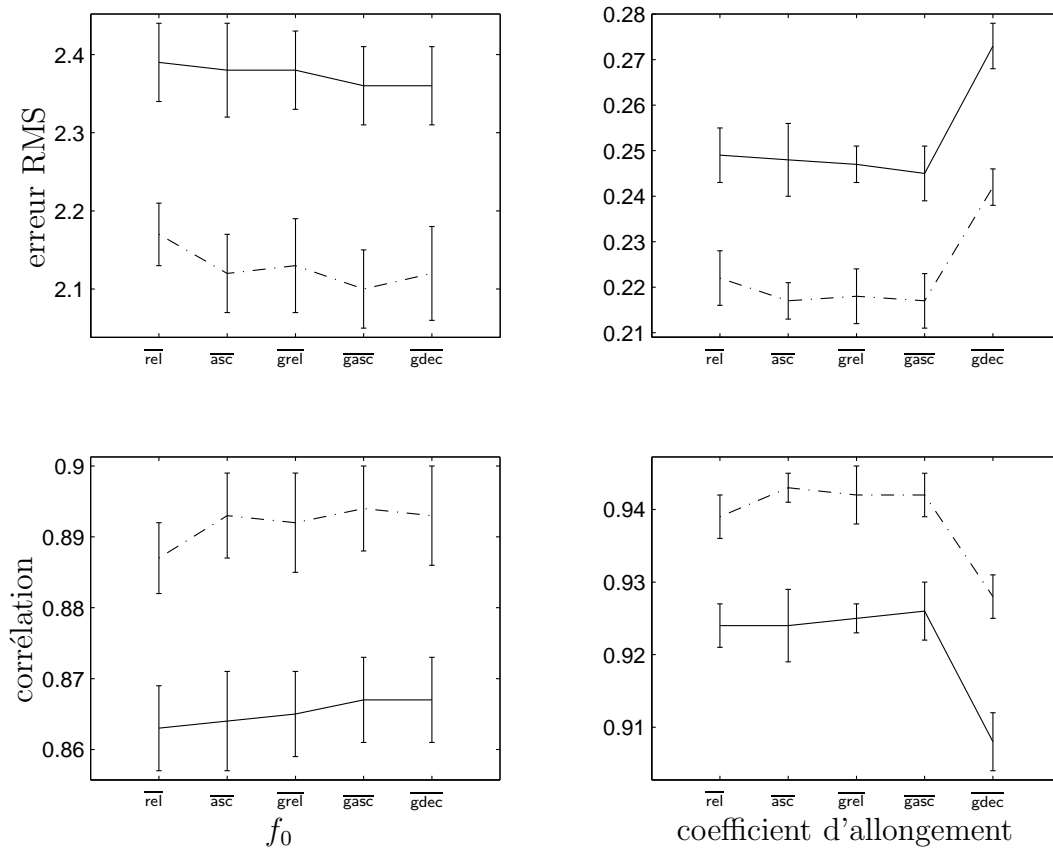


FIG. 3.13: Erreurs de prédiction vs. types de rampe d'entrée (5 entrées). L'entrée indiquée en abscisse est celle omise sur les six proposées.

Paramètres des RN. Concernant les RN en eux-même, deux paramètres influent principalement sur leur performance : la fonction d'activation des nœuds et le nombre de nœuds cachés. Après plusieurs essais, une fonction logistique⁷ semble donner les meilleurs résultats tant par rapport aux erreurs qu'à la rapidité de la convergence.

Quant aux unités cachées, nous nous proposons ici de suivre l'évolution des mesures d'erreurs en fonction de leur nombre (entre 5 et 25). La figure 3.14 montre les résultats de cette enquête sous la forme à laquelle nous nous sommes déjà habitués. Il n'est guère surprenant que l'erreur de prédiction pour le corpus d'apprentissage baisse de manière monotone en augmentant le nombre d'unités car les RN disposent simplement de plus de paramètres libres pour s'adapter aux contours vus pendant l'apprentissage. L'erreur sur le corpus de test est plus intéressante parce qu'elle exprime la capacité des RN à généraliser. Elle indique si les paramètres supplémentaires sont mis au profit des formes de contours représentatifs ou s'ils induisent plutôt un sur-apprentissage.

Compte tenu du fait qu'un nombre élevé d'unités ralentit l'apprentissage et qu'un sur-apprentissage diminuerait l'interprétabilité des familles de contours, le meilleur choix se situe entre 15 et 20 nœuds cachés (cf. surtout la partie f_0 de la figure 3.14). Nous avons retenu le nombre 17 pour la suite.

Résumé. Dans ce chapitre, nous avons suivi l'influence des paramètres de l'implémentation du SFC sur la qualité numérique des prédictions. Bien entendu, il ne s'agit là pas des seuls choix qui ont dû être faits. La fonction d'apprentissage⁸ des RN, par exemple, ou le paramétrage du cycle d'analyse par la synthèse (critères d'arrêt, paramètres de distribution de l'erreur etc.) ne sont pas sans importance.

Cependant, il me semblait particulièrement important de montrer que le codage des configurations phonotactiques pour leur faire associer des contours multi-paramétriques à travers un RN n'est pas anodin. La présentation de l'influence de la taille de la couche cachée poursuit surtout le but de nous assurer que les RN opèrent dans leur « zone verte ».

⁷fonction logistique : $a_j(t) = (1 + e^{\text{net}_j(t)+t_j})^{-1}$, où $\text{net}_j(t)$ est la somme pondérée des sorties des unités en amont.

⁸Notre choix a porté sur la « resilient backpropagation ».

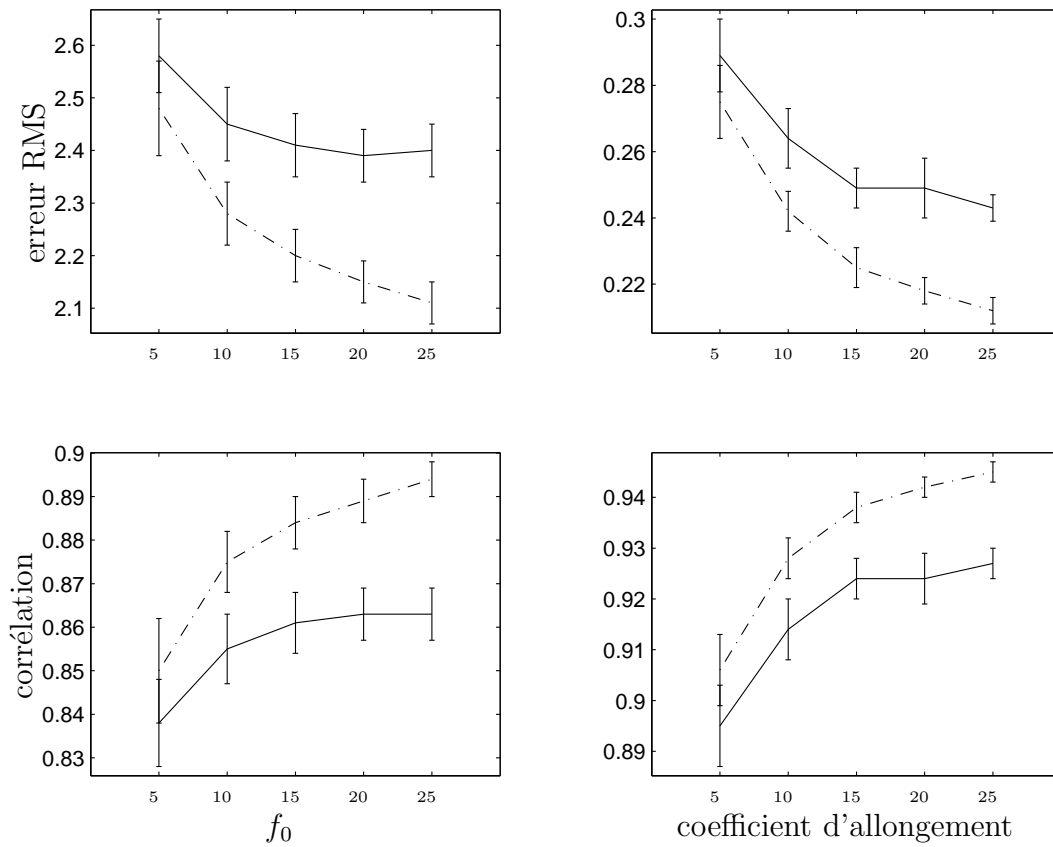


FIG. 3.14: **Erreurs de prédiction vs. nombre d'unités cachées.** L'utilisation de plus que 20 unités cachées n'améliore plus les performances par rapport au corpus de test (trait plein).

3.3 Évolution du modèle : exemple pas à pas

Après avoir suivi dans le chapitre 3.2 l'influence de quelques paramètres techniques de l'implémentation sur l'erreur objective de prédiction du modèle, je présenterai ici une approche plus descriptive des résultats fournis par le SFC.

Pour ce faire, nous allons faire un pas en arrière par rapport à la description du modèle dans le chapitre 2.3 où les fonctions communicatives que nous identifions habituellement ont été présentées. À partir du corpus des « 1001 phrases » (cf. page 117), nous allons voir comment le SFC permet de tester des hypothèses sur l'existence de fonctions différentes dans le corpus. Le modèle offre de part sa conception un outil d'*analyse* : les hypothèses (une fonction identifiée est encodée par des contours d'une même famille) sont confrontées aux données pendant le processus d'apprentissage qui fournit les reflets statistiquement significatifs de ces hypothèses. Les familles de contours obtenues sont toujours interprétables comme suit : « en présence d'une telle fonction le corpus montre une contribution moyenne de telle forme ». Il s'agit bien évidemment d'une contribution au sens du modèle superpositionnel. Nous n'allons pas décider ici si l'encodage prosodique des fonctions se fait « vraiment » par des contours additifs, mais nous insistons sur le fait qu'il s'agit d'une décomposition simple et interprétable.

Plan du chapitre. Partant d'une version très simple qui stipule que la seule fonction représentée prosodiquement dans le corpus est de signaler qu'il s'agit de déclarations, nous allons progressivement tenter d'identifier les manifestations d'autres fonctions (hiérarchisantes et de segmentation). Ainsi, nous allons rencontrer six versions du modèle :

- dc** : modèle simpliste – un seul contour par phrase : le contour porteur DC
- d** : comme dc, avec en plus : toutes les relations de dépendance sont encodées par un seul module supplémentaire D
- dd** : comme d, mais en faisant la distinction entre « dépendance gauche » (DG) et « dépendance droite » (DD)
- x** : comme dd, avec des contours supplémentaires (X) de segmentation des mots de fonction
- dv** : comme x, avec des contours spécialisés (DV) pour la relation entre auxiliaire et infinitif
- dx** : comme dv, mais : identification des contours X et DV

Le chapitre se conclura avec la présentation des erreurs objectives des différentes versions. Les données sont obtenues de la même manière que décrit au début du chapitre 3.2, mis à part le fait que nous nous contentons d'une paire de partage du corpus en parties apprentissage et test, soit un total de six apprentissages par version. Les

deux phrases utilisées pour les exemples de décomposition proviennent du corpus de test, afin d'éviter de discuter des phénomènes dus à un éventuel sur-apprentissage.

3.3.1 Un début très simple : la version dc

Cette version fait office de point de référence. Elle utilise pour chaque phrase un seul contour qui a comme empan la phrase entière. Un tel contour est semblable à une moyenne de tous les contours du corpus. Il est capable d'encoder l'attitude de l'énoncé (cf. [Morlec, 1997]). Puisqu'il s'agit d'un corpus de déclarations, nous appelons DC la famille de ces contours porteurs⁹.

L'exemple de la figure 3.15 montre le comportement typique des déclarations : une baisse de f_0 et un allongement à la fin de l'énoncé. Le creux de f_0 au début de l'énoncé peut être vu comme un artefact du corpus d'apprentissage. Mais cette interprétation impliquerait une idée *a priori* – certainement défendable dans ce cas par l'expérience avec d'autres corpora – de la forme qu'un contour de déclaration devrait avoir. L'interprétation neutre est : sous l'hypothèse qu'aucune fonction (para-)linguistique à part l'attitude « déclaration » n'est encodée prosodiquement (par des contours portés) dans le corpus, cette attitude se manifeste (entre autres) par un creux de f_0 en début d'énoncé. Nous verrons plus loin que ce creux peut s'expliquer de manière bien plus satisfaisante en tant que manifestation d'une autre fonction.

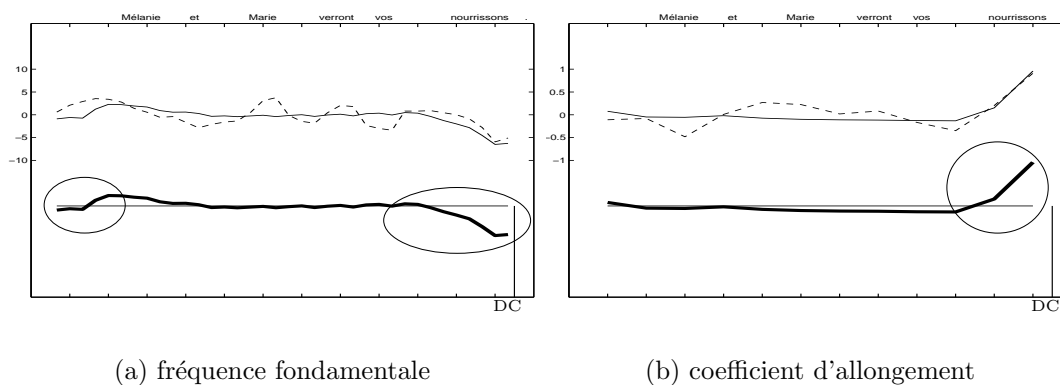


FIG. 3.15: **Exemple avec la version dc.** On observe le comportement typique de déclaration : baisse de f_0 et allongement en fin d'énoncé, mais également un comportement « surprenant » : un creux de f_0 en début d'énoncé.

⁹La notion de « porteur » n'a évidemment un sens seulement dès que nous ajoutons des contours « portés » dans les versions suivantes du modèle.

3.3.2 Ajouter une relation de dépendance : la version d

Par rapport à la précédente, la version d ajoute des contours D encodant les relations de dépendance. Là où nous identifions deux groupes syntaxiques adjoints dont un domine l'autre nous prévoyons un contour porté qui est coextensif aux groupes en question et qui est ancré à leur frontière (cf. page 84). Nous ne distinguons pas ici le type de dépendance, c'est-à-dire lequel des deux groupes domine l'autre. Cette version offre la possibilité de capturer des moyens prosodiques de structuration, mais – comme nous allons voir – d'une manière trop grossière.

La figure 3.16 montre le progrès de prédiction fait par rapport à la version dc, notamment à la frontière entre le sujet « Mélanie et Marie » et le groupe verbal « verront vos nourrissons » (cf. encerclé). Néanmoins, l'accent appris entre ces deux groupes semble – au moins pour cet exemple – trop faible. Avec la version suivante, nous tenterons de savoir si la situation s'améliore en supposant qu'il y a deux fonctions distinctes selon le type de dépendance.

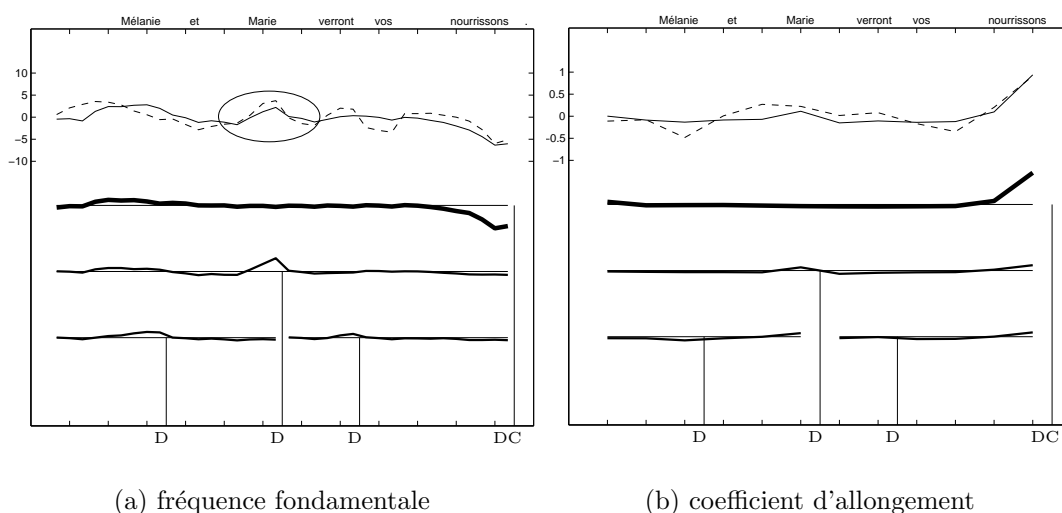


FIG. 3.16: **Exemple avec la version d.** La capture de la structuration prosodique est assez grossière.

3.3.3 Diversifier les dépendances : la version dd

Il s'agit ici de la version « standard ». En plus de la fonction « signaler l'attitude de déclaration » (DC), nous supposons qu'il y a une fonction DG signalant que de deux groupes adjoints celui de *gauche* dépend de celui de droite, et une fonction DD de dépendance *droite*. D'un point de vue décompositionnel, ceci revient à dire que la famille des contours D est séparée en deux : DG et DD.

Comme le montre la figure 3.17, l'accent à la fin de « Marie » est maintenant bien appris. Nous en concluons que la distinction entre les types de dépendance est bénéfique, c'est-à-dire les deux fonctions ont des manifestations prosodiques distinguables. Autrement dit, la version *d* empêche l'émergence de contours caractéristiques en regroupant les deux fonctions.

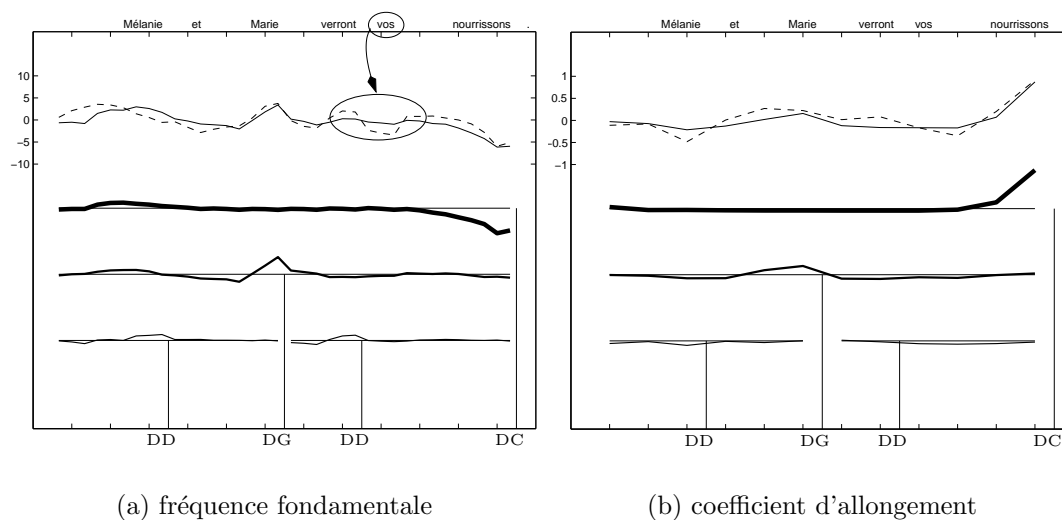


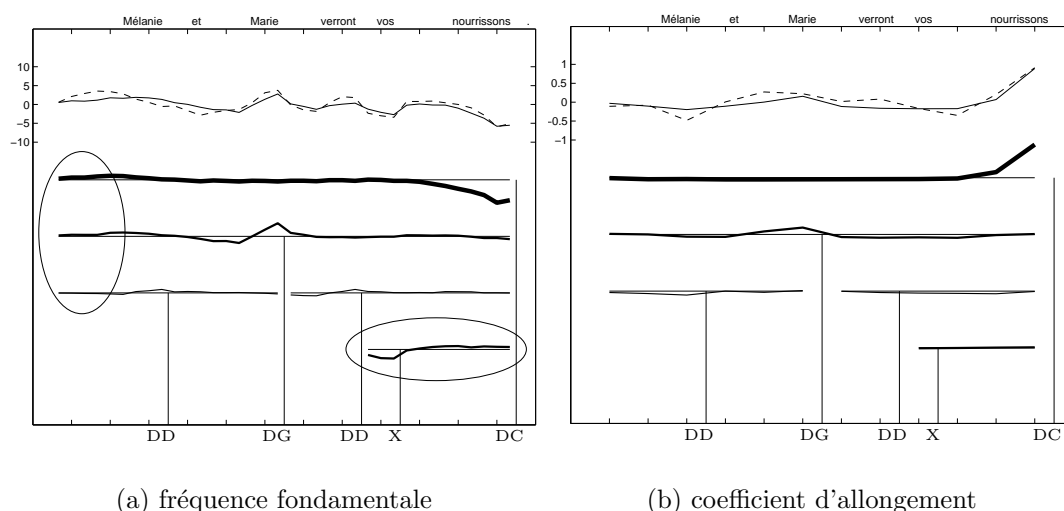
FIG. 3.17: **Exemple avec la version dd.** En distinguant entre dépendance *droite* et *gauche* les contours s'améliorent. Cette version n'est pas capable de générer correctement la désaccentuation de « vos ».

Il va de soi que de telles conclusions ne peuvent pas être tirées à partir d'un seul exemple. Le corpus en entier doit être pris en considération – nous le ferons sommairement un peu plus loin en discutant les erreurs de prédiction (cf. page 141). Pour suivons pour l'instant l'analyse de notre exemple. Il a été choisi parce qu'il montre un phénomène très intéressant que le modèle ne peut pas encore traiter. Il s'agit de la segmentation entre un article ou pronom et le mot suivant. Nous observons que le modèle échoue systématiquement à générer le creux mélodique sur l'article ou le pronom – comme encadré dans la figure 3.17.

3.3.4 Une fonction de segmentation : la version x

Nous supposons ici qu'il existe une fonction supplémentaire qui sert à segmenter un article ou un pronom du mot auquel il est attaché. Nous ajoutons donc une famille de contours, appelée X, au modèle. Il s'agit d'une fonction « locale » : elle n'agit que sur les deux mots en question qui formeront donc l'empan des contours. Le point d'ancrage est la frontière entre les mots.

Les bénéfices du contour supplémentaire dans la figure 3.18 pour l'évolution de f_0 sur « vos » sont évidents. Il faut noter également que l'ajout des contours X améliore aussi les autres familles de contours : le creux de f_0 en début d'énoncé auquel nous avons fait allusion en 3.3.1 et qui persistait pour les versions précédentes dans les contours de DC, DG et DD, s'est résorbé. En effet, le corpus contient un grand nombre de phrases commençant par un article. L'exemple étudié ici commence par le mot plein « Mélanie » et ne comporte pas le creux typique que nous attribuons à la segmentation de l'article. En admettant donc l'hypothèse de l'existence de la fonction X, nous pouvons « nettoyer » les autres contours qui avaient auparavant absorbé les reflets « accidentels » de la première.



(a) fréquence fondamentale

(b) coefficient d'allongement

FIG. 3.18: **Exemple avec la version x.** Par l'ajout de la famille de contours X, le modèle est capable de générer correctement la mélodie de « vos ». Les autres contours se trouvent « nettoyés » (voir : le « cercle » gauche).

L'exemple de la figure 3.19 montre cependant un problème de la version x. La phrase a été choisie parce qu'elle contient deux contours DD avec les mêmes configurations phonotactiques, c'est-à-dire liant dans les deux cas un groupe de deux syllabes et un groupe de trois syllabes – il s'agit donc de deux contours identiques. Or, la courbe observée de f_0 semble avoir un comportement bien différent aux deux occurrences : une montée sur « loup » et un creux à la fin de « pourra ». Sont-ce deux manifestations variantes de la même fonction ou est-ce qu'il convient de distinguer une nouvelle fonction ?

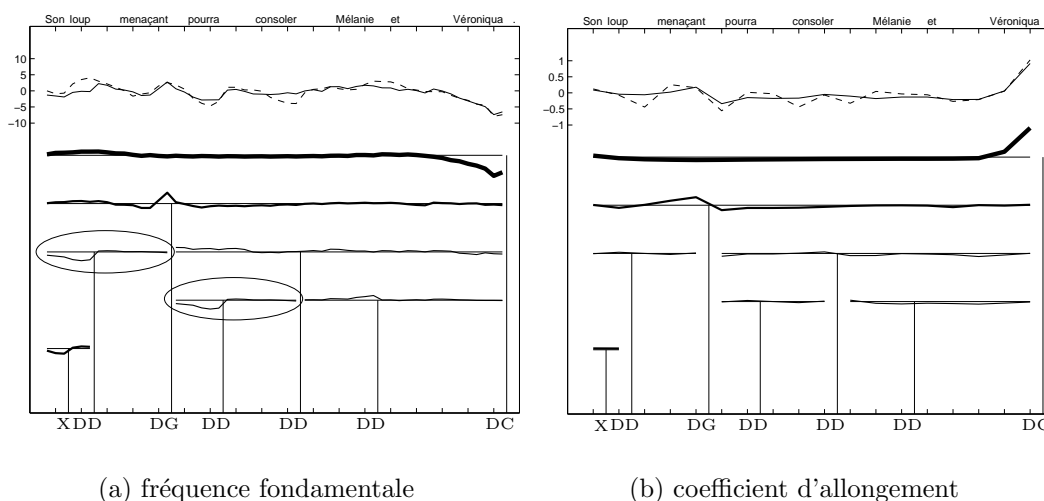


FIG. 3.19: **Exemple avec la version x.** Les deux contours encadrés sont identiques (même fonction, même empan, même ancrage). Est-il raisonnable de soutenir qu'ils encodent la même fonction ?

3.3.5 Identifier une fonction spécialisée : la version dv

Dans ce paragraphe nous allons tester l'hypothèse que notre locuteur exprime une fonction supplémentaire qui servirait à indiquer la relation particulière entre un verbe auxiliaire et un infinitif. Nous l'appelons DV. Dans la version précédente nous avons présumé qu'il s'agissait simplement d'une relation de dépendance droite.

Le résultat pour la même phrase après l'apprentissage des désormais cinq modules (DC, DG, DD, X, DV) montre une nette amélioration de la prédiction (cf. figure 3.20).

Afin d'apprécier le changement provoqué par la prise en compte de la fonction DV, il est intéressant de regarder les familles de contours concernées. La figure 3.22(a) illustre un extrait de la famille DD sous l'hypothèse que la relation entre auxiliaire et infinitif appartient à la fonction « dépendance droite » (version x). Sur les neuf contours on fait varier de gauche à droite le nombre de syllabes du groupe gauche (de 1 à 3), et de haut en bas le nombre de syllabes du groupe droit (de 2 à 3). Dans le cadre de notre modèle, ce genre de famille est *la* manifestation prosodique d'une fonction (para-)linguistique. La représentation graphique permet d'observer la morphologie de la fonction selon les contraintes phonotactiques.

Dans la version dv, cette famille est séparée en deux : DD et DV (cf. les figures 3.22(b) et 3.22(c)). Rappelons ici le principe de l'inversion du modèle (apprentissage) : la procédure cherche pour chaque catégorie (famille) les meilleurs contours paramétrisés par les caractéristiques phonotactiques. Si l'attribution des contours aux catégories est mauvaise (comme c'était le cas dans la version x avec l'identification des

catégories DD et DV, ou dans la version **d**, où nous n'avons pas distingué DG et DD), les « meilleurs » contours ne peuvent être que des compromis. Et c'est exactement ce que l'on peut observer par exemple pour les contours 2+2 (encadrés dans les figures) : le contour DD en version **dv** (figure 3.22(b)) montre une légère montée en f_0 avant la frontière de groupes et le contour DV (figure 3.22(c)) est caractérisé par un saut de f_0 à la frontière. Le contour DD en version **x** (figure 3.22(a)) est le compromis attendu. Il n'est donc pas étonnant qu'il soit mal adapté – aussi bien pour les instances que nous associons maintenant à la nouvelle fonction DV que pour celles où nous continuons de supposer l'action de la fonction DD.

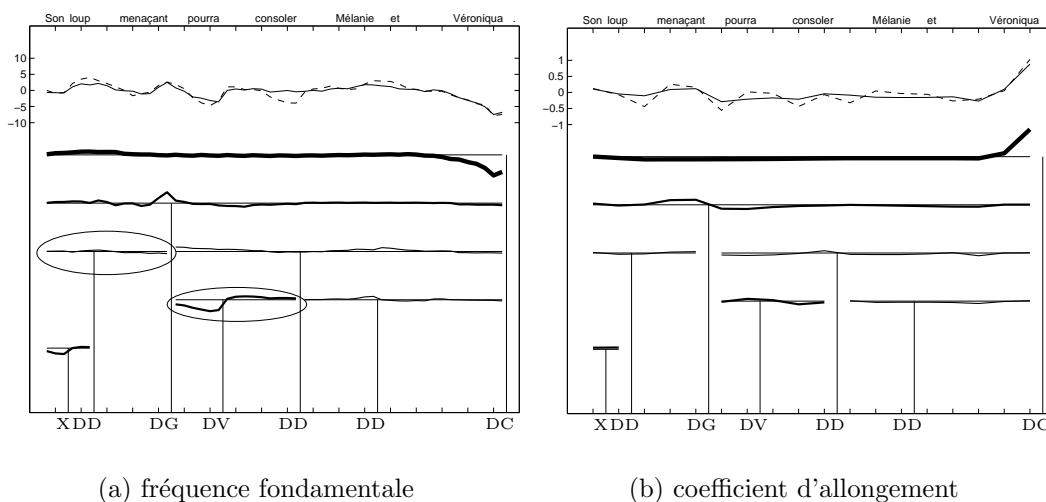


FIG. 3.20: **Exemple avec la version **dv****. L'hypothèse d'une fonction spécialisée pour exprimer la relation entre verbe auxiliaire et infinitif améliore la prédiction.

3.3.6 Réinterprétation unifiante : la version **dx**

Jusqu'ici nous avons ajouté avec chaque nouvelle version une nouvelle famille de contours. Il peut sembler évident que ceci améliore la prédiction prosodique au moins pour les phrases qui ont incité par le résultat de leur analyse le changement du modèle : avec chaque module de génération ajouté, nous augmentons le nombre de paramètres disponibles pour approximer les contours observés. Il est hors de question de procéder ainsi à l'infini et d'ajouter des petits contours de ci et de là afin de satisfaire le moindre cas particulier. Si jusqu'ici la discussion de l'évolution du modèle a peut-être donné cette impression, il s'agit d'un défaut de présentation. L'utilisation de seulement deux phrases d'exemple me semblait préférable pour mieux suivre les changements provoqués. Néanmoins, la décision d'ajouter ou non un type de contours doit toujours

être prise en vue du corpus entier et sans négliger la vocation des contours : encoder prosodiquement des fonctions (para-)linguistiques *distinctes*.¹⁰

La version dx que nous présentons dans ce paragraphe emprunte en quelque sorte un chemin inverse : incités par la ressemblance morphologique des familles DV et X (cf. les figures 3.22(c) et 3.23(a)), nous testons l'hypothèse d'une possible unification des deux familles. Ceci revient à réinterpréter la fonction DV comme une « simple » segmentation entre auxiliaire et infinitif qui serait alors du même type que la segmentation des articles et pronoms. La figure 3.23(b) montre la nouvelle famille X : les deux anciennes familles s'y trouvent réunies sans (mauvaise) surprise. L'examen de notre phrase d'exemple (cf. la figure 3.21) s'avère également réconfortante : les contours ne semblent pas dégradés. Cependant, le creux de f_0 à la fin de « consoler » pourrait faire l'objet de futures investigations.

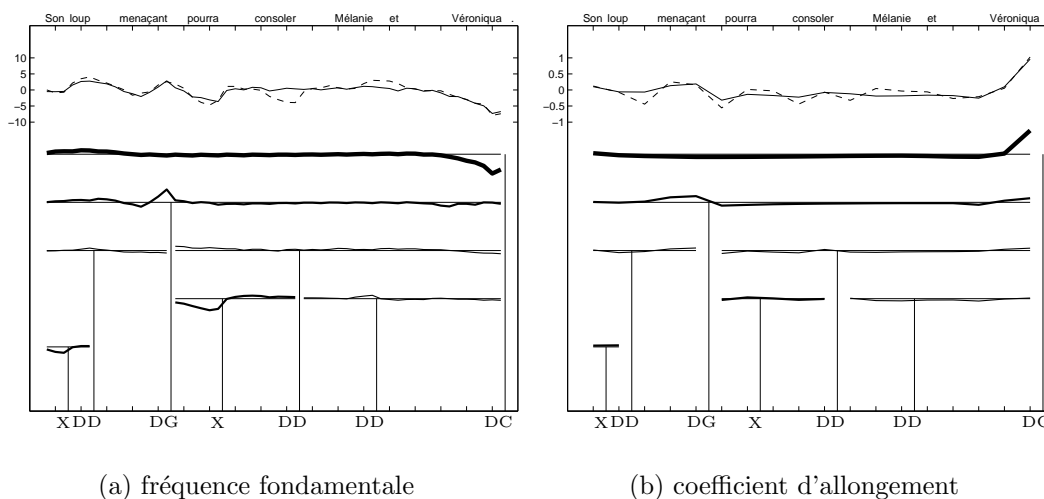


FIG. 3.21: Exemple avec la version dx.

¹⁰Une vocation qui se cache vraisemblablement trop bien derrière les acronymes comme DC, DD, X, etc.

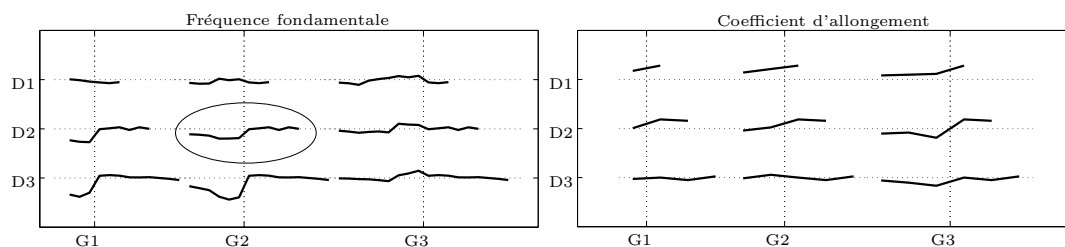
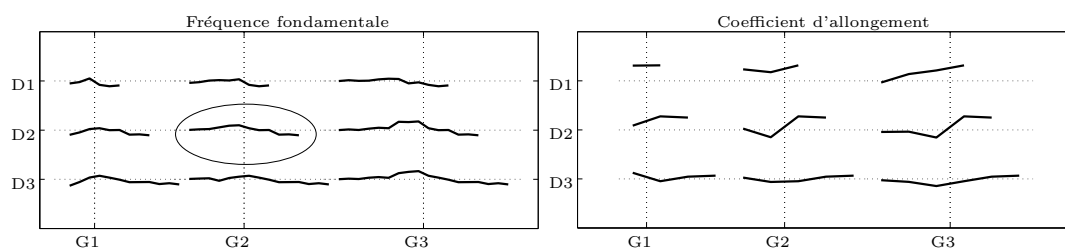
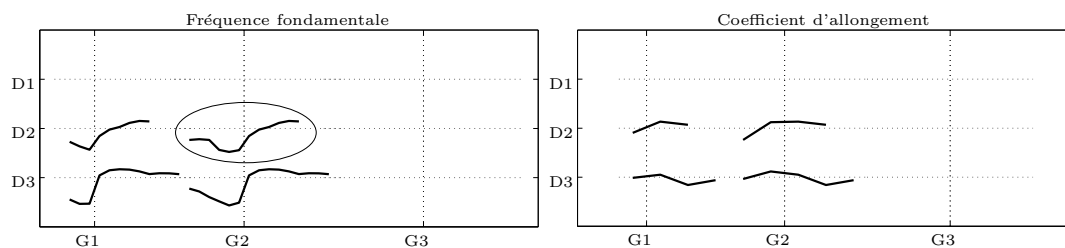
(a) Famille de contours DD (version x).(b) Famille de contours DD (version dv).(c) Famille de contours DV (version dv).

FIG. 3.22: **Familles de contours.** Les contours sont représentés ici en fonction des contraintes phonotactiques : la longueur du groupe gauche varie entre 1 et 3 syllabes (G1–G3) ainsi que celle du groupe droite (D1–D3). Seuls les contours ayant des occurrences dans le corpus sont montrés : le corpus ne contient aucun contour DV avec un groupe droit d’une seule syllabe ni avec un groupe gauche de trois syllabes. Les contours DD appris en version x (cf. (a)) sont des « compromis » entre les contours DD et DV en version dv (cf. (b) et (c))

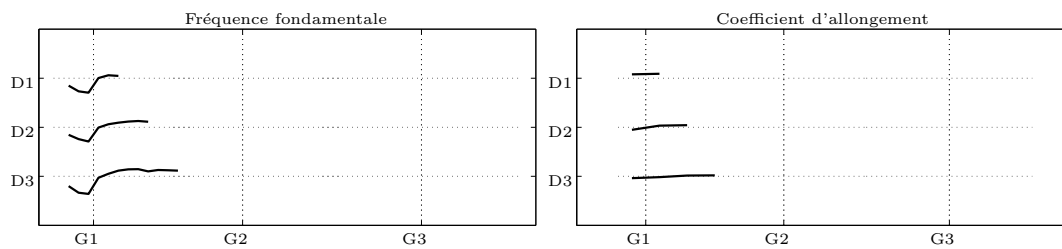
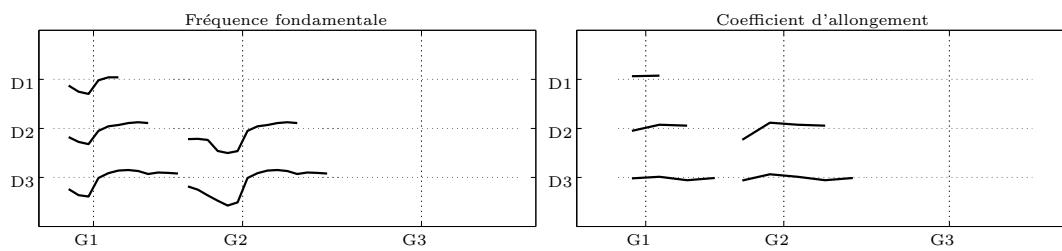
(a) version **dv**(b) version **dx**

FIG. 3.23: **Familles de contours X.** La famille X en version dx (cf. (b)) unit les familles X et DV en version dv (cf. (a) et figure 3.22(c).)

3.3.7 Les erreurs de prédiction

Comme annoncé précédemment, nous allons maintenant essayer d’obtenir une vue plus globale des effets provoqués par les changements du modèle : comment évolue l’erreur de prédiction ? Les courbes de la figure 3.24 sont du même type que celles du chapitre 3.2 sur l’impact de l’implémentation des générateurs, à la différence près que les barres d’erreur ne correspondent ici qu’à six apprentissages différents et que nous avons ajouté la courbe correspondant au corpus entier (celle du milieu).

On peut observer que les erreurs du corpus de test ne sont que légèrement supérieures à celle du corpus d’apprentissage – et ceci pour toutes les versions du modèle. Cela montre

- que nous avons évité un sur-apprentissage, et
- que l’augmentation du nombre de familles disponibles n’a pas entamé la faculté de généralisation du modèle.

La baisse particulièrement importante de l’erreur de la version **dc** vers la version **d** était attendue en vue de la simplicité de la première. Mais l’ajout des familles **DD**, **X** et **DV** n’a pas été vain non plus. Les erreurs des versions **dv** et **dx** sont comparables – l’unification des deux familles semble donc bien justifiée par le gain en simplicité du modèle.

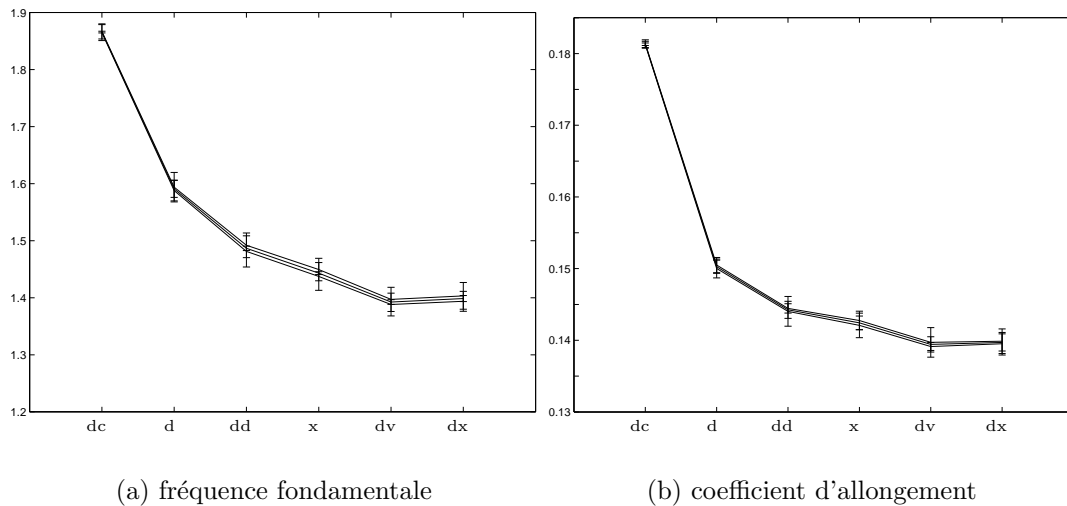


FIG. 3.24: **Les erreurs de prédiction en fonction de la version du modèle.** Pour chaque paramètre (f_0 , coefficient d’allongement) les trois courbes montrent l’évolution de l’erreur pour les différentes versions du modèle. La courbe du milieu a été obtenue sur le corpus entier. Celles de bas et de haut correspondent respectivement au corpus d’apprentissage et au corpus de test. Les barres d’erreurs indiquent les écart-types des erreurs entre les 6 répétitions d’apprentissage.

3.3.8 Résumé

Dans ce chapitre, nous avons suivi l'évolution du modèle dans le sens où nous avons successivement affiné nos hypothèses concernant les fonctions réalisées dans le corpus par notre locutrice. La présentation a été volontairement exagérée en commençant l'analyse sous l'hypothèse que seule la modalité des énoncés aurait été signalée. En effet, la préparation d'un corpus est guidée par les fonctions que nous souhaitons analyser afin d'équilibrer la couverture des différentes configurations phonotactiques. Nous allons revenir sous peu (dans le chapitre 3.5) sur ce même corpus en proposant une analyse plus détaillée d'une fonction.

Le but ici était de souligner l'interprétabilité des analyses fournies par le SFC. Le cycle d'analyse par la synthèse extrait à partir des observations les contributions des différentes fonctions dans le cadre du SFC, qui est la superposition de contours. Les familles de contours sont le reflet direct et interprétable de cette décomposition.

Mis à part ces considérations sur la méthode, le résultat le plus marquant de cette analyse concerne la fonction de segmentation, appelée ici X. Mon interprétation n'est pas qu'il *existe* une telle fonction, mais que supposer son existence *parvient à mieux expliquer les observations*. Bien qu'avancée prudemment, cette interprétation reste également valable pour les autres corpora de phrases isolées en français que nous avons examinés. C'est sous cet angle que nous rejetons l'hypothèse d'une fonction spécifique indiquant la relation entre auxiliaire et infinitif : elle n'est pas nécessaire à l'explication des données.

3.4 Expansion du mouvement

Le modèle SFC est un successeur du modèle proposé par Morlec [1997]. Un des points forts de ce modèle était la mise en évidence de l'expansion de mouvement des contours porteurs. Rappelons que chez Morlec l'apprentissage est hiérarchique : les contours porteurs sont d'abord appris en l'absence d'information sur les éventuels contours portés. L'expansion des mouvements montrée par Morlec se base sur l'apprentissage avec des phrases « monomots » seulement. Il évite ainsi l'influence des contours portés. Ici, les contours sont issus d'un apprentissage avec une moitié, déterminée aléatoirement, du corpus – les monomots en constituent alors une partie très petite. Pourtant, comme nous allons voir, les résultats ne s'en voient pas dégradés, car le SFC incorpore dès le départ la négociation entre contours de types différents. Ceci conforte les arguments en faveur de cette négociation, discutés dans le chapitre précédent, p.e. l'effacement du creux mélodique en début du contour porteur grâce à l'introduction d'une nouvelle famille de contours, appelée X, qui l'absorbe de manière plus satisfaisante.

Les figures 3.25 et 3.26 montrent que les observations de Morlec [1997, pp. 86 à 89] restent globalement valables dans le cadre du SFC ; par exemple pour le doute-incrédulité : la mise en relief de l'avant-dernière syllabe, le mouvement montant sur la dernière et l'allongement plus fort en fin d'énoncé que pour DC.¹¹

Afin de donner une idée de l'expansion à plus grande échelle, nous avons recours au corpus « 1001 phrases ». La figure 3.27 montre que les contours de déclaration pour cette locutrice sont semblables à ceux du corpus « attitudes ». Il est à noter que les contours de 21 à 30 syllabes dépassent le cadre des configurations présentes dans le corpus. D'une manière générale, il ne faut pas s'attendre à des miracles en ce qui concerne l'extrapolation par les générateurs de contours, mais ici la généralisation se passe plutôt bien, parce que les contours sont simples et parce que le corpus fournit une base d'apprentissage très solide jusqu'à 20 syllabes.

¹¹Le coefficient d'allongement du dernier énoncé dépend artificiellement de la pause suivante qui est sans signification puisqu'il s'agit d'enregistrements de phrases isolées. Nous nous sommes assurés que l'effet observé *n'est pas* dû à cette pause.

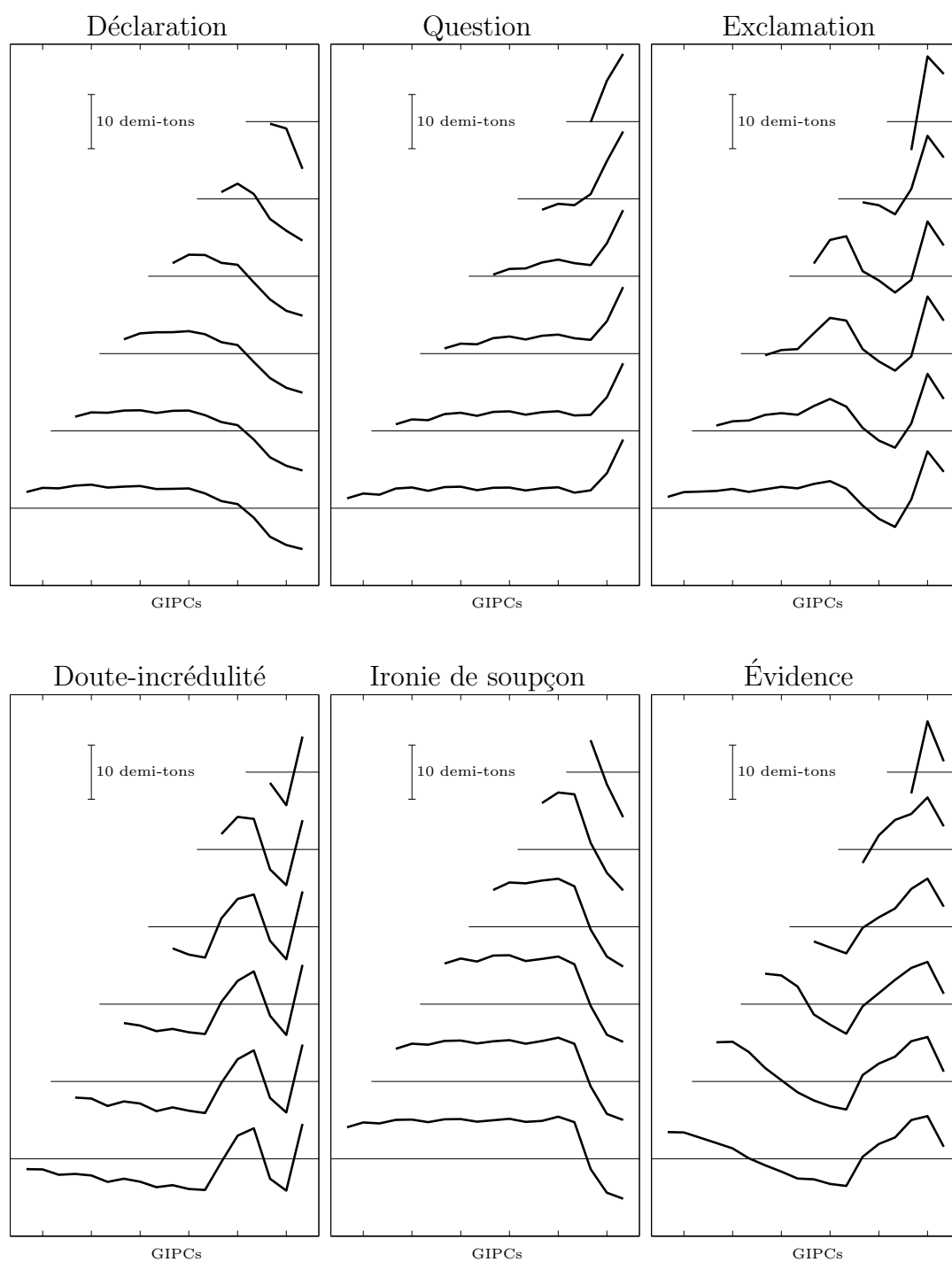


FIG. 3.25: **Expansion de mouvements porteurs. Fréquence fondamentale.** L'évolution des contours porteurs pour des phrases de 1 à 6 syllabes (de haut en bas) est montrée par modalité. Les lignes horizontales correspondent à la fréquence de référence (126 Hz). L'échelle (logarithmique) est indiquée en haut à gauche de chaque sous-figure.

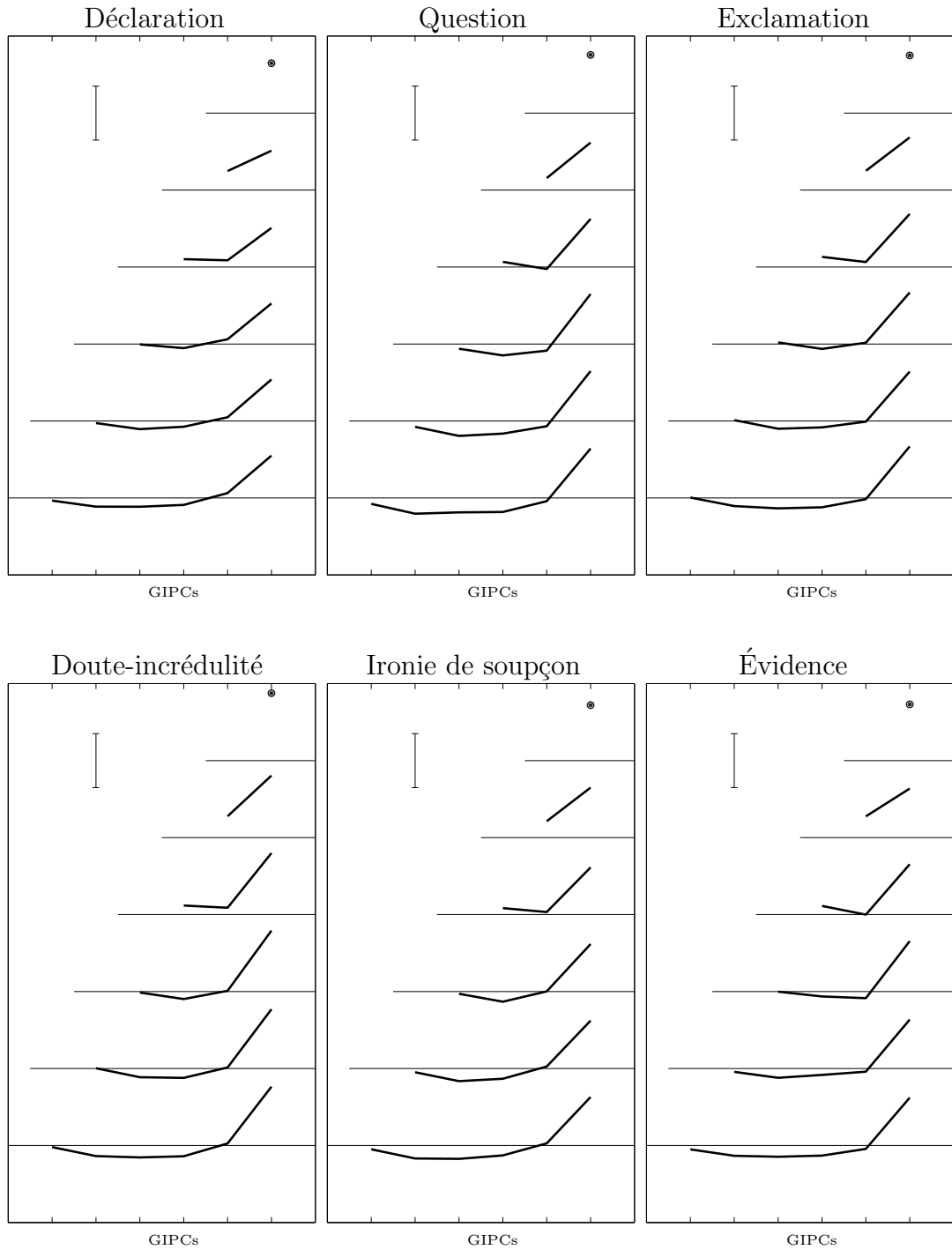


FIG. 3.26: **Expansion de mouvements porteurs. Coefficient d'allongement.** L'indicateur d'échelle en haut à gauche correspond à une unité du coefficient d'allongement. Les contours se ressemblent beaucoup entre modalités. On peut noter un débit légèrement plus élevé (i.e. allongement plus faible) pour les questions.

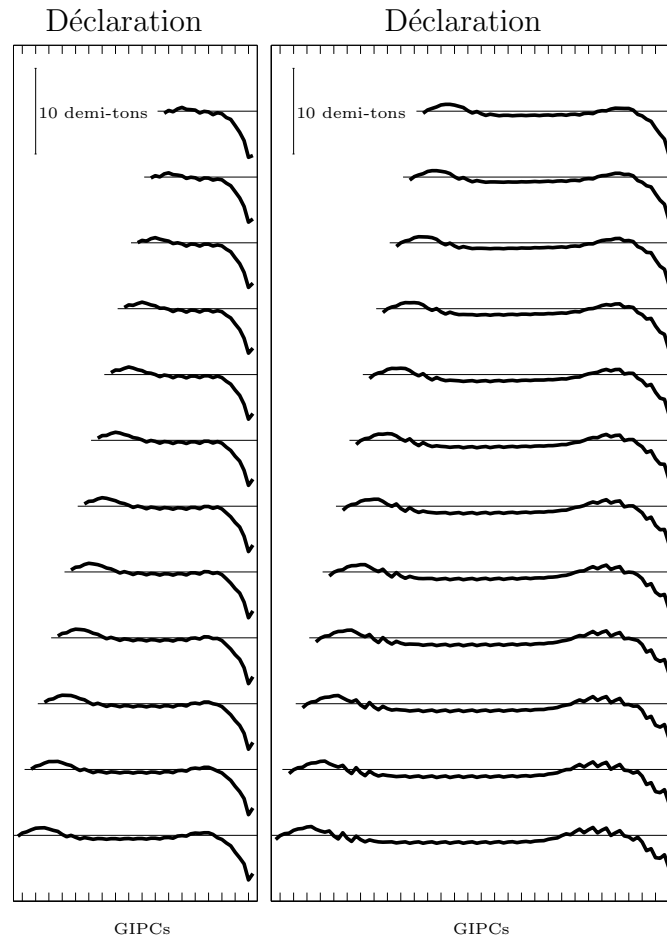


FIG. 3.27: **Extrapolation de mouvements porteurs. Fréquence fondamentale.** Les contours présentés ici sont issus du corpus « 1001 phrases ». Les dix derniers contours sont une « pure » extrapolation par rapport aux configurations du corpus. On voit apparaître des artefacts (modulations rapides), mais ceci reste tolérable.

3.5 Fonction de qualification et nature du qualifiant

Dans ce paragraphe, nous allons investiguer si et dans quelle mesure la nature d'un groupe qualificatif influe sur les contours observés. Plus précisément, nous allons regarder les syntagmes nominaux en position initiale (donc de sujet) qui se composent d'un substantif (avec article ou pronom possessif) qualifié par un adjectif, un complément avec préposition ou une subordonnée relative.

Le SFC dans la forme sous laquelle nous l'avons présenté émet l'hypothèse que la relation de qualification (à droite) est encodée par des contours identiques (modulo ses caractéristiques phonotactiques, c.-à.-d. le nombre de syllabes de qualifié et qualifiant). Bien que les contours doivent toujours être considérés dans leur intégralité, la discussion des différences porte ici principalement sur le comportement du GIPC préfrontalier. Il n'est guère surprenant que c'est à cet endroit que des éventuelles différences soient le plus visibles.

Le corpus. Le corpus qui a été construit pour étudier, entre autre, cette hypothèse, est celui des « 1001 phrases » (cf. 3.1.3). D'une manière générale, les contours sont seulement comparables s'ils ont les mêmes empan. Ceci pose une grande exigence au corpus : pour une configuration phonotactique d'empan donnée, il doit contenir assez d'exemples pour *toutes* les alternatives de l'opposition qui nous intéresse. Malheureusement, cette condition n'est pas tout à fait remplie : pour la relation de qualification à droite, seules les configurations 2+3 et 4+2 sont représentées pour les trois alternatives. De plus, la subordonnée relative n'existe qu'en un seul exemple pour chacune de ces configurations ce qui ne permettrait pas de tirer une conclusion appuyée. Or, j'espère pouvoir montrer que la confrontation paire par paire – c.-à.-d. « adjectif » vs. « complément » et « complément » vs. « subordonnée » – offre une interprétation cohérente des différences observées. Les contours prédits par le SFC nous seront d'un grand recours en tant que référence.

La présentation. À l'opposé de l'analyse proposée dans le chapitre 3.3, où nous avons pu illustrer nos propos avec deux phrases typiques, il me semble ici important de fournir une vue de l'ensemble des contours caractéristiques. Les figures des pages 149 à 157 proposent donc un tour guidé (sous forme de « bandes dessinées ») des contours de la fréquence fondamentale ainsi que du coefficient d'allongement. Outre les deux configurations 2+3 et 4+2 déjà citées, j'ai choisi toutes les configurations représentées dans le corpus qui ont des empan de 7 syllabes, car elles couvrent le mieux *toutes* trois alternatives.

Les figures sont triées par configuration et les GIPCs préfrontaliers sont encadrés. Les contours observés sont en traits pointillés et les contours prédits par le SFC en trait plein. Ces derniers fournissent la référence de comparaison car ils se basent sur l'hypothèse que les contours associés à la relation de qualification sont identiques. Ils

incorporent évidemment les contours porteurs et, le cas échéant, des contours portés d'empans plus petits. Ainsi, ils permettent de mettre en évidence des différences attribuables directement à l'opposition qui nous intéresse : la nature du qualifiant. Évidemment, chaque observation unique peut contenir des variations diverses ; néanmoins, les contours prédits incorporent dans la mesure du possible – c'est-à-dire : compte tenu du corpus entier – les influences présumées connues.

Notez enfin que le dernier GIPC des exemples n'appartient plus au groupe considéré. Nous l'avons inclus dans les graphiques pour montrer que le contour de groupe est bien accordé au contour global de phrase.

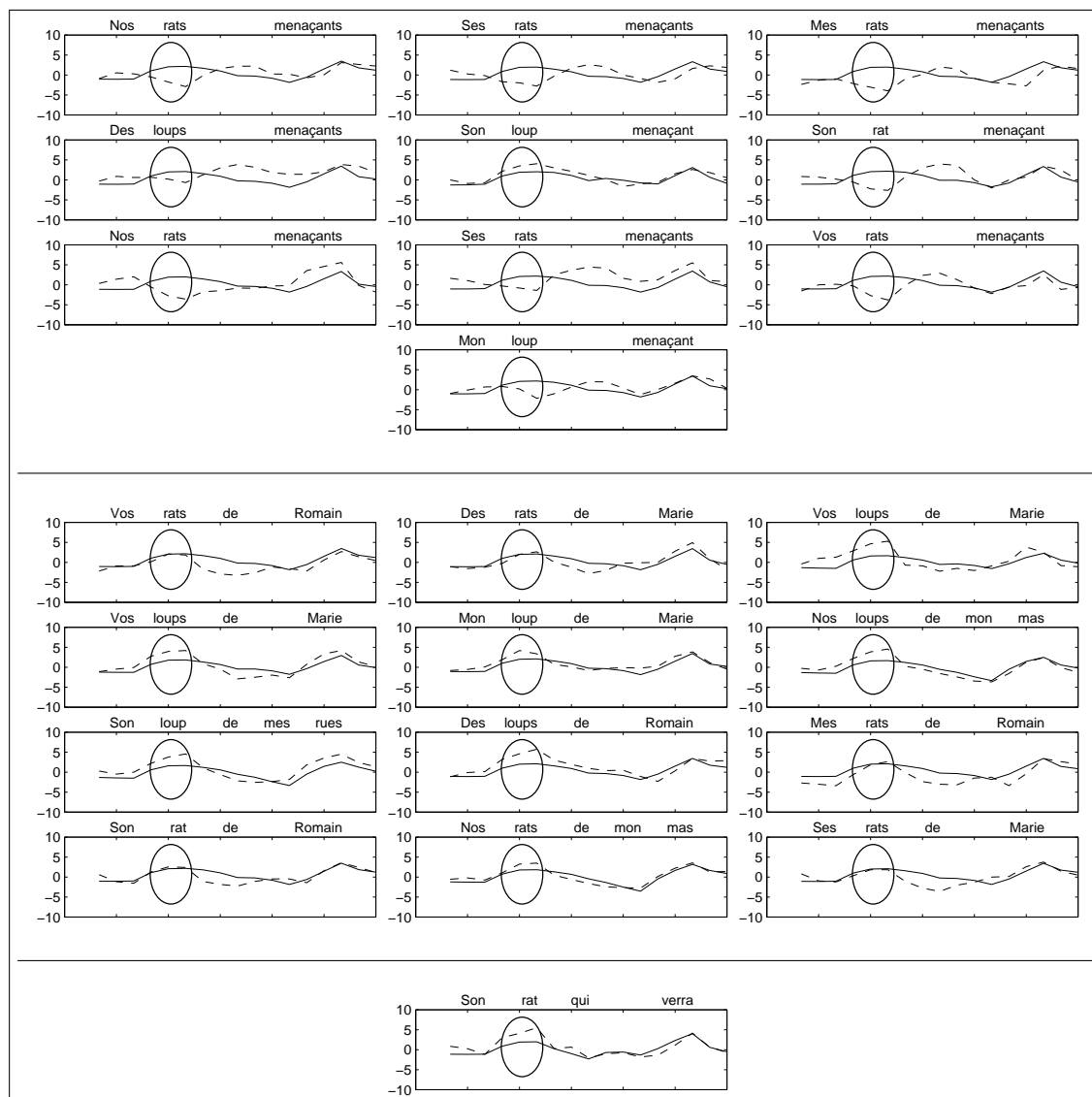


FIG. 3.28: 2+3 syllables. **Fréquence fondamentale.** On observe une grande différence entre les contours (pour rappel : les contours observés sont en pointillé ; les traits pleins correspondent aux prédictions) avec un adjectif qualifiant (en haut) et les contours avec complément (au milieu) : les premiers montrent un mouvement descendant tandis que les derniers suivent le mouvement montant de la prédiction. Cependant, le contour montant au milieu de la deuxième ligne indique que cette option n'est pas exclue pour les adjectifs. Le seul exemple avec une subordonnée (en bas) s'aligne dans le rang des compléments – peut-être avec une montée encore plus prononcée.

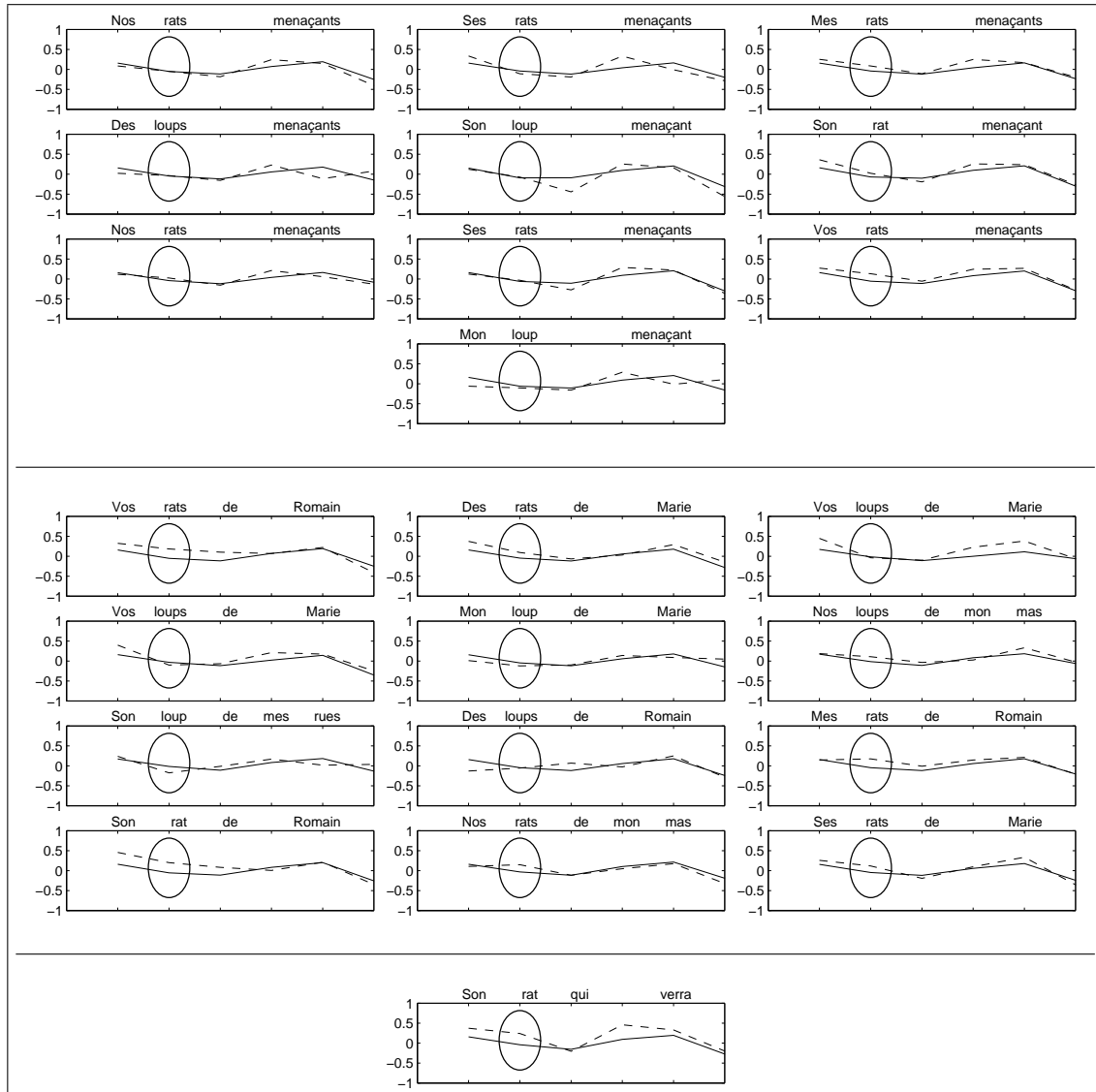


FIG. 3.29: **2+3 syllables. Coefficient d'allongement.** Pour l'allongement on remarque qu'adjectif et complément se comportent d'une manière homogène et conforme aux prédictions. En revanche, la subordonnée semble caractérisée par un plus fort allongement avant la frontière. Avec un seul exemple dans cette configuration, il pourrait bien s'agir d'un cas particulier. Une affaire à suivre...

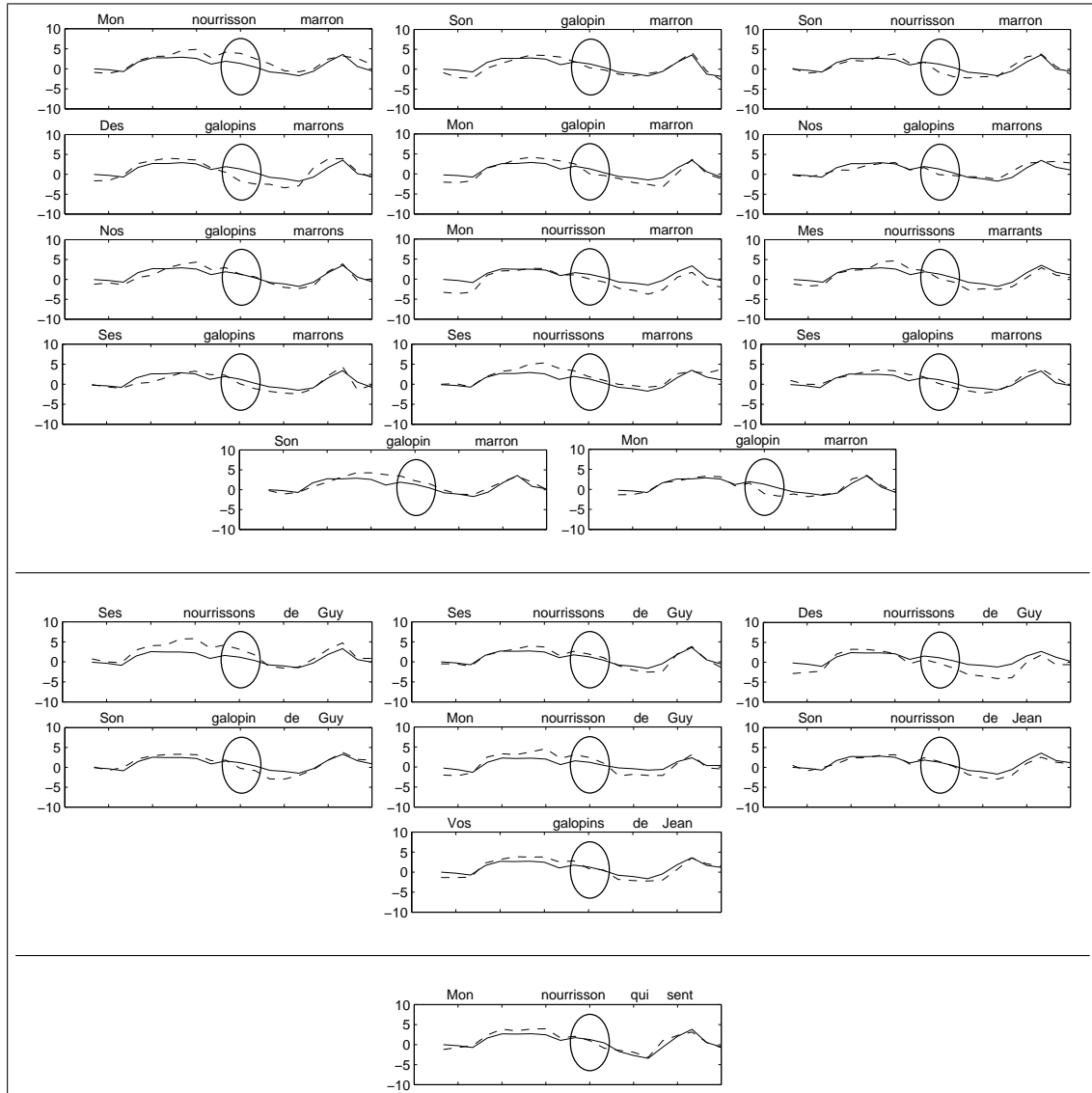


FIG. 3.30: 4+2 syllables. Fréquence fondamentale. Dans cette configuration le SFC prédit un contour relativement plat qui semble valable pour chacune des trois alternatives.

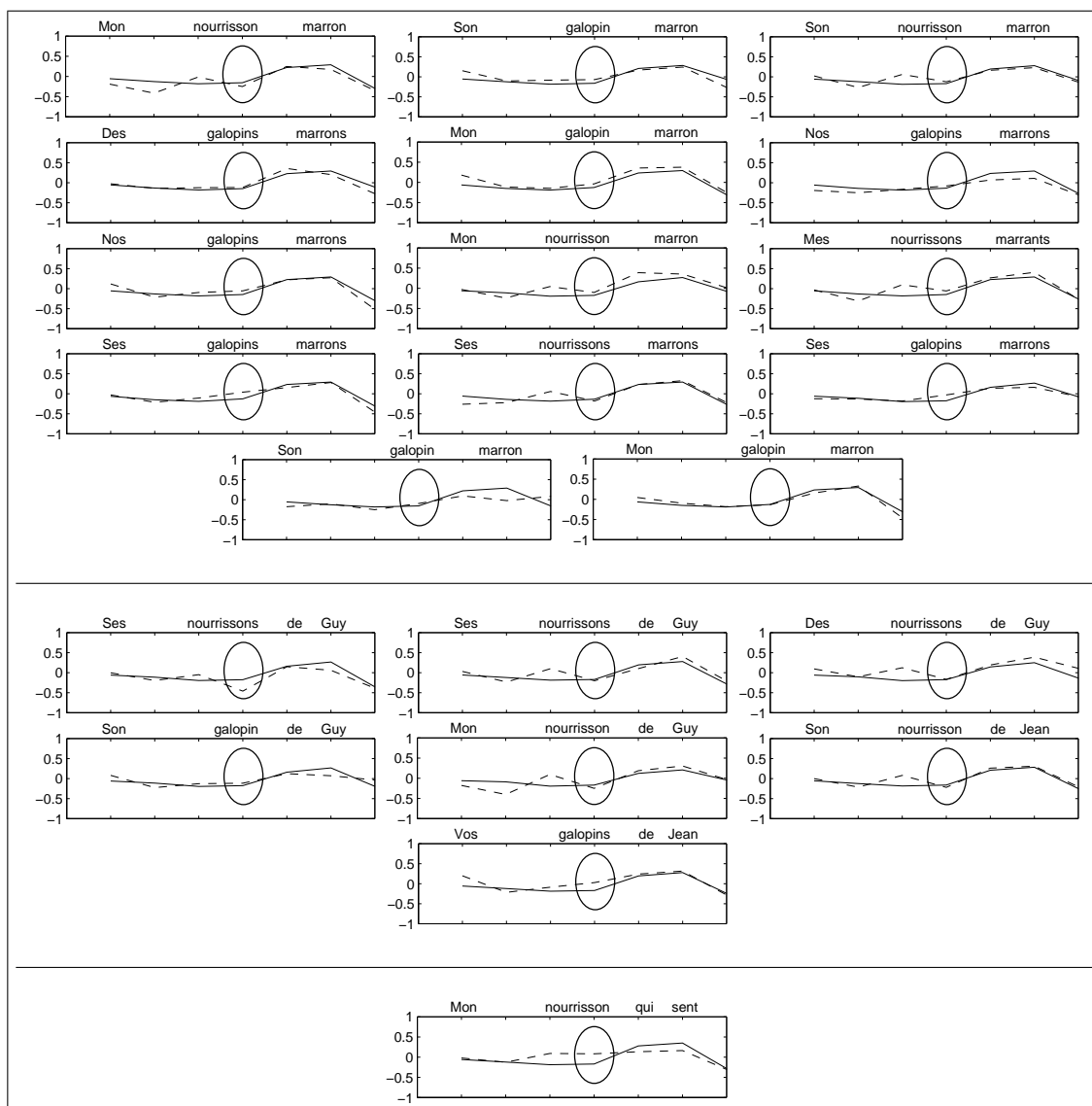


FIG. 3.31: 4+2 syllables. Coefficient d'allongement. Rien de spectaculaire non plus au niveau de la durée si ce n'est que notre soupçon par rapport aux subordonnées n'est pas contredit : l'allongement observé avant la frontière est plus grand que celui prédit.

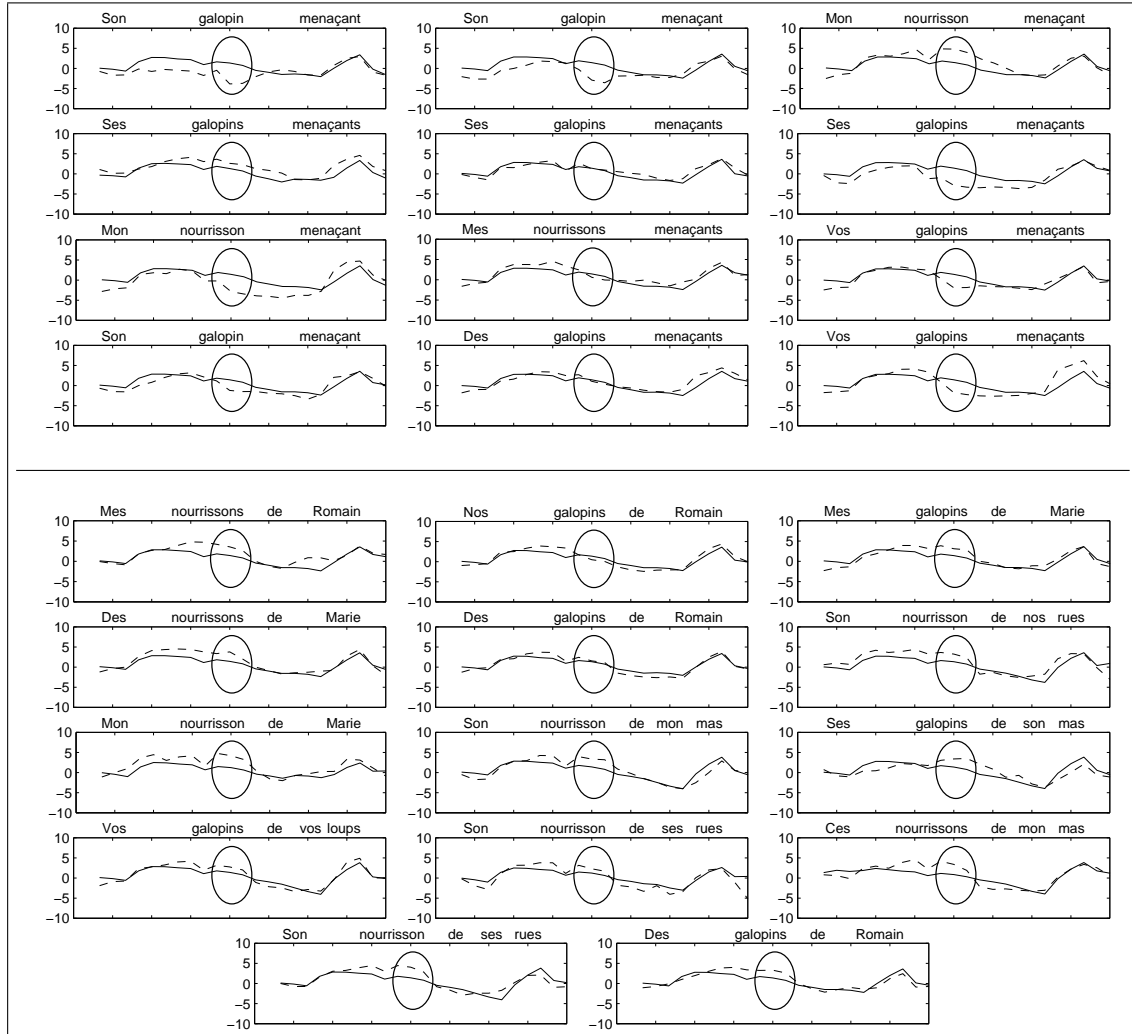


FIG. 3.32: 4+3 syllables. **Fréquence fondamentale.** Dans cette configuration qui n'est pas représentée par des subordonnées, on retrouve la même tendance que dans la première (figure 3.28) : les contours « adjectif » dévient vers le bas par rapport aux prédictions, bien que l'effet semble plus léger. Les contours « complément » se conforment pour la plupart aux prédictions avec une tendance aux mouvements montants plus forts.

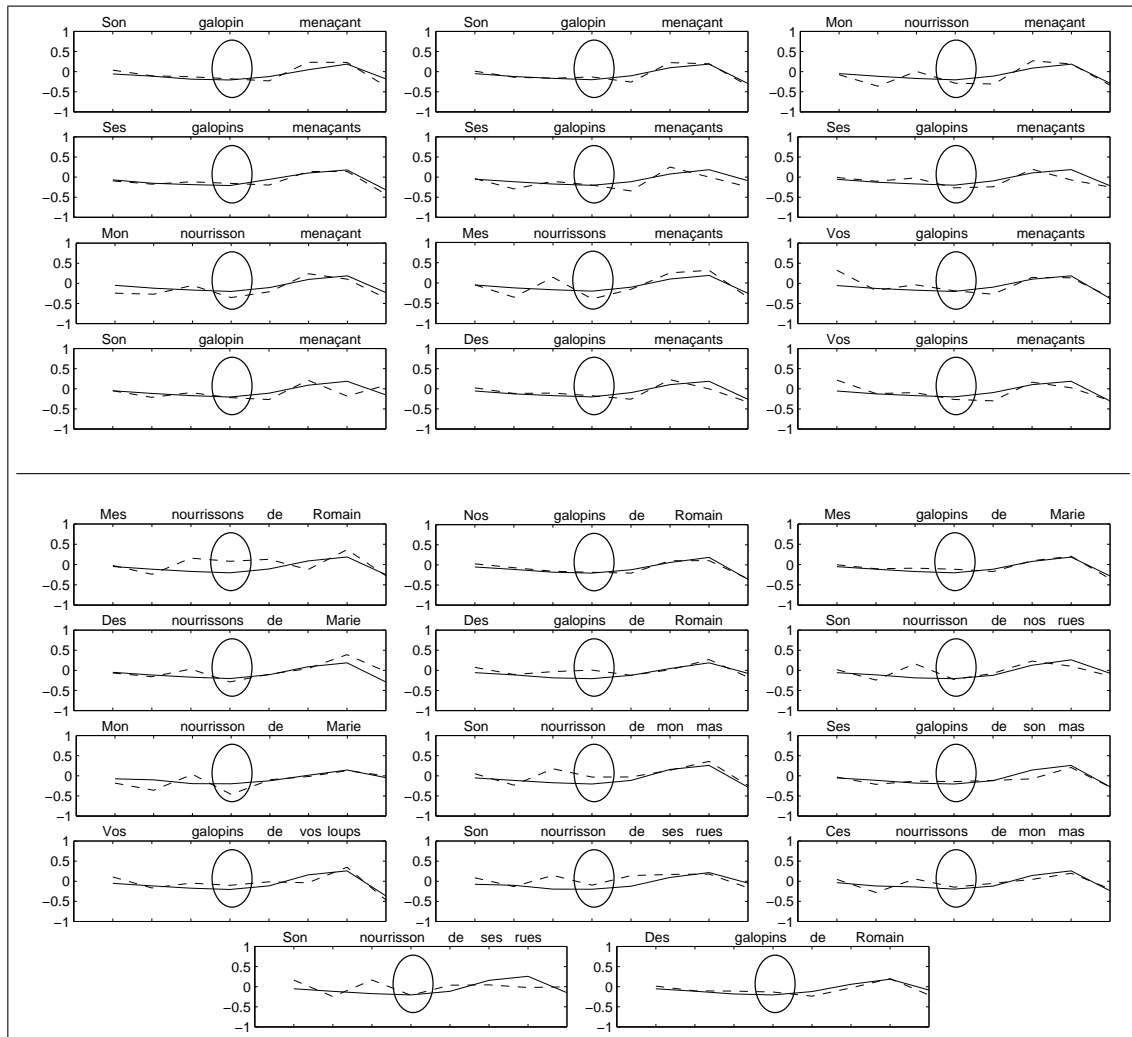


FIG. 3.33: 4+3 syllables. Coefficient d'allongement. Au premier regard : rien de nouveau – sauf que l'on trouve quelques contours « complément » ayant un allongement préfrontalier plus important, ce qui ne semble pas exister pour les adjectifs.

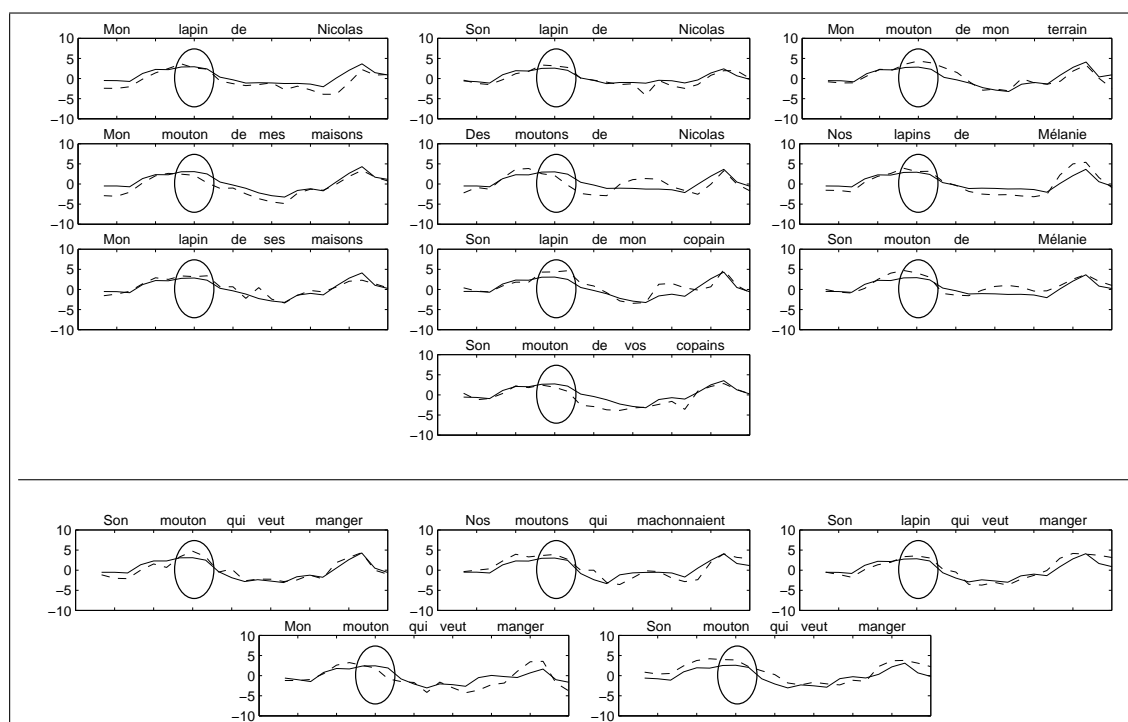


FIG. 3.34: **3+4 syllables. Fréquence fondamentale.** Le corpus ne contenant pas d'adjectifs de plus de 3 syllables, nous voilà seuls avec « compléments » et « subordonnées ». Prédications et observations sont très proches l'une de l'autre – difficile de voir une différence de comportement entre les deux alternatives.

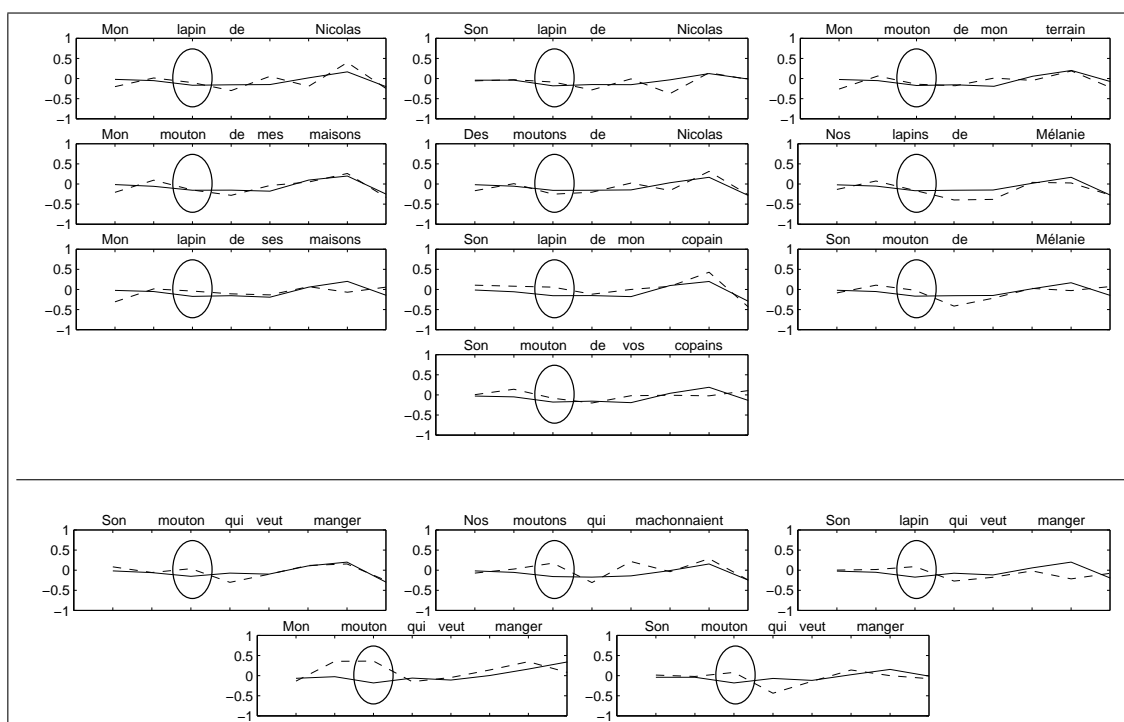


FIG. 3.35: **3+4 syllables. Coefficient d'allongement.** Heureusement, nous disposons maintenant d'un peu plus d'exemples pour les subordonnées car ils nous permettent d'établir cinq sur cinq ce qui n'était qu'un soupçon jusqu'ici : la qualification par subordonnée relative est caractérisée par un allongement plus fort.

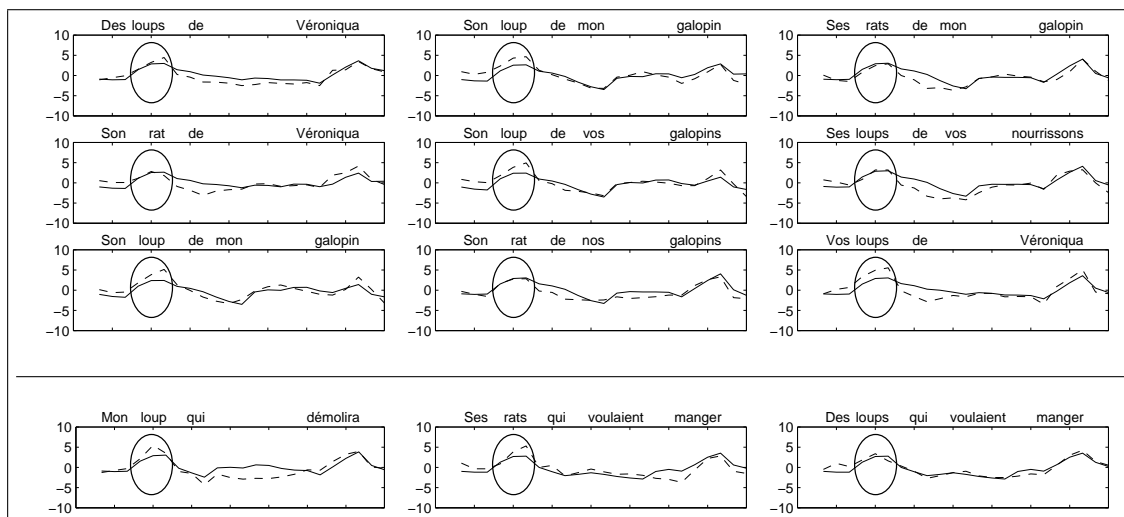


FIG. 3.36: **2+5 syllables. Fréquence fondamentale.** La dernière configuration montre des mouvements montants forts. La prédiction suit, mais avec une dynamique légèrement plus petite.

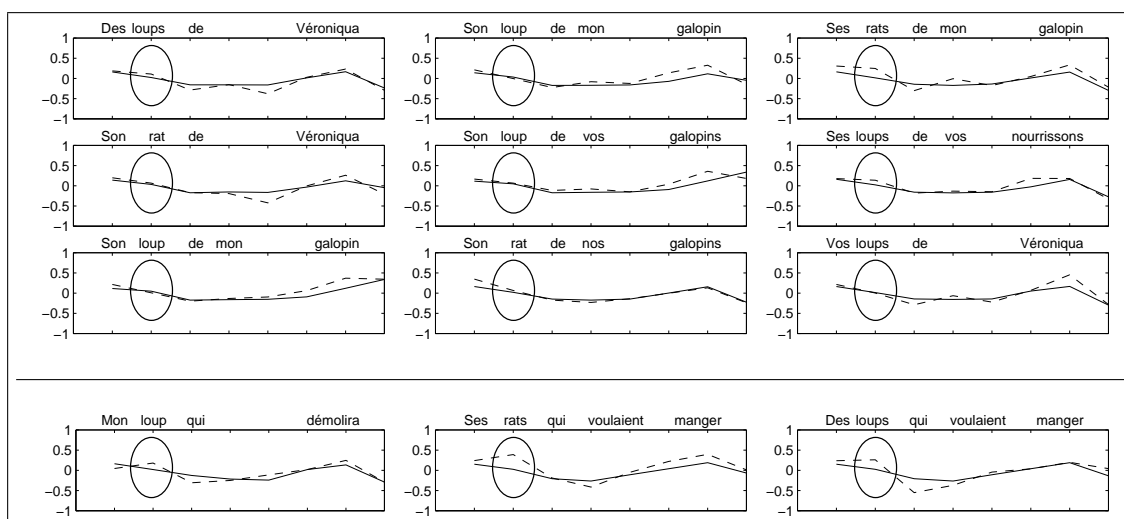


FIG. 3.37: **2+5 syllables. Coefficient d'allongement.** La différence d'allongement du GIPC préfrontalier mise en évidence dans la figure 3.35, se confirme dans cette configuration. Puis, comme déjà observé dans la figure 3.33, l'option allongement est réalisée par quelques « compléments ».

3.5.1 Résumé

Quelle image émerge de cet ensemble de contours pour l'influence de la nature du groupe qualifiant ? Si un mouvement mélodique montant et un allongement prononcé sont deux moyens de marquer une frontière, on peut conclure que de la qualification par un *adjectif* en passant par le *complément* jusqu'à la *subordonnée relative*, les contours mobilisent davantage ces deux moyens pour une même configuration d'empan. Les changements entre configurations évoluent parallèlement aux flèches de la figure 3.38 : si pour l'adjectif la pente bascule de descendante vers légèrement montante, le complément présente un renforcement du mouvement montant et une tendance légèrement accrue à mobiliser un allongement préfrontalier. Entre complément et subordonnée le comportement mélodique change en parallèle.

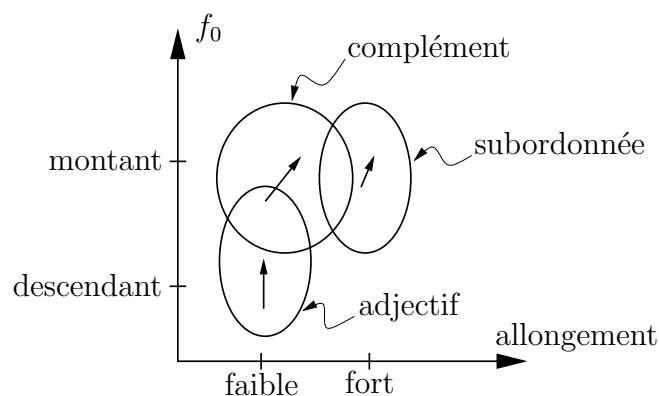


FIG. 3.38: **Résumé.** Pour une même configuration, la nature du qualifiant entraîne une mobilisation accrue de deux moyens de marquage (accent et allongement) dans l'ordre suivant : adjectif \rightarrow complément \rightarrow subordonnée. Entre configurations, les contours évoluent parallèlement comme indiqué par les flèches.

Interprétation. Est-ce que cette analyse invalide le SFC ? Oui, *mais...* Oui, parce que la morphologie des contours observés d'un groupe de qualifié et qualifiant n'est pas indépendante de la nature du qualifiant – ce qui était une des hypothèses très fortes du départ. Pour l'interprétation, il s'offre une voie rapide dans le cadre du SFC : nous nous sommes trompés sur l'hypothèse que la fonction de qualification est la même dans les trois cas – en réalité, il s'agirait de *trois fonctions distinctes* dont les contours associés auraient du être appris et générés par trois modules différents. Cette interprétation est schématisé par la figure 3.39(a).

Notons cependant que, bien que les différences soient assez systématiques, les contours majoritairement adoptés pour signaler une qualification par un groupe donné peuvent aussi minoritairement être utilisés par un autre groupe. Cette observation autorise l'interprétation qu'il s'agit simplement de trois *variantes* de réalisation d'une

même fonction. La nature du qualifiant conditionnerait – éventuellement de manière probabiliste – le choix de la variante, comme proposé par la figure 3.39(b). Il faut avouer que le SFC en l'état ne peut apporter aucun appui pour trancher entre ces deux interprétations. Dans les deux cas, nous sommes en présence de trois générateurs qui apprendront à générer les mêmes contours indépendamment du fait qu'on les appelle « générateur fonction A » ou « générateur de variante A de la fonction ».

Rappelons finalement que l'évolution de la morphologie à travers différentes configurations d'empan semble – indépendamment de la nature du qualifiant – mettre en œuvre d'une manière cohérente deux moyens dans l'ordre suivant : création/renforcement d'accent et augmentation de l'allongement. Cette cohérence me semble justifier de considérer les trois qualifications comme appartenant à une même catégorie de relation de dépendance et de générer les contours associés par un même module (cf. figure 3.39(c)). Le SFC dans sa forme actuelle en serait d'après cette lecture une première approximation et pourrait évoluer en incorporant un paramètre d'entrée supplémentaire qui coderait la « force » du qualifiant. Cette piste affaiblit l'hypothèse que la morphologie des contours associés à une même fonction est *uniquement* déterminée par des caractéristiques phonotactiques. Le paramètre d'entrée supplémentaire pourrait être cependant interprété comme une paramétrisation de la fonction en question (cf. la notion de gradience discutée par Ladd [1993]).

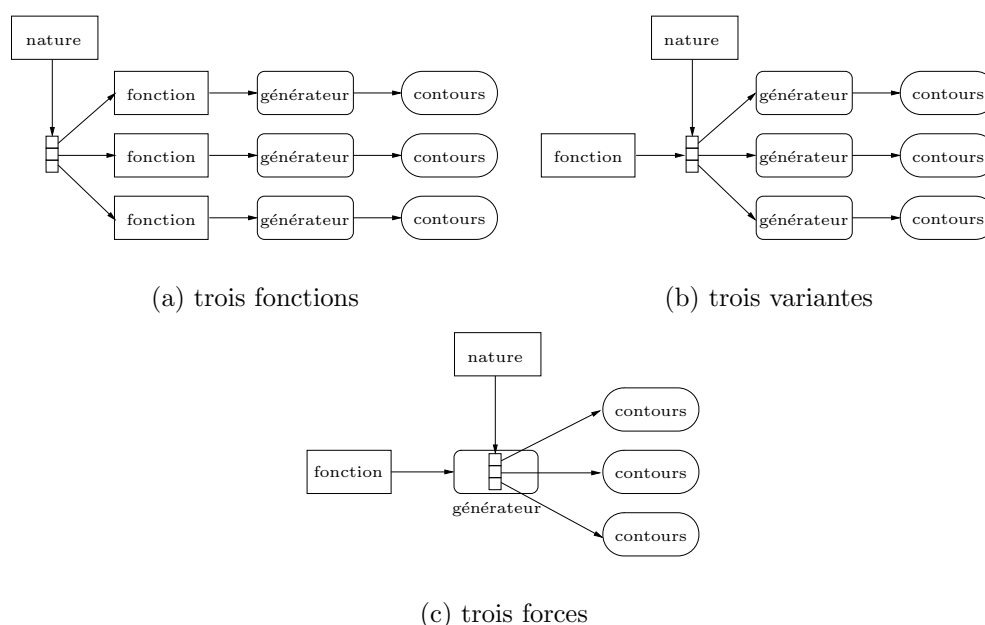


FIG. 3.39: Trois interprétations.

3.6 Formules mathématiques – français

Arbres de performance. Dans le chapitre 2.2.2 (p. 77) nous avons présenté les arbres de performance. Leur relation étroite avec les arbres syntaxiques et leur déviation par rapport à ces derniers, qui semblait principalement due aux poids syllabiques des groupements syntaxiques, ont fourni un argument important pour le choix des paramètres d'entrée des générateurs : la configuration phonotactique. Il est temps maintenant de vérifier si le SFC prédit des structures rythmiques semblables aux observations.

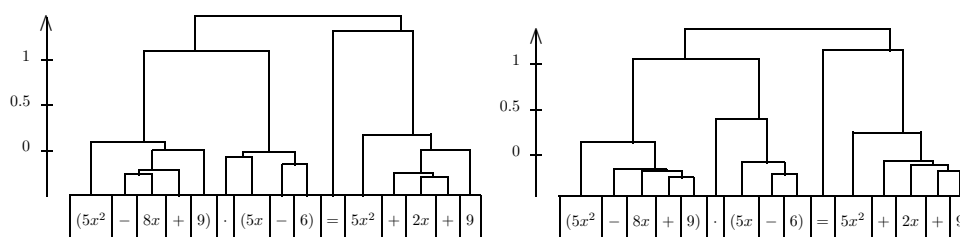


FIG. 3.40: **Structure de performance : observée vs. prédite.** La branche droite de l'opérateur multiplicatif est trop haute. Globalement, les structures sont très proches.

La figure 3.40 reprend l'exemple de la page 80 et montre que la prédiction (à droite) suit de très près l'original. La seule différence majeure est la branche droite de l'opérateur de multiplication : elle est plus haute en prédiction. L'opérateur se trouve ainsi rythmiquement un peu plus isolé. Une explication possible est, qu'à cet endroit, l'information sur le poids syllabique de l'opérande gauche du produit manque au modèle : le contour qui génère principalement l'allongement ici est celui qui est associé à la relation de dépendance entre opérateur et opérande droite. Ce contour est généré sans référence à ce qui se trouve avant l'opérateur. Poursuivre cette explication jusqu'au bout amènerait à revoir l'analyse fonctionnelle des formules mathématiques que nous avons proposée : au lieu de décomposer les opérations binaires en deux relations de dépendance avec l'opérateur comme tête, on pourrait l'analyser en tant que relation ternaire, le contour associé ayant quatre points d'ancrage : début de première opérande, début et fin de l'opérateur et fin de la deuxième opérande. Nous n'avons pas exploré cette piste au cours de cette thèse,¹² car l'allongement généré semble satisfaisante.

Un phénomène que nous avons soulevé en 2.2.2 était le basculement de l'attachement d'un opérateur en fonction des poids syllabiques de ses arguments. La figure 3.41 montre que le SFC reproduit ce comportement. Dans le deuxième exemple de cette figure on observe un changement de hiérarchie entre le « - » et le « < ». Les deux nœuds étant quasiment de même hauteur, ceci est certainement sans importance. Il

¹²Avis aux amateurs...

s'agit plutôt d'un exemple de mise en garde contre une sur-interprétation des arbres de performance. Comme déjà souligné plus haut, les arbres ne sont qu'une visualisation de la structure rythmique qui est de nature continue. L'interprétation des arbres doit donc tenir compte de la « fausse discrétisation » qu'ils suggèrent.

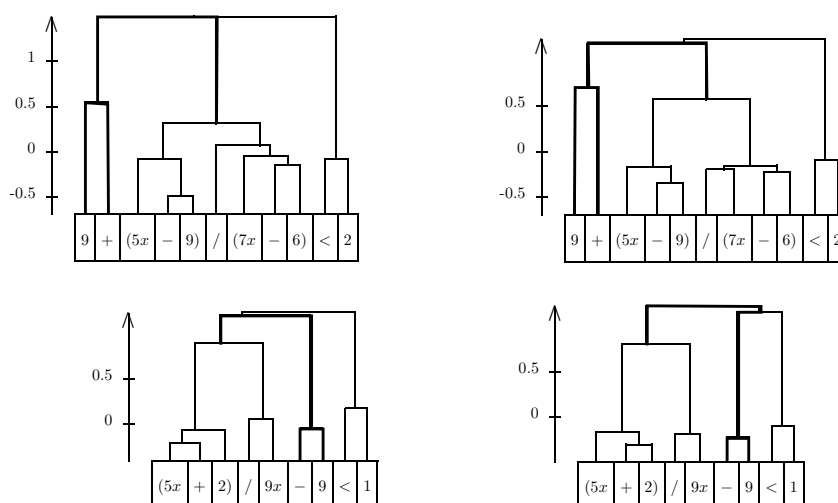


FIG. 3.41: **L'attachement : observé vs. prédit.** Le SFC reproduit le phénomène que l'attachement de l'opérateur est réalisé de préférence avec l'opérande qui est plus courte.

Après ces deux exemples individuels, vérifions le comportement global de l'attachement : la figure 3.42 reprend l'observation que l'attachement d'un opérateur mathématique est fortement corrélé avec la balance syllabique de ses arguments. Le SFC reproduit très fidèlement ce comportement. Il n'est guère étonnant que la dispersion des valeurs prédites soit plus petite que pour les observations – le modèle contrairement à notre locuteur est un système déterministe.

La figure 3.43, finalement, montre que le SFC produit un bon prédicteur pour l'attachement.

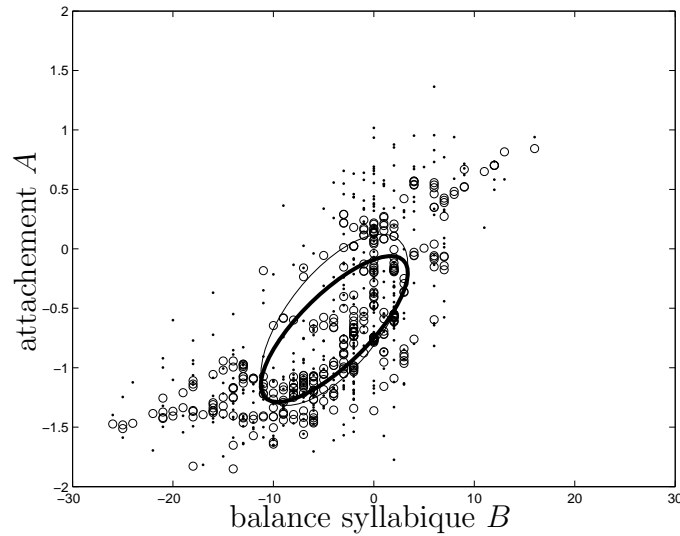


FIG. 3.42: **Attachement vs. balance syllabique : observations et prédictions.** Les valeurs observées sont représentées par des points ; les prédictions par les cercles. La dispersion des valeurs prédites est plus petite.

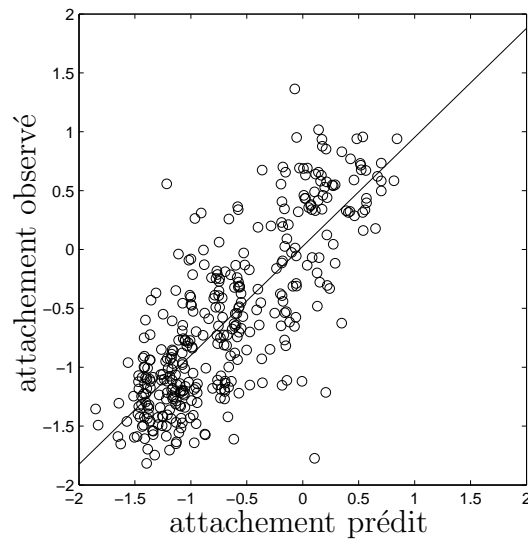


FIG. 3.43: **Attachement : observations vs. prédictions.** Le coefficient de corrélation entre prédiction et observation de l'attachement est 0,79. La droite de régression a pour équation $A_{\text{obs}} = 0,93A_{\text{pred}} + 0,03$.

Fréquence fondamentale. Ce n'est pas par hasard que la présentation des résultats pour le corpus des formules mathématiques a commencé avec les données sur la structure temporelle : la particularité de ce corpus est sa structuration riche et profonde qui engendre une réalisation prosodique particulièrement intéressante au niveau rythmique. Cependant, l'évolution de la fréquence fondamentale présente aussi de l'intérêt.

La figure 3.44 montre deux exemples de prédiction de f_0 . Le premier consiste en une formule relativement simple. On remarque un accent mélodique (encerclé) qui est (correctement) généré par le contour associé à la relation « dépendance droite » de l'opérateur « > » et le côté droit de l'équation. Il est situé sur l'avant-dernière syllabe de la partie gauche de l'empan ce qui est un comportement inhabituel. L'explication est la suivante : l'oralisation de l'opérateur « > » est « strictement supérieur à » ; l'accent est donc sur la dernière syllabe de « supérieur ». Il aurait été probablement plus convenable de modéliser ceci par un contour porté supplémentaire, mais le SFC, en tant qu'extracteur de « manifestations typiques » des fonctions (para-)linguistiques dans le corpus, génère cet accent comme étant associé directement à la relation entre « > » et « $6x$ ». Il s'agit ici d'une observation comparable à celle qui nous a amené à introduire une fonction de segmentation dans le corpus des « 1001 phrases »¹³ (cf. page 131). Ceci montre également que le SFC est tout à fait capable d'apprendre des mouvements non-triviaux qui ne sont pas localisés directement au point d'ancrage. Si toute fois c'est le cas, on est souvent mis sur une piste de réinterprétation des fonctions réalisées.

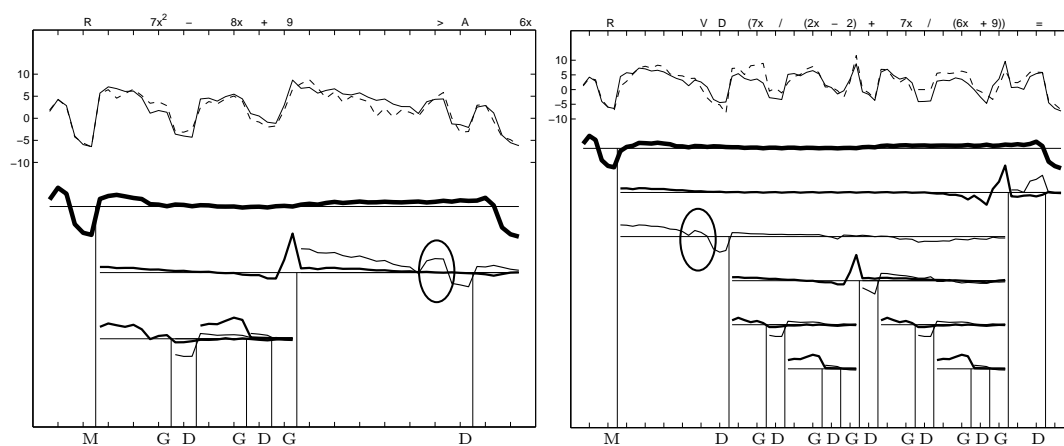


FIG. 3.44: **Prédiction de f_0 : deux exemples.** Les étiquettes des contours fonctionnels ont été abrégées.

¹³et également dans le corpus des modalités

Le second exemple, $\left| \frac{7x}{2x-2} + \frac{7x}{6x+9} \right| = 1$, est une des formules les plus complexes du corpus. La figure 3.44 montre les 14 contours se superposant. On aperçoit le même phénomène que discuté plus haut concernant une légère montée apprise sur la dernière syllabe de « absolue » dans l'opérateur « valeur absolue de » (voir encadré). Cette fois-ci, la concordance avec l'observation est moins bonne et l'accent appris est plus petit. Si l'on penche pour l'ajout d'un contour porté, ceci s'expliquerait par sa réalisation non-systématique dans le corpus par le locuteur : le SFC apprend un compromis.

Après ces exemples où le SFC réussit à produire une prédiction satisfaisante, la figure 3.45 montre les difficultés auxquelles le modèle (et tout modèle) est confronté face à ce corpus : les accents sur « 6 » et le premier « 2 » sont particulièrement forts et sur le dernier « + » le registre reste très élevé tout comme sur la première syllabe de « = ». Cette formule est effectivement celle du corpus qui présente la plus grande erreur moyenne de la prédiction de f_0 .

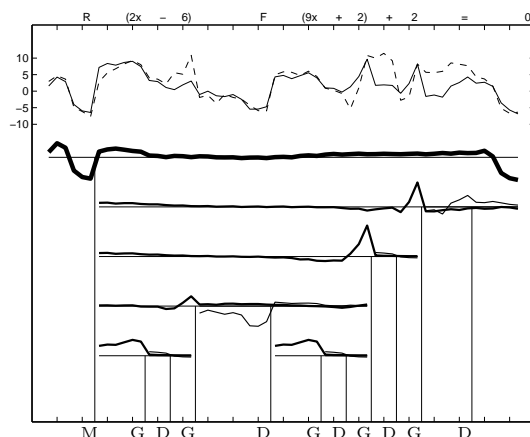


FIG. 3.45: Prédiction de f_0 : dernier exemple.

La figure 3.46 finalement, représente un extrait de la famille des contours DG : elle montre leur évolution pour les configurations phonotactiques de 2+2 jusqu'à 9+5 syllabes. On s'aperçoit clairement de la croissance de l'accent avec le nombre croissant de syllabes du groupe gauche.

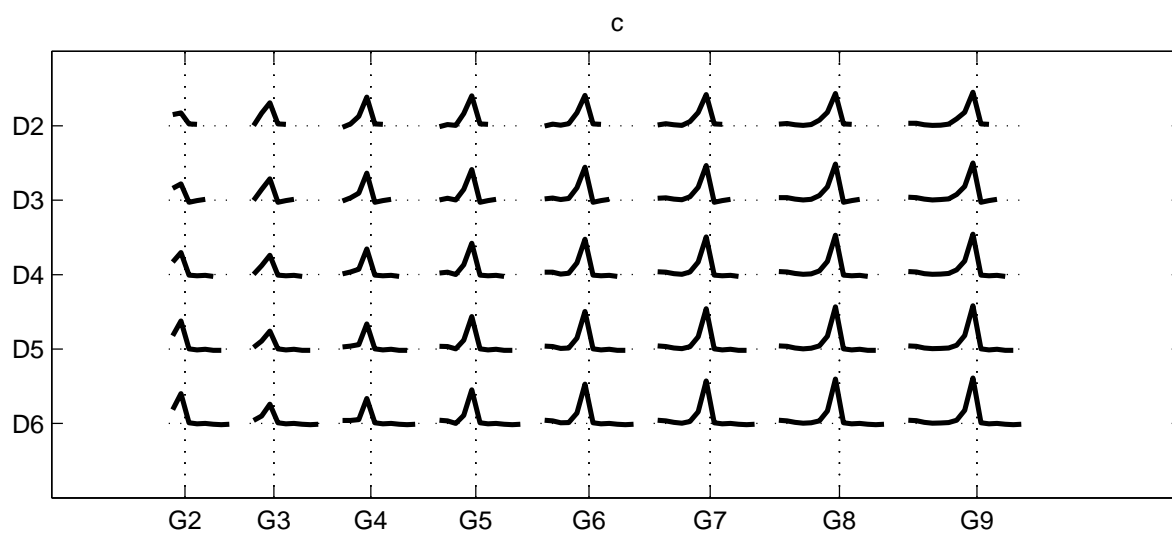


FIG. 3.46: **Famille de contours : DG.** La figure montre les contours fonctionnels associés à la relation de « dépendance gauche ». Chaque contour correspond à une configuration phonotactique : le nombre de syllabes du groupe gauche augmente de gauche à droite ; de haut en bas c'est le nombre de syllabes du groupe droit qui augmente.

3.7 Formules mathématiques – allemand

Comme pour le corpus français, la présentation des résultats sur les formules mathématiques commence avec un regard sur leur structure rythmique. Précisons d'abord que la définition du coefficient d'allongement pour l'allemand est moins bien établie que pour le français :

- nous ne disposons pas d'un corpus indépendant qui permettrait d'établir les moyennes de durées des phonèmes,
- le statut des GIPCs par rapport aux syllabes est moins bien appuyé, et
- le traitement des diphtongues au même titre que les voyelles nécessiterait plus d'investigations.

Il s'en suit que l'interprétation des structures fines des arbres de performance doit être entreprise avec encore plus de précaution que pour le français. La figure 3.47 montre à gauche une observation typique et à droite la prédiction correspondante. Le premier point à soulever est que la similitude avec la structure syntaxique est indéniable. Une deuxième observation est que les nœuds ont tendance à montrer des branchement à droite. La question de savoir s'il s'agit là d'un phénomène de groupement hiérarchique, ou plutôt d'un phénomène continu d'accélération au sein du groupe avant le ralentissement final, doit ici rester ouverte. Quant à la prédiction, on remarque une inversion de la hiérarchie entre le « > » et le « + » principal.

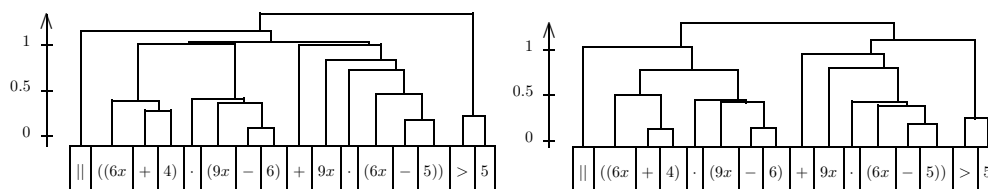


FIG. 3.47: **Structure de performance : observée vs. prédite (allemand).**

La correspondance étroite entre arbres syntaxiques et structures de performances n'implique en aucun cas les observations que nous avons faites pour le corpus français quant à l'attachement des opérateurs et sa relation avec la balance syllabique des opérandes. Ce n'était donc pas sans une grande curiosité que j'ai extrait les données correspondantes du corpus allemand. D'autant plus que le corpus a été équilibré par rapport au corpus français où il manque de données pour des balances syllabiques élevées. Bref, les figures 3.48 et 3.49 montrent un comportement tout à fait semblable au français bien que légèrement moins prononcé : la préférence de l'attachement à l'opérande la plus courte.

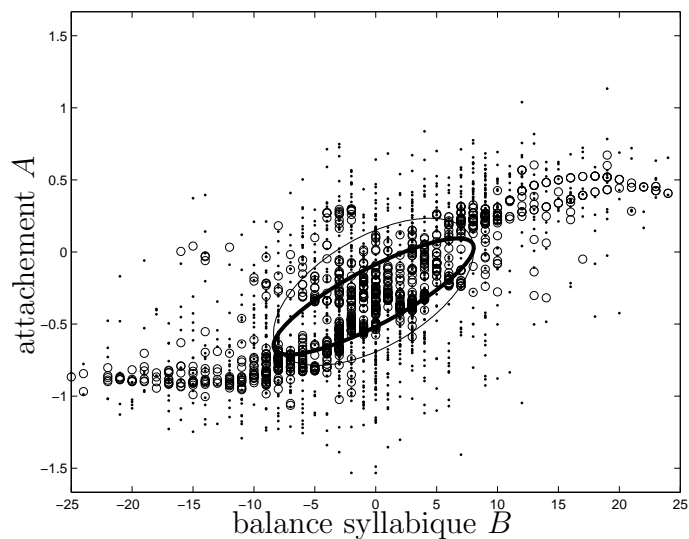


FIG. 3.48: **Attachement vs. balance syllabique (allemand)**. Les attachements prédits (cercles) suivent la tendance des attachements observés (points) à augmenter avec la balance syllabique.

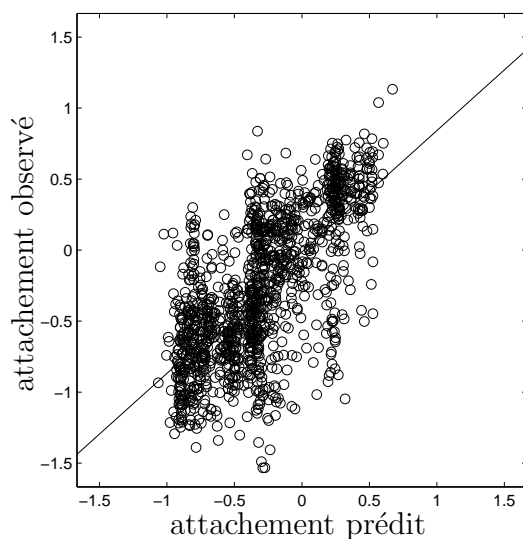


FIG. 3.49: **Attachement : observations vs. prédictions (allemand)**. Le coefficient de corrélation entre prédiction et observation de l'attachement est 0,67. La droite de régression a pour équation $A_{\text{obs}} = 0,85A_{\text{pred}} - 0,01$.

Fréquence fondamentale. Examinons, pour conclure, les prédictions de quelques contours de la fréquence fondamentale. La figure 3.50 montre qu'elles sont globalement très satisfaisantes. Remarquons que l'exemple en bas à droite a été choisi parce qu'il s'agit de la formule présentant la plus grande erreur de prédiction de tout le corpus. Cette erreur provient évidemment du plateau de f_0 sur « $4x - 9$ » que le SFC ne veut pas générer. Compte tenu du fait que le contour « proposé » par le modèle pour ce passage semble raisonnable et que le reste de l'énoncé est prédit très convenablement, ceci est plutôt une illustration de la difficulté à mesurer la qualité d'une prédiction de manière objective.

Les deux endroits encerclés montrent que l'on trouve des accents omis par le SFC et également des accents ajoutés par rapport à l'observation. En comparaison avec les contours français, on peut remarquer que nous trouvons dans le corpus allemand plus de plateaux de f_0 (haut – comme discuté avant ; ou dans un registre bas – comme au début de l'exemple en bas à gauche sur « V », prononcé « Betrag aus »). Cette observation n'est évidemment valable que pour nos deux locuteurs et ne prétend pas décrire une différence fondamentale entre les deux langues.

Notons finalement qu'il peut surprendre que le SFC produise des résultats si satisfaisants pour l'allemand tant qu'au niveau rythmique que mélodique, puisqu'il n'a pas été conçu pour gérer une langue à accents libres. Je pense que ce succès est, au moins en partie, dû à la particularité du corpus des formules mathématiques : d'une part, le locuteur est très contraint par la tâche ardue consistant à véhiculer la structure complexe des formules, et d'autre part, les mots représentés dans le corpus sont relativement courts et d'une variabilité très restreinte ce qui permet au SFC de capturer leur régularité sans être informé de l'emplacement des accents lexicaux.

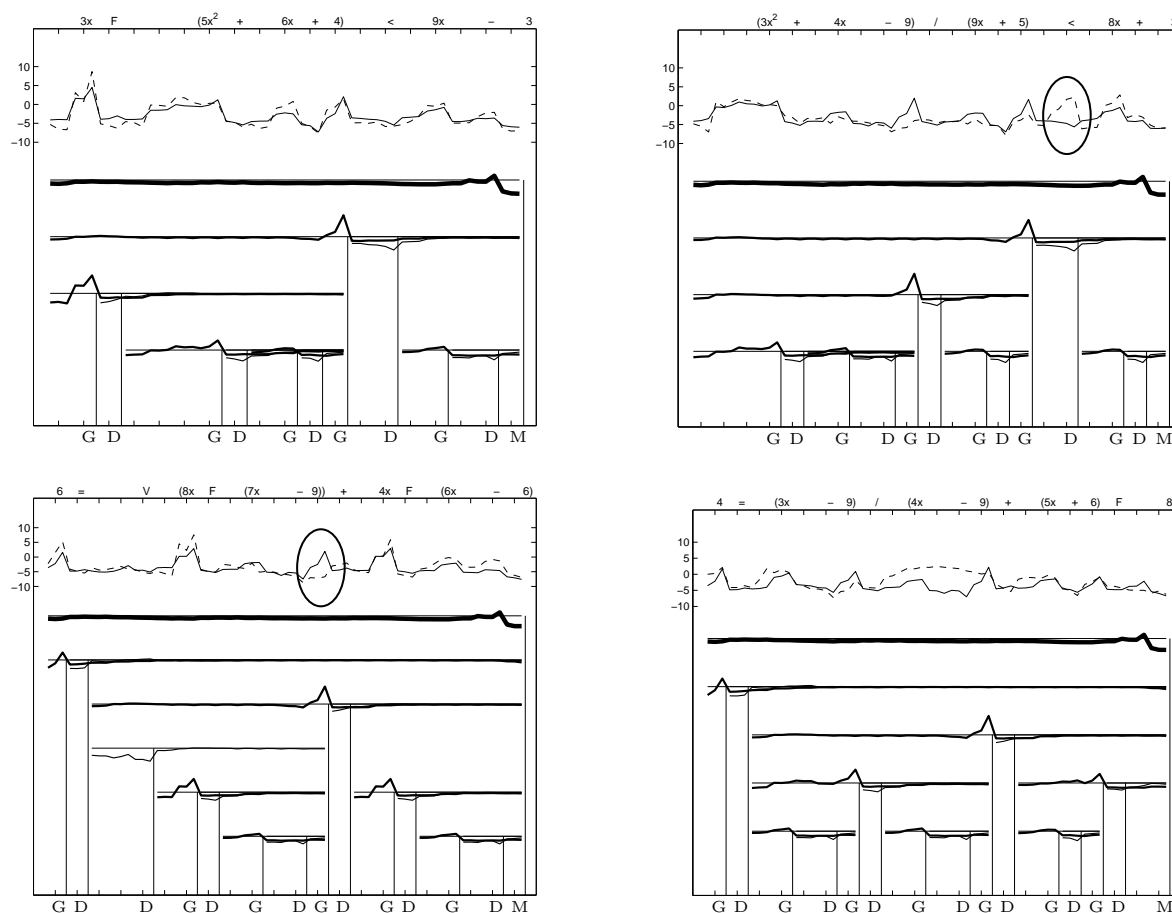


FIG. 3.50: **Prédiction de f_0 : formules en allemand.** Sont encadrés : un accent omis par le SFC (en haut à droite) et un accent ajouté (en bas à gauche). À noter également : les plateaux de f_0 (en registre bas : au début de l'exemple en bas à gauche ; en registre haut : au milieu de l'exemple en bas à droite).

3.8 Corpus de Vigo

Les résultats présentés ici pour le corpus en galicien sont à considérer comme préliminaires.

Une caractéristique de ce corpus est que la structuration mélodique par rapport à la décomposition en fonctions de dépendance qui a été directement transposée du français, semble très réduite. Les contours fonctionnels plats en témoignent (à gauche dans la figure 3.51).

Une première hypothèse est que les accents lexicaux empêchent l'émergence des contours. Nous avons donc entrepris de les modéliser en tant que contours portés ayant l'empan du mot et la syllabe accentuée comme point d'ancrage. Dans la figure 3.51, ces contours sont représentés en dernière ligne, étiquetés d'un « a ». Comme attendu, la syllabe accentuée se voit attribuer un f_0 légèrement plus élevé. Néanmoins, la prédiction globale reste peu satisfaisante.

Une des raisons de cet échec est certainement la dynamique très réduite des enregistrements, aggravée par des passages relativement longs où la détection de f_0 n'a pas été possible à cause du fort dévoisement. Il est ici cependant impossible de décider, s'il s'agit là d'une caractéristique du galicien ou du locuteur.

Pour le rythme, nous étions confronté à des voyelles fortement réduites ce qui aurait certainement nécessité une révision approfondie de notre définition du coefficient d'allongement. Les effets sont clairement visibles sous forme de creux profonds du contour observé dans la figure 3.51 en bas à droite.

Remarquons finalement que le manque d'équilibre au niveau des configurations phonotactiques (cf. la figure 3.9 et les commentaires sur la page 117) laisse des traces dans les contours fonctionnels appris (encerclé).

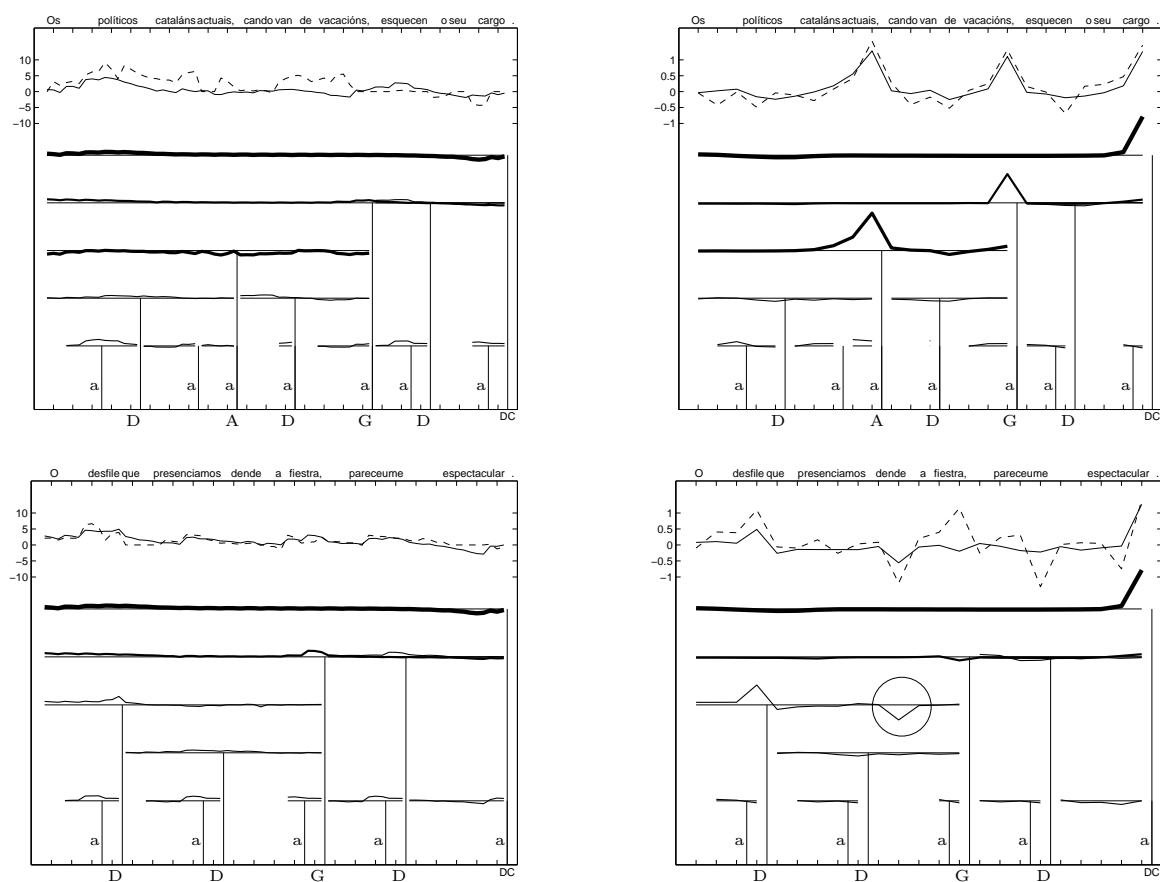


FIG. 3.51: Deux exemples pour le galicien. À gauche : fréquence fondamentale ; à droite : coefficient d'allongement.

Évaluation perceptive

La dernière partie de ce manuscrit est consacrée aux tentatives d'évaluation perceptive du SFC que nous avons entreprises. L'importance d'une telle évaluation est incontestable étant donné qu'un modèle de génération de la prosodie est conçu pour produire de la parole qui va être écoutée par des humains. Juger de la *qualité perceptive* d'un modèle en ne se fiant qu'aux erreurs de prédiction mesurées par rapport à une observation « cible » est hasardeux. En effet, la perception d'un stimulus de parole est un processus à dimensions multiples : sa compréhension ou son appréciation dépendent de son intelligibilité, de son naturel ainsi que du cadre pragmatique. Toute tâche que peut effectuer un auditeur en face d'un stimulus est ainsi le résultat complexe d'une interaction entre qualité des signaux et investissement cognitif de l'auditeur.

Plan de la partie. Nous proposons d'abord un dispositif original permettant de placer un auditeur dans une situation bien définie (une dictée) et sous observation rapprochée. Nous espérons accéder ainsi à travers la performance de l'auditeur aux reflets d'une charge cognitive plus ou moins accrue que nous supposons influencée directement par la qualité (ou : l'adéquation à la tâche) de la prosodie.

La deuxième partie présentera un test de préférence beaucoup plus classique. Il mettra en oeuvre les prédictions prosodiques de deux modèles différents appliqués tous les deux au français et à l'allemand. Le domaine visé pour ce test est, comme pour la tâche de dictée, l'énonciation de formules mathématiques.

4.1 Une tâche de dictée

La première évaluation perceptive que nous décrivons dans ce chapitre a pour objectif ambitieux d'évaluer la charge cognitive imposée à un auditeur écoutant l'énonciation de formules mathématiques. En tant qu'évaluation d'un modèle de génération de paramètres prosodiques, l'idée est simple et séduisante : une « bonne » prosodie devrait permettre un accès rapide et facile à l'information véhiculée par le locuteur.

Le succès d'une telle méthode aurait l'avantage certain d'avoir une mesure objective de l'impact de la prosodie au décodage de l'information. Un autre avantage serait que ce que l'on mesurerait comme « qualité » serait clairement défini par le cadre de l'expérience – évitant ainsi toutes les difficultés que l'on rencontre habituellement en scrutant l'opinion d'un ensemble de sujets et qui se résume par : est-ce que les sujets répondent réellement à la question que l'on croyait poser ?

4.1.1 Mise en place du test

Disposition du test. La tâche que nous avons choisie est celle d'une dictée. Les sujets écoutent avec un casque une formule mathématique et sont priés de la noter sur une feuille qui est posée sur une tablette graphique. Ce dispositif nous permet d'informatiser les tracés ainsi que leur déroulement dans le temps. L'acquisition de ces données a été possible grâce à un logiciel écrit par Tahar Lallouache pour le LPNC (Université Mendès France) gracieusement mis à notre disposition par Jean-Pierre Orliaguet que je tiens à remercier. Ce logiciel a été adapté au laboratoire à l'acquisition de formules.

Après chaque formule, deux questions sont posées aux participants (sous forme de boîte de dialogue à l'écran d'ordinateur) demandant : « Êtes-vous sûr de ce que vous avez écrit ? » et – après avoir vu la formule cible à l'écran – de donner une appréciation en cinq niveaux de l'adéquation entre prosodie et formule. L'intérêt de ces questions est double : recueillir en passant de l'information exploitable en tant que MOS¹ et d'alléger le rythme de la tâche de dictée qui est quelque peu stressant.

Les stimuli. Le but de ce test était en premier lieu d'explorer ce paradigme d'évaluation. À terme, nous espérons pouvoir l'utiliser pour évaluer des modèles différents (ou différentes versions du SFC) dans cette tâche particulière qui est la dictée. Nous avons choisi 24 formules de notre corpus (cf. page 109) afin de couvrir au mieux les différents degrés de complexité qui y sont représentés. Trois phrases ont été retenues pour la courte² phase d'échauffement au début de chaque test. Le test a consisté en deux parties : une contenant les stimuli originaux (ci-après : version D), l'autre des stimuli synthétisés en trois versions différentes. Sur les 15 volontaires qui ont participé

¹anglais: "mean opinion score"

²Le déroulement du test ne posait aucun problème aux sujets – une phase plus longue d'acclimatation à la tâche ne semblait donc pas nécessaire.

au test, 8 ont d'abord effectué le test avec les stimuli synthétiques suivi de l'autre partie. L'ordre des parties était inversé pour les 7 autres sujets.

Les trois versions synthétiques de ce test ont été obtenus de la manière suivante :

Version C : paramètres prosodiques extraits de l'enregistrement. Nous appelons également cette version la « resynthèse ».

Version B : prédiction des paramètres par le SFC et synthèse avec le même dictionnaire de diphone que précédemment.

Version A : paramètres obtenus grâce à une version volontairement dégradée du SFC.

Les empan des sous-contours ont été coupés afin de ne jamais excéder trois syllabes à gauche ou à droite d'un marqueur. Mis à part sa fonction en tant que référence « vers le bas » par rapport au SFC complet, cette version a été conçue dans l'idée de fournir des éléments de réponse concernant la réalité cognitive de ces contours. Il va de soi qu'un éventuel « échec » de cette version devrait être interprété avec beaucoup de précaution.

Afin de ne pas alourdir le test, qui dure environ deux fois 15 minutes, la partie « synthétique » du test ne comportait qu'une version de chaque formule, soit 7 phrases par version. L'attribution des versions aux phrases a été contrebalancée à travers les sujets.

Les données disponibles. La figure 4.1 montre les informations que la disposition expérimentale permet d'acquérir. En haut est reproduite la transcription par un des sujets de la formule $\sqrt{2/9x} = 1$. La courbe du milieu représente la pression exercée avec le stylo sur la tablette graphique en fonction du temps. Elle permet de récupérer avec précision les débuts d'écriture de chaque « mot ». L'information temporelle est bien entendu synchronisée avec le stimulus auditif qui trouve sa représentation (alignée avec sa segmentation en phonèmes³) en bas de la figure. Les flèches « penchées » indiquent ainsi le délai entre l'écriture et le stimulus auditif.

4.1.2 Les erreurs

Disposer des tracés des transcriptions permet de vérifier facilement si les sujets ont réussi la tâche. Globalement, la réponse est : oui. Seulement, 2 transcriptions sur 630 n'était pas exploitables parce que le sujet avait renoncé. 489 des 628 formules restantes sont transcrites correctement. Compte tenu de la complexité des formules et du fait que l'enregistrement du corpus n'était pas fait dans la perspective d'une dictée, ceci doit être considéré comme une proportion importante de « sans faute ».

Au delà de ces résultats globaux, nous cherchons évidemment à savoir si la version du stimulus a une influence sur la capacité de transcription. Nous tenterons donc de

³voir remarque 1 sur p. 22

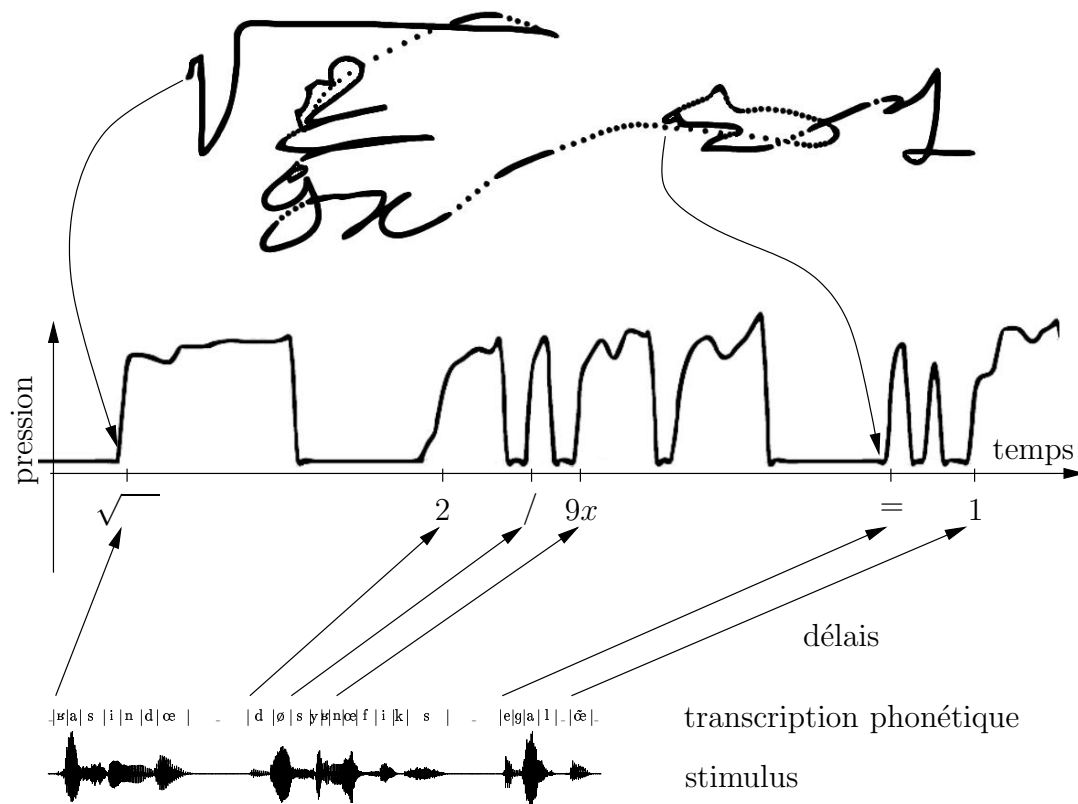


FIG. 4.1: **Acquisition des données.** Le niveau de pression du stylo sur la tablette graphique permet de récupérer l'organisation temporelle de l'écriture qui peut ainsi être mise en relation avec le stimulus auditif.

rejeter l'hypothèse nulle qui stipule l'indépendance des résultats par rapport au facteur « version du stimulus ». Dans les paragraphes suivants, nous présenterons l'analyse sommaire de quatre types d'erreur différents.

Omissions. Pour chaque transcription, nous avons compté le nombre de « mots » manquants par rapport au stimulus écouté. Le tableau suivant donne donc le nombre moyen d'omissions par phrase selon les versions :

A	B	C	D
0,47	0,30	0,30	0,12

Globalement, ces résultats sont conformes à nos attentes : il semble que la version dégradée induise plus d'omissions que les stimuli obtenus grâce au SFC ou en version resynthétisée. La voix naturelle de notre locuteur obtient les meilleures performances. Cependant, des tests de rang selon Wilcoxon révèlent que seule cette dernière différence est statistiquement significative ($p < 0,04$ par rapport à la version resynthétisée). Si

l'on veut interpréter ce résultat, cela revient à dire que la tâche de dictée est plus facile à accomplir avec une parfaite qualité de voix. Mais, nous n'apprenons rien sur l'influence des paramètres prosodiques.

Insertions. Un autre type d'erreur, l'insertion dans la transcription de « mots » non présents dans le stimulus, est très rare. Nous comptons ses occurrences moyennes suivantes :

A	B	C	D
0,04	0,00	0,00	0,01

Bien que le plus mauvais score soit obtenu par les paramètres prosodiques altérés, rien ne permet de rejeter l'hypothèse nulle. Ce type d'erreur ne semble donc pas pertinent pour la recherche de l'impact d'une prosodie bien-formée.

Erreurs lexicales. Par erreur lexicale, j'entends le désaccord entre stimulus et tracé écrit sur un mot donné, p.e. $6x \rightarrow 9x$. Le comptage :

A	B	C	D
0,10	0,10	0,12	0,15

révèle que c'est la version originale qui obtient les plus mauvais résultats, mais la différence n'est pas statistiquement significative, ce qui nous laisse une fois de plus sans indices en vue de l'évaluation des paramètres prosodiques.

Erreurs structurales. Il s'agit ici des erreurs sur la structure hiérarchique des formules. Pour les comptabiliser, nous avons établi à partir de la disposition en deux dimensions des tracés le parenthésage sous-entendu par le sujet, et nous avons repéré le nombre de parenthèses qui ne coïncident pas avec celles qui étaient à l'origine des oralisations. On peut s'attendre à ce qu'une « mauvaise » prosodie ait le plus d'impact sur ce type d'erreur, puisqu'il réfère au désaccord entre groupements/dépendances émis et perçus. Voici, les occurrences moyennes :

A	B	C	D
0,36	0,20	0,22	0,14

La tendance allant dans le sens de nos attentes, permet d'établir des différences significatives (Wilcoxon) entre la voix originale et la version altérée ($p < 0,002$) et entre l'original et la resynthèse ($p < 0,04$). Force est donc de constater que nous ne pouvons pas ici non plus établir de conclusions quant à la qualité des paramètres prosodiques.

4.1.3 Les mesures temporelles

Jusqu'ici nous n'avons pas encore exploité l'information temporelle acquise grâce à la tablette graphique et qui a principalement motivé la mise en place de cette évaluation. Est-ce que l'on peut mesurer la charge cognitive pesant sur les sujets en observant de près leur écriture ? Notre hypothèse de base est évidemment que la prosodie a un impact sur cette charge et que cette charge a un impact sur la performance de la transcription.

Parmi les nombreuses traitements que l'on pourrait imaginer, nous allons nous intéresser dans la suite à trois mesures. Elles ont été choisies dans l'éventail de nos tentatives non pas pour leurs résultats, mais pour représenter quelques approches différentes possibles. Toutes les données se basent uniquement sur les formules transcrites dans le même ordre que l'oralisation. Nous avons également exclu les transcriptions comportant des omissions, insertions ou erreurs structurales. Les erreurs lexicales ont été jugées acceptables pour l'analyse de phasage entre stimulus et écriture. Nous aboutissons ainsi à 533 transcriptions soumises aux analyses temporelles.

Délai. Cette mesure est fondée sur l'idée que des difficultés à accomplir la tâche de dictée pourraient avoir des reflets dans le retard accumulé par rapport au stimulus écouté. Nous avons donc déterminé la différence (toujours positive) entre la transcription d'un mot et son énonciation. Nous obtenons ainsi un délai moyen pour chaque phrase, dont les moyennes par version se présentent comme suit :

A	B	C	D
1,58 s	1,66 s	1,75 s	1,75 s

Ici nous attend la première surprise : les retards accumulés en écoutant la voix naturelle sont plus grands que ceux pour l'écoute de la version avec des paramètres prosodiques altérés. Les sujets suivraient-ils mieux une mauvaise prosodie ? Soyons rassurés : les différences ne sont pas significatives. De plus, notons que cette mesure, bien que simple et intuitive, comporte le risque de renseigner plus sur la différence entre les vitesses d'élocution et d'écriture que sur la charge cognitive : un débit très lent permettrait au sujet de réfléchir « longuement » à ce qu'il doit écrire.

Pause et écriture. Si l'on a du mal à « comprendre » une formule, ceci pourrait se refléter dans le ratio entre la durée de pauses entre les mots et le temps consacré à l'écriture des mots en eux-mêmes. En cas de difficulté, les sujets hésiteraient beaucoup *entre* les mots et rattraperaient le temps perdu une fois qu'ils auraient saisi (ou décidé) quoi écrire : le ratio devrait être plus grand. Les moyennes ci-dessous montrent que cela n'est pas le cas (ou au moins que la prosodie n'a aucune influence sur cette mesure) :

A	B	C	D
0,50	0,50	0,50	0,55

Durée totale. La mesure qui est le sujet de ce paragraphe est dans un sens plus simple que celles discutées ci-dessus. Il s'agit de la durée totale de l'écriture. Or, la longueur des formules étant très variable, il semble normal que les distributions de ces durées ne se distinguent pas de manière significative en fonction des versions. Pour cette raison, nous avons tenté de les normaliser en retranchant pour les versions A, B et C le temps pris par le même sujet pour écrire la même formule en version D, profitant ainsi du fait que *toutes* les phrases ont été écoutées avec cette dernière version par *tous* les sujets. Cette mesure se rapproche donc quelque peu du « délai moyen » avec la différence importante que la référence est ici le temps nécessaire à l'écriture et non pas le temps de l'oralisation.

C'est avec beaucoup de regret que je dois annoncer que cette mesure non plus n'était pas à la hauteur de nos espérances :

A	B	C
-0,16	-0,28	-0.08

Notons le fait intéressant (et surprenant) que toutes les versions en qualité de voix synthétique étaient en moyenne écrites plus rapidement que les formules sous la dictée de la voix humaine – ceci (bien sûr) sans pouvoir en faire une affirmation statistiquement significative.

4.1.4 Le questionnaire

Venons maintenant aux réponses aux deux questions recueillies après l'écriture de chaque formule.

Confiance. La première question demandait au sujet s'il était, oui ou non, confiant dans l'exactitude de la transcription qu'il venait de donner. À part une bonne corrélation avec les erreurs discutées plus haut, les réponses ne fournissent pas de résultats exploitables.

Notes (MOS). Après la présentation visuelle de la formule cible, une seconde question a été posée : « Est-ce que la prosodie était adaptée à la formule ? » en laissant le choix entre les réponses « pas du tout », « mal », « moyennement », « bien » et « parfaitement ». La figure 4.2 montre les résultats de cette enquête selon les versions en attribuant les notes 1 à 5 aux réponses ci-dessus.

La première constatation est que les sujets étaient d'une manière générale très tolérants : l'échelle semble saturée. Je pense que ceci est dû au contexte de dictée effectuée car il donne aux sujets pour la plus part du temps l'impression de succès. Cependant, les moyennes (données en dessous des histogrammes) indiquent une préférence suivant l'ordre attendu. Quant à leur caractère significatif, nous devons assumer un dernier échec en ne pouvant établir aucune différence entre les versions A, B et C. Seule la voix naturelle est mieux notée que toutes les autres ($p < 0,02$ par rapport à C).

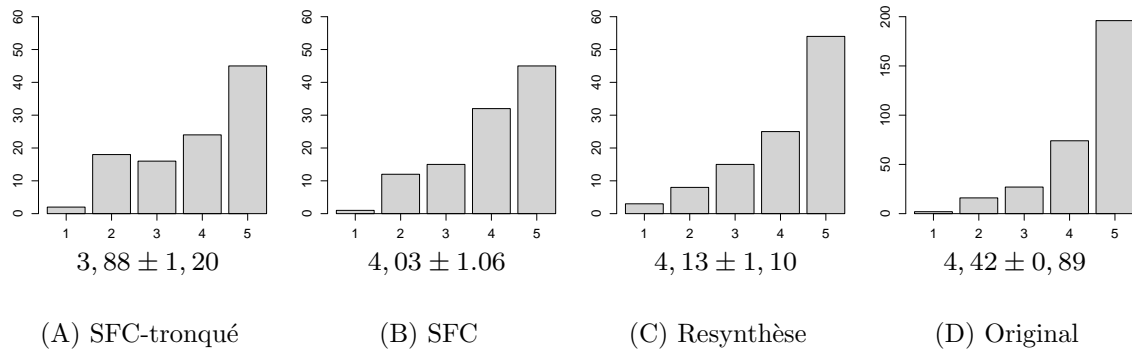


FIG. 4.2: **MOS – résultats selon les versions.** Sous les histogrammes sont inscrites les moyennes des distributions avec leurs écart-types.

4.1.5 Conclusions

Nous avons voulu bénéficier de la relative simplicité des matériaux linguistiques énoncés pour évaluer de manière globale la performance du modèle de génération prosodique proposé. Force est de constater que malgré les handicaps sévères que l'on peut imposer au stimuli acoustiques, les sujets ne laissent pas de traces évidentes de charge cognitive accrue, ce qui ne veut pas dire qu'elle est inexistante. D'autres paradigmes expérimentaux restent à mettre en œuvre, consistant notamment à limiter les capacités cognitives des sujets. Imposer une tâche secondaire pourrait ainsi être un moyen de faire ressortir les traces de charge différente selon la qualité prosodique du stimulus.

4.2 Évaluation comparative

Le test de préférence qui est le sujet de ce chapitre porte sur deux modèles appliqués à la génération de la prosodie des formules mathématiques : le SFC et une adaptation de l'IGM⁴ ; il porte également sur deux langues : le français et l'allemand. La transposition de l'IGM qui a permis son application aux formules mathématiques et les tests ont été réalisés dans le cadre d'un stage d'étude de Raidt [2002] que je voudrais remercier ici pour une collaboration agréable et efficace. Le but général de cette étude est de mettre en œuvre une méthodologie de comparaison de modèles d'apprentissage automatique de la prosodie sur des corpora identiques pour l'apprentissage et le test.

4.2.1 Adaptation de l'IGM

Ce paragraphe ne se veut pas une présentation de l'IGM que nous avons déjà rencontré brièvement dans l'état de l'art (cf. page 27 et 44) et dont on trouve les références dans la bibliographie. Il s'agit plutôt de signaler les principaux changements apportés à ce modèle de type Fujisaki. Ceci est nécessaire pour pouvoir apprécier les résultats du test de préférence : il est évident que nous ne disposons pas pour l'IGM du même temps (ni des mêmes compétences) que pour le SFC afin de le rendre apte aux particularités des formules mathématiques. Pour cette raison, j'utiliserai le sigle « igm » quand il s'agit de notre adaptation de l'IGM.

L'IGM prédit des paramètres Fujisaki sur la base de 20 paramètres d'entrée par syllabe à l'aide d'un réseau de neurones avec feed-back. Suite à un manque de temps pour son implémentation, le réseau que nous utilisons fait l'impasse sur le retour des sorties en entrée ; il s'agit donc d'un simple réseau « feed-forward » avec une couche cachée de 20 unités.

Cependant, les différences les plus importantes portent sur les paramètres d'entrée. Certains d'entre eux n'ont pas pu être transposés à l'application des formules.

- Ainsi, le paramètre associé aux débuts de paragraphe a été simplement omis, puisque nous nous intéressons ici seulement à la synthèse de phrases isolées.
- L'indice de rupture⁵ entre 0 et 4 a été réduit à trois niveaux (dont deux correspondant à l'importance syntaxique des opérateurs mathématiques et un marquant les frontières intra-mot).
- La structure complexe des formules a également rendu impossible l'établissement de règles simples quant au positionnement des frontières de groupes intonatifs⁶. Les entrées relatives à celles-ci ont été remplacées par des informations phonotactiques inspirées de l'expérience avec le SFC. En effet, ces nouvelles entrées sont plus riches en ce qui concerne l'association des poids syllabiques et des frontières

⁴Integrated Model of German Prosody [Mixdorff, 1998, 2002a; Mixdorff et Jokisch, 2001]

⁵anglais: “break index”

⁶anglais: “phrases”

de mots, mais elles laissent à la charge du RN la décision de déterminer si on est réellement en présence d'une frontière de groupe intonatif.

- Finalement, nous ne fournissons pas d'information sur la position des accents lexicaux. Cette réduction qui pourrait sembler très importante est cependant contrebalancée par le fait que le corpus ciblé ne contient que peu de mots différents, et de plus très courts. À cela s'ajoute le fait que le SFC ne dispose pas non plus de cette information.

Nous nous retrouvons donc avec une version réduite à 14 entrées de l'IGM qui permet néanmoins la prédiction d'une paramétrisation Fujisaki très acceptable pour les formules en allemand. Quant au français, la difficulté principale résidait dans l'étiquetage des commandes d'accent et de groupe. Peut-être l'utilisation de commandes d'accent négatives aurait-elle permis une paramétrisation plus économique et plus cohérente à mettre en relation avec les paramètres d'entrée du RN. Les quelques contours montrés dans la figure C.1 de l'annexe soulignent le caractère préliminaire de notre application de l'igm à l'énonciation des formules mathématiques en français.

4.2.2 Méthodologie

La comparaison porte sur les deux modèles (SFC et l'igm) appliqués au corpus des formules mathématiques en deux langues (français et allemand). Nous avons restreint les corpora aux formules dont la structure est présente dans les *deux* corpora.⁷ Des 134 phrases restantes nous choisissons de manière aléatoire 104 phrases pour l'apprentissage des modèles – en réservant ainsi 30 pour l'évaluation.

Afin d'améliorer la sélectivité du test de comparaison, nous ne présentons aux sujets que les 10 formules dont les différences (en terme d'erreur RMS de la fréquence fondamentale) sont les plus grandes entre les deux prédictions. Il est important de noter que ce processus de sélection entre les phrases réservées à l'évaluation ne porte pas sur la qualité des prédictions par rapport à l'original : nous choisissons simplement les versions synthétiques qui sont numériquement les plus *différentes*.

Les dix phrases sont synthétisées en trois versions : avec les paramètres prosodiques prédits par les deux modèles et avec les paramètres extraits des enregistrements. Puisque le corpus allemand constituait également la source du dictionnaire de diphtongues, nous avons volontairement renoncé à la possibilité offerte par le système de synthèse de choisir entre plusieurs alternatives pour chaque diphtongue. Sinon, la resynthèse aurait trop profité du fait de toujours disposer des segments originaux : non seulement la modification des diphtongues devient plus petite parce que l'on peut utiliser des unités plus proches des cibles – ce qui améliore *toutes* les versions ; mais l'enchaînement de segments dans leur ordre d'origine améliore significativement la qualité de voix – masquant ainsi la qualité des paramètres prosodiques.

⁷Essentiellement, ceci revient à écarter pour ce test les extensions apportées aux corpus allemand par rapport à celui enregistré en français précédemment.

Nous obtenons ainsi 30 paires de stimuli qui sont présentés deux fois (en inversant l'ordre : AB BA). Les sujets sont placés dans un environnement calme devant un écran d'ordinateur qui affiche la forme écrite de la formule à venir. Une interface graphique simple leur permet d'activer l'écoute (éventuellement répétée) de chaque stimulus avec un casque. Dès que les deux stimuli de la paire ont été écoutés au moins une fois, deux boutons apparaissent invitant le sujet à exprimer sa préférence (cf. figure 4.3). Nous avons demandé aux sujets de tenir compte de l'adéquation de la mélodie et du rythme avec la structure de la formule ainsi que de leur impression sur le naturel de la prosodie.

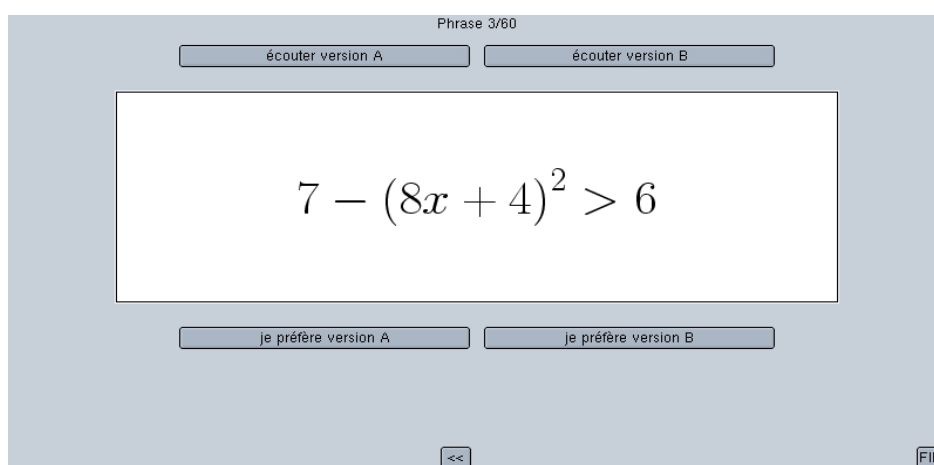


FIG. 4.3: Capture d'écran.

Les sujets étaient âgés de 20 à 30 ans et ne présentaient en l'occurrence pas de troubles auditifs. Leur participation au test qui durait environ 20 minutes n'était pas rémunérée. Les sujets ont évalué les modèles dans leur langue maternelle : 11 en allemand et 10 en français.

4.2.3 Les résultats

Résultats globaux. Avant de dépouiller les résultats plus en détail, je propose un bref regard sur l'ensemble des réponses données pour chacune des paires SFC-igm, Org-SFC et Org-igm. Celles-ci sont représentés dans la figure 4.4. Le nombre de réponses par paire est de 200 pour le test français et de 220 pour celui en allemand.

En allemand, il n'y a pas de nette préférence d'une version synthétique envers l'autre : le SFC n'obtient des résultats que légèrement supérieurs à l'igm, mais cette différence n'est pas significative. La version resynthétisée est largement préférée. Le SFC est néanmoins un peu plus souvent préféré à l'original que l'igm, ce qui confirme son léger avantage en comparaison directe.

En français, les différences entre les deux modèles sont très prononcées : le SFC est préféré dans plus de 90% des cas. Ici aussi, la comparaison avec la version resynthétisée confirme le résultat : le score de SFC est de 31% de préférence versus 4% pour l'igm.

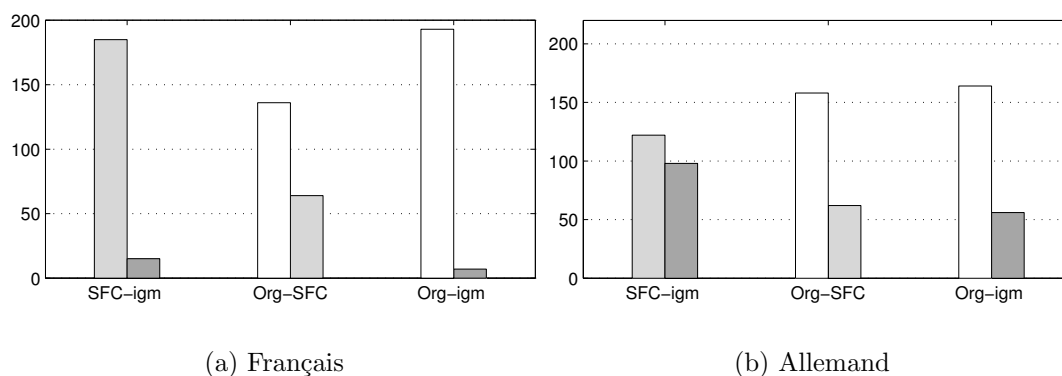


FIG. 4.4: **Préférences totales en comparaison directe.**

Fouille des données. Le but de ce paragraphe est de pouvoir se faire une idée de l'homogénéité des résultats. Est-ce que les sujets ont des avis très divergents ? Et, est-ce que l'on peut déceler des formules « problématiques » pour l'un ou l'autre modèle (ou pour la resynthèse des paramètres originaux) ? La discussion se basera essentiellement sur des représentations graphiques des données. Le lecteur intéressé peut trouver les matrices détaillées des réponses dans l'annexe C.1.

Une précision avant de commencer : les « votes » pour les stimuli resynthétisés (Org) seront toujours comptés comme « succès », et en comparaison entre le SFC et l'igm, on compte ceux pour le SFC. Ce choix tout à fait arbitraire trouve son reflet dans l'ordre dans lequel nous nommons les paires : **Org-SFC**, **Org-igm** et **SFC-igm**.

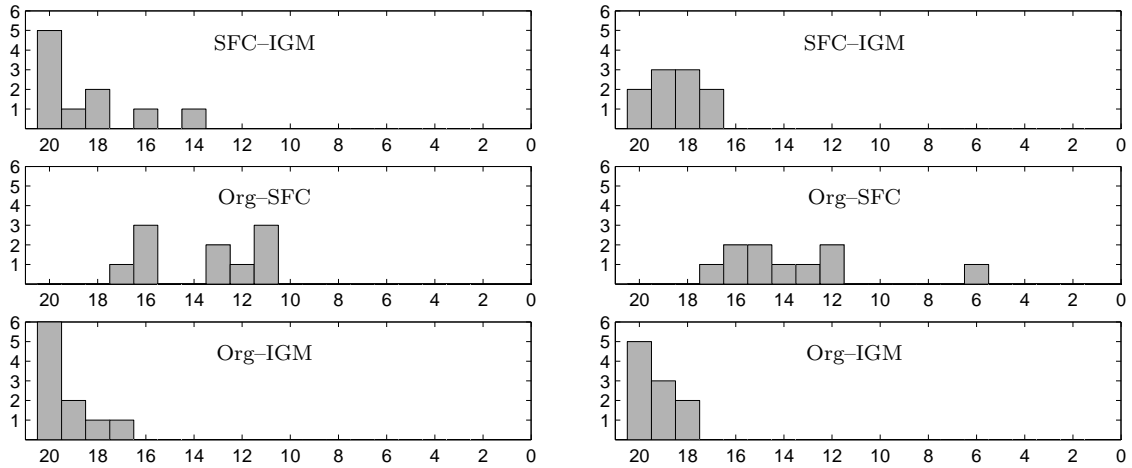
Dans la figure 4.5 sont représentés les cumuls des « succès » par numéros et par sujets. Commençons dans l'ordre : dans la figure 4.5(a), on trouve pour chacune de nos trois paires la distribution du nombre de « succès » cumulé pour toutes les phrases. La première sous-figure (SFC-igm) se lit donc comme suit : 5 sujets français ont voté 20 fois pour le SFC, 1 sujet 19 fois, 2 sujets 18 fois et ainsi de suite. Le nombre maximal de votes est ici de 20 (10 phrases en ordre AB et BA).⁸ L'interprétation est simple : plus la distribution est décalée vers la gauche, plus les sujets préfèrent la version nommée en premier dans la paire.

La figure 4.5(b) est construite de manière analogue, mais les cumuls sont effectués par rapport aux sujets.⁹ Elle permet donc des constats tels que : « dans les comparaisons entre l'original et le SFC, il y avait 2 phrases qui ont réuni 12 votes en faveur de

⁸Cet histogramme correspond à la colonne « \sum » dans le tableau C.1 en annexe.

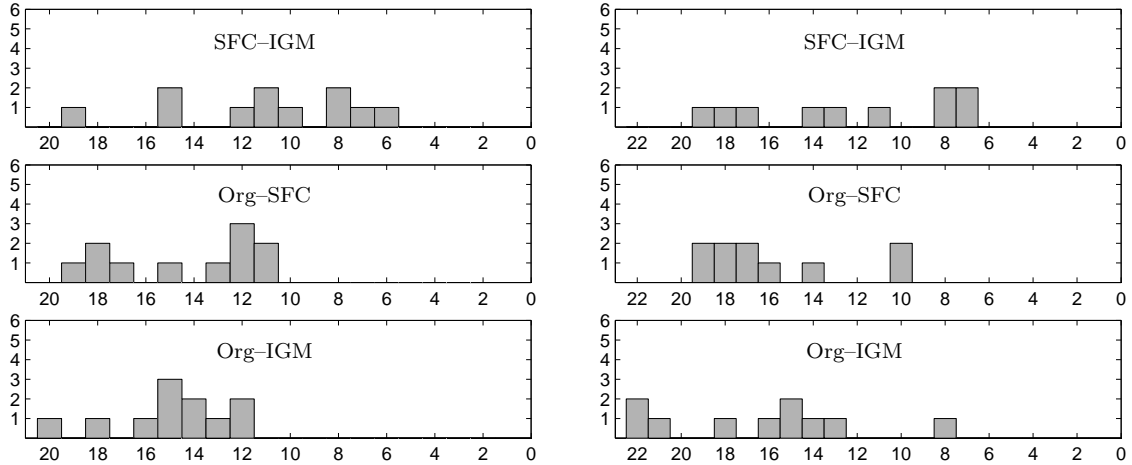
⁹Elle est la représentation de la ligne « \sum » dans le tableau C.1.

l'original » (sous-figure du milieu : 6^{ème} barre de gauche). Ici aussi, le nombre maximal de votes est 20, puisqu'il y a 10 sujets qui écoutent chaque paire deux fois. Les figures 4.5(c) et 4.5(d) donnent les mêmes renseignements pour le test en allemand.



(a) Français. Sujets

(b) Français. Phrases



(c) Allemand. Sujets

(d) Allemand. Phrases

FIG. 4.5: Fouille des données – histogrammes.

Ces histogrammes, bien qu'ils soient déjà plus informatifs que les résultats globaux de la figure 4.4, ont néanmoins deux désavantages :

- ils ne permettent pas de mettre en relation les résultats entre les paires de comparaison directe ;
- on perd l'information de *qui* a donné combien de voix pour une version des stimuli (ce qui est d'un intérêt très secondaire), mais surtout : *quelle phrase* est préférée en quelle version ?

Les figures 4.6 et 4.7 tenteront de répondre à ces deux problèmes. Dans la première, chaque paire de versions des stimuli est représenté par un des côtés du grand triangle. Les versions en question sont indiquées près des sommets de ce triangle. À l'intérieur, on trouve, pour les figures 4.6(a) et 4.6(c), un petit triangle pour chaque *sujet*. Les sommets de ces petits triangles sont placés sur les côtés du grand selon la préférence exprimée par le sujet dans les comparaisons associées aux côtés. Par exemple, un sujet (pas très sélectif) qui n'aurait exprimé aucune préférence dans les trois paires de comparaison, serait représenté par un triangle liant les milieux des trois côté du grand triangle. Plus il exprime sa préférence pour une des trois versions, plus les sommets qui correspondent migrent vers la version préférée. Enfin, dans les figures 4.6(b) et 4.6(d), un petit triangle correspond à une *phrase* donnée.

On doit reconnaître que la lecture de ces dessins n'est pas particulièrement aisée, et certains préféreront sans doute l'étude des chiffres dans les tableaux de l'annexe. De plus, l'identification des phrases n'est toujours pas possible. D'où l'idée de représenter les petits triangles par leurs centres de gravité.¹⁰ C'est ce que propose la figure 4.7. Le prix à payer pour cette simplification est évidemment une perte d'information : le sujet indécis que nous avons pris pour exemple plus haut sera représenté par le même point au centre qu'un sujet qui choisit p.e. entre Org et igm : l'original ; entre SFC et Org : le SFC et entre SFC et igm : l'igm. Ce sujet serait « indécis » d'une autre manière. L'inspection des triangles de la figure 4.6 montre cependant que ce type de sujet était absent de notre ensemble d'auditeurs.

Ajoutons deux remarques : la première pour préciser que les centres de gravité ne peuvent pas atteindre les sommets du grand triangle de la figure 4.6 – raison pour laquelle les dessins se présentent sous forme d'hexagones¹¹.

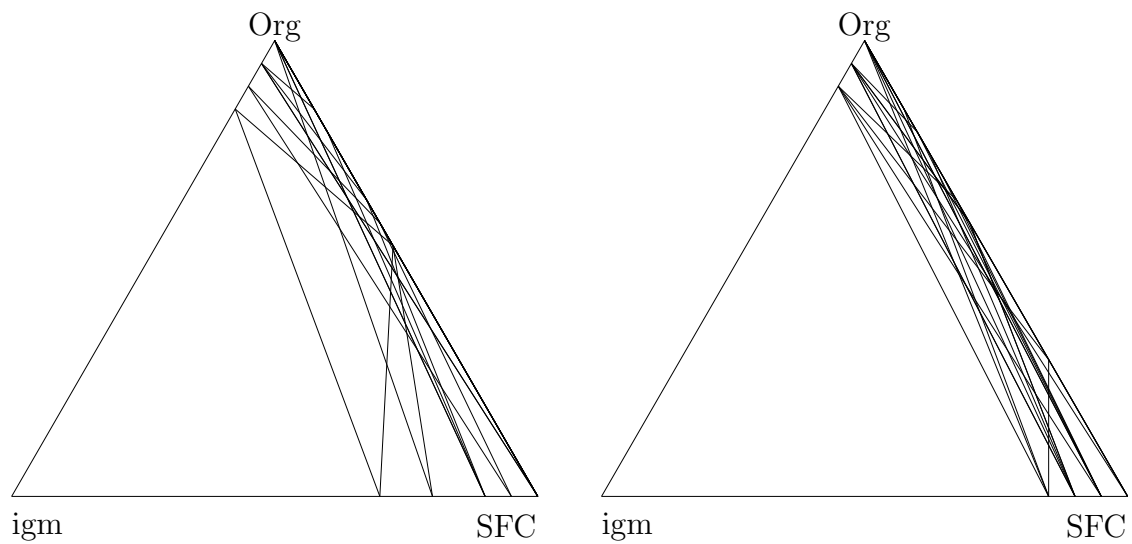
La deuxième remarque tente d'expliquer l'origine des formes « étranges » des centres de gravité : en effet, il ne s'agit pas exactement des centres des triangles de la figure précédente. Prenons l'exemple de la phrase 306 dans la figure 4.7(d) car le triangle correspondant dans la figure 4.6(d) est facilement identifiable (en gras) : de tous les votes pour cette phrase en comparaison SFC–igm, nous avons enlevé tour à tour ceux

¹⁰Les centres de gravité des triangles peuvent être interprétés d'une autre manière : leurs emplacements correspondent à la somme des chemins parcourus en partant du centre du grand triangle en direction de ses sommets, en pondérant ces déplacements par le nombre total de votes pour les versions correspondantes.

¹¹La zone du SFC se trouve tout naturellement au sud-est...

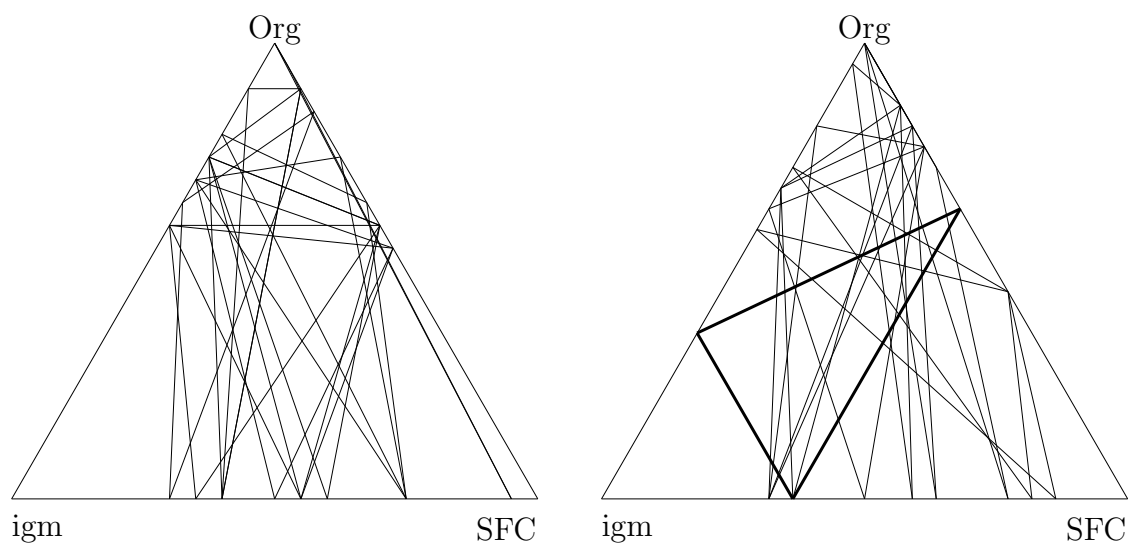
d'un sujet (le vote maximal se réduisant alors à $18 : 9 \text{ sujets} \times 2 \text{ écoutes}$). Nous obtenons ainsi 10 scores différents dont nous calculons la moyenne et l'écart-type. Ensuite, nous retenons les deux valeurs (moyenne – écart-type) et (moyenne + écart-type) que nous marquons sur le côté SFC–igm du grand triangle. En procédant de la même manière pour les paires Org–SFC et Org–igm, nous nous retrouvons avec trois fois deux points qui donnent lieu à six petits triangles. Les centres de gravité de ces triangles forment les sommets des petits hexagones irréguliers de la figure 4.7.

Précisons que le but de cette manipulation n'est pas de visualiser l'exactitude des jugements pour une phrase donnée ou la « fiabilité » d'un sujet, mais plutôt de donner une idée de l'influence qu'ont les sélections (bien qu'aléatoires) de nos sujets et des formules qui leur ont été présentées durant ce test.



(a) Français. Sujets

(b) Français. Phrases



(c) Allemand. Sujets

(d) Allemand. Phrases

FIG. 4.6: **Mettre les votes en relation.** Chaque petit triangle représente un sujet (à gauche) ou une phrase (à droite). Les sommets des petits triangles sont placés sur les côtés du grand triangle en fonction de la relative préférence en comparaison directe entre les deux versions liées par ce côté.

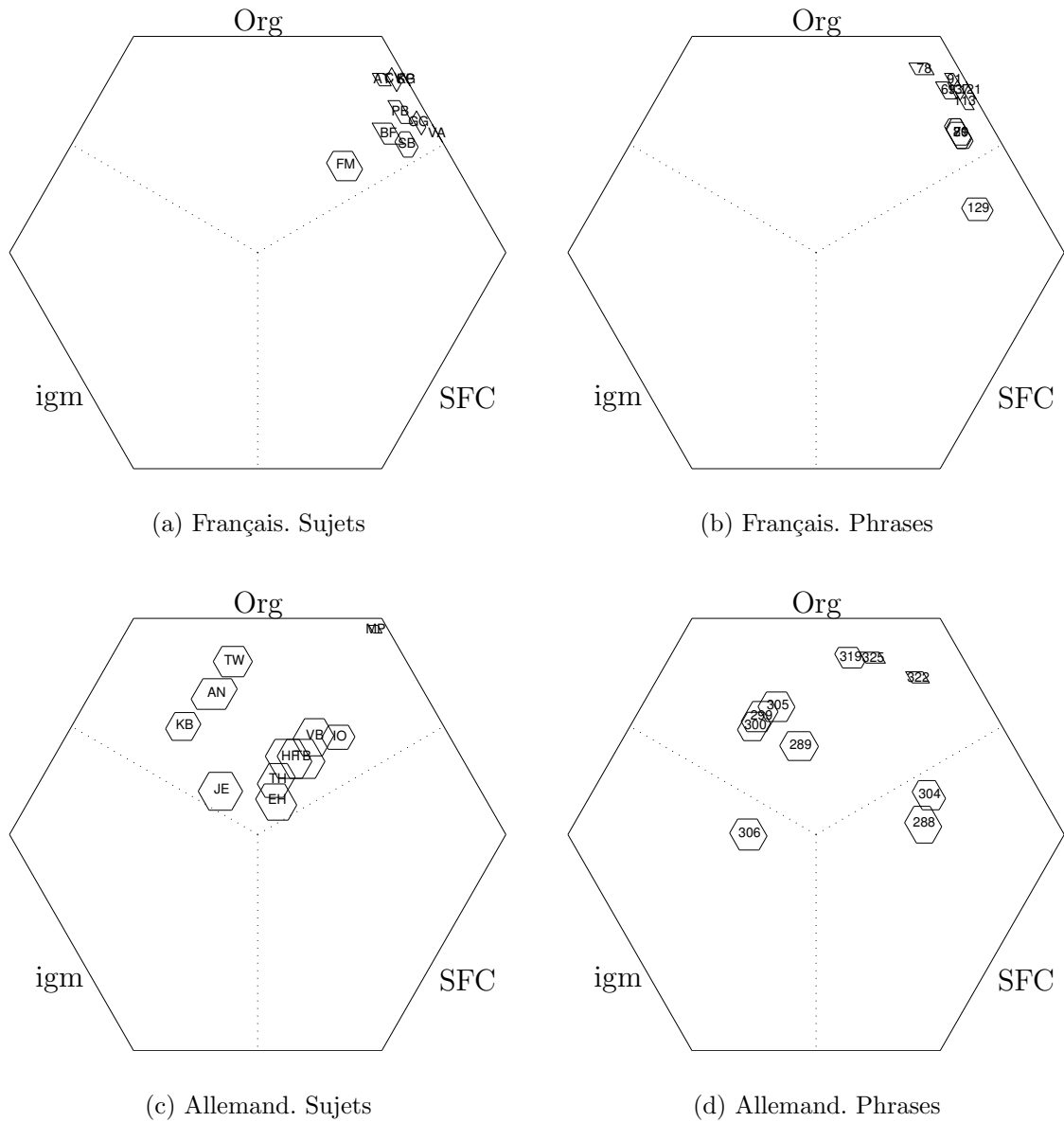


FIG. 4.7: **Centres de gravité.** Les « points » à l'intérieur de l'hexagone correspondent (avec un grain de sel – voir explications dans le texte) aux centres de gravité des triangles de la figure 4.6. L'hexagone même est la zone de déplacements possibles de ces centres à l'intérieur du grand triangle.

Discussion – test allemand. Que peut-on conclure de ces représentations graphiques des résultats? Pour les sujets qui ont participé au test en allemand, on s’aperçoit tout de suite qu’une personne sort du lot : il s’agit de MP avec l’étiquette quelque peu illisible dans l’extrême nord-est de la figure 4.7(d). Sa particularité est de rejeter quasiment tous les stimuli synthétisés avec les paramètres prédits par l’igm et ceci tant en comparaison avec l’original (100%) que confrontés à la version SFC (95%). Entre les stimuli Org et SFC, sa préférence est aussi unanime en faveur de l’original (95%). Ce comportement, établissant une hiérarchie stricte $\text{Org} \searrow \text{SFC} \searrow \text{igm}$, est intéressant, car il montre que les trois versions ont des caractéristiques *typiques* qui déclenche chez ce sujet une acceptation ou un rejet systématique.

Les sujets TW et AN, préférant l’igm légèrement au SFC, apprécient surtout les paramètres originaux ce qui les rapproche du comportement de MP – mis à part le fait que leur préférence pour une des versions synthétiques est inversée et nettement moins tranchée. KB et JE s’opposent par le fait que KB exprime surtout son rejet du SFC. Les autres sujets se caractérisent par une relative grande acceptation du SFC comparé à l’original tout en préférant le SFC légèrement à l’igm.

Pour les phrases, nous remarquons que seul¹² le numéro 306 « réussit » à se faire préférer en comparaison directe avec l’original. Le contour de f_0 dans la figure C.1 de l’annexe montre qu’il est très bien corrélé avec l’original en ayant une dynamique plus forte. Les deux phrases se situant dans la zone SFC doivent leur « succès » à leur plus grande acceptabilité face aux versions igm. Le groupe des trois phrases situé le plus en haut à gauche consiste en des formules relativement longues avec une structure semblable : la prédiction de l’igm montre des faiblesses dans la deuxième partie de ces énoncés. Des formules restantes ayant obtenu des résultats légèrement en faveur de l’igm, il s’agit majoritairement de cas où le SFC manque de dynamique ; l’exception étant le numéro 299 : sa génération implique de toute évidence des contours se superposant dont la configuration phonotactique n’est pas assez représentée dans le corpus d’apprentissage – ce qui se traduit généralement par une oscillation rapide des contours de f_0 (cf. figure C.1 de l’annexe).

En résumé, on peut constater que les deux modèles fournissent des prédictions globalement acceptables. Les erreurs de l’igm semblent plus localisées (omission ou décalage d’accent), tandis que le SFC – quelque peu plus monotone pour l’allemand – semble plus passe-partout.

Discussion – test français. Pour le français, les lacunes de notre implémentation actuelle de l’igm rendent l’analyse des résultats difficile. On s’aperçoit ainsi de la valeur de disposer de deux modèles « compétitifs », comme c’était le cas pour l’allemand, car ceci permet d’identifier plus clairement des problèmes pour chacun des modèles. Mis à part le sujet FM, clairement moins sélectif, tout le monde rejette cette version. Les

¹²cf. le triangle gras de la figure 4.6(d)

résultats pour le SFC comparés à l'original sont meilleurs qu'en allemand.

Terminons cette fouille des données avec une mise en garde : la phrase 129 est clairement préférée quand elle est synthétisée avec les paramètres du SFC. Ceci semble fournir la preuve que le modèle est capable de générer une réalisation prosodique meilleure que le locuteur humain – ce cas de figure étant tout à fait plausible pour des énoncés individuels. Mais prenons le temps d'une inspection approfondie de la phrase en question. Il s'avère que la différence la plus perceptible entre les deux versions est très localisée : il s'agit du GIPC autour de la frontière entre « six » et « facteur de » dans $\sqrt{6} \cdot (9x^2 - 5x + 6) = 0$. Le [s] de l'original est, avec 300 ms, exceptionnellement long ce qui n'est pas gênant dans l'enregistrement, puisque il est de faible intensité. Or, la version *resynthétisée* emploie un diphone très énergétique à la fin du [s] qui supporte en occurrence mal son allongement par le traitement de signal : le pic d'énergie devient perceptivement gênant. La différence des jugements pour cette phrase n'est donc vraisemblablement pas due à la qualité des paramètres prosodiques, mais plutôt liée à l'interaction des diphones disponibles avec ces paramètres.

Préférences subjectives et mesures objectives. Parce qu'elle seule permet de juger de la qualité des modèles, l'évaluation perceptive de la prosodie synthétique est indispensable. Les mesures objectives (différents types d'erreurs, corrélations etc.) doivent, en général, faire impasse sur l'impact réel des déviations entre prédictions et observations. Elles sont néanmoins elles-aussi indispensables, car facilement et rapidement accessibles, afin de guider le développement de tout modèle. On se demande donc naturellement quelles relations entretiennent ces mesures avec les résultats subjectifs discutés ci-dessus.

Avant de fournir des éléments de réponse à cette question, faisons une parenthèse non sans importance sur la manière dont le test de préférence a été conçu (cf. page 182) : quelle est la relation entre les *erreurs* objectives des modèles et les *différences* entre leurs prédictions ? Je rappelle que, dans l'idée de rendre le test plus sélectif, les stimuli ont été choisis afin de maximiser la différence entre les versions synthétiques.¹³ Il n'a pas été question de privilégier *a priori* l'un ou l'autre modèle, mais sous l'hypothèse que les deux modèles fournissent des prédictions majoritairement correctes, cette procédure sélectionne les stimuli qui sont mauvais pour *un* des deux modèles. Pour les 30 phrases disponibles pour le test en allemand, la corrélation entre la différence des versions synthétiques et l'erreur par rapport à l'observation est de 0,39 pour le SFC et de 0,64 pour l'igm. Autrement dit : plus les modèles sont en désaccord, plus ils ont tendance à différer beaucoup de l'observation. Les formules sélectionnées ont donc été plutôt celles qui posent le plus de difficultés aux *deux* modèles (cette tendance étant plus forte pour l'igm).

Compte tenu des problèmes pour le français avec notre implémentation actuelle de

¹³Un choix qui n'a probablement pas été judicieux pour le français.

l'igm, il n'est pas surprenant que la procédure de sélection le pénalise beaucoup plus : la corrélation de 0,88 entre différence inter-modèle et erreur de l'igm souligne cet effet. De côté du SFC, la corrélation devient même négative : $-0,19$. Il profiterait donc, bien que très légèrement, de la sélection.

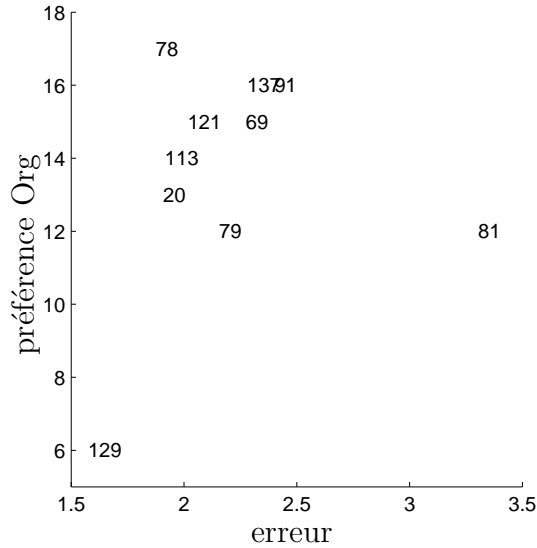
Revenons aux résultats subjectifs : la figure 4.8 montre les relations entre les erreurs objectives et les votes obtenus par la version originale en comparaison directe. Les corrélations sont indiquées sous les graphiques, mais, au vue du faible nombre de données, il me semble préférable d'apprécier leur distribution « à l'œil nu ». Mon interprétation est que le seul cas où on peut identifier une véritable relation est celui du SFC en allemand : plus l'erreur est grande, plus l'original est préféré. Ceci semble cohérent avec l'observation faite plus haut que ses prédictions suivent bien les observations mais qu'il manque pour certains énoncés de dynamique – la relation entre ce type d'erreur et la qualité perçue semble plus linéaire que les erreurs incidentes que nous avons décrites pour l'igm.

Notons finalement qu'en réunissant les données des deux modèles dans un même graphique, on aurait obtenu l'impression de corrélations plus fortes – évidemment surtout pour le français. Mais, la relation intra-modèle me semble ici plus importante, et ne devrait pas être masquée par une observation triviale tel que « un modèle produisant de très grandes erreurs obtient des mauvais scores ». Retenons plutôt : certains types de défauts entraînent des erreurs objectives qui sont en relation plus ou moins directe avec la qualité perçue ; d'autres nécessitent une évaluation perceptive pour être mis en lumière du jour.

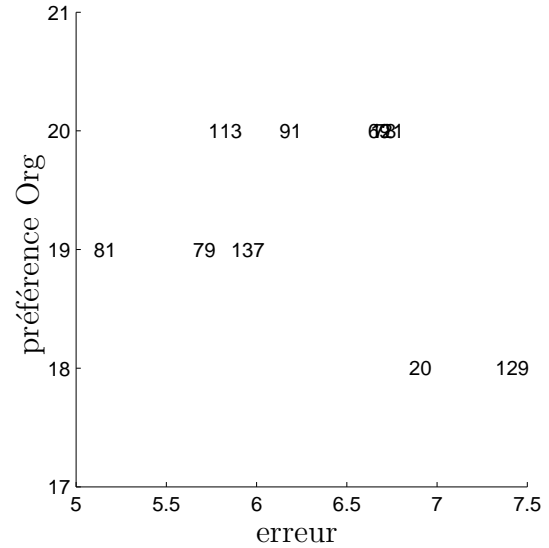
4.2.4 Conclusions

Bien que cette évaluation comparative soit de faible envergure aussi bien en nombre de sujet qu'en nombre de phrases différentes testées, elle semble refléter une image plutôt fidèle de l'état actuel des deux modèles. À aucun moment il n'a été question de juger de la supériorité d'un modèle sur l'autre : le temps disponible pour l'adaptation de l'igm à l'énonciation des formules mathématiques a été relativement restreint et n'a pas permis une recherche très approfondie sur la meilleure représentation phonologique à fournir en entrée du RN. L'extraction des paramètres Fujisaki pour le français s'est avérée également plus difficile que pour l'allemand ce qui donne un élément supplémentaire pour expliquer les résultats négatifs pour le français.

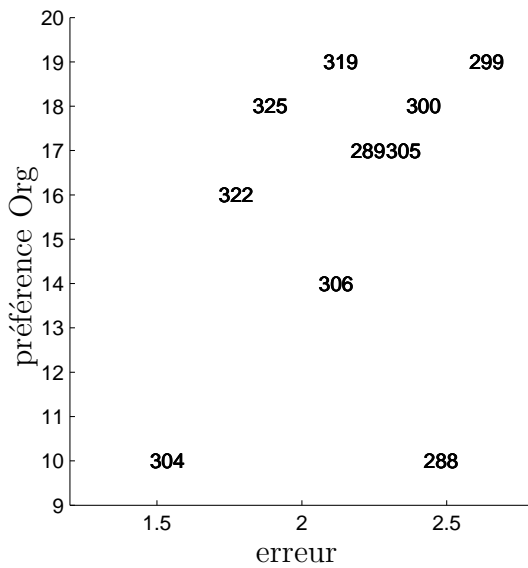
D'un autre côté, l'application du SFC à l'allemand s'est déroulée d'une manière étonnamment facile. Rappelons qu'il a été transposé tel quel et que l'on aurait pu s'attendre à une plus forte dégradation par rapport au français puisqu'il ne dispose d'aucune information sur les emplacements des accents lexicaux. La nature du corpus avec ses mots courts et répétés à l'intérieur de structures semblables facilite probablement la tâche pour le SFC.



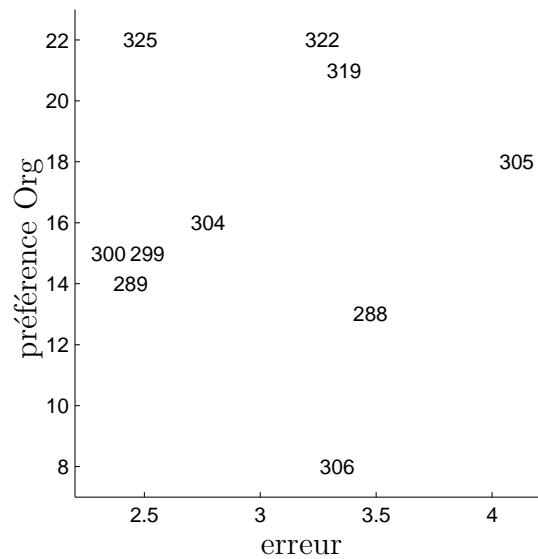
(a) SFC – français (corrélation : 0,19)



(b) igm – français (corrélation : -0,23)



(c) SFC – allemand (corrélation : 0,33)



(d) igm – allemand (corrélation : 0,03)

FIG. 4.8: **Erreurs objectives vs. préférences subjectives.** L'ordonnée représente les scores de préférence de l'original par rapport aux modèles. L'abscisse indique l'erreur RMS des prédictions. Les nombres correspondent aux numéros de phrases. Attention : par souci de lisibilité nous avons été contraint d'utiliser des échelles différentes pour les quatre graphiques.

Conclusions et perspectives

La prosodie est un phénomène complexe. Si l'on considère qu'elle encode d'une manière ou d'une autre un certain nombre d'informations en vue de leur transmission à travers le signal acoustique, la tâche d'un modèle de génération de la prosodie est de proposer une méthode qui réalise cette transition entre informations et leurs réalisations.

On fait souvent la distinction entre des modèles dits compréhensifs qui tenteraient d'identifier les *vrais* processus cognitifs mis en œuvre et les modèles imitatifs qui se contentent de produire un bon résultat – en exagérant : peu importe les moyens. De mon point de vue, la présente étude ne s'intègre ni dans l'une ni dans l'autre de ces approches : je ne prétend donc pas la *réalité cognitive* des contours associés aux fonctions communicatives et, en même temps, je suis convaincu que le modèle proposé est plus qu'une approximation numérique des contours prosodiques observés.

En quoi consiste donc la contribution de cette étude ? L'essentiel d'un modèle qui se veut plus qu'une boîte noire est de proposer une factorisation du processus de génération en différents éléments dont l'interaction reste la plus simple possible. Ces éléments peuvent être des tons que l'on identifie dans la réalisation prosodique. Le rôle du modèle serait alors de décrire leur relation avec les informations (para-)linguistiques et leur interaction entre eux. Ces éléments peuvent également être des contours d'accents d'inspiration physiologique avec un modèle associé qui les lie à des exigences de haut niveau.

L'intérêt du modèle exposé ici est qu'il propose une *méthode itérative* permettant de faire émerger des objets interagissants. Au cœur de notre approche est l'idée qu'on ne peut observer des faits prosodiques que le résultat d'un processus plus ou moins complexe. Le SFC repose sur l'hypothèse que les réalisations prosodiques observées peuvent être décomposées en une simple superposition de contours qui sont directement associés à des fonctions de haut niveau. Ces contours sont à l'état initial – c.-à.-d. avant la confrontation du modèle aux observations – plats et égaux à zéro, ce qui correspond à la connaissance *a priori* sur la forme que ces contours pourraient prendre : elle est nulle. Ce qui n'est pas nul est le rôle (ou le sens) que nous attribuons aux contours. C'est

l'exploitation systématique de la différence entre contours synthétiques et contours observés qui permet de donner au sens des contours une forme. Le cycle de l'analyse par la synthèse est ainsi contraint par le sens des objets (les contours) et par une efficacité maximale dans le cadre du mécanisme de leur interaction (superposition additive).

Je tiens à souligner que toute analyse que le SFC peut fournir doit s'appuyer sur de *multiples observations*. La décomposition qu'il propose n'a un sens que par rapport à un ensemble de données (corpus). Il s'agit là d'une différence importante avec d'autres modèles superpositionnels comme ceux de type Fujisaki : leurs contraintes permettant l'inversion peuvent opérer sur la base d'un seul énoncé ; mais ils n'établissent pas de relation entre fonctions communicatives et objets prosodiques comme le SFC le fait.

Les résultats obtenus grâce à cette approche montrent qu'une factorisation de la réalisation prosodique par rapport aux fonctions de haut niveau est capable de générer une prosodie riche et proche de l'original.

Le SFC implique une autre factorisation : au niveau de la composante rythmique, nous tentons d'isoler micro- et macro-rythme en utilisant le coefficient d'allongement qui joue le rôle d'un indicateur de vitesse d'élocution locale. Sa définition repose sur des observations obtenues à l'aide de corpora qui sont, par rapport au macro-rythme, le plus neutres possible. Or, il me semble intéressant de considérer la présente définition plutôt comme une première approximation : la prédiction du SFC basée sur cette définition pourrait être utilisée afin de « soustraire » la composante macro-rythmique d'un corpus qui ne serait ainsi plus soumis à la contrainte d'être « neutre » – une contrainte qui n'est de fait jamais satisfaite. Les données ainsi « nettoyées » pourront alors servir à aboutir à une deuxième approximation de la composante micro-rythmique. Et, on entrera dans une boucle d'analyse par la synthèse qui vise une factorisation optimale entre micro- et macro-rythme en s'appuyant sur des observations multiples de la manifestation nécessairement conjointe de ces deux composantes.

Une telle approche n'est-elle pas auto-justifiante et difficilement falsifiable ? Oui et non. Oui : elle comporte le risque d'attacher aux objets identifiés plus de signification qu'ils n'en ont. Dire que les familles de contours que nous associons avec une telle ou telle fonction sont *la* manifestation prosodique de cette fonction va certainement trop loin – pourtant elles le sont : mais uniquement dans le cadre des hypothèses, c.-à.-d. si l'on accepte que leur interaction se fait par superposition additive. Et non : cette approche n'est pas auto-justifiante puisque ses prédictions sont constamment confrontées aux observations. Au contraire, je pense que l'approche permet particulièrement bien de tester la valeur de différentes hypothèses. Ainsi, je considère les fonctions implémentées par le SFC au cours de cette étude non pas comme figées mais sujettes à des modifications au fil de leurs confrontations avec de nouvelles données.

Perspectives. J'espère que la partie de ce manuscrit sur les résultats et applications du modèle aura montré à quel point je conçois le SFC en tant qu'outil : en s'appuyant sur des fonctions « comprises », chaque analyse effectuée ouvre des pistes à de nouvelles

aventures. Ainsi, réaliser un modèle de variantes (comme seulement esquissé sur la page 158) serait de grand intérêt.

Dans le chapitre sur l'expansion de mouvements, nous avons observé des difficultés liées à l'extrapolation des contours. L'intégration du SFC dans un système de synthèse « grandeur nature » rend une réponse à ce problème indispensable. Elle devrait passer par des contraintes imposées aux entrées des générateurs ou des paramétrisations de leurs sorties (cf. page 106).

Implémenter des nouvelles fonctions, comme par exemple la focalisation, ne serait qu'un autre challenge.

Quelle est la relation entre familles de contours d'une même fonction mais de différents locuteurs ?

Nous n'avons considéré au cours de cette étude que deux dimensions de la prosodie : la fréquence fondamentale et le rythme. Étendre l'approche d'analyse par la synthèse aux dimensions multi-modaux de la prosodie (mouvements de la tête etc.) me semble également une voie prometteuse.

Corpora

A.1 Les formules mathématiques – français

$6x < 1$	(A.1.1)	$2x \times 9 < 1$	(A.1.19)
$5x - 9 = 0$	(A.1.2)	$2x \cdot (8x + 6) = 0$	(A.1.20)
$2x^2 + 5x - 2 = 2$	(A.1.3)	$8x \cdot (6x^2 - 9x + 6) > 2$	(A.1.21)
$9 = 7x$	(A.1.4)	$\frac{5x - 9}{6} = 0$	(A.1.22)
$7x = 6x$	(A.1.5)	$(9x + 6) \times 7x = 0$	(A.1.23)
$8x - 2 = 2x$	(A.1.6)	$\frac{5x - 2}{8x - 2} > 1$	(A.1.24)
$7x^2 - 8x + 9 > 6x$	(A.1.7)	$\frac{9x - 9}{7x^2 - 9x + 6} = 2$	(A.1.25)
$6 > 9x + 6$	(A.1.8)	$\frac{5x^2 - 7x - 6}{6} = 0$	(A.1.26)
$7x = 6x + 2$	(A.1.9)	$(8x^2 + 8x + 2) \times 5x > 2$	(A.1.27)
$7x - 9 = 9x + 2$	(A.1.10)	$\frac{5x^2 + 8x - 9}{6x + 2} = 2$	(A.1.28)
$7x^2 + 8x - 2 < 2x - 9$	(A.1.11)	$\frac{9x^2 - 6x - 2}{7x^2 - 8x - 6} = 2$	(A.1.29)
$6 = 5x - 2x + 2$	(A.1.12)	$(2x - 6)^2 = 0$	(A.1.30)
$2x > 9x^2 + 8x + 9$	(A.1.13)	$6 \times 7x = 2x$	(A.1.31)
$2x + 2 < 5x^2 - 2x - 2$	(A.1.14)	$\frac{2}{2x - 6} = 8x$	(A.1.32)
$5x^2 + 8x - 6 = 2x^2 - 6x - 2$	(A.1.15)	$\frac{9}{6x^2 + 6x + 2} = 6x$	(A.1.33)
$\frac{2}{5x} = 0$	(A.1.16)	$\frac{8x}{2} < 8x$	(A.1.34)
$2 \cdot (6x - 2) = 0$	(A.1.17)	$\frac{6x}{7x + 2} > 5x$	(A.1.35)
$2 \cdot (2x^2 - 6x - 9) < 2$	(A.1.18)		

$$\frac{6x}{8x^2 - 8x + 9} = 2x \quad (\text{A.1.36})$$

$$(7x - 9) \times 9 = 2x \quad (\text{A.1.37})$$

$$\frac{2x - 6}{2x} < 2x \quad (\text{A.1.38})$$

$$(9x + 9) \cdot (5x - 6) < 6x \quad (\text{A.1.39})$$

$$(9x + 6) \cdot (7x^2 - 8x + 9) < 5x \quad (\text{A.1.40})$$

$$(7x^2 - 7x + 2) \times 9 = 9x \quad (\text{A.1.41})$$

$$\frac{6x^2 + 9x + 9}{7x} > 5x \quad (\text{A.1.42})$$

$$(8x^2 + 9x - 2) \cdot (2x + 2) = 9x \quad (\text{A.1.43})$$

$$(5x^2 + 7x - 2) \cdot (5x^2 + 7x - 9) < 6x \quad (\text{A.1.44})$$

$$(6x - 2)^2 = 5x \quad (\text{A.1.45})$$

$$\frac{9}{6x} = 8x - 6 \quad (\text{A.1.46})$$

$$6 \cdot (6x - 2) = 7x - 9 \quad (\text{A.1.47})$$

$$9 \cdot (8x^2 + 7x - 6) = 9x - 9 \quad (\text{A.1.48})$$

$$8x \times 9 = 9x + 9 \quad (\text{A.1.49})$$

$$8x \cdot (8x - 6) = 2x - 6 \quad (\text{A.1.50})$$

$$5x \cdot (9x^2 + 8x + 2) = 8x + 6 \quad (\text{A.1.51})$$

$$\frac{6x + 6}{2} = 2x - 2 \quad (\text{A.1.52})$$

$$(9x - 6) \times 2x = 9x - 2 \quad (\text{A.1.53})$$

$$\frac{6x + 6}{8x + 9} = 5x + 9 \quad (\text{A.1.54})$$

$$\frac{5x - 9}{8x^2 + 6x + 6} = 2x - 2 \quad (\text{A.1.55})$$

$$\frac{8x^2 - 2x + 9}{2} = 9x - 9 \quad (\text{A.1.56})$$

$$(5x^2 - 2x - 2) \times 8x = 8x - 6 \quad (\text{A.1.57})$$

$$\frac{8x^2 - 5x - 6}{8x + 9} > 5x - 9 \quad (\text{A.1.58})$$

$$\frac{7x^2 - 7x - 9}{9x^2 - 9x - 2} > 9x - 9 \quad (\text{A.1.59})$$

$$(9x - 6)^2 = 2x + 9 \quad (\text{A.1.60})$$

$$2 \times 2x = 6x^2 + 7x - 9 \quad (\text{A.1.61})$$

$$\frac{9}{5x - 9} < 5x^2 - 2x - 9 \quad (\text{A.1.62})$$

$$\frac{6}{5x^2 - 2x - 6} = 6x^2 + 5x + 6 \quad (\text{A.1.63})$$

$$\frac{8x}{9} = 5x^2 - 2x + 6 \quad (\text{A.1.64})$$

$$\frac{2x}{2x + 2} = 5x^2 + 6x + 2 \quad (\text{A.1.65})$$

$$\frac{5x}{9x^2 - 9x + 2} < 5x^2 + 9x + 2 \quad (\text{A.1.66})$$

$$(9x - 6) \times 2 = 8x^2 + 9x - 6 \quad (\text{A.1.67})$$

$$\frac{8x + 9}{8x} > 8x^2 + 7x + 6 \quad (\text{A.1.68})$$

$$(9x + 6) \cdot (6x - 9) < 8x^2 - 8x + 9 \quad (\text{A.1.69})$$

$$(2x + 6) \cdot (8x^2 + 9x + 2) = 5x^2 - 9x + 2 \quad (\text{A.1.70})$$

$$(9x^2 - 5x + 9) \times 2 = 5x^2 - 7x - 2 \quad (\text{A.1.71})$$

$$\frac{9x^2 - 6x - 2}{5x} = 5x^2 - 7x + 2 \quad (\text{A.1.72})$$

$$(5x^2 - 8x + 9) \cdot (5x - 6) = 5x^2 + 2x + 9 \quad (\text{A.1.73})$$

$$(9x^2 + 6x - 2) \cdot (2x^2 + 7x - 9) = 7x^2 - 8x + 6 \quad (\text{A.1.74})$$

$$(6x + 9)^2 = 7x^2 + 2x + 9 \quad (\text{A.1.75})$$

$$(2x + 9)^2 > (5x + 6)^2 \quad (\text{A.1.76})$$

$$\frac{2 - 6}{7x} = 2 \quad (\text{A.1.77})$$

$$2 - 9 \cdot (5x + 9) > 2 \quad (\text{A.1.78})$$

$$9 - 6 \cdot (5x^2 - 6x - 2) = 0 \quad (\text{A.1.79})$$

$$6 + 9x \times 6 = 1 \quad (\text{A.1.80})$$

$$6 + 2x \cdot (9x - 6) = 0 \quad (\text{A.1.81})$$

$$6 - 7x \cdot (2x^2 + 9x - 6) = 1 \quad (\text{A.1.82})$$

$$9 - \frac{2x + 6}{6} > 1 \quad (\text{A.1.83})$$

$$6 + (5x + 6) \times 7x = 1 \quad (\text{A.1.84})$$

$$9 + \frac{5x - 9}{7x - 6} < 2 \quad (\text{A.1.85})$$

$$6 - \frac{8x - 6}{7x^2 - 6x + 9} = 2 \quad (\text{A.1.86})$$

$$2 + \frac{7x^2 + 8x + 2}{6} = 0 \quad (\text{A.1.87})$$

$$2 - (5x^2 + 8x - 6) \times 9x < 1 \quad (\text{A.1.88})$$

$$2 + \frac{2x^2 - 9x - 9}{6x - 2} = 2 \quad (\text{A.1.89})$$

$$9 + \frac{2x^2 - 8x - 2}{6x^2 - 8x + 6} = 1 \quad (\text{A.1.90})$$

$$6 + (9x - 9)^2 = 0 \quad (\text{A.1.91})$$

$$2 \times 9x + 9 = 0 \quad (\text{A.1.92})$$

$$\frac{6}{9x - 6} - 6 = 0 \quad (\text{A.1.93})$$

$$\frac{9}{2x^2 - 2x - 9} - 9 = 0 \quad (\text{A.1.94})$$

$$\frac{6x}{2} + 9 = 0 \quad (\text{A.1.95})$$

$$\frac{5x}{9x + 6} + 9 = 2 \quad (\text{A.1.96})$$

$$\frac{9x}{7x^2 + 5x - 6} + 9 > 2 \quad (\text{A.1.97})$$

$$(2x - 2) \times 9 - 6 < 1 \quad (\text{A.1.98})$$

$$\frac{5x + 2}{9x} - 9 < 1 \quad (\text{A.1.99})$$

$$(2x - 6) \cdot (9x + 2) + 2 = 0 \quad (\text{A.1.100})$$

$$(5x + 2) \cdot (8x^2 + 8x - 9) - 9 > 2 \quad (\text{A.1.101})$$

$$(6x^2 - 9x + 2) \times 2 + 6 = 2 \quad (\text{A.1.102})$$

$$\frac{7x^2 + 9x + 2}{8x} + 9 = 0 \quad (\text{A.1.103})$$

$$(5x^2 - 2x - 9) \cdot (6x + 6) + 9 = 0 \quad (\text{A.1.104})$$

$$(2x^2 + 6x - 2) \cdot (2x^2 + 8x - 6) + 2 = 0 \quad (\text{A.1.105})$$

$$(9x - 9)^2 + 2 = 2 \quad (\text{A.1.106})$$

$$\frac{5x}{8x - 9} + \frac{2x}{8x - 9} = 0 \quad (\text{A.1.107})$$

$$6x \cdot (9x - 2) + (5x + 9) \times 7x = 0 \quad (\text{A.1.108})$$

$$\frac{7x}{6x + 9} + \frac{8x + 2}{2x + 9} = 0 \quad (\text{A.1.109})$$

$$\frac{6x}{5x + 9} - (9x + 6)^2 < 1 \quad (\text{A.1.110})$$

$$(6x - 2) \times 8x + 2x \cdot (2x - 6) = 0 \quad (\text{A.1.111})$$

$$\frac{7x + 9}{2x} + \frac{8x + 2}{7x} = 0 \quad (\text{A.1.112})$$

$$(2x - 9) \times 5x - (9x + 6) \cdot (7x + 9) = 2 \quad (\text{A.1.113})$$

$$(5x - 2) \times 6x + (2x + 6)^2 = 0 \quad (\text{A.1.114})$$

$$\frac{8x + 6}{7x - 6} - \frac{7x}{2x - 2} = 1 \quad (\text{A.1.115})$$

$$(9x + 2) \cdot (8x - 6) + (9x - 9) \times 2x = 1 \quad (\text{A.1.116})$$

$$\frac{9x - 2}{9x + 6} - \frac{6x + 6}{9x + 2} = 1 \quad (\text{A.1.117})$$

$$\frac{2x + 9}{5x + 6} - (8x + 6)^2 = 2 \quad (\text{A.1.118})$$

$$(7x - 6)^2 + 6x \cdot (7x - 9) = 1 \quad (\text{A.1.119})$$

$$(5x - 6)^2 - \frac{7x + 2}{5x} = 0 \quad (\text{A.1.120})$$

$$(5x + 6)^2 - \frac{6x - 9}{2x + 2} = 1 \quad (\text{A.1.121})$$

$$(7x - 9)^2 - (8x + 6)^2 = 2 \quad (\text{A.1.122})$$

$$\sqrt{x} = 1 \quad (\text{A.1.123})$$

$$\sqrt{5x} = 0 \quad (\text{A.1.124})$$

$$|7x - 9| = 0 \quad (\text{A.1.125})$$

$$\sqrt{8x^2 - 9x - 2} < 2 \quad (\text{A.1.126})$$

$$\sqrt{\frac{2}{9x}} = 1 \quad (\text{A.1.127})$$

$$\sqrt{6 \cdot (7x + 6)} = 2 \quad (\text{A.1.128})$$

$$\sqrt{6 \cdot (9x^2 - 5x + 6)} = 0 \quad (\text{A.1.129})$$

$$\sqrt{8x \times 6} < 1 \quad (\text{A.1.130})$$

$$\sqrt{5x \cdot (7x - 9)} = 1 \quad (\text{A.1.131})$$

$$\sqrt{7x \cdot (7x^2 - 8x + 6)} = 2 \quad (\text{A.1.132})$$

$$\sqrt{\frac{5x + 2}{2}} = 0 \quad (\text{A.1.133})$$

$$|(7x - 9) \times 9x| = 0 \quad (\text{A.1.134})$$

$$\left| \frac{7x - 6}{9x - 9} \right| = 0 \quad (\text{A.1.135})$$

$$\left| \frac{6x + 6}{9x^2 - 9x - 2} \right| = 1 \quad (\text{A.1.136})$$

$$\left| \frac{8x^2 + 5x - 9}{6} \right| < 2 \quad (\text{A.1.137})$$

$$|(9x^2 + 7x + 9) \times 2x| = 0 \quad (\text{A.1.138})$$

$$\sqrt{\frac{6x^2 + 5x - 6}{2x + 6}} = 0 \quad (\text{A.1.139})$$

$$\sqrt{\frac{5x^2 + 2x + 6}{8x^2 + 5x + 6}} < 2 \quad (\text{A.1.140})$$

$$\left| 9 - \frac{9x}{7x + 2} \right| = 0 \quad (\text{A.1.141})$$

$$\sqrt{9 + (8x - 9) \times 7x} = 0 \quad (\text{A.1.142})$$

$$\sqrt{2x \cdot (6x - 6) - 2} = 2 \quad (\text{A.1.143})$$

$$\begin{aligned} \sqrt{\frac{6x-2}{9x}} - 6 < 1 & \quad (\text{A.1.144}) & \left| \frac{5x+2}{8x-6} + \frac{9x}{6x+6} \right| > 2 & \quad (\text{A.1.151}) \\ \left| \frac{7x}{2x-2} + \frac{7x}{6x+9} \right| = 1 & \quad (\text{A.1.145}) & \sqrt{(6x-9) \cdot (8x-9) + (6x-2) \times 8x} = 1 & \quad (\text{A.1.152}) \\ |7x \cdot (5x+9) + (7x+2) \times 2x| > 2 & \quad (\text{A.1.146}) & \sqrt{\frac{8x+9}{9x-9} - \frac{6x+9}{6x-6}} = 1 & \quad (\text{A.1.153}) \\ \left| \frac{2x}{6x-9} - \frac{5x+2}{7x-2} \right| = 0 & \quad (\text{A.1.147}) & \sqrt{8x} = \sqrt{6x} & \quad (\text{A.1.154}) \\ |(2x-9) \times 9x - 7x \cdot (6x+9)| = 1 & \quad (\text{A.1.148}) & \sqrt{9x} = \sqrt{2x+2} & \quad (\text{A.1.155}) \\ \left| \frac{5x-9}{8x} - \frac{5x-2}{5x} \right| < 1 & \quad (\text{A.1.149}) & \sqrt{7x+2} < \sqrt{6x} & \quad (\text{A.1.156}) \\ |(7x+6) \times 5x - (8x-6) \cdot (6x+6)| = 0 & \quad (\text{A.1.150}) & \sqrt{6x-2} = \sqrt{8x-6} & \quad (\text{A.1.157}) \end{aligned}$$

A.2 Les formules mathématiques – allemand

$$\begin{aligned} 4x > 3 & \quad (\text{A.2.1}) & 3x - 7 > 6x & \quad (\text{A.2.18}) \\ 9x - 7 > 5 & \quad (\text{A.2.2}) & 9x^2 - 6x + 5 < 9x & \quad (\text{A.2.19}) \\ 9x^2 - 6x + 7 = 5 & \quad (\text{A.2.3}) & 6 \cdot (9x + 7) > 9x & \quad (\text{A.2.20}) \\ \frac{4}{7x-3} < 6 & \quad (\text{A.2.4}) & \frac{6}{8x^2 - 9x + 7} = 9x & \quad (\text{A.2.21}) \\ 6 \cdot (6x^2 - 9x + 4) = 9 & \quad (\text{A.2.5}) & 7x \cdot (9x - 5) = 8x & \quad (\text{A.2.22}) \\ \frac{6x}{3x+9} = 3 & \quad (\text{A.2.6}) & \frac{8x}{3x^2 + 8x + 6} < 3x & \quad (\text{A.2.23}) \\ 5x \cdot (6x^2 - 6x + 4) > 3 & \quad (\text{A.2.7}) & (5x + 6) \cdot 7 < 4x & \quad (\text{A.2.24}) \\ \frac{8x-4}{5} < 9 & \quad (\text{A.2.8}) & (8x - 9) \cdot 6x < 3x & \quad (\text{A.2.25}) \\ \frac{6x-5}{7x} < 5 & \quad (\text{A.2.9}) & \frac{8x+3}{7x-6} < 5x & \quad (\text{A.2.26}) \\ (7x+9) \cdot (6x+9) < 5 & \quad (\text{A.2.10}) & \frac{4x+3}{5x^2 + 8x - 7} > 6x & \quad (\text{A.2.27}) \\ \frac{3x-9}{5x^2 + 8x - 3} > 7 & \quad (\text{A.2.11}) & (3x^2 - 4x + 9) \cdot 7 < 6 & \quad (\text{A.2.28}) \\ \frac{8x^2 - 6x + 9}{5} < 9 & \quad (\text{A.2.12}) & \frac{3x^2 + 7x - 5}{6x} > 6x & \quad (\text{A.2.29}) \\ \frac{3x^2 - 8x - 5}{8x} = 3 & \quad (\text{A.2.13}) & \frac{4x^2 + 6x + 9}{5x - 7} < 3x & \quad (\text{A.2.30}) \\ \frac{8x^2 - 3x + 3}{5x - 4} = 3 & \quad (\text{A.2.14}) & (7x^2 + 8x + 7) \cdot (9x^2 - 9x - 3) = 7x & \quad (\text{A.2.31}) \\ (8x^2 - 5x - 5) \cdot (8x^2 + 7x + 9) > 3 & \quad (\text{A.2.15}) & (3x - 4)^2 < 9x & \quad (\text{A.2.32}) \\ (3x + 9)^2 > 3 & \quad (\text{A.2.16}) & 6 < 6x + 4 & \quad (\text{A.2.33}) \\ 9 > 7x & \quad (\text{A.2.17}) & 9x < 9x + 7 & \quad (\text{A.2.34}) \end{aligned}$$

$8x + 6 = 4x - 6$	(A.2.35)	$(8x^2 + 7x - 5) \cdot 7x = 4x^2 - 9x + 9$	(A.2.63)
$9x^2 + 4x + 7 = 4x + 5$	(A.2.36)	$\frac{3x^2 + 6x + 9}{6x + 5} > 8x^2 - 8x - 3$	(A.2.64)
$4 \cdot (3x + 6) > 8x + 7$	(A.2.37)	$\frac{6x^2 + 5x + 3}{7x^2 + 6x - 4} < 3x^2 + 9x - 7$	(A.2.65)
$5 \cdot (7x^2 - 9x - 9) = 4x - 7$	(A.2.38)	$(8x - 5)^2 > 4x^2 - 8x + 4$	(A.2.66)
$\frac{7x}{8x - 4} < 7x - 5$	(A.2.39)	$5 < 6 \cdot (6x + 4)$	(A.2.67)
$3x \cdot (5x^2 + 6x + 4) < 9x - 3$	(A.2.40)	$6x = 6 \cdot (9x - 4)$	(A.2.68)
$(8x - 4) \cdot 5 < 4x - 6$	(A.2.41)	$6x + 6 = \frac{7}{5x + 7}$	(A.2.69)
$\frac{5x - 7}{3x} = 8x + 7$	(A.2.42)	$6x^2 - 8x - 5 > 3 \cdot (7x - 9)$	(A.2.70)
$\frac{3x - 5}{8x - 4} > 3x + 3$	(A.2.43)	$\frac{5}{8x - 4} < 4 \cdot (5x - 3)$	(A.2.71)
$\frac{4x + 6}{6x^2 - 3x + 9} = 6x + 5$	(A.2.44)	$\frac{7}{5x^2 - 8x + 9} > 7 \cdot (4x - 4)$	(A.2.72)
$(4x^2 - 3x + 7) \cdot 7 < 9x - 3$	(A.2.45)	$\frac{9x}{7x + 5} = \frac{6}{3x + 4}$	(A.2.73)
$(5x^2 - 8x + 9) \cdot 5x > 6x - 3$	(A.2.46)	$\frac{9x}{7x^2 + 9x + 9} = \frac{4}{3x - 5}$	(A.2.74)
$\frac{3x^2 + 4x - 9}{9x + 5} < 8x + 3$	(A.2.47)	$\frac{4x + 4}{7} < \frac{6}{7x + 6}$	(A.2.75)
$(7x^2 + 6x + 3) \cdot (8x^2 - 3x + 6) = 4x + 3$	(A.2.48)	$(5x - 3) \cdot 7x < \frac{9}{8x - 6}$	(A.2.76)
$(3x + 5)^2 > 6x - 5$	(A.2.49)	$(5x - 3) \cdot (7x + 5) < 6 \cdot (5x + 7)$	(A.2.77)
$3 = 5x^2 + 3x + 4$	(A.2.50)	$\frac{5x - 9}{3x^2 - 7x + 6} = 3 \cdot (3x + 7)$	(A.2.78)
$3x = 8x^2 - 9x - 5$	(A.2.51)	$\frac{6x^2 + 8x + 4}{3} = 9 \cdot (9x - 6)$	(A.2.79)
$7x - 7 > 6x^2 - 9x - 7$	(A.2.52)	$\frac{5x^2 + 9x + 3}{5x} = 5 \cdot (9x - 7)$	(A.2.80)
$8x^2 + 4x + 5 < 7x^2 - 3x + 6$	(A.2.53)	$\frac{5x^2 - 6x - 3}{8x + 9} = \frac{3}{4x - 5}$	(A.2.81)
$6 \cdot (6x - 3) = 9x^2 + 4x - 7$	(A.2.54)	$\frac{3x^2 - 7x + 5}{6x^2 + 4x - 7} = 4 \cdot (9x - 4)$	(A.2.82)
$\frac{7}{7x^2 - 4x + 5} > 4x^2 - 3x + 3$	(A.2.55)	$(7x + 3)^2 > \frac{5}{4x + 7}$	(A.2.83)
$6x \cdot (7x + 5) < 4x^2 - 5x + 9$	(A.2.56)	$4 < \frac{9}{8x^2 + 4x - 3}$	(A.2.84)
$\frac{3x}{8x^2 + 9x + 5} > 3x^2 - 4x - 7$	(A.2.57)	$7x < 5 \cdot (7x^2 - 8x + 4)$	(A.2.85)
$(4x + 9) \cdot 5 = 5x^2 - 9x - 3$	(A.2.58)	$3x - 6 < 5 \cdot (8x^2 + 4x - 3)$	(A.2.86)
$\frac{4x - 7}{8x} = 6x^2 + 3x + 5$	(A.2.59)	$3x^2 + 9x - 6 = \frac{9}{8x^2 + 6x + 5}$	(A.2.87)
$(8x - 5) \cdot (9x + 7) > 3x^2 - 5x - 9$	(A.2.60)	$3 \cdot (9x - 6) > \frac{3}{3x^2 + 6x - 7}$	(A.2.88)
$(7x - 9) \cdot (3x^2 - 5x - 3) < 5x^2 - 5x - 3$	(A.2.61)		
$(8x^2 - 3x - 5) \cdot 4 = 3x^2 + 6x - 4$	(A.2.62)		

$$6 \cdot (4x^2 + 5x - 9) > \frac{3}{5x^2 + 4x - 7} \quad (\text{A.2.89})$$

$$6x \cdot (4x + 7) < 9 \cdot (7x^2 + 8x + 3) \quad (\text{A.2.90})$$

$$6x \cdot (7x^2 + 6x - 9) > \frac{4}{5x^2 + 7x - 3} \quad (\text{A.2.91})$$

$$\frac{8x + 9}{5} < \frac{5}{9x^2 - 8x - 5} \quad (\text{A.2.92})$$

$$\frac{5x - 3}{4x} > 7 \cdot (8x^2 + 4x - 5) \quad (\text{A.2.93})$$

$$\frac{8x + 7}{8x - 4} < 3 \cdot (4x^2 - 4x - 9) \quad (\text{A.2.94})$$

$$(4x - 5) \cdot (4x^2 + 7x - 3) = \frac{7}{7x^2 - 3x - 5} \quad (\text{A.2.95})$$

$$(8x^2 - 8x + 4) \cdot 7 > 7 \cdot (8x^2 + 4x + 6) \quad (\text{A.2.96})$$

$$\frac{6x^2 - 6x - 5}{7x} > 3 \cdot (8x^2 + 5x + 6) \quad (\text{A.2.97})$$

$$(9x^2 + 9x - 3) \cdot (8x + 6) = 5 \cdot (3x^2 + 9x + 7) \quad (\text{A.2.98})$$

$$(8x^2 + 3x + 9) \cdot (7x^2 + 5x + 3) = \frac{5}{9x^2 - 6x + 3} \quad (\text{A.2.99})$$

$$(6x + 4)^2 < \frac{5}{5x^2 - 7x + 5} \quad (\text{A.2.100})$$

$$3 = 9x \cdot (3x - 6) \quad (\text{A.2.101})$$

$$8x = 5x \cdot (5x + 5) \quad (\text{A.2.102})$$

$$9x + 9 = 4x \cdot (7x - 4) \quad (\text{A.2.103})$$

$$6x^2 - 9x + 5 < \frac{6x}{9x - 3} \quad (\text{A.2.104})$$

$$\frac{6}{3x - 3} = \frac{9x}{5x + 7} \quad (\text{A.2.105})$$

$$\frac{3}{9x^2 - 6x - 7} < \frac{4x}{8x + 7} \quad (\text{A.2.106})$$

$$\frac{8x}{5x + 4} < \frac{7x}{3x + 4} \quad (\text{A.2.107})$$

$$\frac{7x}{8x^2 + 3x - 6} < 6x \cdot (8x + 6) \quad (\text{A.2.108})$$

$$(4x + 5) \cdot 6 > 8x \cdot (3x - 7) \quad (\text{A.2.109})$$

$$\frac{5x + 4}{8x} > \frac{7x}{5x + 7} \quad (\text{A.2.110})$$

$$(3x + 3) \cdot (4x + 4) > 5x \cdot (6x + 7) \quad (\text{A.2.111})$$

$$(7x + 9) \cdot (8x^2 - 9x + 7) < 3x \cdot (6x + 9) \quad (\text{A.2.112})$$

$$\frac{5x^2 - 7x - 6}{9} > \frac{6x}{6x + 4} \quad (\text{A.2.113})$$

$$(6x^2 + 5x + 6) \cdot 5x < \frac{5x}{4x - 7} \quad (\text{A.2.114})$$

$$\frac{7x^2 - 6x - 6}{4x - 9} < \frac{7x}{4x + 9} \quad (\text{A.2.115})$$

$$\frac{5x^2 - 3x + 5}{7x^2 - 5x + 9} > \frac{4x}{3x + 3} \quad (\text{A.2.116})$$

$$(9x - 9)^2 = \frac{7x}{6x + 9} \quad (\text{A.2.117})$$

$$6 < \frac{3x}{5x^2 + 5x - 9} \quad (\text{A.2.118})$$

$$6x < \frac{7x}{7x^2 - 9x - 5} \quad (\text{A.2.119})$$

$$8x - 4 < 4x \cdot (3x^2 + 5x + 5) \quad (\text{A.2.120})$$

$$7x^2 - 5x + 3 < \frac{4x}{9x^2 - 8x + 7} \quad (\text{A.2.121})$$

$$9 \cdot (9x + 4) > 5x \cdot (8x^2 - 7x + 9) \quad (\text{A.2.122})$$

$$\frac{4}{6x^2 - 7x - 7} > 9x \cdot (5x^2 - 9x + 5) \quad (\text{A.2.123})$$

$$7x \cdot (3x + 6) = \frac{3x}{8x^2 - 8x - 6} \quad (\text{A.2.124})$$

$$\frac{6x}{9x^2 - 3x - 9} < \frac{6x}{4x^2 + 6x + 7} \quad (\text{A.2.125})$$

$$\frac{3x - 7}{9} = 5x \cdot (7x^2 - 9x + 9) \quad (\text{A.2.126})$$

$$\frac{4x + 7}{5x} < 3x \cdot (5x^2 + 4x - 6) \quad (\text{A.2.127})$$

$$(5x + 3) \cdot (8x + 9) > 4x \cdot (7x^2 + 5x - 6) \quad (\text{A.2.128})$$

$$(3x - 7) \cdot (7x^2 + 6x + 3) = \frac{8x}{5x^2 + 4x + 9} \quad (\text{A.2.129})$$

$$(8x^2 - 3x - 4) \cdot 4 > 5x \cdot (3x^2 + 9x + 5) \quad (\text{A.2.130})$$

$$\frac{4x^2 - 5x - 3}{9x} = \frac{3x}{7x^2 - 9x + 3} \quad (\text{A.2.131})$$

$$(7x^2 + 3x + 5) \cdot (6x - 6) > 7x \cdot (4x^2 + 9x + 3) \quad (\text{A.2.132})$$

$$(6x^2 + 4x + 3) \cdot (9x^2 + 6x - 5) > \frac{4x}{5x^2 + 7x + 5} \quad (\text{A.2.133})$$

$$(6x - 4)^2 > \frac{8x}{7x^2 + 3x - 5} \quad (\text{A.2.134})$$

$$5 = \frac{8x + 6}{7} \quad (\text{A.2.135})$$

$$3x < \frac{4x + 4}{9} \quad (\text{A.2.136})$$

$$4x - 6 = (6x + 5) \cdot 9 \quad (\text{A.2.137})$$

$$9x^2 - 3x + 4 > (7x - 5) \cdot 9 \quad (\text{A.2.138})$$

$$\frac{4}{3x - 9} > \frac{8x - 4}{4} \quad (\text{A.2.139})$$

$$\frac{5}{3x^2 + 7x - 9} > \frac{5x - 6}{4} \quad (\text{A.2.140})$$

$$\frac{9x}{7x + 5} = (3x + 6) \cdot 7 \quad (\text{A.2.141})$$

$$9x \cdot (5x^2 + 6x + 7) > (8x + 6) \cdot 6 \quad (\text{A.2.142})$$

$$\frac{3x + 5}{4} > \frac{8x + 6}{9} \quad (\text{A.2.143})$$

$$\frac{7x + 9}{9x} = \frac{9x + 7}{4} \quad (\text{A.2.144})$$

$$(4x + 6) \cdot (9x + 7) > \frac{6x - 9}{9} \quad (\text{A.2.145})$$

$$\frac{8x - 3}{4x^2 + 9x - 3} > \frac{8x - 9}{5} \quad (\text{A.2.146})$$

$$(9x^2 + 7x + 5) \cdot 3 > \frac{7x - 4}{6} \quad (\text{A.2.147})$$

$$(8x^2 - 4x + 7) \cdot 3x > \frac{5x - 6}{9} \quad (\text{A.2.148})$$

$$(3x^2 + 3x + 3) \cdot (5x + 7) = (5x + 4) \cdot 5 \quad (\text{A.2.149})$$

$$\frac{5x^2 - 5x - 3}{7x^2 - 8x - 4} > \frac{6x + 9}{7} \quad (\text{A.2.150})$$

$$(9x - 3)^2 = (6x + 7) \cdot 3 \quad (\text{A.2.151})$$

$$7 > (4x - 7) \cdot 4x \quad (\text{A.2.152})$$

$$8x < (8x + 5) \cdot 6x \quad (\text{A.2.153})$$

$$3x - 4 = \frac{7x + 3}{7x} \quad (\text{A.2.154})$$

$$5x^2 - 5x + 3 > \frac{7x - 6}{8x} \quad (\text{A.2.155})$$

$$\frac{5}{7x - 7} = \frac{4x - 4}{8x} \quad (\text{A.2.156})$$

$$\frac{7}{7x^2 - 9x + 3} > \frac{8x + 3}{7x} \quad (\text{A.2.157})$$

$$\frac{3x}{5x + 5} = (3x - 3) \cdot 6x \quad (\text{A.2.158})$$

$$\frac{7x}{9x^2 - 8x - 7} = \frac{4x + 5}{9x} \quad (\text{A.2.159})$$

$$(8x + 5) \cdot 7 = (8x - 3) \cdot 4x \quad (\text{A.2.160})$$

$$\frac{3x - 6}{3x} < \frac{9x - 5}{6x} \quad (\text{A.2.161})$$

$$\frac{4x + 5}{5x + 6} < (6x + 9) \cdot 9x \quad (\text{A.2.162})$$

$$\frac{7x - 5}{8x^2 + 9x - 3} < (7x + 5) \cdot 5x \quad (\text{A.2.163})$$

$$\frac{9x^2 - 3x + 7}{4} > \frac{4x + 4}{9x} \quad (\text{A.2.164})$$

$$(8x^2 + 6x - 6) \cdot 3x > \frac{4x + 6}{4x} \quad (\text{A.2.165})$$

$$\frac{6x^2 + 9x - 6}{8x + 9} > (7x + 4) \cdot 6x \quad (\text{A.2.166})$$

$$\frac{7x^2 + 5x + 5}{4x^2 - 3x - 3} = (7x + 4) \cdot 8x \quad (\text{A.2.167})$$

$$(7x - 3)^2 > (5x - 5) \cdot 8x \quad (\text{A.2.168})$$

$$7 = (3x + 6) \cdot (8x - 9) \quad (\text{A.2.169})$$

$$5x = (9x + 3) \cdot (4x + 6) \quad (\text{A.2.170})$$

$$3x - 5 = \frac{7x + 4}{9x + 6} \quad (\text{A.2.171})$$

$$3x^2 - 7x - 4 > \frac{5x + 6}{3x + 7} \quad (\text{A.2.172})$$

$$7 \cdot (4x + 6) = (7x + 7) \cdot (9x + 6) \quad (\text{A.2.173})$$

$$4 \cdot (3x^2 - 6x - 7) < (5x - 3) \cdot (9x - 7) \quad (\text{A.2.174})$$

$$6x \cdot (3x - 9) = (7x + 5) \cdot (5x - 6) \quad (\text{A.2.175})$$

$$9x \cdot (3x^2 - 7x - 5) < \frac{8x - 4}{3x - 3} \quad (\text{A.2.176})$$

$$(9x + 5) \cdot 7 > \frac{6x + 4}{7x - 7} \quad (\text{A.2.177})$$

$$(5x + 4) \cdot 4x > (8x - 7) \cdot (3x - 9) \quad (\text{A.2.178})$$

$$\frac{4x - 7}{8x - 4} > \frac{6x - 9}{7x - 7} \quad (\text{A.2.179})$$

$$\frac{9x + 7}{6x^2 + 7x - 4} < \frac{8x - 7}{8x + 6} \quad (\text{A.2.180})$$

$$(6x^2 - 3x + 7) \cdot 6 < \frac{4x - 9}{6x - 5} \quad (\text{A.2.181})$$

$$\frac{8x^2 + 7x - 5}{8x} < (5x + 7) \cdot (8x - 9) \quad (\text{A.2.182})$$

$$\frac{8x^2 - 9x - 3}{9x + 6} < (9x - 3) \cdot (9x + 4) \quad (\text{A.2.183})$$

$$\frac{9x^2 - 4x - 5}{3x^2 + 6x - 4} < (8x + 4) \cdot (7x - 7) \quad (\text{A.2.184})$$

$$(5x - 4)^2 = (4x - 4) \cdot (9x + 7) \quad (\text{A.2.185})$$

$$3 = \frac{8x + 9}{3x^2 - 5x + 5} \quad (\text{A.2.186})$$

$$9x = (6x + 3) \cdot (8x^2 + 3x + 3) \quad (\text{A.2.187})$$

$$4x + 9 = (3x + 6) \cdot (4x^2 + 4x + 7) \quad (\text{A.2.188})$$

$$8x^2 - 9x + 4 = \frac{7x + 4}{5x^2 - 5x - 9} \quad (\text{A.2.189})$$

$$6 \cdot (5x - 4) = (4x - 5) \cdot (6x^2 - 5x + 6) \quad (\text{A.2.190})$$

$$6 \cdot (5x^2 - 8x + 6) = (6x + 4) \cdot (5x^2 - 4x + 4) \quad (\text{A.2.191})$$

$$5x \cdot (3x + 3) = (9x - 6) \cdot (5x^2 - 6x - 7) \quad (\text{A.2.192})$$

$$\frac{5x}{4x^2 + 8x + 7} = \frac{5x + 7}{7x^2 - 8x + 9} \quad (\text{A.2.193})$$

$$\frac{4x - 5}{9} = (4x + 3) \cdot (7x^2 + 7x - 4) \quad (\text{A.2.194})$$

$$\frac{3x+7}{5x} > (5x-7) \cdot (6x^2+6x-3) \quad (\text{A.2.195})$$

$$(3x-4) \cdot (5x-9) > (3x+3) \cdot (9x^2+3x+4) \quad (\text{A.2.196})$$

$$(9x+6) \cdot (3x^2-7x+5) = \frac{9x+5}{7x^2+5x-3} \quad (\text{A.2.197})$$

$$(3x^2-3x+5) \cdot 7 = (4x+5) \cdot (8x^2-9x-7) \quad (\text{A.2.198})$$

$$\frac{5x^2-7x-6}{9x} = \frac{5x-7}{4x^2-3x-5} \quad (\text{A.2.199})$$

$$\frac{5x^2+9x-5}{4x+4} < (5x-4) \cdot (9x^2+5x+6) \quad (\text{A.2.200})$$

$$\frac{6x^2-8x-7}{3x^2+5x+7} = (5x+9) \cdot (4x^2-6x+6) \quad (\text{A.2.201})$$

$$(4x+6)^2 = (7x+5) \cdot (9x^2-4x-4) \quad (\text{A.2.202})$$

$$6 > (3x^2-4x+4) \cdot 3 \quad (\text{A.2.203})$$

$$5x = (5x^2-8x-4) \cdot 5 \quad (\text{A.2.204})$$

$$3x-7 > (7x^2-6x+4) \cdot 6 \quad (\text{A.2.205})$$

$$4x^2+8x+9 < (3x^2-6x+6) \cdot 9 \quad (\text{A.2.206})$$

$$6 \cdot (7x+9) = \frac{3x^2-4x-9}{7} \quad (\text{A.2.207})$$

$$9 \cdot (3x^2-3x-6) < \frac{3x^2-8x-5}{7} \quad (\text{A.2.208})$$

$$\frac{7x}{9x+6} < (9x^2+3x+6) \cdot 3 \quad (\text{A.2.209})$$

$$5x \cdot (4x^2-3x-9) < \frac{5x^2+9x+7}{9} \quad (\text{A.2.210})$$

$$(6x-4) \cdot 5 = (8x^2+9x+6) \cdot 7 \quad (\text{A.2.211})$$

$$(4x-6) \cdot 3x < (4x^2+8x+6) \cdot 5 \quad (\text{A.2.212})$$

$$(4x-6) \cdot (7x-4) > (6x^2+7x-4) \cdot 7 \quad (\text{A.2.213})$$

$$(7x+7) \cdot (4x^2+9x+5) < \frac{7x^2+5x-6}{4} \quad (\text{A.2.214})$$

$$(5x^2-3x-9) \cdot 9 < (3x^2-3x-5) \cdot 4 \quad (\text{A.2.215})$$

$$\frac{6x^2+5x+6}{8x} = \frac{4x^2+5x+4}{6} \quad (\text{A.2.216})$$

$$(6x^2-4x-5) \cdot (9x-5) < \frac{4x^2-8x-6}{5} \quad (\text{A.2.217})$$

$$(3x^2-7x+6) \cdot (8x^2-4x+7) > (9x^2-3x-7) \cdot 5 \quad (\text{A.2.218})$$

$$(9x-7)^2 < (5x^2+9x+5) \cdot 7 \quad (\text{A.2.219})$$

$$3 = (6x^2-7x+3) \cdot 4x \quad (\text{A.2.220})$$

$$4x > (4x^2-6x-6) \cdot 5x \quad (\text{A.2.221})$$

$$3x-6 = \frac{6x^2-5x-4}{4x} \quad (\text{A.2.222})$$

$$7x^2-7x+4 < \frac{5x^2+9x-3}{4x} \quad (\text{A.2.223})$$

$$9 \cdot (6x-5) = (7x^2-4x+3) \cdot 8x \quad (\text{A.2.224})$$

$$3 \cdot (5x^2+6x-7) < (4x^2-5x+3) \cdot 6x \quad (\text{A.2.225})$$

$$5x \cdot (6x-9) < (8x^2+7x-3) \cdot 5x \quad (\text{A.2.226})$$

$$\frac{3x}{6x^2-7x-5} < (8x^2-4x+5) \cdot 5x \quad (\text{A.2.227})$$

$$\frac{4x+9}{3} = (8x^2+3x+7) \cdot 5x \quad (\text{A.2.228})$$

$$\frac{7x-7}{3x} = \frac{5x^2+3x+9}{7x} \quad (\text{A.2.229})$$

$$\frac{3x-3}{8x+3} < (7x^2-6x+5) \cdot 7x \quad (\text{A.2.230})$$

$$\frac{4x-6}{7x^2+7x-4} < (8x^2+8x+3) \cdot 4x \quad (\text{A.2.231})$$

$$(8x^2+9x-9) \cdot 3 < (4x^2-5x+5) \cdot 6x \quad (\text{A.2.232})$$

$$(8x^2+8x-6) \cdot 6x < (3x^2+4x-5) \cdot 6x \quad (\text{A.2.233})$$

$$(5x^2+8x-7) \cdot (6x-9) < (4x^2-6x+3) \cdot 7x \quad (\text{A.2.234})$$

$$(9x^2+8x+6) \cdot (8x^2+7x+6) < (4x^2+4x+9) \cdot 3x \quad (\text{A.2.235})$$

$$(3x+4)^2 = \frac{6x^2+3x-7}{4x} \quad (\text{A.2.236})$$

$$9 = \frac{3x^2-4x-6}{3x+7} \quad (\text{A.2.237})$$

$$6x < \frac{9x^2+3x+3}{9x-7} \quad (\text{A.2.238})$$

$$3x+4 < \frac{5x^2-4x-3}{7x+7} \quad (\text{A.2.239})$$

$$4x^2-3x+4 > (5x^2+7x-4) \cdot (3x+6) \quad (\text{A.2.240})$$

$$9 \cdot (3x+6) < (8x^2-8x+9) \cdot (9x-5) \quad (\text{A.2.241})$$

$$\frac{7}{8x^2-3x+5} > (5x^2-3x+4) \cdot (4x-6) \quad (\text{A.2.242})$$

$$\frac{7x}{3x-7} < \frac{5x^2-8x-7}{7x-5} \quad (\text{A.2.243})$$

$$\frac{5x}{6x^2+3x-4} > (7x^2-5x-7) \cdot (8x+3) \quad (\text{A.2.244})$$

$$\frac{6x-5}{4} < \frac{5x^2-6x+9}{3x+7} \quad (\text{A.2.245})$$

$$\frac{8x+3}{6x} < (7x^2 - 6x - 3) \cdot (9x - 7) \quad (\text{A.2.246})$$

$$\frac{3x+5}{7x+6} < \frac{8x^2 - 3x - 4}{5x+9} \quad (\text{A.2.247})$$

$$\frac{7x+3}{9x^2 - 3x - 7} < (5x^2 - 9x - 3) \cdot (7x + 9) \quad (\text{A.2.248})$$

$$\frac{3x^2 + 5x + 4}{6} > (4x^2 - 3x + 6) \cdot (3x + 7) \quad (\text{A.2.249})$$

$$(5x^2 - 7x + 7) \cdot 3x = \frac{9x^2 - 7x - 6}{5x+9} \quad (\text{A.2.250})$$

$$\frac{5x^2 + 9x + 6}{8x+3} < \frac{7x^2 - 8x - 4}{9x+5} \quad (\text{A.2.251})$$

$$\frac{7x^2 + 4x + 5}{9x^2 - 7x + 6} > (7x^2 + 3x - 5) \cdot (9x + 7) \quad (\text{A.2.252})$$

$$(6x+5)^2 > \frac{8x^2 + 3x - 5}{7x+4} \quad (\text{A.2.253})$$

$$7 < \frac{5x^2 + 4x - 6}{6x^2 - 4x + 4} \quad (\text{A.2.254})$$

$$8x = (7x^2 + 5x + 9) \cdot (3x^2 - 4x - 9) \quad (\text{A.2.255})$$

$$6x - 9 = (4x^2 - 8x + 4) \cdot (4x^2 + 6x - 4) \quad (\text{A.2.256})$$

$$3x^2 + 7x + 5 < (3x^2 - 4x - 3) \cdot (9x^2 - 9x - 5) \quad (\text{A.2.257})$$

$$6 \cdot (5x - 4) = (7x^2 - 6x - 7) \cdot (3x^2 - 4x - 6) \quad (\text{A.2.258})$$

$$3 \cdot (9x^2 - 4x + 9) > (9x^2 + 5x - 3) \cdot (8x^2 - 8x + 6) \quad (\text{A.2.259})$$

$$\frac{8x}{5x-5} = \frac{5x^2 - 6x + 4}{9x^2 + 5x - 5} \quad (\text{A.2.260})$$

$$7x \cdot (5x^2 - 8x + 6) > \frac{4x^2 - 9x - 3}{6x^2 - 3x - 5} \quad (\text{A.2.261})$$

$$\frac{8x+3}{7} > (3x^2 + 6x + 4) \cdot (7x^2 + 9x + 7) \quad (\text{A.2.262})$$

$$(7x - 6) \cdot 8x > (4x^2 + 9x + 7) \cdot (8x^2 + 4x + 7) \quad (\text{A.2.263})$$

$$(4x - 3) \cdot (7x + 9) < (8x^2 + 9x + 4) \cdot (5x^2 + 5x - 9) \quad (\text{A.2.264})$$

$$(4x - 3) \cdot (7x^2 - 8x - 9) < \frac{4x^2 - 3x + 7}{8x^2 + 3x - 4} \quad (\text{A.2.265})$$

$$\frac{4x^2 - 5x + 7}{7} > (8x^2 + 8x + 7) \cdot (4x^2 + 3x + 9) \quad (\text{A.2.266})$$

$$(9x^2 - 6x - 7) \cdot 3x > (3x^2 - 3x - 4) \cdot (9x^2 - 8x + 4) \quad (\text{A.2.267})$$

$$(4x^2 + 4x - 6) \cdot (5x - 9) = \frac{4x^2 + 5x - 5}{3x^2 - 8x - 5} \quad (\text{A.2.268})$$

$$\frac{5x^2 - 9x - 5}{3x^2 + 7x + 7} < (5x^2 - 4x - 3) \cdot (4x^2 + 6x + 4) \quad (\text{A.2.269})$$

$$(3x+9)^2 > \frac{6x^2 - 8x - 5}{3x^2 + 9x + 6} \quad (\text{A.2.270})$$

$$6 = (7x - 4)^2 \quad (\text{A.2.271})$$

$$6x = (3x + 7)^2 \quad (\text{A.2.272})$$

$$8x - 7 < (9x - 4)^2 \quad (\text{A.2.273})$$

$$7x^2 + 8x + 9 > (4x - 7)^2 \quad (\text{A.2.274})$$

$$\frac{3}{4x-9} = (8x - 4)^2 \quad (\text{A.2.275})$$

$$7 \cdot (8x^2 - 5x + 4) < (4x - 5)^2 \quad (\text{A.2.276})$$

$$3x \cdot (4x - 7) < (9x - 7)^2 \quad (\text{A.2.277})$$

$$4x \cdot (7x^2 + 4x - 9) > (8x - 5)^2 \quad (\text{A.2.278})$$

$$\frac{4x-4}{7} > (4x+9)^2 \quad (\text{A.2.279})$$

$$(3x+7) \cdot 9x > (6x+7)^2 \quad (\text{A.2.280})$$

$$\frac{7x+3}{3x-6} < (3x-5)^2 \quad (\text{A.2.281})$$

$$\frac{9x-7}{6x^2-9x+6} > (6x+3)^2 \quad (\text{A.2.282})$$

$$\frac{3x^2-8x+3}{6} > (3x+3)^2 \quad (\text{A.2.283})$$

$$\frac{7x^2-8x+9}{9x} > (4x-6)^2 \quad (\text{A.2.284})$$

$$(7x^2 - 5x - 4) \cdot (8x + 6) < (6x - 5)^2 \quad (\text{A.2.285})$$

$$\frac{5x^2 - 8x - 4}{9x^2 - 6x - 7} = (4x - 5)^2 \quad (\text{A.2.286})$$

$$(6x+6)^2 = (5x+3)^2 \quad (\text{A.2.287})$$

$$7 - \frac{7}{6x+3} > 6 \quad (\text{A.2.288})$$

$$9 + 9 \cdot (4x^2 + 9x + 3) > 4 \quad (\text{A.2.289})$$

$$5 - \frac{9x}{8x+9} > 5 \quad (\text{A.2.290})$$

$$4 - \frac{7x}{6x^2 - 4x - 5} < 7 \quad (\text{A.2.291})$$

$$4 - \frac{6x-5}{9} = 6 \quad (\text{A.2.292})$$

$$9 - (7x - 6) \cdot 9x > 4 \quad (\text{A.2.293})$$

$$9 + (5x - 9) \cdot (3x + 4) = 7 \quad (\text{A.2.294})$$

$$4 + \frac{3x-6}{5x^2 + 7x + 6} = 4 \quad (\text{A.2.295})$$

$5 + \frac{4x^2 + 6x - 6}{3} > 5$	(A.2.296)	$\frac{4x - 5}{7x + 5} + \frac{4x + 6}{6x} < 9$	(A.2.323)
$5 + (9x^2 - 4x + 7) \cdot 7x > 3$	(A.2.297)	$(6x + 5) \cdot (3x - 9) + \frac{7x - 6}{9x - 3} > 6$	(A.2.324)
$4 + (6x^2 - 4x - 7) \cdot (8x + 4) = 3$	(A.2.298)	$(3x - 3) \cdot (4x - 9) - (8x + 9)^2 = 4$	(A.2.325)
$7 - \frac{9x^2 - 3x - 9}{4x^2 + 5x + 4} > 5$	(A.2.299)	$(3x - 4)^2 - \frac{6x}{4x - 7} < 9$	(A.2.326)
$7 - (8x + 4)^2 > 6$	(A.2.300)	$(8x + 7)^2 - (6x - 7) \cdot 4x < 3$	(A.2.327)
$\frac{7}{3x - 5} - 5 = 6$	(A.2.301)	$(7x + 9)^2 - \frac{8x + 9}{9x + 6} < 9$	(A.2.328)
$\frac{3}{3x^2 + 5x - 9} - 9 = 5$	(A.2.302)	$(8x + 6)^2 - (3x - 6)^2 > 5$	(A.2.329)
$6x \cdot (7x + 7) + 4 = 6$	(A.2.303)	$3 = 7 + 9 \cdot (8x - 7)$	(A.2.330)
$\frac{5x}{6x^2 + 5x - 7} + 5 < 4$	(A.2.304)	$9 < 4 + 6 \cdot (3x^2 + 8x + 7)$	(A.2.331)
$(5x - 6) \cdot 3 - 5 < 7$	(A.2.305)	$3 = 5 + 9x \cdot (6x + 7)$	(A.2.332)
$(3x - 6) \cdot 7x + 5 > 9$	(A.2.306)	$4 = 9 - 8x \cdot (6x^2 - 7x + 6)$	(A.2.333)
$\frac{7x + 7}{3x - 4} + 7 = 7$	(A.2.307)	$4 > 7 + \frac{8x + 4}{4}$	(A.2.334)
$(5x + 6) \cdot (7x^2 - 6x - 5) + 3 > 5$	(A.2.308)	$6 < 4 - (9x - 6) \cdot 5x$	(A.2.335)
$(5x^2 + 8x + 6) \cdot 6 + 4 < 9$	(A.2.309)	$5 = 4 + (8x - 6) \cdot (8x - 4)$	(A.2.336)
$\frac{4x^2 + 9x + 7}{7x} - 7 < 4$	(A.2.310)	$3 = 4 + (5x - 5) \cdot (7x^2 - 6x - 4)$	(A.2.337)
$(4x^2 - 9x + 5) \cdot (3x + 7) + 4 = 6$	(A.2.311)	$6 < 5 - \frac{5x^2 + 4x + 7}{5}$	(A.2.338)
$\frac{9x^2 + 8x + 7}{5x^2 + 5x + 4} + 6 < 6$	(A.2.312)	$9 = 5 - \frac{6x^2 + 3x + 6}{4x}$	(A.2.339)
$(5x + 7)^2 + 9 = 3$	(A.2.313)	$7 > 7 - (3x^2 - 8x + 9) \cdot (5x - 6)$	(A.2.340)
$\frac{7x}{3x - 4} + 3x \cdot (7x + 6) = 9$	(A.2.314)	$6 < 9 - (3x^2 - 7x - 7) \cdot (8x^2 - 4x + 5)$	(A.2.341)
$7x \cdot (3x - 7) + (7x - 5) \cdot 6x = 5$	(A.2.315)	$6 < 6 - (3x + 5)^2$	(A.2.342)
$4x \cdot (6x - 9) + (7x + 5) \cdot (6x + 4) < 3$	(A.2.316)	$9 = \frac{9}{6x - 9} + 7$	(A.2.343)
$7x \cdot (5x - 9) + (5x + 5)^2 = 7$	(A.2.317)	$7 > \frac{6}{8x^2 - 7x + 9} + 4$	(A.2.344)
$\frac{6x + 7}{8x} + 5x \cdot (6x + 7) = 9$	(A.2.318)	$7 < \frac{5x}{5x - 9} + 6$	(A.2.345)
$\frac{8x + 9}{8x} - \frac{4x + 6}{4x} > 4$	(A.2.319)	$5 < \frac{9x}{8x^2 - 3x - 4} + 3$	(A.2.346)
$(7x + 9) \cdot 4x + (7x + 6) \cdot (9x - 5) > 4$	(A.2.320)	$4 < (8x + 9) \cdot 9 + 3$	(A.2.347)
$(7x - 6) \cdot 7x + (6x - 5)^2 < 7$	(A.2.321)	$6 < (9x + 9) \cdot 8x + 5$	(A.2.348)
$(8x - 3) \cdot (6x - 7) + \frac{5x}{5x - 6} > 4$	(A.2.322)	$6 < \frac{9x + 3}{6x + 7} + 9$	(A.2.349)
		$3 < \frac{6x + 9}{8x^2 - 5x - 6} - 7$	(A.2.350)

$$7 < \frac{4x^2 + 3x + 3}{7} + 3 \quad (\text{A.2.351})$$

$$4 = (8x^2 - 7x - 7) \cdot 8x - 3 \quad (\text{A.2.352})$$

$$7 > \frac{6x^2 + 8x - 4}{4x - 9} - 6 \quad (\text{A.2.353})$$

$$7 = \frac{8x^2 - 7x + 7}{5x^2 + 5x - 6} + 9 \quad (\text{A.2.354})$$

$$5 = (7x - 3)^2 - 9 \quad (\text{A.2.355})$$

$$9 > \frac{4x}{3x + 3} - 8x \cdot (3x + 3) \quad (\text{A.2.356})$$

$$6 = 7x \cdot (5x + 7) - (9x - 3) \cdot 8x \quad (\text{A.2.357})$$

$$9 > \frac{3x}{7x - 5} + \frac{6x + 7}{6x + 6} \quad (\text{A.2.358})$$

$$6 > 4x \cdot (5x + 7) + (9x - 5)^2 \quad (\text{A.2.359})$$

$$4 = \frac{4x - 9}{9x} + 7x \cdot (3x + 4) \quad (\text{A.2.360})$$

$$9 > \frac{5x - 9}{4x} + \frac{7x - 6}{8x} \quad (\text{A.2.361})$$

$$3 < \frac{7x + 6}{4x} - (5x - 9) \cdot (7x - 4) \quad (\text{A.2.362})$$

$$4 = (7x - 6) \cdot 3x - (6x - 4)^2 \quad (\text{A.2.363})$$

$$7 = (9x + 4) \cdot (8x + 5) + \frac{6x}{9x + 6} \quad (\text{A.2.364})$$

$$4 = \frac{3x - 9}{4x - 9} + (5x + 6) \cdot 8x \quad (\text{A.2.365})$$

$$9 > \frac{4x - 6}{3x + 6} - \frac{8x + 7}{5x - 4} \quad (\text{A.2.366})$$

$$6 = (9x - 3) \cdot (9x - 7) - (4x + 6)^2 \quad (\text{A.2.367})$$

$$6 < (5x + 6)^2 - \frac{4x}{7x + 9} \quad (\text{A.2.368})$$

$$3 > (9x + 6)^2 + \frac{5x - 9}{9x} \quad (\text{A.2.369})$$

$$9 < (4x - 7)^2 + \frac{4x - 3}{8x - 9} \quad (\text{A.2.370})$$

$$9 = (5x - 7)^2 - (3x + 4)^2 \quad (\text{A.2.371})$$

$$|x| = 6 \quad (\text{A.2.372})$$

$$\sqrt{8x} < 9 \quad (\text{A.2.373})$$

$$\sqrt{3x + 4} = 9 \quad (\text{A.2.374})$$

$$|5x^2 + 9x - 4| > 7 \quad (\text{A.2.375})$$

$$\sqrt{\frac{4}{7x + 6}} > 3 \quad (\text{A.2.376})$$

$$\left| \frac{9}{3x^2 - 5x - 3} \right| = 4 \quad (\text{A.2.377})$$

$$\sqrt{7x \cdot (9x - 9)} < 3 \quad (\text{A.2.378})$$

$$|5x \cdot (9x^2 + 3x - 6)| < 4 \quad (\text{A.2.379})$$

$$\sqrt{\frac{9x - 6}{6}} < 5 \quad (\text{A.2.380})$$

$$|(7x + 4) \cdot 4x| > 5 \quad (\text{A.2.381})$$

$$\sqrt{(4x - 9) \cdot (9x - 7)} = 5 \quad (\text{A.2.382})$$

$$\sqrt{(7x - 9) \cdot (5x^2 + 3x - 7)} < 7 \quad (\text{A.2.383})$$

$$\sqrt{(5x^2 - 3x + 3) \cdot 6} = 3 \quad (\text{A.2.384})$$

$$\sqrt{\frac{5x^2 + 7x - 4}{8x}} = 4 \quad (\text{A.2.385})$$

$$\sqrt{(5x^2 - 7x + 3) \cdot (5x - 9)} > 4 \quad (\text{A.2.386})$$

$$\sqrt{(6x^2 - 3x + 4) \cdot (5x^2 + 9x - 5)} < 3 \quad (\text{A.2.387})$$

$$|5 + 6x \cdot (8x + 4)| > 6 \quad (\text{A.2.388})$$

$$\sqrt{7 + (9x - 6) \cdot 5x} = 4 \quad (\text{A.2.389})$$

$$\sqrt{4 - (5x - 6) \cdot (5x - 4)} < 9 \quad (\text{A.2.390})$$

$$\sqrt{7x \cdot (6x + 7) - 3} < 9 \quad (\text{A.2.391})$$

$$|(7x + 3) \cdot 5x - 9| > 5 \quad (\text{A.2.392})$$

$$\sqrt{\frac{9x - 6}{4x - 6} + 6} < 5 \quad (\text{A.2.393})$$

$$\sqrt{\frac{9x}{3x - 7} + \frac{8x}{5x - 9}} < 6 \quad (\text{A.2.394})$$

$$\sqrt{4x \cdot (8x + 3) - (8x + 4) \cdot 4x} < 7 \quad (\text{A.2.395})$$

$$\sqrt{4x \cdot (9x + 6) - \frac{3x - 9}{7x - 9}} = 4 \quad (\text{A.2.396})$$

$$\left| \frac{8x - 9}{5x} - \frac{7x}{3x + 4} \right| > 4 \quad (\text{A.2.397})$$

$$\sqrt{(5x - 7) \cdot 5x + (8x + 4) \cdot 6x} = 7 \quad (\text{A.2.398})$$

$$\left| \frac{5x + 3}{6x} - \frac{5x - 3}{4x + 9} \right| < 4 \quad (\text{A.2.399})$$

$$|(6x + 4) \cdot (9x - 6) + 9x \cdot (6x - 5)| > 5 \quad (\text{A.2.400})$$

$$\left| \frac{7x - 4}{5x + 4} - \frac{3x - 9}{7x} \right| = 9 \quad (\text{A.2.401})$$

$$\left| \frac{7x + 9}{5x - 4} - (9x - 7) \cdot (9x - 5) \right| = 4 \quad (\text{A.2.402})$$

$$4 > |x| \quad (\text{A.2.403})$$

$6 = \sqrt{3x}$	(A.2.404)	$9 < \left \frac{8x-4}{6x} - \frac{9x+7}{3x} \right $	(A.2.429)
$4 > \sqrt{7x+3}$	(A.2.405)	$4 > (6x-6) \cdot 3x - (5x-6) \cdot (7x-9) $	(A.2.430)
$6 < \sqrt{6x^2+3x-7}$	(A.2.406)	$9 = (9x+4) \cdot (3x+9) - 7x \cdot (7x-5) $	(A.2.431)
$3 = 7 \cdot (4x-9) $	(A.2.407)	$3 < \left \frac{9x+3}{5x-4} - \frac{8x-5}{7x} \right $	(A.2.432)
$5 < \sqrt{\frac{7}{5x^2+3x+4}}$	(A.2.408)	$9 > \left \frac{4x-7}{3x-3} + (8x-4) \cdot (6x+9) \right $	(A.2.433)
$7 < \sqrt{\frac{7x}{9x-4}}$	(A.2.409)	$ 8x = 5x $	(A.2.434)
$6 < 7x \cdot (9x^2+6x+3) $	(A.2.410)	$\sqrt{9x-4} < \sqrt{5x}$	(A.2.435)
$7 < \sqrt{\frac{5x+7}{4}}$	(A.2.411)	$ 6x < 3x+7 $	(A.2.436)
$9 = \sqrt{\frac{7x+4}{3x}}$	(A.2.412)	$\sqrt{5x+9} > \sqrt{8x+9}$	(A.2.437)
$5 > \sqrt{(5x+3) \cdot (6x-9)}$	(A.2.413)	$ 3x \cdot (5x-5) - \frac{8x}{6x-9} < 3$	(A.2.438)
$3 < (3x-5) \cdot (7x^2+9x-9) $	(A.2.414)	$ 5x \cdot (9x-4) - (6x-4) \cdot 3x < 9$	(A.2.439)
$3 = (6x^2-8x-4) \cdot 3 $	(A.2.415)	$\left \frac{7x}{9x-5} \right - \frac{8x+4}{8x-5} = 4$	(A.2.440)
$9 = \sqrt{(4x^2-5x+6) \cdot 6x}$	(A.2.416)	$\left \frac{7x}{4x+5} \right - 3 = 4$	(A.2.441)
$6 < \sqrt{\frac{3x^2-8x-4}{4x+9}}$	(A.2.417)	$\left \frac{3x-9}{4x} \right - \frac{6x}{9x-4} > 9$	(A.2.442)
$4 < \left \frac{9x^2+4x+5}{3x^2-5x-9} \right $	(A.2.418)	$ (7x+5) \cdot 4x + (9x-4) \cdot 3x = 9$	(A.2.443)
$3 < \left 5 + \frac{4x}{8x-6} \right $	(A.2.419)	$\left \frac{8x+6}{3x} \right - (7x+5) \cdot (8x+6) > 3$	(A.2.444)
$9 = 4 - (3x+5) \cdot 9x $	(A.2.420)	$\sqrt{(7x-9) \cdot 5x} + 3 = 4$	(A.2.445)
$4 < \sqrt{9 - \frac{6x+5}{5x-5}}$	(A.2.421)	$\left \frac{8x-4}{9x-4} \right + \frac{6x}{3x-3} > 4$	(A.2.446)
$7 > \sqrt{\frac{6x}{5x+7} - 4}$	(A.2.422)	$\sqrt{\frac{4x+4}{6x-9}} + (8x+3) \cdot 5x < 5$	(A.2.447)
$9 = \left \frac{3x-6}{5x} + 7 \right $	(A.2.423)	$\sqrt{(9x-5) \cdot (3x+7)} - (3x+9) \cdot (8x+3) = 7$	(A.2.448)
$7 < \left \frac{9x-3}{7x-9} - 4 \right $	(A.2.424)	$\sqrt{(4x-6) \cdot (5x+3)} - 6 > 3$	(A.2.449)
$3 = \sqrt{6x \cdot (7x+5) + \frac{3x}{6x-7}}$	(A.2.425)	$6 = 8x \cdot (7x-9) + 4x \cdot (6x-6)$	(A.2.450)
$6 > \left 8x \cdot (9x-3) + \frac{5x+7}{4x} \right $	(A.2.426)	$7 < 5x \cdot (4x-7) - (8x-5) \cdot 4x$	(A.2.451)
$3 < \left 8x \cdot (9x-7) - \frac{9x-5}{5x+4} \right $	(A.2.427)	$9 > 9x \cdot (3x-4) + (9x-7) \cdot (4x+9)$	(A.2.452)
$5 = \left (3x+5) \cdot 5x - \frac{4x}{6x-6} \right $	(A.2.428)	$4 < \sqrt{\frac{9x}{3x+7}} + 7$	(A.2.453)
		$7 = (9x-9) \cdot 5x + \frac{5x}{7x-4}$	(A.2.454)
		$7 = \left \frac{7x-3}{3x} \right + (8x-6) \cdot 7x$	(A.2.455)

$3 > \sqrt{(3x+5) \cdot 5x + (7x+7) \cdot (6x-4)}$	(A.2.456)	$\frac{ 3x^2 - 4x + 9 }{9x} < 3$	(A.2.473)
$4 = \sqrt{\frac{5x+3}{6x}} + 4$	(A.2.457)	$\sqrt{7x^2 - 6x + 3} \cdot (7x+3) > 3$	(A.2.474)
$7 = \left \frac{7x-3}{8x+5} \right + 4x \cdot (7x-4)$	(A.2.458)	$\sqrt{7x^2 - 8x - 3} \cdot (8x^2 - 7x - 9) > 6$	(A.2.475)
$4 > \sqrt{\frac{8x-5}{4x-6} + \frac{7x-3}{7x}}$	(A.2.459)	$4 = \sqrt{3} \cdot 7x$	(A.2.476)
$9 > \sqrt{(9x+4) \cdot (3x+5)} - \frac{8x+7}{7x+5}$	(A.2.460)	$3 < \sqrt{5} \cdot (7x-4)$	(A.2.477)
$3 > \sqrt{(4x+5) \cdot (7x-3)} + 5$	(A.2.461)	$7 > \frac{ 5x }{3}$	(A.2.478)
$\frac{\sqrt{3}}{9x} > 3$	(A.2.462)	$7 = 8x \cdot 6x$	(A.2.479)
$\sqrt{3} \cdot (6x+9) < 4$	(A.2.463)	$4 = \sqrt{5x} \cdot (7x-6)$	(A.2.480)
$\frac{\sqrt{7x}}{5} < 4$	(A.2.464)	$7 > 9x \cdot (4x^2 + 5x + 5)$	(A.2.481)
$ 7x \cdot 4x > 9$	(A.2.465)	$7 < \frac{\sqrt{4x-7}}{3}$	(A.2.482)
$\frac{ 8x }{6x-7} = 5$	(A.2.466)	$6 < \frac{\sqrt{8x-9}}{5x}$	(A.2.483)
$\frac{\sqrt{9x}}{4x^2 + 5x + 3} < 5$	(A.2.467)	$6 > \frac{\sqrt{7x-6}}{4x-9}$	(A.2.484)
$ 3x-5 \cdot 7 < 9$	(A.2.468)	$7 = \frac{\sqrt{9x-3}}{6x^2 + 4x - 9}$	(A.2.485)
$\sqrt{4x+5} \cdot 9x > 5$	(A.2.469)	$9 < \frac{ 8x^2 + 8x - 7 }{4}$	(A.2.486)
$ 4x-3 \cdot (6x-4) < 7$	(A.2.470)	$9 = \frac{\sqrt{9x^2 + 7x + 4}}{5x}$	(A.2.487)
$\frac{ 6x-9 }{9x^2 + 5x - 6} < 4$	(A.2.471)	$4 < 8x^2 - 3x - 3 \cdot (7x-4)$	(A.2.488)
$\frac{ 9x^2 + 4x - 7 }{5} = 5$	(A.2.472)	$4 = \frac{\sqrt{8x^2 - 7x - 9}}{5x^2 + 5x + 7}$	(A.2.489)

A.3 Les 1001 phrases

1. Véronique et Marie voulaient manger Guy.
2. Marie et Jean redémoliront Guy.
3. Mélanie et Jean pourront consoler Véronique.
4. Mélanie et Véronique redémoliront Guy.
5. Jean et Mélanie verront Romain.
6. Romain et Nicolas démoliront Romain.
7. Nicolas et Véronique redémoliront Guy.
8. Véronique et Mélanie machonnaient Nicolas.
9. Marie et Nicolas voulaient manger Mélanie.
10. Nous démolirons Mélanie et Nicolas.
11. Romain et Nicolas redémoliront Mélanie.
12. Véronique et Guy démoliront Nicolas.
13. Jean et Romain pourront consoler Nicolas.
14. Marie et Mélanie démoliront Jean.
15. Marie et Nicolas sauront dévorer Marie.
16. Romain et Véronique redécouvriront Jean.
17. Véronique et Marie pourront consoler Jean.
18. Romain et Guy verront Nicolas.
19. Jean et Romain pourraient dévorer Guy.
20. Guy et Mélanie redécouvriront Marie.
21. Mélanie et Nicolas pourront consoler Mélanie.
22. Jean et Nicolas sauront dévorer Mélanie.
23. Jean et Nicolas démoliront Véronique.
24. Mélanie et Romain sauront dévorer Marie.
25. Romain démolira Marie.
26. Romain redémolira Marie.
27. Véronique verra Guy.
28. Jean pourra consoler Mélanie.
29. Jean vend Mélanie.

30. Véronique veut manger Marie.
31. Marie pourra consoler Nicolas.
32. Guy veut dévorer Jean.
33. Marie vend Mélanie.
34. Véronique redécouvrira Jean.
35. Jean veut dévorer Mélanie.
36. Mélanie machonnait Nicolas.
37. Romain verra Jean.
38. Romain saura dévorer Jean.
39. Véronique pourra consoler Mélanie.
40. Jean redécouvrira Nicolas.
41. Nicolas redémolira Marie.
42. Romain vend Guy.
43. Romain redémolira Mélanie.
44. Marie pourra dévorer Guy.
45. Jean démolira Romain.
46. Jean sent Romain.
47. Marie pourra dévorer Véronique et Jean.
48. Romain pourra dévorer Rémi et Jean.
49. Jean veut dévorer Guy et Véronique.
50. Marie sent Nicolas et Guy.
51. Wer diesen Satz liest, gewinnt ein Schokoladenei.
51. Nicolas redécouvrira Jean et Romain.
52. Nos nourrissons roux verront mon mouton roux.
53. Mélanie veut dévorer Jean et Véronique.
54. Guy pourra consoler Romain et Mélanie.
55. Guy veut manger Jean et Nicolas.
56. Nicolas machonnait Romain et Véronique.
57. Marie et Guy voulaient manger des lapins.
58. Guy démolira Marie et Romain.
59. Jean redémolira Romain et Jean.
60. Véronique redécouvrira Nicolas et Guy.
61. Nicolas veut dévorer Mélanie et Marie.
62. Nicolas redémolira Romain et Nicolas.
63. Nicolas machonnait Mélanie et Marie.
64. Romain saura dévorer Nicolas et Véronique.
65. Guy verra Nicolas et Marie.
66. Véronique veut manger Romain et Carolina.
67. Véronique verra Jean et Guy.
68. Nicolas verra Mélanie et Marie.
69. Mélanie redécouvrira Véronique et Romain.
70. Marie verra Véronique et Guy.
71. Jean pourra consoler Jean et Marie.
72. Nicolas pourra consoler Véronique et Mélanie.
73. Marie saura dévorer Guy et Mélanie.
74. Romain veut dévorer Mélanie et Nicolas.
75. Véronique démolira Nicolas et Marie.
76. Guy redémolira Marie et Nicolas.
77. Nicolas et Véronique voulaient manger Mélanie et Véronique.
78. Véronique et Marie démoliront Nicolas et Véronique.
79. Mélanie et Nicolas sauront dévorer Monica et Romain.
80. Nicolas et Mélanie voulaient manger Jean et Véronique.
81. Marie et Jean pourraient dévorer Romain et Mélanie.
82. Nicolas et Mélanie machonnaient Marie et Mélanie.
83. Jean verra Nicolas et Jean.
84. Mélanie et Romain machonnaient Mélanie et Jean.
85. Véronique et Romain pourront consoler Marie et Guy.
86. Mélanie et Marie redécouvriront Nicolas et Marie.
87. Guy et Romain pourront consoler Marie et Véronique.
88. Romain et Mélanie sauront dévorer Romain et Véronique.
89. Jean et Romain voulaient manger Marie et Véronique.
90. Mélanie et Véronique sauront dévorer Véronique et Guy.
91. Jean et Véronique sauront dévorer Mélanie et Jean.
92. Romain et Nicolas pourraient dévorer Mélanie et Romain.
93. Mélanie et Marie pourront consoler Romain et Nicolas.
94. Mélanie et Romain démoliront Guy et Mélanie.
95. Guy et Jean machonnaient Véronique et Romain.
96. Jean et Nicolas sauront dévorer Véronique et Jean.
97. Véronique et Nicolas verront Romain et Nicolas.
98. Mélanie et Guy pourront consoler Jean et Nicolas.
99. Véronique et Mélanie machonnaient Romain et Jean.
100. Romain et Marie sauront dévorer Marie et Nicolas.
101. Guy machonnait mon nourrisson.
102. Marie veut manger mes rats.
103. Véronique sent vos rats.
104. Véronique démolira ses nourrissons.
105. Guy sent des loups.
106. Jean pourra dévorer des lapins.
107. Romain machonnait mon rat.
108. Guy vend ses galopins.
109. Romain veut dévorer mon mouton.
110. Romain vend ses galopins.
111. Romain démolira nos loups.
112. Romain redémolira mes nourrissons.
113. Son long nourrisson pourra consoler nos dangereux loups.
114. Romain veut manger ses loups.
115. Mélanie démolira son galopin.
116. Guy veut dévorer nos lapins.
117. Nicolas verra mes lapins.
118. Mélanie verra son lapin.
119. Véronique démolira mon mouton.
120. Mélanie machonnait des lapins.
121. Nicolas saura dévorer mon galopin.
122. Mélanie pourra consoler mon galopin.
123. Véronique vend son nourrisson.
124. Mélanie pourra dévorer mes rats.
125. Romain veut dévorer des moutons.
126. Mélanie redémolira vos moutons.
127. Nicolas démolira mes lapins.
128. Nicolas et Mélanie voulaient manger mon mouton.
129. Mélanie et Jean machonnaient vos rats.
130. Guy et Véronique sauront dévorer son rat.
131. Mélanie et Marie verront vos nourrissons.
132. Véronique et Nicolas verront ses rats.
133. Véronique et Marie démoliront vos nourrissons.
134. Guy et Marie redécouvriront mes loups.
135. Guy et Romain pourront consoler ses nourrissons.
136. Ses nourrissons de Guy voulaient manger Guy.
137. Marie et Guy redémoliront nos rats.
138. Romain et Guy voulaient manger nos galopins.
139. Véronique et Guy pourront consoler des lapins.
140. Nicolas et Romain pourraient dévorer son lapin.
141. Marie et Mélanie pourront consoler mon galopin.
142. Véronique et Guy verront mon rat.
143. Mélanie et Guy voulaient manger nos moutons.
144. Nicolas et Véronique redécouvriront ses moutons.
145. Romain et Jean machonnaient des loups.
146. Marie et Mélanie machonnaient mes galopins.
147. Marie et Véronique pourront consoler nos galopins.
148. Romain et Marie démoliront vos lapins.
149. Marie et Véronique machonnaient mon lapin.
150. Vos moutons verront Guy.
151. Son rat saura dévorer Guy.
152. Son mouton machonnait Jean.
153. Mon nourrisson veut manger Jean.
154. Des rats redécouvriront Nicolas.
155. Mes lapins voulaient manger Véronique.
156. Des moutons machonnaient Marie.
157. Mes lapins redémoliront Nicolas.
158. Son nourrisson démolira Jean.
159. Son mouton machonnait Nicolas.

160. Mon loup redécouvrira Guy.
 161. Ses loups pourraient dévorer Véronique.
 162. Son nourrisson démolira Mélanie.
 163. Ses loups pourraient dévorer Romain.
 164. Mes moutons verront Marie.
 165. Mon nourrisson sent Marie.
 166. Mon mouton redémolira Romain.
 167. Des loups verront Romain.
 168. Mon nourrisson sent Marie.
 169. Mes lapins redémoliront Véronique.
 170. Ses lapins redécouvriront Marie.
 171. Mes lapins sauront dévorer Guy.
 172. Ses loups pourraient dévorer Romain.
 173. Mes galopins redécouvriront Romain.
 174. Des moutons pourront consoler Guy.
 175. Des rats pourront consoler Jean.
 176. Nos lapins voulaient manger Nicolas.
 177. Mes nourrissons voulaient manger Romain.
 178. Des galopins pourraient dévorer Guy.
 179. Ses moutons redémoliront Mélanie et Véronique.
 180. Ses rats redécouvriront Romain et Mélanie.
 181. Nos galopins pourront consoler Mélanie et Véronique.
 182. Vos galopins démoliront Romain et Jean.
 183. Mon lapin démolira Guy et Romain.
 184. Ses nourrissons redémoliront Romain et Guy.
 185. Mon loup redémolira Marie et Romain.
 186. Vos moutons verront Marie et Véronique.
 187. Son rat redécouvrira Mélanie et Romain.
 188. Vos lapins pourront consoler Marie et Nicolas.
 189. Nos loups pourraient dévorer Guy et Nicolas.
 190. Mon galopin redémolira Jean et Marie.
 191. Mes galopins démoliront Romain et Jean.
 192. Des lapins sauront dévorer Marie et Mélanie.
 193. Ses moutons démoliront Nicolas et Marie.
 194. Son galopin pourra dévorer Marie et Guy.
 195. Son lapin redémolira Nicolas et Romain.
 196. Ses moutons verront Mélanie et Guy.
 197. Vos loups pourraient dévorer Marie et Romain.
 198. Des galopins redécouvriront Jean et Véronique.
 199. Son mouton redémolira Mélanie et Guy.
 200. Son nourrisson démolira Nicolas et Jean.
 201. Mes loups machonnaient Mélanie et Véronique.
 202. Son mouton redécouvrira des moutons.
 203. Ses galopins voulaient manger nos nourrissons.
 204. Des loups redécouvriront nos galopins.
 205. Des loups redécouvriront des nourrissons.
 206. Son mouton machonnait mon loup.
 207. Vos nourrissons voulaient manger nos loups.
 208. Son rat saura dévorer nos loups.
 209. Mon galopin démolira des nourrissons.
 210. Vos nourrissons voulaient manger nos nourrissons.
 211. Son loup redécouvrira nos lapins.
 212. Vos loups redémoliront ses rats.
 213. Mes lapins verront ses nourrissons.
 214. Mon lapin vend nos lapins.
 215. Son galopin veut dévorer ses nourrissons.
 216. Son rat verra mes loups.
 217. Mes moutons pourront consoler nos galopins.
 218. Ses lapins machonnaient son nourrisson.
 219. Vos galopins redémoliront nos nourrissons.
 220. Son nourrisson vend mon loup.
 221. Ses moutons pourront consoler vos moutons.
 222. Mes moutons sauront dévorer mon nourrisson.
 223. Mes moutons pourraient dévorer ses galopins.
 224. Des galopins redécouvriront son nourrisson.
 225. Vos lapins redécouvriront son nourrisson.
 226. Des galopins démoliront des galopins.
 227. Son galopin veut manger vos rats.
 228. Mes nourrissons verront ses lapins.
 229. Mon lapin verra des moutons.
 230. Véronique verra son beau mouton.
 231. Jean redémolira ses longs lapins.
 232. Marie démolira ses longs rats.
 233. Nicolas veut manger nos jolis moutons.
 234. Marie pourra consoler mes jolis galopins.
 235. Jean veut dévorer vos beaux galopins.
 236. Romain machonnait mes beaux nourrissons.
 237. Jean redémolira son beau loup.
 238. Jean sent vos longs moutons.
 239. Véronique veut dévorer mes beaux rats.
 240. Mélanie pourra dévorer nos longs loups.
 241. Mélanie vend ses beaux galopins.
 242. Marie redécouvrira vos dangereux loups.
 243. Mélanie saura dévorer mes jolis galopins.
 244. Nicolas sent son beau rat.
 245. Guy pourra consoler ses longs rats.
 246. Nicolas sent nos jolis moutons.
 247. Mélanie redémolira des dangereux rats.
 248. Romain démolira ses jolis nourrissons.
 249. Romain veut manger des dangereux loups.
 250. Romain démolira des jolis moutons.
 251. Mélanie redécouvrira vos longs lapins.
 252. Mélanie verra nos longs nourrissons.
 253. Jean vend ses dangereux moutons.
 254. Jean veut manger vos dangereux loups.
 255. Jean veut dévorer nos longs moutons.
 256. Nicolas verra mon joli mouton.
 257. Jean démolira des beaux rats.
 258. Nicolas machonnait ses beaux moutons.
 259. Véronique veut manger mes longs nourrissons.
 260. Nicolas et Guy pourraient dévorer des dangereux moutons.
 261. Jean et Nicolas redémoliront des beaux moutons.
 262. Marie et Véronique machonnaient son beau loup.
 263. Mélanie et Guy redécouvriront mes jolis moutons.
 264. Jean et Nicolas redémoliront ses jolis galopins.
 265. Mélanie et Guy voulaient manger mes longs galopins.
 266. Mélanie et Guy verront nos dangereux rats.
 267. Mélanie et Véronique sauront dévorer mes dangereux loups.
 268. Guy et Jean pourront consoler ses dangereux rats.
 269. Véronique et Guy pourraient dévorer son joli galopin.
 270. Guy et Nicolas pourront consoler mes dangereux nourrissons.
 271. Jean et Marie pourront consoler ses dangereux nourrissons.
 272. Marie et Véronique démoliront son beau nourrisson.
 273. Romain et Jean verront mon long rat.
 274. Nicolas et Véronique sauront dévorer ses jolis rats.
 275. Guy et Véronique démoliront mes longs lapins.
 276. Marie et Mélanie démoliront ses beaux rats.
 277. Véronique et Mélanie redémoliront mon joli nourrisson.
 278. Romain et Jean verront des longs loups.
 279. Jean et Guy démoliront mon beau loup.
 280. Véronique et Jean sauront dévorer mes dangereux nourrissons.
 281. Jean et Guy machonnaient nos beaux lapins.
 282. Romain et Véronique redécouvriront mes dangereux galopins.
 283. Marie et Nicolas pourront consoler des beaux nourrissons.
 284. Mélanie et Nicolas pourraient dévorer mes jolis rats.
 285. Marie et Nicolas sauront dévorer mes dangereux galopins.
 286. Véronique et Marie machonnaient vos longs moutons.
 287. Véronique et Romain machonnaient des dangereux galopins.

288. Guy et Romain sauront dévorer vos jolis nourrissons.
 289. Guy et Nicolas démoliront des dangereux loups.
 290. Véronique et Romain redémoliront vos beaux moutons.
 291. Nicolas et Jean pourront consoler mes jolis galopins.
 292. Ses jolis loups redécouvriront Jean.
 293. Mes jolis galopins machonnaient Jean.
 294. Son joli galopin machonnait Véronique.
 295. Mon dangereux rat machonnait Guy.
 296. Mon joli galopin démolira Mélanie.
 297. Son joli rat sent Marie.
 298. Son dangereux mouton machonnait Marie.
 299. Mon joli mouton veut manger Marie.
 300. Nos dangereux rats machonnaient Romain.
 301. Nos longs loups voulaient manger Romain.
 302. Nos longs moutons pourraient dévorer Marie.
 303. Mon beau rat machonnait Romain.
 304. Mes jolis moutons verront Guy.
 305. Nos longs loups sauront dévorer Véronique.
 306. Mes dangereux galopins voulaient manger Jean.
 307. Ses dangereux galopins pourraient dévorer Jean.
 308. Vos dangereux lapins pourraient dévorer Romain.
 309. Nos dangereux galopins voulaient manger Romain.
 310. Ses beaux galopins redécouvriront Nicolas.
 311. Vos longs galopins verront Véronique.
 312. Mon long galopin sent Guy.
 313. Des jolis nourrissons redécouvriront Jean.
 314. Ses dangereux rats machonnaient Mélanie.
 315. Mon beau galopin verra Véronique.
 316. Ses beaux moutons voulaient manger Guy.
 317. Vos beaux galopins voulaient manger Véronique.
 318. Vos jolis lapins voulaient manger Nicolas.
 319. Ses dangereux moutons voulaient manger Marie.
 320. Vos longs lapins verront Guy et Marie.
 321. Son joli loup démolira Mélanie et Romain.
 322. Vos longs moutons sauront dévorer Mélanie et Guy.
 323. Des jolis galopins verront Marie et Mélanie.
 324. Vos moutons pourront consoler des loups menaçants.
 325. Vos dangereux loups redécouvriront Mélanie et Marie.
 326. Mes jolis galopins sauront dévorer Romain et Véronique.
 327. Vos beaux rats machonnaient Véronique et Romain.
 328. Nos jolis galopins verront Véronique et Romain.
 329. Mes dangereux lapins voulaient manger Jean et Guy.
 330. Son dangereux nourrisson pourra dévorer Guy et Véronique.
 331. Mon joli lapin démolira Nicolas et Jean.
 332. Son beau nourrisson machonnait Véronique et Mélanie.
 333. Mon long loup pourra consoler nos longs lapins.
 334. Mon long lapin sent Véronique et Romain.
 335. Ses jolis rats pourront consoler Jean et Mélanie.
 336. Mes jolis galopins démoliront Marie et Jean.
 337. Vos jolis lapins verront Guy et Jean.
 338. Nos longs lapins redémoliront Véronique et Mélanie.
 339. Nos dangereux loups redécouvriront Véronique et Mélanie.
 340. Son dangereux galopin redémolira Mélanie et Romain.
 341. Vos beaux galopins redémoliront Véronique et Romain.
 342. Mes dangereux nourrissons voulaient manger Mélanie et Véronique.
 343. Des jolis loups redémoliront Jean et Véronique.
 344. Ses beaux nourrissons sauront dévorer Jean et Romain.
 345. Nos jolis loups pourraient dévorer Mélanie et Guy.
 346. Vos jolis moutons sauront dévorer mon lapin.
 347. Mon beau loup saura dévorer mon lapin.
 348. Vos dangereux loups redémoliront vos moutons.
 349. Son dangereux loup veut dévorer mon rat.
 350. Mon long galopin vend mes galopins.
 351. Des jolis loups pourraient dévorer ses moutons.
 352. Nos dangereux nourrissons pourront consoler mon loup.
 353. Mes longs loups verront des galopins.
 354. Mon beau rat redécouvrira des lapins.
 355. Mon joli nourrisson veut dévorer vos nourrissons.
 356. Ses longs moutons verront mon loup.
 357. Son beau loup verra vos lapins.
 358. Vos dangereux nourrissons pourront consoler mon mouton.
 359. Son dangereux nourrisson verra nos galopins.
 360. Son joli nourrisson saura dévorer des galopins.
 361. Mon joli lapin sent mes lapins.
 362. Nos dangereux galopins machonnaient vos galopins.
 363. Nos dangereux nourrissons redécouvriront ses nourrissons.
 364. Nos beaux loups démoliront son nourrisson.
 365. Mes jolis loups démoliront mes galopins.
 366. Des longs galopins redémoliront son lapin.
 367. Mon dangereux galopin pourra dévorer des nourrissons.
 368. Son dangereux nourrisson vend ses nourrissons.
 369. Son joli lapin veut dévorer mes moutons.
 370. Mes beaux lapins redécouvriront vos loups.
 371. Des dangereux loups pourront consoler son nourrisson.
 372. Des longs loups pourraient dévorer mon galopin.
 373. Ses dangereux lapins démoliront vos galopins.
 374. Romain et Marie voulaient manger mon nourrisson marron.
 375. Jean et Véronique pourront consoler ses rats marrons.
 376. Mélanie et Jean machonnaient mon lapin marron.
 377. Romain et Nicolas pourront consoler nos loups menaçants.
 378. Mélanie et Romain sauront dévorer mon lapin marron.
 379. Nicolas et Mélanie sauront dévorer nos moutons marrons.
 380. Guy et Nicolas sauront dévorer nos loups roux.
 381. Nicolas et Marie verront nos rats menaçants.
 382. Véronique et Romain sauront dévorer vos nourrissons roux.
 383. Jean et Guy démoliront son rat roux.
 384. Marie et Romain verront mon rat marron.
 385. Marie et Guy démoliront ses loups roux.
 386. Jean et Véronique machonnaient son nourrisson menaçant.
 387. Marie et Véronique pourront consoler nos rats roux.
 388. Guy et Nicolas redémoliront des lapins roux.
 389. Marie et Véronique redémoliront mon galopin roux.
 390. Romain et Jean voulaient manger des lapins menaçants.
 391. Romain et Marie verront des loups menaçants.
 392. Véronique et Marie démoliront son rat marron.
 393. Jean et Nicolas redécouvriront son nourrisson menaçant.
 394. Mélanie et Véronique voulaient manger mes rats menaçants.
 395. Mélanie et Romain démoliront nos nourrissons marrons.
 396. Marie et Romain pourraient dévorer son lapin roux.
 397. Romain et Nicolas pourront consoler vos lapins roux.
 398. Nos jolis loups voulaient manger mes galopins roux.
 399. Nos nourrissons roux voulaient manger Jean.
 400. Son galopin roux vend Guy.
 401. Son mouton menaçant veut dévorer Nicolas.
 402. Ses loups roux voulaient manger Mélanie.
 403. Son galopin menaçant veut manger Marie.
 404. Des moutons roux redécouvriront Romain.
 405. Mon loup marron veut manger Marie.
 406. Vos moutons marrons pourraient dévorer Véronique.
 407. Mes moutons roux redémoliront Nicolas.
 408. Nos rats menaçants redécouvriront Mélanie.
 409. Son galopin menaçant saura dévorer Véronique.
 410. Ses rats menaçants démoliront Romain.
 411. Mes moutons marrons démoliront Véronique.
 412. Mes rats menaçants redémoliront Jean.
 413. Des loups menaçants machonnaient Véronique.
 414. Des rats marrons démoliront Romain.
 415. Mon nourrisson menaçant saura dévorer Mélanie.
 416. Vos galopins roux verront Mélanie.

417. Des rats marrons démoliront Guy.
 418. Mes lapins roux pourront consoler Nicolas.
 419. Mes rats marrons pourront consoler Mélanie.
 420. Son lapin menaçant saura dévorer Romain.
 421. Mon nourrisson marron pourra dévorer Véronique.
 422. Mon lapin roux machonnait Véronique.
 423. Vos rats qui pourraient dévorer voulaient manger Véronique.
 424. Mon loup qui démolira saura dévorer Nicolas.
 425. Mon nourrisson qui sent saura dévorer Marie.
 426. Mon rat qui pourra consoler redécouvrira Mélanie.
 427. Des lapins qui redémoliront verront Véronique.
 428. Son loup qui pourra dévorer redécouvrira Véronique.
 429. Ses rats qui voulaient manger verront Marie.
 430. Son mouton qui veut manger pourra dévorer Véronique.
 431. Nos moutons qui pourront consoler sauront dévorer Romain.
 432. Mes rats qui machonnaient démoliront Jean.
 433. Je veux manger des galopins.
 434. Vos rats qui redémoliront pourront consoler Romain.
 435. Nos moutons qui machonnaient redécouvriront Marie.
 436. Son rat qui veut manger saura dévorer Nicolas.
 437. Nos nourrissons qui voulaient manger verront Romain.
 438. Ses rats qui sauront dévorer pourraient dévorer Mélanie.
 439. Son rat qui pourra dévorer verra Guy.
 440. Son rat qui verra saura dévorer Jean.
 441. Son galopin qui machonnait démolira Mélanie.
 442. Son mouton qui pourra consoler redécouvrira Marie.
 443. Mon nourrisson qui veut manger démolira Jean.
 444. Mes rats qui pourront consoler voulaient manger Mélanie.
 445. Son lapin qui redémolira saura dévorer Jean.
 446. Mon loup qui saura dévorer pourra consoler Jean.
 447. Ses galopins qui machonnaient sauront dévorer Mélanie.
 448. Des loups qui voulaient manger machonnaient Guy.
 449. Mon lapin qui veut dévorer veut manger Marie.
 450. Son lapin qui veut manger démolira Guy.
 451. Mon mouton qui veut manger sent Guy.
 452. Des moutons menaçants voulaient manger Guy et Romain.
 453. Des lapins menaçants redémoliront Nicolas et Véronique.
 454. Mes rats marrons redécouvriront Guy et Véronique.
 455. Des lapins roux verront Marie et Guy.
 456. Ses galopins menaçants pourraient dévorer Nicolas et Véronique.
 457. Des loups roux redécouvriront Nicolas et Marie.
 458. Son mouton qui veut manger verra Gui.
 459. Son loup menaçant pourra consoler Mélanie et Véronique.
 460. Son galopin marron sent Véronique et Marie.
 461. Ses rats roux redémoliront Mélanie et Véronique.
 462. Ses galopins menaçants verront Guy et Mélanie.
 463. Mes jolis loups voulaient manger Mélanie et Marie.
 464. Son rat marron veut dévorer Romain et Guy.
 465. Son mouton menaçant saura dévorer Jean et Marie.
 466. Ses galopins menaçants démoliront Nicolas et Romain.
 467. Nos nourrissons roux redécouvriront Jean et Véronique.
 468. Mes nourrissons roux machonnaient Marie et Nicolas.
 469. Son nourrisson marron verra Guy et Mélanie.
 470. Des galopins marrons verront Véronique et Mélanie.
 471. Mon galopin marron saura dévorer Nicolas et Mélanie.
 472. Mon lapin menaçant veut manger Jean et Véronique.
 473. Mon loup marron pourra consoler Romain et Guy.
 474. Nos galopins marrons redémoliront Véronique et Guy.
 475. Mon nourrisson menaçant veut dévorer Mélanie et Jean.
 476. Mon loup roux veut manger Guy et Mélanie.
 477. Des nourrissons roux pourront consoler Romain et Nicolas.
 478. Des loups marrons démoliront Nicolas et Guy.
 479. Ses moutons roux verront Romain et Marie.
 480. Nos galopins marrons pourront consoler Jean et Mélanie.
 481. Des lapins verront mon mouton roux.
 482. Mon galopin pourra consoler des loups marrons.
 483. Mes beaux nourrissons voulaient manger Nicolas et Guy.
 484. Mon rat redécouvrira son mouton menaçant.
 485. Mes nourrissons pourraient dévorer vos rats marrons.
 486. Son loup démolira des nourrissons menaçants.
 487. Nos lapins redécouvriront son galopin menaçant.
 488. Ses rats démoliront mes rats marrons.
 489. Vos loups voulaient manger ses galopins menaçants.
 490. Je saurai dévorer vos longs nourrissons menaçants.
 491. Des rats pourront consoler mes rats menaçants.
 492. Des nourrissons voulaient manger nos nourrissons roux.
 493. Son galopin sent mon loup menaçant.
 494. Vos nourrissons redémoliront nos moutons menaçants.
 495. Vos loups voulaient manger mes nourrissons marrons.
 496. Son mouton redémolira vos galopins marrons.
 497. Son mouton verra mon mouton roux.
 498. Ses loups redémoliront vos loups roux.
 499. Ses rats voulaient manger son rat menaçant.
 500. Mes galopins machonnaient mon galopin roux.
 501. Ses lapins voulaient manger nos rats roux.
 502. Mes galopins machonnaient ses rats marrons.
 503. Son loup redémolira nos nourrissons marrons.
 504. Mon nourrisson sent nos loups marrons.
 505. Vos rats redémoliront des moutons roux.
 506. Mes rats redécouvriront mon lapin roux.
 507. Vos nourrissons machonnaient son loup roux.
 508. Nos moutons sauront dévorer des galopins marrons.
 509. Mon loup vend mon mouton menaçant.
 510. Mes nourrissons menaçants redémoliront ses loups marrons.
 511. Mon loup marron saura dévorer ses lapins marrons.
 512. Son rat menaçant saura dévorer nos moutons menaçants.
 513. Vos lapins menaçants démoliront son loup marron.
 514. Nos rats menaçants verront ses loups marrons.
 515. Ses rats menaçants pourraient dévorer son rat menaçant.
 516. Mes rats roux verront son nourrisson roux.
 517. Mon méchant nourrisson veut manger mes gentils nourrissons.
 518. Mon nourrisson marron verra mon rat roux.
 519. Mes nourrissons marrants redécouvriront nos moutons marrons.
 520. Mes lapins menaçants machonnaient mes lapins marrons.
 521. Mon lapin roux démolira son nourrisson roux.
 522. Mes lapins roux verront mes galopins roux.
 523. Mon rat marron machonnait des nourrissons roux.
 524. Des moutons marrons verront vos nourrissons roux.
 525. Guy et Nicolas sauront dévorer Guy.
 526. Ses galopins marrons pourront consoler mes nourrissons menaçants.
 527. Mon galopin roux pourra consoler mon loup menaçant.
 528. Vos galopins menaçants voulaient manger nos galopins marrons.
 529. Son nourrisson roux sent des loups roux.
 530. Des moutons marrons sauront dévorer vos rats menaçants.
 531. Son galopin menaçant machonnait mes lapins marrons.
 532. Son loup roux redécouvrira ses moutons roux.
 533. Mon rat marron veut dévorer des lapins menaçants.
 534. Ses nourrissons marrons redémoliront son nourrisson marron.
 535. Nos lapins roux pourront consoler vos moutons menaçants.
 536. Mon dangereux rat veut manger nos beaux rats.
 537. Vos beaux lapins voulaient manger vos beaux nourrissons.
 538. Son beau loup démolira son beau nourrisson.
 539. Nos jolis lapins pourront consoler vos longs rats.
 540. Nos longs lapins sauront dévorer mon beau loup.
 541. Nos beaux moutons démoliront son long lapin.

542. Nos dangereux rats redémoliront mes jolis moutons.
 543. Des longs lapins sauront dévorer ses jolis lapins.
 544. Son joli galopin pourra consoler mes beaux moutons.
 545. Des dangereux nourrissons démoliront nos jolis lapins.
 546. Son long mouton vend des jolis nourrissons.
 547. Vos jolis nourrissons pourraient dévorer nos beaux loups.
 548. Nos longs rats verront mes beaux loups.
 549. Nos jolis rats verront vos dangereux galopins.
 550. Ses beaux loups sauront dévorer vos jolis galopins.
 551. Son dangereux galopin pourra consoler mon beau nourrisson.
 552. Mes beaux loups sauront dévorer des longs nourrissons.
 553. Vos beaux loups pourront consoler mes beaux loups.
 554. Son long nourrisson pourra consoler mes dangereux nourrissons.
 555. Ses dangereux nourrissons pourront consoler mes beaux rats.
 556. Nos longs nourrissons voulaient manger mes jolis nourrissons.
 557. Des jolis galopins machonnaient nos beaux rats.
 558. Son long lapin pourra consoler vos dangereux nourrissons.
 559. Son dangereux galopin redémolira nos dangereux galopins.
 560. Nos beaux loups verront mon beau galopin.
 561. Nos jolis lapins voulaient manger nos beaux rats.
 562. Nos longs loups redémoliront mon dangereux lapin.
 563. Mes moutons de Marie démoliront Romain.
 564. Vos rats de Romain redécouvriront Mélanie.
 565. Ses nourrissons de Guy sauront dévorer Marie.
 566. Vos moutons de Marie verront Jean.
 567. Ses galopins de Véronique démoliront Marie.
 568. Son galopin de Véronique démolira Marie.
 569. Des rats de Marie sauront dévorer Romain.
 570. Vos loups de Marie verront Jean.
 571. Mon lapin de Nicolas démolira Marie.
 572. Mes nourrissons de Romain redémoliront Nicolas.
 573. Ses loups de Guy pourront consoler Véronique.
 574. Nos galopins de Romain voulaient manger Romain.
 575. Vos loups de Marie voulaient manger Jean.
 576. Mes loups de Mélanie voulaient manger Véronique.
 577. Mes galopins de Marie pourront consoler Marie.
 578. Son lapin de Nicolas verra Guy.
 579. Des loups de Véronique redémoliront Jean.
 580. Ses loups de Mélanie pourraient dévorer Jean.
 581. Des rats de Guy verront Mélanie.
 582. Des rats de Nicolas redémoliront Mélanie.
 583. Mon mouton de Romain sent Mélanie.
 584. Des nourrissons de Marie sauront dévorer Mélanie.
 585. Mon lapin de Romain saura dévorer Marie.
 586. Des galopins de Romain verront Nicolas.
 587. Mes rats de Guy sauront dévorer Romain.
 588. Des lapins de nos mas démoliront Marie.
 589. Ses galopins de son galopin pourraient dévorer Marie.
 590. Mes lapins de son nourrisson sauront dévorer Jean.
 591. Son lapin de son mas pourra consoler Romain.
 592. Mon mouton de mon terrain sent Véronique.
 593. Des galopins de ses terrains démoliront Véronique.
 594. Nos moutons de mon galopin pourraient dévorer Jean.
 595. Son lapin de mes mas vend Marie.
 596. Son mouton de mes nourrissons saura dévorer Romain.
 597. Son nourrisson de nos rues sent Véronique.
 598. Nos rats de mes parents démoliront Nicolas.
 599. Mes rats de mes maisons voulaient manger Guy.
 600. Mon mouton de mes maisons redécouvrira Mélanie.
 601. Son galopin de ses maisonnées veut manger Guy.
 602. Vos nourrissons de nos maisonnées redécouvriront Nicolas.
 603. Son galopin de vos maisonnées veut dévorer Guy.
 604. Vous pourrez consoler Nicolas.
 605. Nous redécouvrirons Marie.
 606. Vous démolirez Mélanie.
 607. Je verrai Véronique.
 608. Je veux dévorer Romain.
 609. Nous redécouvrirons Guy.
 610. Je veux manger Mélanie.
 611. Je saurai dévorer Guy.
 612. Nous redécouvrirons Jean.
 613. Vous saurez dévorer Nicolas.
 614. Je sens Marie.
 615. Vous machonnez Marie.
 616. Je saurai dévorer Jean.
 617. Je veux manger Jean.
 618. Nous redémolirons Mélanie.
 619. Nous démolirons Nicolas.
 620. Je verrai Véronique.
 621. Nous redécouvrirons Guy.
 622. Je veux manger Véronique.
 623. Vous machonnez Mélanie.
 624. Nous pourrions consoler Marie.
 625. Je veux manger Véronique.
 626. Je pourrai consoler Romain.
 627. Vous redécouvrirez Nicolas et Mélanie.
 628. Vous pourrez consoler Guy et Romain.
 629. Je pourrai consoler Jean et Guy.
 630. Je sens Marie et Mélanie.
 631. Nous pourrions consoler Véronique et Jean.
 632. Je veux manger Jean et Nicolas.
 633. Je pourrai consoler Guy et Mélanie.
 634. Vous machonnez Guy et Véronique.
 635. Je démolirai Nicolas et Mélanie.
 636. Je sens Marie et Guy.
 637. Je veux dévorer Nicolas et Guy.
 638. Vous verrez Marie et Jean.
 639. Nous pourrions consoler Jean et Guy.
 640. Je veux dévorer Nicolas et Guy.
 641. Je machonnais Véronique et Romain.
 642. Je pourrai dévorer Jean et Romain.
 643. Vous saurez dévorer Véronique et Mélanie.
 644. Je veux manger Mélanie et Marie.
 645. Je saurai dévorer Véronique et Mélanie.
 646. Vous pourrez consoler Véronique et Nicolas.
 647. Mes rats roux pourront consoler Nicolas et Véronique.
 648. Je machonnais Mélanie et Véronique.
 649. Nous pourrions dévorer mes rats.
 650. Vous voulez manger ses nourrissons.
 651. Vos beaux moutons pourraient dévorer mon rat marron.
 652. Vous redémolirez vos loups.
 653. Vous pourrez dévorer nos moutons.
 654. Vous machonnez son mouton.
 655. Je machonnais nos lapins.
 656. Je veux dévorer vos nourrissons.
 657. Je démolirai nos loups.
 658. Je redémolirai vos nourrissons.
 659. Je veux dévorer mes nourrissons.
 660. Nous pourrions dévorer ses moutons.
 661. Je saurai dévorer vos lapins.
 662. Nous saurons dévorer des galopins.
 663. Je veux manger nos rats.
 664. Je veux manger son nourrisson.
 665. Je machonnais mes galopins.
 666. Je saurai dévorer nos loups.
 667. Je pourrai consoler nos loups.
 668. Je vends ses nourrissons.
 669. Je redécouvrirai mes moutons.
 670. Mon lapin de son galopin redécouvrira Marie.

671. Vous voulez manger son mouton.
672. Je redécouvrirai nos longs loups.
673. Vous démolirez ses beaux galopins.
674. Je verrai vos longs loups.
675. Nous démolirons mes dangereux lapins.
676. Je veux manger nos jolis moutons.
677. Je saurai dévorer mon beau nourrisson.
678. Je vends vos beaux moutons.
679. Je machonnais mes dangereux rats.
680. Vous démolirez son dangereux lapin.
681. Je saurai dévorer son dangereux rat.
682. Vous pourrez dévorer son dangereux galopin.
683. Vous redécouvrirez nos beaux loups.
684. Je pourrai consoler des longs rats.
685. Je verrai vos longs galopins.
686. Vous saurez dévorer ses longs loups.
687. Je verrai des longs nourrissons.
688. Nous saurons dévorer nos dangereux galopins.
689. Je saurai dévorer son joli lapin.
690. Je saurai dévorer vos dangereux rats.
691. Je démolirai nos jolis loups.
692. Je démolirai ses jolis loups.
693. Je pourrai consoler son galopin menaçant.
694. Je pourrai consoler ses galopins menaçants.
695. Je redémolirai son mouton menaçant.
696. Je pourrai consoler ses nourrissons marrons.
697. Je pourrai consoler vos loups roux.
698. Vous démolirez mon rat roux.
699. Nous démolirons ses nourrisson menaçants.
700. Je veux dévorer mes lapins menaçants.
701. Je veux manger nos rats roux.
702. Je veux manger vos nourrissons menaçants.
703. Je veux dévorer ses loups menaçants.
704. Je redécouvrirai mon nourrisson menaçant.
705. Je saurai dévorer mon loup roux.
706. Vous saurez dévorer mes loups marrons.
707. Nous démolirons ses nourrisson menaçant.
708. Je vends ses nourrissons menaçants.
709. Je veux dévorer nos loups roux.
710. Je veux dévorer vos galopins marrons.
711. Je pourrai dévorer mon mouton marron.
712. Vous voulez manger vos moutons menaçants.
713. Vous pourrez consoler son galopin marron.
714. Je pourrai consoler nos lapins roux.
715. Je pourrai consoler nos lapins roux.
716. Nos rats roux redémoliront mon lapin marron.
717. Vos rats menaçants pourraient dévorer ses lapins roux.
718. Des galopins menaçants redémoliront ses nourrissons roux.
719. Ses moutons roux pourront consoler mes moutons marrons.
720. Mes lapins roux pourront consoler ses loups menaçants.
721. Ses galopins marrons verront ses loups roux.
722. Mon loup menaçant machonnait des galopins menaçants.
723. Vos moutons menaçants verront son galopin menaçant.
724. Ses lapins menaçants pourront consoler nos moutons menaçants.
725. Son galopin marron machonnait son lapin menaçant.
726. Mon galopin roux sent des galopins marrons.
727. Mon galopin marron saura dévorer son galopin menaçant.
728. Vos lapins menaçants démoliront mes rats menaçants.
729. Nos nourrissons roux redécouvriront mes moutons marrons.
730. Mes galopins roux démoliront vos moutons menaçants.
731. Ses moutons menaçants redécouvriront mes nourrissons roux.
732. Ses moutons roux redécouvriront ses rats marrons.
733. Son galopin roux machonnait mes lapins marrons.
734. Ses lapins roux pourront consoler nos lapins roux.
735. Nos nourrissons roux machonnaient vos nourrissons marrons.
736. Vos moutons marrons pourront consoler mon loup menaçant.
737. Son beau rat redécouvrira mes dangereux lapins.
738. Vos jolis loups démoliront des dangereux galopins.
739. Son long lapin redémolira des beaux moutons.
740. Son dangereux lapin démolira mon dangereux galopin.
741. Mon rat de Guy machonnait son lapin.
742. Des jolis loups redécouvriront vos longs galopins.
743. Son joli nourrisson veut dévorer des jolis rats.
744. Des jolis lapins pourront consoler vos jolis galopins.
745. Mes beaux loups machonnaient mon long rat.
746. Vos beaux lapins machonnaient mes longs galopins.
747. Son joli mouton redémolira mon dangereux lapin.
748. Son dangereux lapin veut manger vos beaux galopins.
749. Mon beau mouton saura dévorer ses dangereux galopins.
750. Son joli lapin pourra dévorer mes dangereux galopins.
751. Ses longs loups redécouvriront son dangereux lapin.
752. Nos beaux loups pourront consoler vos longs lapins.
753. Nos longs nourrissons pourraient dévorer mes beaux galopins.
754. Mon beau loup veut dévorer des dangereux nourrissons.
755. Mes longs lapins redémoliront son dangereux galopin.
756. Nos longs nourrissons pourraient dévorer vos dangereux moutons.
757. Son long nourrisson sent des jolis galopins.
758. Son dangereux rat redémolira mes beaux moutons.
759. Ses longs loups verront son joli lapin.
760. Mon mouton de Marie pourra dévorer Jean.
761. Des lapins de Marie redémoliront Jean.
762. Des nourrissons de Guy machonnaient Guy.
763. Son galopin de Guy démolira Jean.
764. Vos rats de Jean redécouvriront Guy.
765. Des moutons de Marie verront Jean.
766. Ses rats de Mélanie redémoliront Romain.
767. Son lapin de Marie saura dévorer Nicolas.
768. Vos galopins menaçants sauront dévorer Guy et Mélanie.
769. Mon nourrisson de Mélanie redémolira Guy.
770. Mon loup de Marie machonnait Véronique.
771. Des moutons de Nicolas sauront dévorer Romain.
772. Son nourrisson de Véronique verra Marie.
773. Son loup de Jean machonnait Romain.
774. Nos galopins de Mélanie machonnaient Véronique.
775. Nos nourrissons de Nicolas redémoliront Mélanie.
776. Nos galopins de Véronique redécouvriront Nicolas.
777. Mon nourrisson de Marie sent Romain.
778. Mon rat de Jean pourra dévorer Romain.
779. Des nourrissons de Mélanie pourraient dévorer Guy.
780. Nos lapins de Mélanie sauront dévorer Jean.
781. Mes galopins de mes maisons pourraient dévorer Jean.
782. Son loup de mon galopin vend Marie.
783. Mon lapin de ses maisons machonnait Romain.
784. Des galopins de son lapin redécouvriront Véronique.
785. Ses rats de mon galopin sauront dévorer Véronique.
786. Nos loups de mon mas verront Marie.
787. Son lapin de mon copain saura dévorer Mélanie.
788. Vos galopins de mes maisons voulaient manger Marie.
789. Son nourrisson de mon mas pourra consoler Romain.
790. Mon nourrisson vend mon loup menaçant.
791. Nos moutons de mon rat démoliront Véronique.
792. Son loup de nos moutons saura dévorer Romain.
793. Ses galopins de son mas pourraient dévorer Jean.
794. Vos galopins de vos loups voulaient manger Mélanie.
795. Nos lapins de ses rues redémoliront Véronique.
796. Son loup de mes rues machonnait Guy.

797. Des loups de vos maisons pourront consoler Guy.
 798. Son nourrisson de ses rues sent Marie.
 799. Son galopin de nos maisonnées veut manger Marie.
 800. Des lapins de Marie voulaient manger mon loup.
 801. Des loups de Romain redécouvriront nos rats.
 802. Mes rats de Romain démoliront mes moutons.
 803. Son rat de Véronique sent des loups.
 804. Son mouton de Mélanie machonnait mes moutons.
 805. Mon loup de Guy mâchonnait des rats.
 806. Mon rat de Nicolas veut dévorer nos loups.
 807. Son rat de Romain redécouvrira nos nourrissons.
 808. Mon nourrisson de Guy vend des loups.
 809. Ses moutons de Jean sauront dévorer des lapins.
 810. Mon lapin de Jean pourra consoler nos loups.
 811. Mon rat de Guy veut manger mes galopins.
 812. Des loups de Jean verront vos nourrissons.
 813. Mon beau galopin vend mes beaux nourrissons.
 814. Vous démolirez des dangereux rats menaçants.
 815. Je veux manger vos longs rats roux.
 816. Vous redécouvrirez vos dangereux rats menaçants.
 817. Nous verrons des dangereux nourrissons marrons.
 818. Vous voulez manger mon joli mouton marron.
 819. Nous machonnons ses jolis nourrissons menaçants.
 820. Je veux dévorer des jolis lapins menaçants.
 821. Je veux manger des jolis nourrissons menaçants.
 822. Vous saurez dévorer ses longs rats marrons.
 823. Je veux manger des beaux lapins roux.
 824. Vous voulez manger ses dangereux moutons menaçants.
 825. Je machonnais des jolis loup roux.
 826. Je saurai dévorer vos longs nourrissons menaçants.
 827. Vous pourrez consoler mon dangereux nourrisson marron.
 828. Je sens nos dangereux moutons marrons.
 829. Vous machonnez ses jolis nourrissons menaçants.
 830. Je veux manger des longs nourrissons roux.
 831. Vous saurez dévorer mes dangereux galopins marrons.
 832. Nous saurons dévorer mon beau mouton marron.
 833. Je redécouvrirai mon long loup marron.
 834. Je saurai dévorer mes longs lapins marrons.
 835. Je machonnais des jolis galopins marrons.
 836. Nous machonnons mes jolis moutons menaçants.
 837. Je démolirai des dangereux galopins marrons.
 838. Je pourrai consoler des longs galopins menaçants.
 839. Nous verrons ses dangereux nourrissons menaçants.
 840. Nous pourrons consoler vos beaux lapins marrons.
 841. Je pourrai consoler des beaux loup marrons.
 842. Nous verrons nos jolis moutons roux.
 843. Mon long lapin pourra dévorer nos nourrissons menaçants.
 844. Mes jolis lapins sauront dévorer ses nourrissons marrons.
 845. Son long mouton verra vos rats menaçants.
 846. Mon joli loup veut manger des moutons menaçants.
 847. Nos jolis rats pourraient dévorer son mouton roux.
 848. Mon long galopin vend son mouton menaçant.
 849. Mon dangereux galopin redécouvrira des loup marrons.
 850. Son dangereux rat démolira son loup roux.
 851. Ses dangereux rats redécouvriront mon mouton roux.
 852. Vos beaux rats pourront consoler son lapin menaçant.
 853. Son dangereux galopin machonnait mon lapin menaçant.
 854. Nos jolis moutons verront vos galopins roux.
 855. Vos dangereux loups pourront consoler vos nourrissons roux.
 856. Mes jolis galopins démoliront des loups menaçants.
 857. Son dangereux mouton redémolira ses moutons roux.
 858. Mes beaux moutons redécouvriront vos nourrissons roux.
 859. Son joli loup redémolira des nourrissons menaçants.
 860. Des longs nourrissons voulaient manger nos lapins menaçants.
 861. Des longs rats redémoliront des nourrissons menaçants.
 862. Nos beaux lapins voulaient manger des moutons roux.
 863. Mon dangereux lapin pourra consoler mes loups roux.
 864. Mélanie et Romain pourront consoler Mélanie et Jean.
 865. Vos dangereux moutons voulaient manger des moutons roux.
 866. Son beau mouton veut manger nos nourrissons marrons.
 867. Mes dangereux moutons redécouvriront ses nourrissons roux.
 868. Je pourrai dévorer mes rats.
 869. Son loup veut manger des rats roux.
 870. Mes galopins machonnaient nos loups menaçants.
 871. Mes lapins verront son mouton marron.
 872. Son rat redémolira des rats menaçants.
 873. Vos moutons sauront dévorer mes moutons roux.
 874. Vos galopins redécouvriront mon rat roux.
 875. Mon loup vend ses galopins roux.
 876. Mes moutons verront son lapin marron.
 877. Mon galopin vend mon lapin roux.
 878. Son mouton veut dévorer mes nourrissons roux.
 879. Son lapin vend des galopins roux.
 880. Nos lapins pourront consoler des nourrissons menaçants.
 881. Son galopin redécouvrira ses moutons marrons.
 882. Des nourrissons redémoliront mon rat roux.
 883. Mon rat redémolira ses rats marrons.
 884. Ses moutons pourraient dévorer son rat marron.
 885. Nos loups pourront consoler mon lapin menaçant.
 886. Vos nourrissons redécouvriront mes rats menaçants.
 887. Des nourrissons pourraient dévorer nos moutons roux.
 888. Ses galopins sauront dévorer mon rat menaçant.
 889. Des moutons redémoliront mon mouton menaçant.
 890. Vos rats démoliront mes galopins roux.
 891. Son loup démolira nos moutons marrons.
 892. Mes moutons pourront consoler mon mouton roux.
 893. Son mouton démolira son rat roux.
 894. Des loups verront nos moutons roux.
 895. Son mouton saura dévorer ses lapins roux.
 896. Mon nourrisson démolira des loup roux.
 897. Nos rats machonnaient nos loups menaçants.
 898. Mes moutons voulaient manger mes rats marrons.
 899. Nos rats de mon mas démoliront Jean.
 900. Mon rat de mes parents sent Guy.
 901. Nos rats de mes parents machonnaient Marie.
 902. Des loups de nos lapins voulaient manger Véronique.
 903. Son mouton de vos copains machonnait Guy.
 904. Ces nourrissons de mon mas redémoliront Jean.
 905. Ses loups de mes maisons démoliront Guy.
 906. Son loup de vos galopins sent Guy.
 907. Son galopin de nos maisonnées redécouvrira Mélanie.
 908. Mes lapins de mes nourrissons pourront consoler Romain.
 909. Mon loup de mes maisons sent Jean.
 910. Nos moutons de vos nourrissons redécouvriront Véronique.
 911. Ses loups de vos nourrissons pourront consoler Jean.
 912. Son loup de mon galopin sent Guy.
 913. Son rat de nos galopins veut dévorer Véronique.
 914. Mon rat de ses maisons démolira Véronique.
 915. Des rats de vos maisons redécouvriront Guy.
 916. Son nourrisson de ses rues machonnait Véronique.
 917. Mon nourrisson de son terrain sent Mélanie.
 918. Vos lapins de vos mas voulaient manger Nicolas.
 919. Son long mouton machonnait ses dangereux galopins.
 920. Mes dangereux rats démoliront des jolis loup.
 921. Nos beaux nourrissons pourront consoler mon joli lapin.
 922. Mon dangereux lapin saura dévorer mon beau nourrisson.
 923. Son long loup sent ses beaux loup.
 924. Vos beaux galopins redécouvriront des beaux galopins.
 925. Nos dangereux rats redécouvriront mon dangereux lapin.

926. Des longs moutons démoliront son long galopin.
927. Son dangereux loup saura dévorer ses dangereux rats.
928. Mon dangereux nourrisson redécouvrira nos beaux lapins.
929. Mon loup de Jean vend ses lapins.
930. Nos longs nourrissons démoliront mon long rat.
931. Nos dangereux nourrissons démoliront des jolis nourrissons.
932. Des longs galopins pourront consoler son beau lapin.
933. Ses beaux moutons sauront dévorer des jolis nourrissons.
934. Mon long nourrisson pourra dévorer vos dangereux nourrissons.
935. Mon dangereux lapin vend nos beaux loups.
936. Mes jolis loups redécouvriront vos longs loups.
937. Mon beau lapin vend mes beaux moutons.
938. Son long lapin saura dévorer mon long loup.
939. Mon dangereux loup démolira ses jolis nourrissons.
940. Ses beaux loups démoliront mes jolis moutons.
941. Mon dangereux loup verra son long loup.
942. Son long rat verra nos longs loups.
943. Mes jolis moutons voulaient manger des jolis rats.
944. Son long nourrisson saura dévorer mon dangereux lapin.
945. Son long mouton veut manger des dangereux rats.
946. Mon lapin de Marie veut manger vos rats.
947. Vos lapins de Jean machonnaient des loups.
948. Son lapin de Guy pourra dévorer mon rat.
949. Son mouton de Marie saura dévorer des moutons.
950. Mon mouton de Jean veut manger mes nourrissons.
951. Des galopins de Romain verront mes nourrissons.
952. Mes moutons de Guy démoliront mon mouton.
953. Nos lapins de Jean pourront consoler mes nourrissons.
954. Nos lapins de Véronique verront vos nourrissons.
955. Vos moutons de Jean pourraient dévorer vos loups.
956. Mon lapin de Guy redécouvrira nos lapins.
957. Nos moutons de Marie voulaient manger mes galopins.
958. Mon mouton de Marie verra vos lapins.
959. Les moutons marrons sauront dévorer Mélanie.
960. Son galopin de Mélanie démolira mon nourrisson.
961. Mon loup de Nicolas veut manger nos nourrissons.
962. Des rats de Nicolas démoliront des lapins.
963. Ses moutons de Marie voulaient manger ses loups.
964. Son nourrisson de Jean redémolira vos moutons.
965. Vos loups de Véronique pourront consoler mes galopins.
966. Vos lapins de Guy démoliront mes moutons.
967. Ses rats de Marie redécouvriront son loup.
968. Nos rats de Nicolas démoliront son rat.
969. Vos galopins de Jean redécouvriront mon loup.
970. Mes lapins de Marie pourraient dévorer mon galopin.
971. Son mouton de Marie saura dévorer vos galopins.
972. Mes nourrissons de Mélanie pourront consoler mes loups.
973. Son mouton dans la maison saura dévorer des lapins.
974. Mon mouton dans la maison veut manger mes nourrissons.
975. Des galopins dans la maison verront mes nourrissons.
976. Nos lapins dans la maison pourront consoler mes nourrissons.
977. Nos lapins dans la maison verront vos nourrissons.
978. Mes galopins dans la maison machonnaient mon lapin.
979. Vos moutons dans la maison pourraient dévorer vos loups.
980. Mon lapin dans la maison redécouvrira nos lapins.
981. Nos moutons dans la maison voulaient manger mes galopins.
982. Ses lapins dans la maison démoliront mon nourrisson.
983. Mon mouton dans la maison verra vos lapins.
984. Son galopin dans la maison démolira mon nourrisson.
985. Mon loup dans la maison veut manger nos nourrissons.
986. Des rats dans la maison démoliront des lapins.
987. Ses moutons dans la maison voulaient manger ses loups.
988. Son nourrisson dans la maison redémolira vos moutons.
989. Des moutons dans la maison redécouvriront mon galopin.
990. Nos nourrissons dans la maison démoliront nos galopins.
991. Vos loups dans la maison pourront consoler mes galopins.
992. Vos lapins dans la maison démoliront mes moutons.
993. Ses rats dans la maison redécouvriront son loup.
994. Nos rats dans la maison démoliront son rat.
995. Vos galopins dans la maison redécouvriront mon loup.
996. Ses loups dans la maison machonnaient mes nourrissons.
997. Mes lapins dans la maison pourraient dévorer mon galopin.
998. Son mouton dans la maison saura dévorer vos galopins.
999. Mes nourrissons dans la maison pourront consoler mes loups.
1000. Je veux manger Carolina.
1001. Vos lapins de vos rues voulaient manger Nicolas.

Erreurs et corrélations

Dans cet annexe, nous présentons l'intégralité des mesures objectives effectuées sur les différents corpora. L'intérêt de cette collection de nombres plutôt difficile à appréhender, est de fournir une référence de l'état actuel du SFC par rapport à ces corpora – ce qui permettra de situer des futures changements apportés au modèle.

Afin d'assurer l'homogénéité des données, les paramètres d'apprentissage ont été choisis identiques pour tous les corpora. Notamment, les coefficients de distribution du résidu (cf. page 92) sont : $\iota = 0,5$, $\kappa = 1$ et $\lambda = 1$. Les critères d'arrêt pour le cycle d'apprentissage par la synthèse restent également inchangés entre les différents corpora, ainsi que le nombre de noeuds (17) utilisé dans l'implémentation des différents générateurs de contours fonctionnels. Seule différence entre les corpora sont alors le nombre de types de générateurs employés.

Pour l'apprentissage nous utilisons toujours la moitié d'un corpus réservant ainsi l'autre moitié comme corpus de test. Afin d'éviter l'incidence du partage aléatoire du corpus en deux parties sur les résultats, nous utilisons trois partages différents par corpus. Nous obtenons ainsi trois paires de sous-corpora A-B, C-D et E-F. Pour chacun de ces 6 sous-corpora l'apprentissage est répété trois fois (versions a, b et c). Par exemple, l'apprentissage « Aa » correspond au premier entraînement sur les phrases de la partie A. Nous calculons l'erreur moyenne (et la corrélation) par rapport aux phrases du sous-corpus A d'entraînement (colonnes « app » dans les tableaux), par rapport au sous-corpus B de test (colonnes « test ») et par rapport au corpus entier (colonnes « tout »). De même, pour les apprentissages « Ab » et « Ac ». Ensuite nous inversons les rôles de A et de B pour les apprentissages « Ba », « Bb » et « Bc ». Les nombres donnés par la suite correspondent aux moyennes (et écart-types) par rapport aux 18(= 6 × 3) apprentissages effectués (cf. aussi 3.2.1, à partir de la page 120).

Détails de présentation. Les étiquettes F1, F2 et F3 correspondent aux trois valeurs de f_0 extraites/prédites par voyelle; F étant la collection de toutes ces valeurs. COEFF évoque le coefficient d'allongement. Les cinq premières lignes réfèrent alors aux

paramètres directement associés aux contours multiparamétriques traités par le SFC.

Les lignes restantes donnent plus de renseignement quant à la prédiction rythmique. ΔGIPC : durées de GIPCs; $\Delta\phi$: durées de phonèmes¹ et ΔP : durées de pauses (s'il y en a dans le corpus). La colonne « pause » précise quels GIPCs ont été pris en compte dans les mesures : ceux *avec* pause, *sans* pause ou *tous* les GIPCs. Dans les mesures relatives aux durées, on a exclu tous les GIPCs qui n'ont pas le même contenu phonémique (erreurs de prédiction de pause), car dans ce cas la comparaison des durées n'a aucun sens. Le taux d'occurrences de pauses étant faible (mis à part les corpora de formules mathématiques) et la génération de pauses plutôt fiable, ces omissions n'ont pratiquement aucune influence sur les résultats.

La partie « nombre » des tableaux indique comme – espérons-le – son nom l'indique, le nombre de données (GIPCs, phonèmes, pauses ou valeurs de f_0) sur lequel se basent les mesures « erreur RMS » et « corrélation ».

Précisons finalement, que les erreurs RMS sont exprimées en demi-tons en ce qui concerne F1, F2, F3 et F. Les données relatives aux durées sont en ms.

¹voir remarque 1 sur p. 22

FA	nombre			erreur RMS			corrélation			
	pause	app	test	tout	app	test	tout	app	test	tout
F1	tout	4800±79	4800±79	9610±0	1,70±0,01	1,74±0,02	1,72±0,01	0,712±0,008	0,694±0,009	0,703±0,005
F2	tout	4800±79	4800±79	9610±0	1,58±0,02	1,63±0,01	1,61±0,01	0,761±0,007	0,744±0,005	0,752±0,004
F3	tout	4800±79	4800±79	9610±0	2,34±0,02	2,42±0,02	2,38±0,01	0,766±0,007	0,749±0,004	0,758±0,004
F	tout	14400±238	14400±238	28800±0	1,91±0,01	1,96±0,01	1,93±0,01	0,754±0,007	0,737±0,004	0,746±0,003
C	tout	4610±85	4610±85	9220±0	0,192±0,002	0,198±0,003	0,195±0,001	0,885±0,003	0,878±0,004	0,882±0,002
ΔGIPC	avec	682±24	681±23	1360±28	44,8±1,1	45,4±1,1	45,1±0,8	0,903±0,004	0,900±0,007	0,901±0,004
	sans	3400±80	3390±78	6780±15	54,6±0,8	55,5±1,1	55,1±0,7	0,865±0,004	0,861±0,005	0,863±0,003
Δφ	avec	4080±97	4070±99	8150±38	53,1±0,6	53,9±0,9	53,5±0,6	0,872±0,003	0,868±0,005	0,870±0,003
	sans	1700±70	1700±60	3410±76	31,2±0,7	31,4±0,8	31,3±0,2	0,887±0,004	0,885±0,004	0,886±0,002
ΔP	avec	11800±261	11800±261	23500±75	25,3±0,3	25,5±0,3	25,4±0,1	0,811±0,004	0,808±0,004	0,809±0,002
	sans	13500±315	13500±311	26900±140	26,1±0,3	26,3±0,3	26,2±0,1	0,854±0,003	0,851±0,002	0,852±0,002
ΔP	avec	682±24	681±23	1360±28	122±4	123±3	122±2	0,341±0,036	0,329±0,025	0,335±0,022

(b) allemand

FF	nombre			erreur RMS			corrélation			
	pause	app	test	tout	app	test	tout	app	test	tout
F1	tout	1540±16	1540±16	3090±0	2,06±0,06	2,28±0,08	2,17±0,03	0,867±0,008	0,838±0,011	0,852±0,004
F2	tout	1540±16	1540±16	3090±0	1,97±0,06	2,19±0,07	2,08±0,03	0,909±0,005	0,888±0,007	0,898±0,003
F3	tout	1540±16	1540±16	3090±0	2,48±0,05	2,72±0,05	2,60±0,03	0,879±0,005	0,853±0,006	0,866±0,003
F	tout	4630±49	4630±49	9260±0	2,18±0,05	2,41±0,06	2,30±0,02	0,886±0,005	0,861±0,006	0,873±0,003
C	tout	1480±17	1480±17	2960±0	0,223±0,004	0,249±0,009	0,236±0,006	0,939±0,002	0,924±0,006	0,931±0,004
ΔGIPC	avec	280±10	273±8	554±6	44,5±1,9	45,1±2,5	44,9±1,2	0,930±0,005	0,929±0,006	0,929±0,004
	sans	1110±12	1100±14	2210±14	37,1±1,2	40,8±2,4	39,0±1,6	0,883±0,006	0,858±0,019	0,870±0,011
Δφ	avec	462±27	450±24	912±10	46,2±1,5	46,3±1,6	46,3±0,4	0,479±0,017	0,478±0,020	0,478±0,011
	sans	3300±37	3270±42	6570±33	23,0±0,3	23,7±0,7	23,4±0,3	0,546±0,019	0,519±0,024	0,533±0,011
ΔP	avec	3760±39	3720±39	7480±30	27,0±0,7	27,4±0,9	27,2±0,3	0,704±0,005	0,691±0,013	0,697±0,006
	sans	280±10	273±8	554±6	171±9	187±19	179±8	0,828±0,013	0,796±0,037	0,811±0,019

(a) français

TAB. B.1: Formules mathématiques.

ERREURS ET CORRÉLATIONS

DI	nombre			erreur RMS			corrélation		
	pause	app	test	app	test	tout	app	test	tout
F1	tout	532±6	532±6	1,48±0,05	1,92±0,12	1,72±0,05	0,796±0,015	0,652±0,041	0,722±0,016
F2	tout	532±6	532±6	1,07±0,03	1,30±0,05	1,20±0,02	0,965±0,003	0,949±0,006	0,957±0,002
F3	tout	532±6	532±6	1,33±0,06	1,51±0,05	1,43±0,02	0,976±0,002	0,970±0,002	0,973±0,001
F	tout	1600±18	1600±18	1,31±0,01	1,60±0,05	1,46±0,03	0,959±0,001	0,939±0,005	0,949±0,002
C	tout	424±6	424±6	0,123±0,008	0,153±0,008	0,139±0,003	0,811±0,019	0,694±0,028	0,751±0,011
ΔGIPC	tout	424±6	424±6	24,1±1,5	29,8±1,6	27,2±0,7	0,908±0,008	0,850±0,016	0,879±0,006
Δφ	tout	982±13	982±13	15,1±0,6	17,0±0,7	16,1±0,2	0,798±0,011	0,737±0,022	0,767±0,008

(c) doute-incrédulité

QS	nombre			erreur RMS			corrélation		
	pause	app	test	app	test	tout	app	test	tout
F1	tout	702±8	702±8	1,23±0,03	1,55±0,10	1,40±0,06	0,753±0,010	0,591±0,056	0,671±0,027
F2	tout	702±8	702±8	1,11±0,01	1,43±0,09	1,28±0,05	0,847±0,007	0,742±0,035	0,793±0,015
F3	tout	702±8	702±8	1,22±0,03	1,52±0,11	1,38±0,05	0,962±0,002	0,942±0,008	0,952±0,003
F	tout	2110±24	2110±24	1,19±0,02	1,50±0,10	1,36±0,05	0,925±0,002	0,881±0,016	0,902±0,007
C	tout	560±8	560±8	0,141±0,005	0,165±0,005	0,154±0,003	0,770±0,016	0,673±0,021	0,721±0,012
ΔGIPC	tout	560±8	560±8	26,6±0,8	30,9±1,0	28,8±0,6	0,888±0,005	0,841±0,010	0,864±0,005
Δφ	tout	1290±15	1290±15	16,0±0,2	17,5±0,3	16,8±0,2	0,770±0,006	0,719±0,011	0,744±0,006

(b) question

DC	nombre			erreur RMS			corrélation		
	pause	app	test	app	test	tout	app	test	tout
F1	tout	1040±22	1040±22	1,58±0,03	1,84±0,07	1,72±0,04	0,909±0,003	0,875±0,010	0,891±0,006
F2	tout	1040±22	1040±22	1,51±0,03	1,81±0,10	1,67±0,06	0,946±0,002	0,922±0,009	0,934±0,005
F3	tout	1040±22	1040±22	1,79±0,03	2,16±0,10	1,99±0,06	0,937±0,002	0,908±0,008	0,922±0,005
F	tout	3120±65	3120±65	1,63±0,03	1,94±0,09	1,80±0,05	0,934±0,002	0,906±0,009	0,920±0,005
C	tout	863±22	863±22	0,154±0,002	0,171±0,004	0,163±0,002	0,803±0,011	0,753±0,019	0,778±0,005
ΔGIPC	avec	1,94±1,26	2,17±1,62	14,8±1,7	20,2±3,2	18,1±2,8	0,755±0,215	0,155±0,626	0,257±0,469
	sans	853±22	853±22	31,6±0,8	34,7±1,7	33,2±0,6	0,884±0,008	0,856±0,020	0,870±0,007
	tout	855±21	855±21	31,5±0,8	34,7±1,7	33,1±0,6	0,884±0,008	0,856±0,020	0,870±0,007
Δφ	avec	1,94±1,26	2,17±1,62	14,8±1,7	20,2±3,2	18,1±2,8	0,755±0,215	0,155±0,626	0,257±0,469
	sans	1920±46	1920±46	20,1±0,2	21,2±0,5	20,6±0,2	0,801±0,006	0,774±0,016	0,788±0,005
	tout	1920±45	1920±45	20,1±0,2	21,2±0,5	20,6±0,2	0,802±0,005	0,776±0,014	0,789±0,005
ΔP	avec	1,94±1,26	2,17±1,62	32,2±11,3	41,6±21,7	44,4±10,3	0,714±0,262	0,0877±0,7273	0,150±0,516

(a) déclaration

TAB. B.2: Attitudes.

SC	nombre			erreur RMS			corrélation			
	pause	app	test	tout	app	test	tout	app	test	tout
F1	tout	622±9	622±9	1240±0	1,25±0,03	1,51±0,06	1,39±0,02	0,932±0,003	0,902±0,007	0,917±0,003
F2	tout	622±9	622±9	1240±0	0,986±0,028	1,22±0,07	1,11±0,03	0,983±0,001	0,974±0,003	0,978±0,001
F3	tout	622±9	622±9	1240±0	1,04±0,02	1,28±0,05	1,16±0,02	0,982±0,001	0,974±0,002	0,978±0,001
F	tout	1870±27	1870±27	3730±0	1,10±0,02	1,33±0,05	1,22±0,02	0,974±0,001	0,962±0,003	0,968±0,001
C	tout	500±9	500±9	1000±0	0,127±0,006	0,154±0,004	0,141±0,001	0,810±0,020	0,708±0,016	0,759±0,005
ΔGIPC	tout	500±9	500±9	1000±0	24,6±1,1	29,5±1,0	27,2±0,4	0,901±0,009	0,850±0,011	0,875±0,003
Δφ	tout	1150±16	1150±16	2300±0	14,8±0,4	16,6±0,4	15,7±0,1	0,804±0,005	0,749±0,005	0,776±0,003

(c) ironie de soupçon

EX	nombre			erreur RMS			corrélation			
	pause	app	test	tout	app	test	tout	app	test	tout
F1	tout	630±7	630±7	1260±0	2,34±0,07	2,79±0,04	2,58±0,02	0,773±0,014	0,665±0,013	0,719±0,006
F2	tout	630±7	630±7	1260±0	2,29±0,07	2,80±0,05	2,56±0,03	0,910±0,005	0,864±0,005	0,887±0,003
F3	tout	630±7	630±7	1260±0	2,99±0,05	3,53±0,06	3,27±0,02	0,838±0,003	0,768±0,007	0,803±0,002
F	tout	1890±20	1890±20	3780±0	2,56±0,06	3,08±0,04	2,82±0,01	0,864±0,005	0,801±0,005	0,832±0,002
C	tout	502±7	502±7	1000±0	0,124±0,006	0,147±0,006	0,136±0,003	0,807±0,018	0,719±0,026	0,762±0,012
ΔGIPC	tout	502±7	502±7	1000±0	25,8±1,5	30,5±1,7	28,3±0,7	0,906±0,008	0,863±0,017	0,885±0,007
Δφ	tout	1160±15	1160±15	2320±0	16,6±0,5	18,2±0,5	17,4±0,2	0,778±0,010	0,730±0,013	0,753±0,007

(b) exclamation

EV	nombre			erreur RMS			corrélation			
	pause	app	test	tout	app	test	tout	app	test	tout
F1	tout	625±8	625±8	1250±0	1,76±0,10	2,25±0,16	2,03±0,05	0,938±0,008	0,897±0,016	0,917±0,004
F2	tout	625±8	625±8	1250±0	1,58±0,11	1,99±0,16	1,80±0,05	0,959±0,006	0,935±0,011	0,947±0,003
F3	tout	625±8	625±8	1250±0	1,76±0,14	2,19±0,20	1,99±0,05	0,938±0,010	0,903±0,018	0,920±0,004
F	tout	1880±23	1880±23	3750±0	1,70±0,12	2,15±0,17	1,94±0,05	0,947±0,008	0,916±0,014	0,931±0,004
C	tout	502±8	502±8	1000±0	0,129±0,004	0,153±0,007	0,142±0,002	0,785±0,010	0,676±0,024	0,730±0,010
ΔGIPC	tout	502±8	502±8	1000±0	25,4±1,0	30,2±1,5	27,9±0,5	0,895±0,007	0,843±0,015	0,869±0,005
Δφ	tout	1150±13	1150±13	2310±0	15,0±0,3	16,7±0,6	15,9±0,2	0,795±0,006	0,740±0,010	0,767±0,007

(a) évidence

TAB. B.3: Attitudes.

	nombre			erreur RMS			corrélation			
	pause	app	test	app	test	tout	app	test	tout	
Vigo	F1	1640±10	1640±10	3290±0	1,62±0,04	1,88±0,07	1,75±0,04	0,726±0,012	0,610±0,032	0,667±0,018
	F2	1640±10	1640±10	3290±0	1,54±0,04	1,80±0,08	1,68±0,06	0,758±0,017	0,647±0,036	0,701±0,022
	F3	1640±10	1640±10	3290±0	1,88±0,04	2,16±0,08	2,03±0,04	0,677±0,009	0,541±0,037	0,608±0,019
C	F	4930±30	4930±30	9860±0	1,69±0,03	1,95±0,07	1,83±0,05	0,721±0,011	0,601±0,034	0,660±0,019
	C	1980±25	1980±25	3970±0	0,326±0,005	0,370±0,009	0,349±0,003	0,670±0,009	0,556±0,017	0,612±0,009
	ΔP	79,9±5,5	75,1±5,3	155±8	42,8±4,3	45,9±5,1	44,5±2,2	0,854±0,021	0,837±0,026	0,842±0,013
ΔGPC	sans	1860±27	1850±29	3700±12	38,3±0,9	43,8±1,7	41,1±0,8	0,734±0,011	0,624±0,031	0,677±0,017
	avec	1940±23	1920±30	3860±18	38,5±0,8	43,9±1,6	41,3±0,8	0,745±0,009	0,645±0,025	0,693±0,013
	Δφ	111±7	104±9	215±15	36,3±3,4	37,8±3,0	37,2±1,5	0,750±0,052	0,719±0,038	0,731±0,020
Δφ	sans	4410±74	4390±73	8810±29	22,0±0,4	24,0±0,8	23,0±0,3	0,663±0,007	0,604±0,015	0,633±0,008
	tout	4530±70	4500±77	9020±41	22,5±0,3	24,4±0,8	23,5±0,3	0,709±0,007	0,655±0,013	0,682±0,007
	ΔP	79,9±5,5	75,1±5,3	155±8	110±7	129±18	120±9	0,408±0,082	0,134±0,100	0,262±0,062

	nombre			erreur RMS			corrélation			
	pause	app	test	app	test	tout	app	test	tout	
1001	F1	6100±47	6100±47	12200±0	1,32±0,01	1,35±0,01	1,34±0,01	0,803±0,005	0,794±0,004	0,799±0,002
	F2	6100±47	6100±47	12200±0	1,36±0,02	1,39±0,02	1,37±0,00	0,869±0,004	0,863±0,003	0,866±0,001
	F3	6100±47	6100±47	12200±0	1,46±0,03	1,49±0,03	1,48±0,01	0,863±0,005	0,856±0,005	0,860±0,001
	F	18300±142	18300±142	36600±0	1,38±0,02	1,41±0,02	1,40±0,01	0,852±0,004	0,845±0,004	0,849±0,001
	C	5600±47	5600±47	11200±0	0,139±0,002	0,141±0,002	0,140±0,001	0,727±0,007	0,716±0,005	0,721±0,004
	ΔGPC	5600±47	5600±47	11200±0	23,2±0,3	23,7±0,3	23,5±0,2	0,835±0,006	0,826±0,006	0,831±0,003
Δφ	tout	11600±94	11600±94	23200±0	15,7±0,1	15,8±0,1	15,7±0,0	0,768±0,002	0,762±0,003	0,765±0,002

(a) 1001 phrases

(b) Corpus de Vigo

TAB. B.4: Phrases déclaratives : français et galicien.

Evaluation

C.1 Test de préférence

C.1.1 Contours des stimuli de test

Sur la page 228, sont représentés les contours de f_0 de tous les stimuli de test : les paramètres originaux en trait plein, la prédiction du SFC en tiret et celle de l'igm en pointillé. Les numéros de phrases se trouvent dans le coin supérieur gauche de chaque sous-figure.

Pour le français, on s'aperçoit d'un très grand écart entre les contours observés et ceux prédits par l'igm : des accents fortement exagérés (phrases 20 et 81 par exemple) et des plateaux en registre bas omettant toute structuration mélodique (p.e. 78,129). Perceptivement très gênants sont quelques mouvements montants en fin d'énoncé qui sont vraisemblablement dus à un décalage d'accent associé à « = » : phrases 91, 113 et 121.

L'image des contours en allemand est plus homogène. Globalement, les prédictions suivent les contours originaux bien qu'avec une dynamique généralement réduite. L'igm est caractérisé par un registre très bas en dernière partie des énoncés. Quant au SFC, il est à noter que la phrase 299 souffre d'oscillations rapides de f_0 qui donnent au stimulus l'impression d'une voix tremblante. Ce phénomène peut être attribué à une mauvaise couverture des configurations phonotactiques dans le corpus d'apprentissage (réduit pour le test de préférence).

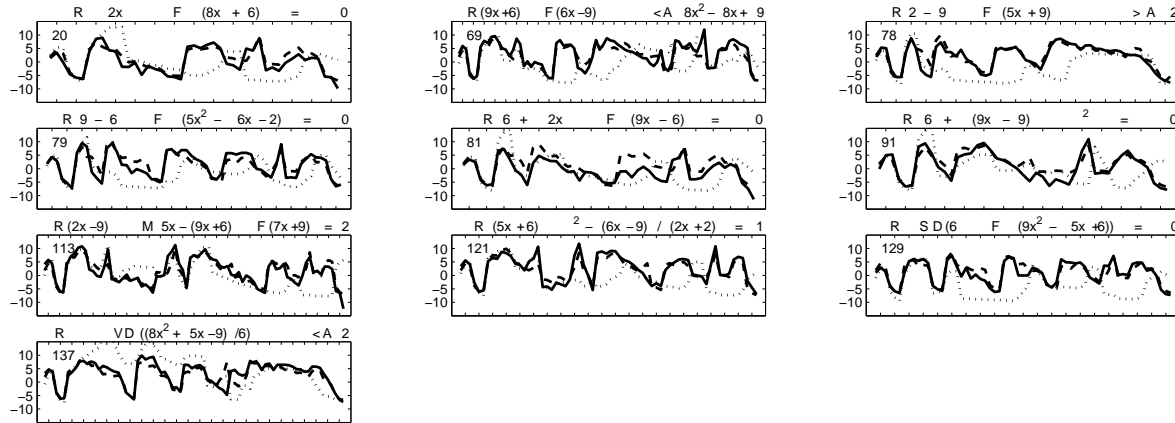


FIG. C.1: Français.

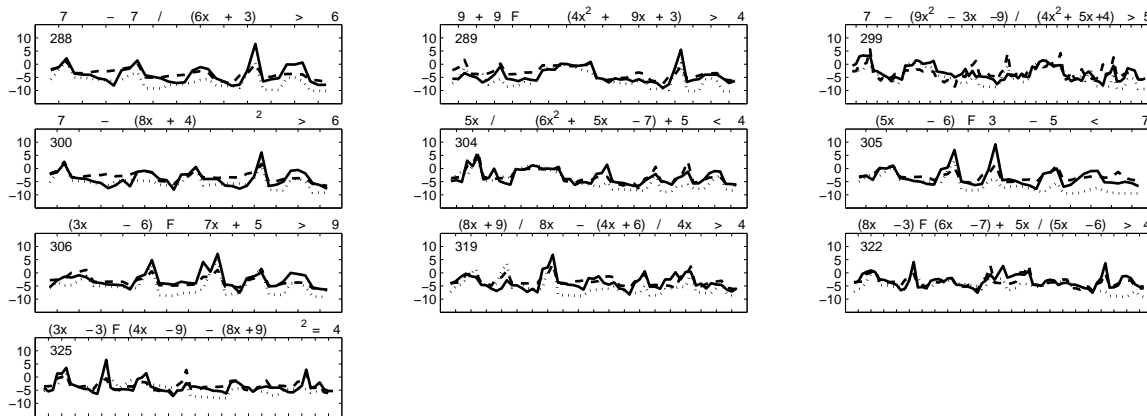


FIG. C.2: Allemand.

C.1.2 Tableaux des réponses – français

	20	69	78	79	81	91	113	121	129	137	Σ
FM	1	1	2	1	1	2	2	2	0	2	14
AT	2	2	2	1	2	1	2	2	2	2	18
GG	2	2	2	2	2	2	2	2	2	2	20
BF	2	1	0	2	1	2	2	2	2	2	16
KP	2	2	2	2	2	2	2	2	2	2	20
PB	2	2	1	2	2	2	2	2	1	2	18
SB	2	2	2	2	2	2	1	2	2	2	19
SG	2	2	2	2	2	2	2	2	2	2	20
CV	2	2	2	2	2	2	2	2	2	2	20
VA	2	2	2	2	2	2	2	2	2	2	20
Σ	19	18	17	18	18	19	19	20	17	20	185

TAB. C.1: SFC–igm

	20	69	78	79	81	91	113	121	129	137	Σ
FM	2	2	1	1	0	1	2	0	1	1	11
AT	1	2	1	2	2	2	2	2	1	1	16
GG	2	1	1	1	2	2	1	2	0	1	13
BF	0	1	2	0	1	2	2	1	0	2	11
KP	2	2	2	1	2	2	1	2	1	1	16
PB	2	0	2	2	0	2	0	2	1	2	13
SB	1	1	2	1	1	1	2	1	0	2	12
SG	1	2	2	1	2	1	2	2	1	2	16
CV	2	2	2	2	2	2	1	2	0	2	17
VA	0	2	2	1	0	1	1	1	1	2	11
Σ	13	15	17	12	12	16	14	15	6	16	136

TAB. C.2: Org–SFC

	20	69	78	79	81	91	113	121	129	137	Σ
FM	1	2	2	1	2	2	2	2	1	2	17
AT	2	2	2	2	2	2	2	2	2	2	20
GG	2	2	2	2	1	2	2	2	2	2	19
BF	2	2	2	2	2	2	2	2	2	2	20
KP	2	2	2	2	2	2	2	2	2	2	20
PB	2	2	2	2	2	2	2	2	2	2	20
SB	1	2	2	2	2	2	2	2	2	1	18
SG	2	2	2	2	2	2	2	2	2	2	20
CV	2	2	2	2	2	2	2	2	1	2	19
VA	2	2	2	2	2	2	2	2	2	2	20
Σ	18	20	20	19	19	20	20	20	18	19	193

TAB. C.3: Org–igm

C.1.3 Tableaux des réponses – allemand

	288	289	299	300	304	305	306	319	322	325	Σ
KB	1	0	0	0	2	0	0	1	1	1	6
AN	2	0	0	0	2	0	0	0	2	2	8
MP	2	2	2	1	2	2	2	2	2	2	19
EH	2	1	1	1	2	0	1	2	0	1	11
TB	2	0	0	1	2	1	0	2	2	2	12
HF	2	2	0	0	2	0	1	1	2	1	11
TH	2	1	1	1	1	1	2	0	1	0	10
IO	2	2	1	1	1	2	0	2	2	2	15
TW	2	1	1	2	0	0	0	1	1	0	8
JE	0	0	0	0	2	0	1	1	2	1	7
VB	2	2	2	0	2	1	1	1	2	2	15
Σ	19	11	8	7	18	7	8	13	17	14	122

TAB. C.4: SFC-igm

	288	289	299	300	304	305	306	319	322	325	Σ
KB	2	1	2	2	2	2	1	2	1	2	17
AN	2	2	2	2	1	2	2	2	2	1	18
MP	2	2	2	2	2	2	2	2	1	2	19
EH	0	0	2	1	2	0	1	2	1	2	11
TB	0	1	2	2	0	2	0	2	1	2	12
HF	0	2	1	2	0	2	0	2	2	1	12
TH	0	1	2	1	1	1	1	2	1	1	11
IO	1	2	1	1	1	1	2	1	1	2	13
TW	2	2	2	2	0	2	2	2	2	2	18
JE	0	2	1	1	1	2	1	0	2	2	12
VB	1	2	2	2	0	1	2	2	2	1	15
Σ	10	17	19	18	10	17	14	19	16	18	158

TAB. C.5: Org-SFC

	288	289	299	300	304	305	306	319	322	325	Σ
KB	1	1	0	1	1	2	1	2	2	2	13
AN	2	1	2	2	2	0	0	2	2	2	15
MP	2	2	2	2	2	2	2	2	2	2	20
EH	1	0	2	1	1	1	1	1	2	2	12
TB	1	1	1	2	2	2	0	2	2	2	15
HF	1	2	0	2	1	2	1	2	2	2	15
TH	0	1	2	0	2	2	1	2	2	2	14
IO	1	2	2	1	1	2	1	2	2	2	16
TW	2	2	2	2	2	2	0	2	2	2	18
JE	0	1	0	2	1	1	1	2	2	2	12
VB	2	1	2	0	1	2	0	2	2	2	14
Σ	13	14	15	15	16	18	8	21	22	22	164

TAB. C.6: Org-igm

Bibliographie

- Mark ANDERSON, Janet PIERREHUMBERT et Mark LIBERMAN. Synthesis by rule of english intonation patterns. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2.8.1–2.8.4, 1984.
- Véronique AUBERGÉ. *La synthèse de la parole : "des règles aux lexiques"*. Thèse de doctorat, Université Pierre Mendès-France, Grenoble – France, 1991. sous la direction de J. Rouault.
- Véronique AUBERGÉ. Developing a structured lexicon for synthesis of prosody. Dans Gérard BAILLY et C. BENOÎT, éditeurs, *Talking Machines : Theories, Models and Designs*, pages 307–321. Elsevier B.V., 1992.
- Véronique AUBERGÉ et Gérard BAILLY. Generation of intonation : a global approach. Dans *Proceedings of the European Conference on Speech Communication and Technology*, pages 2065–2068, Madrid, 1995.
- Gérard BAILLY. Contribution à la détermination automatique de la prosodie du français parlé à partir d'une analyse syntaxique. établissement d'un modèle de génération. Thèse de troisième cycle, Institut National Polytechnique de Grenoble, Grenoble, France, 1983. Sous la direction de M. Rossi et R. Carré.
- Gérard BAILLY et Bleicke HOLM. Learning the hidden structure of speech : from communicative functions to prosody. Dans *Symposium on Prosody and Speech Processing*, pages 113–118, Tokyo, Japan, 2002.
- Plinio BARBOSA. *Caractérisation et génération automatique de la structuration rythmique du français*. Thèse de troisième cycle, Institut National Polytechnique de Grenoble, Grenoble, France, 1994. Thèse de Doctorat Spécialité Sciences Cognitives sous la direction de Gérard Bailly.
- Katarina BARTKOVA et Christel SORIN. A model of segmental duration for speech synthesis in French. *Speech Communication*, 6 : 245–260, 1987.

- Frédéric BEAUGENDRE. Une étude perceptive de l'intonation du français : Développement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte. Thèse de troisième cycle, Université Paris XI, Paris, France, 1994. sous la direction de C. d'Alessandro.
- Alan W. BLACK et Andrew J. HUNT. Generating f0 contours from ToBI labels using linear regression. Dans *Proceedings of the International Conference on Speech and Language Processing*, pages 1385–1388, 1996.
- Cécile BRICHET et Véronique AUBERGÉ. La prosodie de la focalisation en français : faits perceptifs et morphogénétiques. Dans *Journées d'Etudes sur la Parole*, pages 33–36, Nancy – France, 2002.
- Ivan BULYKO et Mari OSTENDORF. Joint prosody prediction and unit selection for concatenative speech synthesis. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, USA, 2001.
- W. Nick CAMPBELL. Syllable-based segmental duration. Dans G. BAILLY et C. BENOÎT, éditeurs, *Talking Machines : Theories, Models and Designs*, pages 211–224. Elsevier B.V., 1992.
- W. Nick CAMPBELL et Alan W. BLACK. Prosody and the selection of source units for concatenative synthesis. Dans Jan P. H. VAN SANTEN, Richard W. SPROAT, Joseph P. OLIVE et Julia HIRSCHBERG, éditeurs, *Progress in Speech Synthesis*, pages 279–292. Springer Verlag, New York, 1997.
- William E. COOPER et Jeanne PACCIA-COOPER. *Syntax and Speech*. Harvard University Press, Cambridge(MA), USA, 1980.
- P. C. DELATTRE. L'intonation par les oppositions. *Le français dans le monde*, 64 : 6–13, 1969.
- R.E. DONOVAN, A. ITTYCHERIAH, M. FRANZ, B. RAMABHADRAN, E. EIDE, M. VISWANATHAN, R. BAKIS, W. HAMZA, M. PICHENY, P. GLEASON, T. RUTHERFORD, P. COX, D. GREEN, E. JANKE, S. REVELIN, C. WAAST, B. ZELLER, C. GUENTHER et J. KUNZMANN. Current status of the IBM trainable speech synthesis system. Dans *4th ISCA Tutorial and Research Workshop on Speech Synthesis*, Perthshire, Scotland, 2001.
- Danièle DUEZ. *Contribution à l'étude de la structuration temporelle de la parole en français*. Thèse de doctorat, Université de Provence, 1987.
- Kurt E. DUSTERHOFF et Alan W. BLACK. Generating F_0 contours for speech synthesis using the tilt intonation theory. Dans *ESCA Workshop on Intonation*, Athens, Greece, 1997.

- Kurt E. DUSTERHOFF, Alan W. BLACK et Paul TAYLOR. Using decision trees within the tilt intonation model to predict f0 contours. Dans *Proceedings of the European Conference on Speech Communication and Technology*, 1999.
- Françoise EMERARD, Laurent MORTAMET et Alain COZANNET. Prosodic processing in a text-to-speech synthesis system using a database and learning procedures. Dans G. BAILLY et C. BENOÎT, éditeurs, *Talking Machines : Theories, Models and Designs*, pages 225–254. Elsevier B.V., 1992.
- Gunnar FANT et Anita KRUCKENBERG. On the quantal nature of speech timing. Dans *Proceedings of the International Conference on Speech and Language Processing*, tome 3, pages 2044–2047, Philadelphia - USA, 1996.
- Ivan FÓNAGY, Eva BÉRARD et Judith FÓNAGY. Clichés mélodiques. *Folia Linguistica*, 17 : 153–185, 1984.
- Hiroya FUJISAKI et Keikichi HIROSE. Analysis of voice fundamental frequency contours for declarative sentences of japanese. *Journal of the Acoustic Society of Japan (E)*, 5(4) : 233–241, 1984.
- Hiroya FUJISAKI et Shuichi NARUSAWA. Automatic extraction of model parameters from fundamental frequency contours. Dans *Symposium on Prosody and Speech Processing*, pages 133–138, Tokyo, Japan, 2002.
- James P. GEE et François GROSJEAN. Performance structures : a psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15 : 411–458, 1983.
- Tuulikki GRÉPILLAT. Perçoit-on, par l'intonation, l'attitude du locuteur avant la fin de l'énoncé? T.E.R. de maîtrise d'anglais mention industrie de la langue, Université Stendhal, Grenoble, France, 1996.
- Martine GRICE, Stefan BAUMANN et Ralf BENZMÜLLER. German intonation within the framework of autosegmental-metrical phonology. Dans Sun-Ah JUN, éditeur, *Prosodic Typology*. Oxford University Press. (à paraître).
- Nina GRØNNUM. *The ground-works of Danish intonation*. Museum Tusulanum Press - Univ. Copenhagen, Copenhagen, 1992.
- François GROSJEAN, Lysiane GROSJEAN et Harlan LANE. The patterns of silence : Performance structures in sentence production. *Cognitive Psychology*, 11 : 58–81, 1979.
- François GROSJEAN, Lysiane GROSJEAN et Harlan LANE. The invariance of sentence performance structures across language modality. *Journal of Experimental Psychology : Human Perception and Performance*, 7 : 58–81, 1981.

- Johan 't HART, René COLLIER et A. COHEN. *A perceptual study of intonation : an experimental-phonetic approach to speech melody*. Cambridge University Press, Cambridge, 1990.
- David G. HAYS. Dependency theory : a formalism and some observations. Dans B. BLOCH, éditeur, *Language*, tome 40, pages 511–525. Kingston Press Services, Hampton, 1964.
- Daniel J. HIRST et Albert Di CRISTO. A survey of intonation systems. Dans Daniel J. HIRST et Albert DI CRISTO, éditeurs, *Intonation systems : a survey of twenty languages*, pages 1–44. Cambridge University Press, Cambridge, 1998.
- Daniel J. HIRST, Albert Di CRISTO et Robert ESPESSER. Levels of representation and levels of analysis for intonation. Dans M. HORNE, éditeur, *Prosody : Theory and Experiment*. Kluwer, Dordrecht, 2000.
- Daniel J. HIRST et Robert ESPESSER. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15 : 71–85, 1993.
- Bleicke HOLM et Gérard BAILLY. Generating prosody by superposing multi-parametric overlapping contours. Dans *Proceedings of the International Conference on Speech and Language Processing*, tome 3, pages 203–206, Beijing, China, 2000a.
- Bleicke HOLM et Gérard BAILLY. Génération de la prosodie par superposition de contours chevauchants : application à l'énonciation de formules mathématiques. Dans *Journées d'Etudes sur la Parole*, pages 113–116, Aussois - France, 2000b.
- Bleicke HOLM et Gérard BAILLY. Learning the hidden structure of intonation : Implementing various functions of prosody. Dans *Proceedings of Speech Prosody*, Aix-en-Provence, France, 2002.
- Bleicke HOLM, Gérard BAILLY et Colette LABORDE. Performance structures of mathematical formulae. Dans *Proceedings of the International Congress of Phonetic Sciences*, tome 2, pages 1297–1300, San Francisco, USA, 1999.
- Alexander V. ISAČENCO et Hans-Joachim SCHÄDLICH. *Untersuchungen über die deutsche Satzintonation*. Akademie-Verlag, 1964.
- N.F. JOHNSON. The psychological reality of phrase structure rules. *Journal of Verbal Learning and Verbal Behavior*, 4 : 469–475, 1965.
- Oliver JOKISCH, Hansjörg MIXDORFF, Hans KRUSCHKE et Ulrich KORDON. Learning the parameters of quantitative prosody models. Dans *Proceedings of the International*

-
- Conference on Speech and Language Processing*, tome 1, pages 645–648, Beijing, China, 2000.
- Adam KENDON. Does gesture communicate? a review. *Research on Language and Social Interaction*, 2(3) : 175–200, 1994.
- Dennis H. KLATT. Synthesis by rule of segmental durations in english sentences. Dans B. LINDBLOM et S. OHLMAN, éditeurs, *Frontiers of Speech Communication Research*, pages 287–300. Academic Press, London, 1979.
- Dennis H. KLATT. Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82(3) : 737–793, 1987.
- Anne LACHERET-DUJOUR et Frédéric BEAUGENDRE. *La prosodie du français*. CNRS ÉDITIONS, Paris, 1999.
- D. Robert LADD. Notes on the phonology of prominence. *Working Papers of Lund University*, 41 : 10–15, 1993.
- A. LJOLJE et Franck FALLSIDE. Synthesis of natural sounding pitch contours in isolated utterances using hidden markov models. *TrASSP*, 34 : 1074–1080, 1986.
- Eduardo LÓPEZ-GONZALO et Jose M. RODRIGUEZ-GARCÍA. Statistical methods in data-driven modeling of spanish for text to speech. Dans *Proceedings of the International Conference on Speech and Language Processing*, pages 1377–1380, Philadelphia, USA, 1996.
- Fabrice MALFRÈRE, Thierry DUTOIT et Piet MERTENS. Automatic prosody generation using suprasegmental unit selection. Dans *ESCA/COCOSDA Workshop on Speech Synthesis*, pages 323–328, 1998.
- Edwin MARTIN. Toward an analysis of subjective phrase structure. *Psychological Bulletin*, 74 : 153–166, 1970.
- Hansjörg MIXDORFF. *Intonation Patterns of German – Model-based Quantitative Analysis and Synthesis of F_0 contours*. Thèse de doctorat, Technische Universität Dresden, Allemagne, 1998.
- Hansjörg MIXDORFF. A novel approach to the fully automatic extraction of Fujisaki model parameters. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, tome 3, pages 1281–1284, Istanbul, 2000.
- Hansjörg MIXDORFF. *An integrated approach to modeling german prosody*. Habilitation, Technische Universität Dresden, 2002a.

- Hansjörg MIXDORFF. Speech technology, ToBI, and making sense of prosody. Dans *Proceedings of the Speech Prosody 2002 conference*, pages 31–38, Aix-en-Provence, France, 2002b.
- Hansjörg MIXDORFF et Oliver JOKISCH. Building an integrated prosodic model of german. Dans *Proceedings of the European Conference on Speech Communication and Technology*, tome 2, pages 947–950, Aalborg, Denmark, 2001.
- Bernd MÖBIUS. Corpus-based speech synthesis : methods and challenges. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ. Stuttgart)*, 6(4) : 87–116, 2000.
- Gregor MÖHLER et Alistair CONKIE. Parametric modeling of intonation using vector quantization. Dans *ESCA Workshop on Speech Synthesis*, Jenolan Caves, Australia, 1998.
- Alex I. C. MONAGHAN. Extracting microprosodic information from diphones – a simple way to model segmental effects on prosody for synthetic speech. Dans *Proceedings of ICSLP*, pages 1159–1162, Banff, Canada, 1992.
- Pascal MONNIN et François GROSJEAN. Les structures de performance en français : caractérisation et prédiction. *L'Année Psychologique*, 93 : 9–30, 1993.
- Yann MORLEC. *Génération multiparamétrique de la prosodie du français par apprentissage automatique*. Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble – France, 1997. Thèse de Doctorat Spécialité Sciences Cognitives sous la direction de Gérard Bailly.
- Yann MORLEC, Gérard BAILLY et Véronique AUBERGÉ. Synthesis and evaluation of intonation with a superposition model. Dans *Proceedings of the European Conference on Speech Communication and Technology*, tome 3, pages 2043–2046, Madrid - Spain, 1995.
- Sven E.G. ÖHMAN. Word and sentence intonation : a quantitative model. Rapport technique 2-3, Speech Transmission Laboratory - Department of Speech Communication and Music Acoustics - KTH, Stockholm - Sweden, 1967.
- Douglas O'SHAUGHNESSY. A study of French vowel and consonant durations. *Journal of Phonetics*, 9 : 385–406, 1981.
- Douglas O'SHAUGHNESSY. A multispeaker analysis of durations in read French paragraphs. *Journal of the Acoustical Society of America*, 76(6) : 1664–1672, 1984.

- Mari OSTENDORF, Patty PRICE et Stefanie SHATTUCK-HUFNAGEL. The Boston University Radio News Corpus. Rapport technique ECS-95-001, Electrical, Computer and Systems Engineering Department, Boston University, Boston, USA, 1995.
- Hartmut R. PFITZINGER. Local speech rate perception in german speech. Dans *Proceedings of the International Congress of Phonetic Sciences*, tome 2, pages 893–896, San Francisco, August 1999.
- Janet PIERREHUMBERT. *The phonetics and phonology of English intonation*. Thèse de doctorat, MIT, 1980.
- Janet PIERREHUMBERT. Synthetizing intonation. *Journal of the Acoustical Society of America*, 70(4) : 985–995, 1981.
- Jan Roelof de PIJPER. *Modelling British English intonation. An analysis by resynthesis of British English intonation*. Thèse de doctorat, Technische Universiteit Eindhoven, Pays-Bas, 1983.
- John F. PITRELLI, Mary E. BECKMAN et Julia HIRSCHBERG. Evaluation of prosodic transcription labeling reliability in the ToBI framework. Dans *Proceedings of the International Conference on Speech and Language Processing*, tome 1, pages 123–126, Yokohama, Japan, 1994.
- Romain PRUDON et Christophe D’ALESSANDRO. A selection/concatenation text-to-speech synthesis system : databases development, system design, comparative evaluation. Dans *4th ISCA Tutorial and Research Workshop on Speech Synthesis*, pages 137–142, Perthshire, Scotland, 2001.
- Stephan RAIDT. Kreuzvergleich : Zwei Modelle zur Generierung von Prosodie. Studienarbeit, Technische Universität Dresden / Institut de la Communication Parlée, Grenoble, France, 2002.
- Mario ROSSI. *L’intonation, le système du français : description et modélisation*. Ophrys, Paris, 1999.
- Yoshinori SAGISAKA. Speech synthesis by rule using an optimal selection of non-uniform synthesis units. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, tome 1, pages 679–682, 1988.
- Yoshinori SAGISAKA. On the prediction of global fo shapes for japanese text-to-speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1 : 325–328, 1990.

- Jan P. H. van SANTEN. Deriving text-to-speech durations from natural speech. Dans G. BAILLY et C. BENOÎT, éditeurs, *Talking Machines : Theories, Models and Designs*, pages 275–285. Elsevier B.V., 1992.
- Jan P. H. van SANTEN. Assignment of segmental duration in text-to-speech synthesis. *Computer Speech and Language*, 8 : 95–128, 1994.
- Jan P. H. van SANTEN. Segmental duration and speech timing. Dans Yoshinori SAGISAKA, Nick CAMPBELL et Norio HIGUCHI, éditeurs, *Computing prosody : Computational models for processing spontaneous speech*, pages 225–249. Springer Verlag, 1997.
- Michael S. SCORDILIS et John N. GOWDY. Neural Network based generation of Fundamental Frequency contours. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 219–222, 1989.
- Sophie K. SCOTT. *Perceptual Centres in Speech - An Acoustic Analysis*. Thèse de doctorat, University College, London, England, 1993.
- Kim SILVERMAN, Mary BECKMAN, John F. PITRELLI, Mari OSTENDORF, Colin W. WIGHTMAN, Patty PRICE, Janet PIERREHUMBERT et Julia HIRSCHBERG. ToBI : a standard for labeling english prosody. *Proceedings of the International Conference on Speech and Language Processing*, 2 : 867–870, 1992.
- Paul TAYLOR. The tilt intonation model. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1383–1386, 1998.
- Paul TAYLOR. Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, 107(3) : 1697–1714, 2000.
- Paul TAYLOR et Alan W. BLACK. Speech synthesis by phonological structure matching. Dans *Proceedings of the European Conference on Speech Communication and Technology*, tome 2, pages 623–626, 1999.
- Nina G. THORSEN. A study of the perception of sentence intonation – evidence from Danish. *Journal of the Acoustical Society of America*, 67(3) : 1014–1030, 1980.
- Stéphanie de TOURNEMIRE. Recherche d’une stylisation extrême des contours de f_0 en vue de leur apprentissage automatique. Dans *Journées d’Etudes sur la Parole*, pages 75–80, Trégastel, France, 1994.
- Stéphanie de TOURNEMIRE. *Identification et génération automatique de contours prosodiques pour la synthèse vocale à partir du texte en français*. Thèse de doctorat, ENST Paris – France, 1998.

- Christof TRABER. Fo generation with a database of natural fo patterns and with a neural network. Dans G. BAILLY et C. BENOÎT, éditeurs, *Talking Machines : Theories, Models and Designs*, pages 287–304. Elsevier B.V., 1992.
- Jean VÉRONIS, Philippe Di CRISTO, Fabienne COURTOIS et Cédric CHAUMETTE. A stochastic model of intonation for text-to-speech synthesis. *Speech Communication*, 26(4) : 233–244, 1998.
- Colin W. WIGHTMAN. ToBI or not ToBI? Dans *Proceedings of the Speech Prosody 2002 conference*, pages 25–30, Aix-en-Provence, France, 2002.
- Colin W. WIGHTMAN, Ann K. SYRDAL, Georg STEMMER et Alistair CONKIE. Perceptually based automatic prosody labeling and prosodically enriched unit selection improve concatenative text-to-speech synthesis. Dans *Proceedings of the International Conference on Speech and Language Processing*, tome 2, pages 71–74, Beijing, China, 2000.
- Hani YEHIA, Takaaki KURATATE et Eric VATIKIOTIS-BATESON. Facial animation and head motion driven by speech acoustics. Dans *5th Seminar on Speech Production : Models and Data & CREST Workshop on Models of Speech Production : Motor Planning and Articulatory Modelling*, pages 265–268, Kloster Seeon, Germany, 2000.
- Brigitte ZELLNER. Prediction of temporal structures for various speech rates. Dans Nick CAMPBELL, éditeur, *Volume on Speech Synthesis*. Springer. (à paraître).
- Brigitte ZELLNER. Pauses and the temporal structure of speech. Dans Eric KELLER, éditeur, *Fundamentals of speech synthesis and speech recognition*, pages 41–62. John Wiley and Sons, Chichester, 1994.