Mémoire présenté pour obtenir

L'Habilitation à Diriger des Recherches
de l'Université Paris-Sud

Spécialité : Informatique

par

Albert Rilliard

Prosodie et Interaction Homme-Machine :

Étude de la variation démarcative, diatopique, diachronique & expressive

Soutenue le 31/10/2014 devant le jury composé de :

J.-F. Bonastre, Professeur, Université d'Avignon (rapporteur)
F. Pellegrino, Directeur de Recherche, DDL (rapporteur)
J. B. Pierrehumbert, Professeure, Northwestern University (rapporteuse)
P. Mertens, Professeur, KU Leuven (examinateur)
A. Vilnat, Professeure, Université Paris Sud (examinatrice)
C. d'Alessandro, Directeur de Recherche, LIMSI-CNRS (parrain)

Habilitation préparée au **LIMSI-CNRS**
Rue John von Neumann
Campus Universitaire d'Orsay
Bât 508
91 403 Orsay CEDEX

Remerciements

Je saisi l'occasion de ce mémoire pour exprimer toute ma gratitude envers Christophe d'Alessandro pour son accueil, son enthousiasme, son érudition et ses conseils toujours très avisés ; ce travail lui doit beaucoup, bien au-delà du parrainage.

Je remercie aussi sincèrement tous les membres du jury de cette habilitation – et particulièrement François Pellegrino, Janet Pierrehumbert et Jean-François Bonastre (rapporteurs) – pour avoir accepté cette charge, ainsi que pour leurs conseils et leurs critiques.

Je souhaite aussi remercier vivement, pour leur aide, leur gentillesse et leur aimable compétence tous mes collègues du LIMSI qui font que travailler ici est un si grand plaisir car ils aplanissent les incongruités et font du LIMSI un lieu bien vivant. Toujours au LIMSI, les collaborations, construites ou décousues, avec Alexandre Allauzen, Boris Doval, Brian Katz, Céline Clavel, Christian Jacquemin, Cong-Thanh Do, David Doukhan, Hélène Maynard, Jean-Claude Martin, Jean-Sylvain Liénard, Laurence Devillers, Laurent Pointal, Lionel Feugère, Marc Évrard, Martine Adda-Decker, Mojtaba Jarrahi, Nathalie Delprat, Nicolas Sturmel, Olivier Perrotin, Sophie Rosset, Sylvain Le Beux, Thi Thu Trang Nguyen et alii, sont très-précieuses et ouvrent l'esprit. Spéciale dédicace à A&A – et à Philippe Boula de Mareüil pour son esprit curieux et critique.

Tout ce travail est grandement redevable, pour leurs avis, les discussions, leur aide, leurs idées, leurs désaccords, leur humour, leurs coups de gueule et/ou leur (in)disponibilité, à – notamment – Antonio Romano, Carmen Muñiz, Dang Khoa Mac, Donna Erickson, Elisa Fernández Rei, Eric Castelli, Hansjörg Mixdorff, Jean-Luc Rouas, Jean-Pierre Lai, João de Moraes, Leticia Rebollo Couto, Lurdes de Castro Moutinho, Mariko Kondo, Marine Guerry, Michel Contini, Nicolas Audibert, Plínio Barbosa, Sandra Madureira, Sylvain Detey, Toshiyuki Sadanobu, Véronique Aubergé, Yan Lu, Zuleica Camargo. Une pensée particulière pour Shochi Sensei –m-(_ _)-m–

Merci, pour leur support, à ma famille et mes amis.

Table des matières

Préambule

Ce document présente un résumé de mes travaux de recherche en vue d'obtenir l'Habilitation à Diriger des Recherches de l'université Paris Sud. Ces travaux portent sur la description de la prosodie, en tant qu'elle intéresse l'interaction homme-homme ou homme-machine ; ils participent donc à l'informatique de la langue parlée. Il s'agit de modéliser la variation prosodique, en identifiant et caractérisant différentes sources de variation.

Ce travail a pour but d'extraire une information fiable et utilisable afin d'améliorer des modèles descriptifs. Ces modèles peuvent servir directement à gérer l'interaction ; ils peuvent aussi avoir un but descriptif, via la représentation graphique (tracés de paramètres, dendrogrammes, représentation géographique, etc.) d'une information destinée à l'analyse linguistique.

Quels que soient les buts, mon travail s'appuie sur des données, essentiellement des corpus de parole ou des corpus multimodaux. L'analyse de ces données se fait grâce à des procédures de réduction de la variation – une stylisation prosodique qui peut se faire sur la base de modèles de perception tonale, de modèles phonétiques ou de contrôle gestuel. Cette stylisation permet l'accès à des informations pertinentes d'un point de vue perceptif et pour la fonction considérée. Ces données doivent être collectées en quantité suffisante, ce qui nécessite un traitement automatique ainsi que la création de bases de données adéquates à une interrogation efficace, y compris par des chercheurs d'autres domaines scientifiques.

Enfin, la qualité de ces données doit être évaluée. Il faut pour cela mettre en place des paradigmes expérimentaux adéquats à la mesure de cette qualité, mais aussi développer des outils de manipulation de la parole qui permettent l'introduction d'une variation contrôlée des données. Des procédures d'analyse et de synthèse de la parole sont donc centrales pour la poursuite de mes travaux. En permettant l'extraction de paramètres, elles permettent la caractérisation de dimensions et donc l'analyse de distances entre différentes performances. En permettant la modification de ces paramètres, ces procédures rendent possible la vérification expérimentale de la validité des modèles théoriques.

J'ai été recruté au CNRS en 2002 et affecté à l'Institut de la Communication Parlée (ICP), où j'ai mené des travaux sur les variations prosodiques liées aux expressions affectives (émotions & attitudes) et sur la variation dialectale dans l'espace roman, dans le cadre du projet AMPER (projet né au Centre de Dialectologie de Grenoble). En 2006, j'ai rejoint le Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) pour les travaux qui y sont poursuivis sur l'analyse et la synthèse de la parole, au sein du groupe Audio & Acoustique (ex-groupe Perception Située). Depuis, j'y mène ces travaux sur la modélisation de la variation prosodique, variation ayant pour origine des différences dialectales ou des stratégies expressives, en contexte, pour différentes langues et cultures. Ces travaux m'ont amené à collaborer avec plusieurs chercheurs du LIMSI et d'autres collègues à Grenoble comme à Bordeaux. L'abord de la variation dialectale et culturelle nécessite aussi la collaboration avec des universitaires de différents pays (notamment : Allemagne, Brésil, Chine, Espagne, États-Unis d'Amérique, Italie, Japon, Portugal, Vietnam).

1 | Introduction

> L'action réside dans l'usage de la voix en fonction de chaque sentiment (*pathos*), et consiste à savoir, par exemple, quand user d'une voix forte, quand user d'une voix faible, quand d'une voix moyenne, et comment se servir des tons, par exemple du ton aigu, du ton grave ou intermédiaire, et à quels rythmes recourir dans chaque cas. Car ce sont trois objectifs qu'ils visent : volume, harmonie, rythme. C'est avec cela, ou quasiment, qu'on remporte les prix dans les concours et de même que les acteurs, dans ces concours, ont un plus grand pouvoir que les poètes, de même ils s'imposent dans les joutes entre citoyens, à cause de la médiocrité de la vie politique. Mais on ne dispose pas d'une technique au point sur ces questions (puisque la préoccupation du style elle-même est intervenue tardivement), et cela paraît, à le bien prendre, quelque chose de vulgaire.
>
> Aristote, *Réthorique*, livre III, chapitre I
> (trad. P. Chiron, 2007)

Le concept de prosodie est traditionnellement relié à la morphosyntaxe et une performance prosodique de qualité est dite nécessaire à la restitution juste et efficace d'un texte. Ce constat a été fait rapidement après l'invention de l'écriture. Ainsi, les grecs anciens développèrent un système de notation *prosodique* destiné à permettre la restitution à voix haute des écrits. Pour noter ces variations prosodiques, ils marquaient les accents mélodiques (voire des tons) et la longueur des voyelles (Koster, 1936). Ces notations se sont aussi développées dans le souci de permettre une lecture juste des poètes (et d'Homère en particulier), dont la langue n'était déjà plus en usage. C'est le même souci de noter aussi justement que possible une langue pour la conserver intacte qui motive les grammairiens hindous (voir les travaux de Pánini, trad. 1897). Les travaux de ces premiers grammairiens vont nourrir toute la tradition « prosodique », notamment en Europe occidentale (mais voir aussi

Garcin de Tassy, 1873), jusqu'à une époque récente. Ils mettent au premier plan l'importance du rythme et de l'intonation pour la réalisation dans la parole des fonctions d'identification lexicale et de segmentation syntaxique : la prosodie réalise une fonction de démarcation prosodique. Mais la littérature rhétorique (voir la citation d'Aristote en exergue) note justement qu'un bon rhéteur doit maitriser la prosodie et plus généralement l'élocution afin d'avoir une parole claire et adéquate à son but dialectique. C'est exactement ce que propose Bell (1849) (voir figure 1.1).

Les recherches actuelles en prosodie mettent toujours au premier plan des fonctions prosodiques son rôle dans la structuration linguistique de la parole. Ainsi, les travaux de référence menés à ce sujet pour le français au LPL (Laboratoire Parole et Langage à Aix-en-Provence) proposent des descriptions traitant essentiellement de ces aspects fonctionnels (cf. Rossi *et al.*, 1981; Hirst et Di Cristo, 1998; Hirst, 2005; Di Cristo, 2013). Aller au-delà soulève des questions difficiles, notamment car l'étendue et la complexité des fonctions remplies par la prosodie (voir Rossi *et al.*, 1981, pour une discussion détaillée) et leur imbrication avec les autres niveaux linguistiques qui participent à la communication, constituent des entraves majeures pour les études en prosodie (Hirst, 2005). Pour le dire autrement, plusieurs fonctions étant encodées dans le même matériau acoustique, il est délicat d'attribuer une variation observée à un signe particulier. La solution à ce problème est toute trouvée dans l'utilisation de parole de laboratoire (Xu, 2010), qui permet d'étudier une fonction, toutes choses égales par ailleurs ; on verra les forces comme les limites de cette approche.

Une solution pour introduire des variations contrôlées d'un paramètre prosodique particulier consiste à utiliser des outils de modification et/ou de génération du signal de parole. Les recherches en synthèse de la parole ont ainsi depuis longtemps accompagné les recherches en phonétique. Que ce soit la machine de von Kempelen et les essais successifs de synthétiseurs de parole contrôlés par le geste (voir les travaux sur la synthèse performative : d'Alessandro, 2011; d'Alessandro *et al.*, 2014), ou des systèmes comme le *Pattern Playback* (Cooper *et al.*, 1951, 1952), les systèmes de synthèse ont grandement contribué à la compréhension de l'importance des différents paramètres impliqués dans la production et la perception de la parole. Depuis le début des années soixante, des systèmes automatiques de synthèse de la parole à partir du texte sont apparus et se sont perfectionnés (voir d'Alessandro, 2001, pour un historique de ces avancées pour le français). L'apparition de la synthèse par concaténation et les possibilités ouvertes par des algorithmes tels que TDPSOLA (Moulines et Charpentier, 1990) et leur popularisation rend la qualité segmentale des voix synthétiques et les possibilités de modification de la fréquence fondamentale et de la durée suffisantes pour étudier

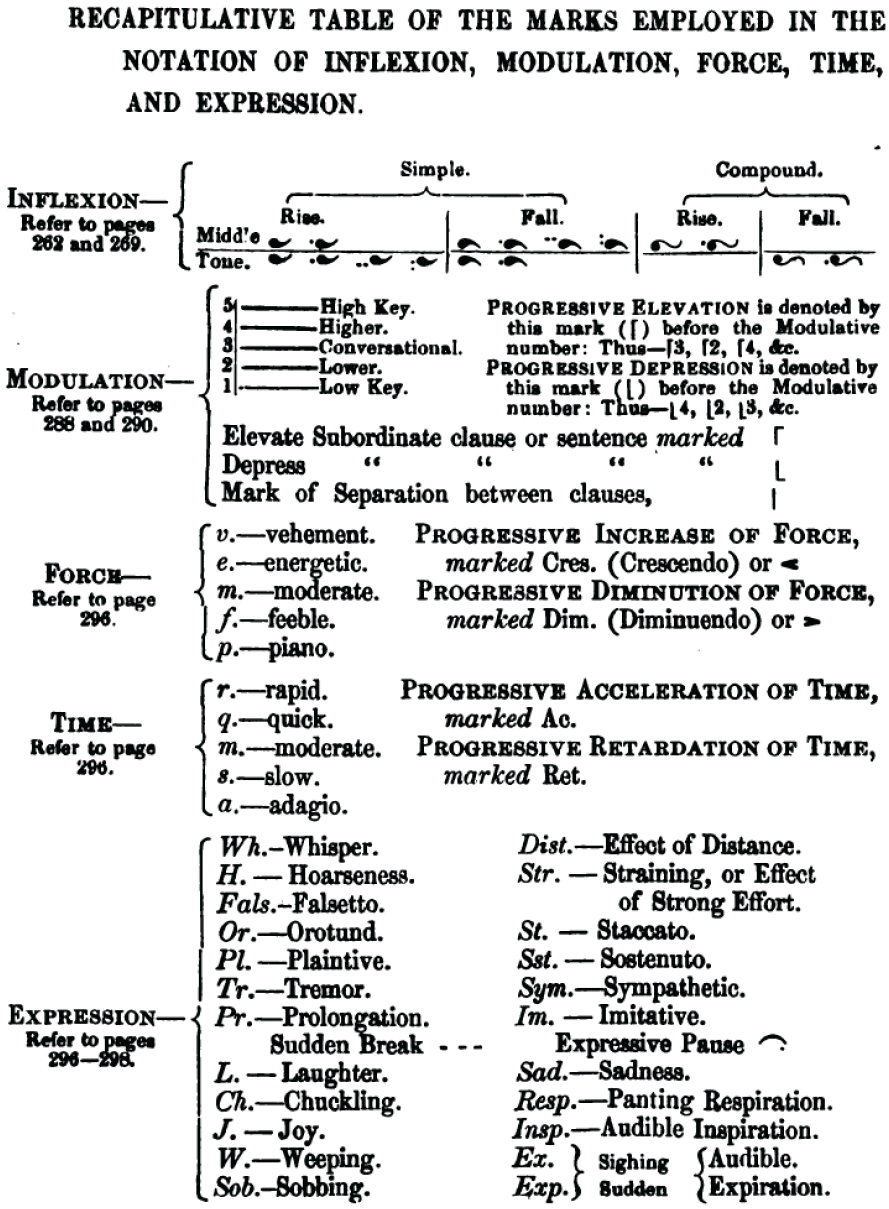


FIGURE *1.1* – *Notations prosodiques proposées par Bell, pour marquer les variations expressives convenables à une élocution expressive (cf. Bell, 1849, p. 299).*

des variations artificielles des ces paramètres de la prosodie dans de bonnes conditions.

C'est cette interaction entre sciences de l'ingénieur et sciences du langage (les premières créant des outils d'analyse et de synthèse de la parole, les se-

condes utilisant ces outils pour tester des modèles théoriques, ce qui concourt à l'amélioration des modèles de synthèse) qui a présidé à mes travaux de thèse (Rilliard, 2000, réalisés sous la direction de V. Aubergé à l'Institut de la Communication Parlée, à Grenoble) qui portaient sur l'évaluation et le diagnostic de la qualité de la fonction de démarcation des systèmes de synthèse de la parole.

C'est aujourd'hui encore la même bipolarité qui guide mes recherches : les problématiques des sciences de l'homme peuvent être abordées avec profit au travers d'outils de modélisation (en ce qui me concerne, la synthèse de parole est l'outil par excellence) qui permettent de tester les hypothèses des modèles afin de les valider expérimentalement – et le LIMSI, que j'ai rejoint en 2006, est un laboratoire pionnier de cette démarche d'interaction entre sciences de l'ingénieur et sciences de l'homme (Leipp *et al.*, 1968; Liénard et Teil, 1970). À l'inverse, une meilleure connaissance des variations pertinentes et une description fidèle et opérationnelle des phénomènes en jeu lors de l'interaction parlée permettent de développer des systèmes de traitement de la parole plus efficaces et adaptés et d'envisager de nouvelles méthodes d'analyses mieux à même de prendre en compte différents phénomènes – en bref, d'améliorer la partie *machine* de l'interaction homme-machine.

Le présent document débute par une description d'un aspect récurrent de mon travail : l'objectivation de la variation prosodique – autrement dit la mesure de « distances » prosodiques. Ce premier chapitre décrit différentes méthodes qui sont employées à cette fin, ainsi que les problèmes que cela pose. On retrouvera par la suite des cas d'application particuliers de ces mesures à une fonction ou un contexte particuliers.

Les premières fonctions abordées traitent des aspects démarcatifs de la prosodie – fonctions d'ordre strictement linguistique qui contribuent à la segmentation de la parole en unités. Il s'agit ici, si l'on reprend la classification proposée par Rossi *et al.* (1981, p. 24), de l'étude des variables non phonématiques contrôlées linguistiquement, pour lesquelles nous étudierons la perception des variations prosodiques associées à ces fonctions linguistiques : en quoi, dans quelle mesure et pour quelles variations syntaxiques, la prosodie – seule – permet-elle de percevoir ces indices morphosyntaxiques ? Cette question est toujours d'actualité pour les systèmes de synthèse de parole (cf. Nguyen *et al.*, 2014). Cependant, ces questions sont peu abordées en perception et c'est à cela que mes premiers travaux dans le domaine se sont attachés. Les résultats de cette validation des capacités distinctives de la prosodie sur les plans syntagmatique et paradigmatique sont résumés dans le chapitre 3.

Par ailleurs, la plupart des travaux en prosodie, et de manière générale en linguistique, s'intéressent à la description d'une forme standard de la langue

(pas toujours clairement identifiée), mettant de côté les sources de variation individuelles ou régionales (les « fonctions indicielles » de Di Cristo, 2013, p. 242). Si les travaux des dialectologues décrivent depuis longtemps cette variation diatopique pour les autres niveaux linguistiques (par exemple phonétique, lexical, etc., cf. Chambers et Trudgill, 1998), il faut attendre les propositions pionnières de Contini (1991) pour voir se mettre en place une méthodologie visant à étudier spécifiquement la variation diatopique des fonctions accentuelle et démarcative de la prosodie. On a donc ici une double description qui se met en place : l'observation de la réalisation prosodique d'une même fonction accentuelle et démarcative, dans sa variation géographique – les fonctions démarcative et diatopique sont abordées de concert. La mise en place de ces travaux, notamment au travers des thèses de Romano (1999) et de Lai (2002), a permis le développement du projet AMPER (Atlas Multimédia Prosodique de l'Espace Roman), qui se donne pour ambition la création d'un atlas dialectologique des variations prosodiques au travers des langues romanes. On s'intéressera donc, dans le cadre du projet AMPER, à la variation diatopique de ces fonctions – et donc aux effets induits par le « conditionnement externe lié aux variations régionales », selon les termes de Rossi *et al.* (1981). Un problème central aux études sur la variation prosodique dialectale vient de la très importante variation prosodique qui est rencontrée. L'abord de cette variabilité nécessite donc des méthodes d'enquête et d'analyse prosodique strictes (on verra le détail de la méthodologie mise en place). Un second problème est lié à la quantité des données nécessaires à l'extraction d'informations fiables pour mesurer des changements pertinents. Tester, grâce à des expériences perceptives, la variation de ces données ne peut être réalisé que sur des sous-ensembles restreints. La mise en place de mesures de distance prosodique a donc été nécessaire ; les résultats déjà obtenus ont permis les premières cartographies de la variation prosodique dialectale : cette avancée permet d'envisager la constitution d'une véritable *géoprosodie* (de Castro Moutinho *et al.*, 2011). Le chapitre 4 présente mes apports à ce projet et des réflexions sur différents aspects de la comparaison diatopique entre réalisations prosodiques.

Mes recherches traitent ensuite de l'expression intentionnelle d'affects, en tant qu'ils modifient le sens du discours, ou le rôle du locuteur lors d'une interaction parlée (cf. chapitre 5). Il s'agit ici des fonctions expressive et illocutoire telles qu'elles sont décrites par Di Cristo (2013, p. 246). La variation interculturelle de ces réalisations expressives constitue un aspect original du regard que je jette sur ces objets : décrire et mesurer la variation des stratégies prosodiques d'une langue ou d'une culture à l'autre (cf. chapitre 6). On pourrait dire qu'il s'agit ici aussi d'une vision diatopique, mais l'espace géographique n'est pas le critère déterminant – la proximité culturelle et lin-

guistique apporte une explication plus pertinente aux variations observées, même si les deux sont bien sûr en partie corrélées. La comparaison des variations linguistique et culturelle ne peut se faire sans résoudre des questions liées à la séparation entre concept et performance lors des tests perceptifs. Une partie des travaux de recherches décrits ici tente donc de résoudre ces problèmes expérimentaux particuliers soulevés par ces comparaisons interculturelles, qui font que deux entrées lexicales similaires (par exemple « ironie » et « irony ») peuvent faire référence à des concepts différents dans deux cultures (voir, pour une argumentation détaillée Wierzbicka, 1985, 2005).

La compréhension des codes prosodiques impliqués dans l'expression affective constitue un enjeu de premier plan pour une modélisation fidèle de la parole naturelle. Les défis sont nombreux et concernent des champs disciplinaires très variés – les sciences du langage et la didactique des langues étrangères, bien sûr. Mais il est impossible de décrire cette variation acoustique sans l'apport de méthodes nouvelles de description de la qualité vocale, notamment ; nous verrons comment ces résultats perceptifs contribuent à la constitution de la notion d'*espace expressif* d'un locuteur. La caractérisation des principales dimensions de cet espace – dimensions acoustiques, articulatoires, conceptuelles, etc. – constitue un cas d'école de la productivité de l'approche pluridisciplinaire décrite ci-dessus. Le développement d'outils d'analyse et de synthèse de paramètres que l'on suppose reliés à ces différentes dimensions en permet un contrôle pour l'évaluation perceptive de leur pertinence (Fourer *et al.*, 2014). L'extraction de paramètres robustes à même de rendre compte de concepts tels que la force de voix (Liénard et Barras, 2013) ou l'apériodicité d'un signal de parole (d'Alessandro *et al.*, 1998; Yegnanarayana *et al.*, 2011) constituent des aspects à développer dans le futur, car ils sont cruciaux tant pour la compréhension et la description des phénomènes que pour leur génération et leur modification par des systèmes de synthèse.

Les fonctions prosodiques démarcative et illocutoire prennent part directement à la gestion des échanges verbaux – et cet intérêt pour la parole dans l'interaction sous-tend mes activités de recherche. La modélisation de cette interaction parlée doit bien sûr se faire avec tous les autres niveaux linguistiques, en production comme en perception – et leurs pendants en traitement automatique de la parole : synthèse et reconnaissance automatiques de la parole. Ajoutons à cela un système de gestion du dialogue et un corps virtuel et l'on obtient le projet « tête parlante » du LIMSI (Martin *et al.*, 2007). Un tel environnement scientifique permet d'envisager le développement de recherches sur des systèmes complexes, seuls à même de modéliser un objet flou comme la prosodie de parole, en interaction.

Déjà en 1970, Liénard et Teil parlent de *traduction* pour évoquer le passage de l'écrit vers l'oral – et il faut effectivement aller bien au-delà de la « simple » conversion graphème-phonème pour produire une synthèse de parole expressive. L'exemple des travaux de Doukhan (2013), qui cherchent à générer la part expressive de la parole, sur la seule base du texte de contes le montre bien : le travail du conteur demande une véritable capacité sémiotique (Rastier, 2001), que la machine doit chercher à modéliser si elle veut savoir *lire*. Quintilien (trad. 1829, chapitre VIII) notait déjà l'importance de cette pratique et en décrivait la principale complexité (du point de vue *machine*) – car on ne saurait lire bien ce que l'on ne comprend pas. Le rapport de la prosodie au sens de la parole est donc crucial.

Dans mes travaux, la réalisation de ces fonctions – de cette sémantique prosodique – dans la matière acoustique est évaluée perceptivement dans un premier temps et cela pose déjà de nombreux problèmes expérimentaux. Dans un second temps, l'importance relative de la prosodie et de ses corrélats acoustiques est mesurée par rapport à l'importance des autres signes qui participent à l'échange langagier (tels que la syntaxe ou le lexique). En résumé, il s'agit de mesurer ce qui, dans cette matérialité de la prosodie, permet la perception – contextualisée – de chacune de ces fonctions. Le présent document a pour ambition de résumer les principaux résultats mis en lumière par ces recherches, puis de montrer autour de quels axes je compte les développer dans le futur.

Il va de soi que l'étude d'un objet complexe comme la prosodie, qui plus est la prosodie de différentes langues, fait appel à des compétences très variées. Il va aussi de soi que je suis loin de maitriser toutes ces compétences. Les travaux qui vont être présentés ici sont l'œuvre de collectifs de chercheurs, venant d'horizons disciplinaires variés, et poursuivant des buts scientifiques différents. Ces travaux ont notamment été réalisés à l'occasion de plusieurs thèses de doctorat ; ils ont aussi permis des collaborations fructueuses avec des chercheurs de cultures scientifiques variées. Ce document présente ma vision de ces travaux, leur articulation autour de mon projet de recherche : comment peut-on modéliser la variation prosodique liée à ces fonctions communicatives, en gardant les pieds ancrés dans le signal et les oreilles collées aux écouteurs, afin d'objectiver ce qu'en pensent les auditeurs.

2 | Objectivation de la variation

$$f_o(t) = F[s(t) + \sum_{k=1}^{m} w_{i_k}(t-t_k)]$$

where the forms of F, s, and w_i are yet to be determined.

Expression de la fréquence fondamentale observée,
d'après Öhman et Lindqvist (1965, p. 1)

Le premier pas de mes travaux pour une description de la variation prosodique consiste donc, ainsi qu'on l'a vu en introduction, à réduire la variation des paramètres prosodiques observés afin de mieux se concentrer sur des différences pertinentes pour une fonction donnée. Cette réduction de la variation constitue l'une des grandes difficultés des travaux en prosodie. Pour certaines fonctions prosodiques – essentiellement les fonctions linguistiques – de nombreuses méthodes ont été proposées, qui permettent une extraction fiable de cette variation. La première section présente une sélection des ces approches.

Dans la seconde section, je détaillerai comment ces méthodes ont été appliquées à la mesure de différentes sources de variation, en renvoyant pour cela à différents travaux que j'ai menés et qui seront décrits de manière plus extensive dans les chapitres suivants. Enfin, je montrerai en quoi ces méthodes rencontrent des difficultés lorsque la variation prosodique se complexifie. Je renverrai alors d'une part à des travaux qui ont pour but de mieux décrire cette variation, afin de mieux la comprendre ; je proposerai aussi quelques pistes préliminaires pour aller plus avant dans la mesure objective de cette variation prosodique.

2.1 Stylisations & distances prosodiques

Les descriptions prosodiques se basent essentiellement sur le paramètre de fréquence fondamentale (F_0) et étudient son évolution en fonction d'unités

linguistiques variées. Il peut s'agir d'étudier l'alignement d'un pic fréquentiel avec une structure syntaxique (Welby, 2006, 2007) ou bien de décrire cette variation en terme de contours (Delattre, 1966). Il est important dans tous les cas d'obtenir une estimation de la courbe de F_0, dans son évolution temporelle. Cette estimation se heurte à de nombreuses difficultés techniques, qui sont à l'origine d'une vaste littérature (voir par exemple Martin, 1982; Hermes, 1988; Boersma, 1993; Kawahara *et al.*, 2005; Liénard *et al.*, 2007). En particulier, la production d'irrégularités dans le signal du fait par exemple de certains modes de vibration des plis vocaux posent des problèmes pour cette estimation (cf. Martin, 2012, pour la détection du *creak*).

Les descriptions phonétiques vont donc considérer la variation dans le temps de ces paramètres, au regard de diverses unités linguistiques. Or, une estimation fine de la périodicité de la vibration des plis vocaux donne un signal de F_0 complexe et porteur d'informations non pertinentes du point de vue perceptif. Les travaux de Rossi (1971, 1978) donnent une indication de notre capacité à percevoir les variations de hauteur de la F_0 en fonction de la dynamique du signal de fréquence fondamentale. À partir de ces travaux, des méthodes de simplification de la courbe de F_0 cherchent à extraire une courbe de la hauteur perçue, et non plus seulement de la période de vibration des plis vocaux. C'est typiquement le cas des travaux de 't Hart *et al.* (1991), puis du modèle de perception tonale proposé par d'Alessandro et Mertens (1995) et implémenté dans le PROSOGRAM (Mertens, 2004). D'autres approches de ce problème de la réduction des données véhiculées par le continuum de F_0 ont été envisagées.

Ainsi, Levitt et Rabiner (1971) ont utilisé une méthode d'approximation des segments de F_0 par une base de polynômes orthogonaux, dont ils utilisent ensuite les coefficients pour mener des comparaisons objectives entre les courbes de plusieurs répétitions de la même phrase. Cette approche se heurte au problème de la taille variable des segments de F_0 et de la prise en compte du temps – et donc de la dynamique. La solution choisie consiste ici a neutraliser les différences de durée en utilisant un nombre prédéfini de points par segment ; cette solution (une approximation, grâce à une combinaison linéaire de polynômes, d'une courbe de F_0 normalisée en durée) est aussi adoptée par des travaux plus récents (par exemple Lai, 2014; Reichel *et al.*, 2014). Ces travaux mettent en lumière des formes de contours caractéristiques de différentes fonctions que les auteurs étudient (questions, accents, etc.). Une fois ces formes de contours identifiées, et avec le postulat que ces différents contours véhiculent des sens différents, il est possible de calculer des *distances* entre eux. Notons que ces mesures objectives ne correspondent pas à proprement parler à des « distances », mais plutôt à des « divergences » prosodiques (voir même à des mesures de similarité) ; le terme de « distance »

sera toutefois utilisé par la suite par commodité. C'est précisément ce que fait Hermes (1998b) en utilisant la corrélation entre les formes des contours observés pour proposer des distances objectives, proches de la distance perçue Hermes (1998a), voir pour ensuite réaliser une classification de ces contours (Klabbers et van Santen, 2004).

Certains travaux (Aston *et al.*, 2010; Evans *et al.*, 2010; Hadjipantelis *et al.*, 2012) se placent dans le cadre mathématique de l'analyse fonctionnelle des données (Ramsay et Silverman, 2005; Ferraty et Vieu, 2006), proposant des outils développés pour spécifiquement étudier la forme des contours, mais proposent une approche somme toute similaire aux précédentes. Tous ces travaux évacuent ainsi la dimension de durée des contours, car ils ont besoin de formes de tailles identiques. En neutralisant la durée, il est possible que la dynamique du contour de F_0 en soit pas conservée ; ceci est sans doute un inconvénient mineur considérant l'étendue restreinte des contours considérés dans ces travaux (de quelques syllabes au maximum).

Ces travaux partagent avec le modèle de Fujisaki (1983, 1988) une approximation basée sur des fonctions, à ceci près que les fonctions de ce dernier modèle sont motivées par des commandes articulatoires initiant les gestes des contours intonatifs. Une implémentation de ce modèle prédictif a été réalisée par Mixdorff (2000), qui permet d'estimer les différentes commandes articulatoires du modèle, et la courbe stylisée du contour de F_0.

Une autre approche classique pour la stylisation du contour de F_0 est proposée par Hirst et Espesser (1993) avec l'algorithme MOMEL. Celui-ci consiste en une approximation de la courbe de F_0 à l'aide de fonctions splines quadratiques entre des points-cibles déterminés automatiquement à l'aide d'un processus de régression. Cette stylisation permet aussi la description du contour de F_0 comme une fonction, mais présente l'avantage de conserver les informations de durée. MOMEL partage cette caractéristique avec l'autre approche de stylisation évoquée ci-dessus (le modèle de perception tonal proposé par d'Alessandro et Mertens, 1995). Les deux approches cherchent à produire une *close-copy stylisation* ('t Hart, 1991) du signal de F_0 original et à supprimer notamment la variation micromélodique. C'est aussi le cas de l'approche proposée par Contini *et al.* (2002), qui proposent de styliser la courbe mélodique en ne retenant que trois points de F_0 par voyelle, jugés suffisants pour conserver intacte l'information prosodique du signal original. L'approche de d'Alessandro et Mertens (1995) va se concentrer sur les noyaux vocaliques comme unités perceptives et discrétiser ainsi le contour de F_0 en tons plats ou dynamiques. À l'inverse, MOMEL va extrapoler une courbe de pitch continue, là où la F_0 présente des discontinuités – pour les phonèmes non voisés, notamment. Ces deux modèles de stylisation de la forme de la F_0

(MOMEL et PROSOGRAM) vont ensuite être utilisés pour extraire un niveau de notation prosodique supérieur, lié à des fonctions – voir Hirst (2005) et Mertens (2006). Un autre modèle d'annotation prosodique est largement utilisé dans la littérature : ToBI (Silverman *et al.*, 1992). Nous le citons ici du fait de sa large utilisation dans la littérature ; cependant, il consiste en une notation directe d'un niveau de description fonctionnel lié à la fonction de segmentation prosodique (d'où son nom) et ne permet pas de représentation intermédiaire comparable au PROSOGRAM ou à MOMEL. Il est donc difficilement utilisable pour observer des variations prosodiques d'ordre expressif et nécessite par ailleurs une adaptation à chaque langue, ce qui le rend délicat à utiliser pour la description objective de la variation prosodique dialectale.

Nous renvoyons le lecteur à l'article de van Santen *et al.* (1998) pour une comparaison de la plupart de ces différents types de modèles prosodiques et une discussion de leurs similarités de fond, au delà des divergences théoriques (voir aussi Grabe *et al.*, 2007).

Ces méthodes de stylisation suppriment les mouvements micromélodiques et beaucoup extrapolent les courbes prosodiques sur les parties non voisées du signal afin d'obtenir une courbe intonative continue, nécessaire pour appliquer par exemple une décomposition polynomiale. Une approche différente de toutes celles-ci a été proposée il y a quelques années (d'Alessandro *et al.*, 2006). S'il s'agit toujours d'arriver à une stylisation de la courbe mélodique, celle-ci est obtenue grâce à un processus d'imitation *chironomique* [1], de reproduction (ou de production tout court), de stimulus de parole à l'aide d'un synthétiseur contrôlé par le geste manuel d'un opérateur. Les paramètres permettant la définition de la fréquence fondamentale en sortie du système de synthèse sont mis en correspondance avec une dimension d'une tablette graphique : la F_0 en sortie du système de synthèse est donc gérée par la position d'un stylet, et donc par les gestes de l'opérateur.

Cette méthode permet de mettre en place un processus d'imitation : l'opérateur peut imiter – ou inventer – à sa guise un contour prosodique en le dessinant sur la tablette. Les méthodes d'imitation de la prosodie ne sont pas nouvelles en soi ; les processus de réitération prosodiques ont ainsi montré la capacité d'un locuteur à reproduire précisément la structure prosodique d'un énoncé qui lui est proposé (Nakatani et Schaffer, 1978; Larkey, 1983). Ce qui diffère ici c'est le transfert de modalité : il s'agit d'un geste manuel qui contrôle le geste prosodique, non plus d'un geste articulatoire. d'Alessan-

[1] Emprunté au grec χειρονομία, « mouvement de pantomime » ; dérivé de χειρονομῶ, « gesticuler » (d'après le TLFi : `http://atilf.atilf.fr/`).

dro *et al.* (2011) ont montré la capacité de locuteurs peu entraînés à mener cette tâche d'imitation avec une précision suffisante pour produire des stimulus perceptivement aussi proches de l'original qu'une imitation vocale. La métaphore gestuelle permet donc bel et bien de *dessiner* le contour mélodique ; la figure 2.1 présente les contours de F_0 de la phrase originale et des meilleures imitations vocale et gestuelle réalisées lors de cette expérience. On remarque sur ces tracés que l'imitation gestuelle fait complètement fi de la micromélodie, bien présente sur l'imitation vocale ; en cela, cette méthode valide les approches de stylisation précédentes qui, elles aussi, suppriment les variations micromélodiques. On peut aussi remarquer que cette courbe chironomique prend des libertés avec la courbe originale, même si les principaux mouvements mélodiques sont précisément suivis, dans leur dynamique comme dans leur temporalité. On verra à la section 8.3.1 ce que cette méthode pourrait apporter de différent des autres processus de stylisation présentés ici pour la compréhension des mouvements mélodiques.

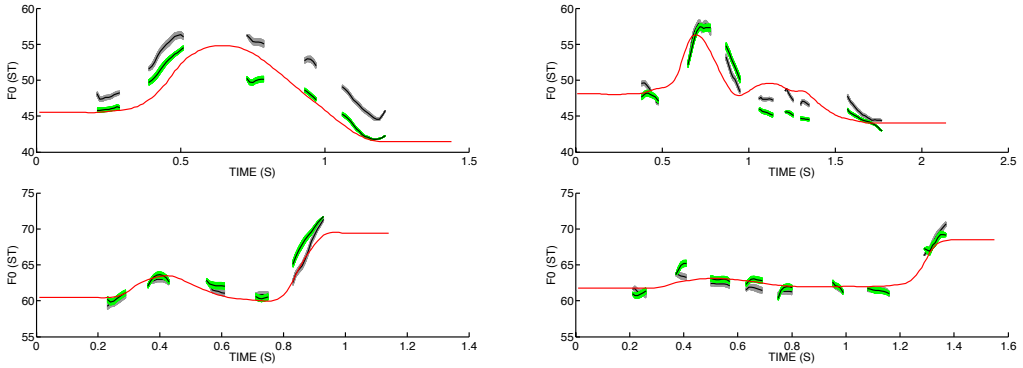


Figure 2.1 – *Tracés de F_0 d'un stimulus original (gris), de son imitation par la voix (vert) et par le geste (rouge) (d'Alessandro* et al.*, 2011).*

2.2 Quelle méthode, pour quel usage ?

Le choix d'une méthode parmi toutes celles qui viennent d'être présentées n'est pas simple ; et rien n'oblige non plus à n'en choisir qu'une seule. Le choix de la méthode dépend en partie de la disponibilité des routines de traitement automatique nécessaires ; en partie aussi des hypothèses à tester. Enfin, un certain nombre de contraintes peuvent rendre telle ou telle méthode plus ou moins souhaitable.

2.2.1 Phrases de durée comparable

L'étude des fonctions prosodiques de segmentation et de hiérarchisation de la parole se fait volontiers sur des phrases construites pour faire varier systématiquement leurs structures morphosyntaxiques. Cela permet notamment la mise en opposition d'un trait particulier entre deux phrases, toutes choses égales par ailleurs. Les structures qui sont comparées sont proches et ont en général le même nombre de syllabes

Les travaux présentés au chapitre 3 cherchent ainsi à évaluer la capacité des informations prosodiques à transmettre, seules, des informations quant à la segmentation de l'énoncé. Pour cela, un paradigme de réitération est mis en place, qui permet de présenter des énoncés véhiculant seulement les informations prosodiques à des auditeurs ayant pour tâche de juger la capacité de cette prosodie à segmenter différentes structures syntaxiques. Les résultats perceptifs montrent que des auditeurs perçoivent certaines distinctions syntaxiques sur la base de ces données prosodiques. Par la suite, les contours mélodiques de ces phrases sont comparés à l'aide de la mesure de corrélation proposée par Hermes (1998b).

Toujours en étudiant les mêmes fonctions morphosyntaxiques, mais dans leur variation diatopique, le chapitre 4 présente ma participation au projet d'étude de la variation prosodique dialectale AMPER. Cette participation consiste notamment à l'étude de l'application de mesures objectives pour aider à la mise en place d'une géoprosodie de l'espace roman. La même mesure de corrélation est appliquée (Hermes, 1998b), puis confrontée à une autre mesure, inspirée des travaux de Martin (1973) et de Contini et Profili (1989). Cette dernière consiste en l'extraction, pour chaque voyelle, d'un ensemble de traits prosodiques analogues aux traits phonétiques du niveau segmental. Cette extraction bénéficie de la stylisation en segments de droites réalisée par le modèle de perception tonale (d'Alessandro et Mertens, 1995) implémenté dans le Prosogram (Mertens, 2004). Une méthode d'extraction automatique de ces traits est proposée, qui permet la création de matrices de traits. Ces matrices sont ensuite comparées à l'aide d'une mesure de distance d'édition (Navarro, 2001), offrant ainsi une alternative nouvelle par rapport aux méthodes détaillées à la section 2.1.

Une évaluation des dispersions prosodiques obtenues entre les différents parlers d'un espace dialectal à l'aide de ces deux mesures de distances prosodiques montre bien sûr des différences liées aux natures très contrastées de ces deux approches, mais le résultat d'un regroupement hiérarchique montre que les mêmes tendances sont à l'œuvre pour les deux types de mesures. Ce résultat soutient la robustesse de ces mesures.

Les mesures de distances présentées dans ces deux chapitres montrent que l'on peut, dans des cas de comparaison de phrases proches, utiliser avec bénéfice les programmes de stylisation de la prosodie afin d'extraire les principales variations pertinentes et ainsi réaliser des analyses à grande échelle sur des ensembles de données importants. Point positif supplémentaire pour ces mesures objectives – leurs résultats sont reproductibles.

2.2.2 Importantes variations de durée

Les mesures utilisées jusqu'à maintenant ont été appliquées sur de la parole très contrôlée, pour laquelle peu de variations de durée sont observées. L'hypothèse selon laquelle ces différences de durée n'induisent pas de variation notable dans les courbes intonatives pouvait donc être soutenue et les données extraites des stimulus comparées à l'aide d'une corrélation. Notons toutefois que la création de matrices de traits ne fait pas cette hypothèse, mais prend en compte la durée parmi les traits prosodiques – on a donc une mesure objective plus complète que la précédente, si ses résultats semblent comparables.

Ces deux mesures sont toutefois problématiques lorsqu'on aborde d'autres aspects de la variation prosodique et en particulier les aspects expressifs. Un premier problème vient des importantes variations de durée qui peuvent être introduites en parole expressive – variations qui vont avoir un effet majeur sur la dynamique de la F_0. Une normalisation telle que celle utilisée précédemment n'est donc plus envisageable car elle modifierait cette dynamique. Le second problème est lié au domaine d'application des contours intonatifs considérés. Pour les fonctions tonales (Evans *et al.*, 2010) ou syntaxiques (Reichel *et al.*, 2014), on peut considérer des segments de parole de taille réduite (1 à quelques syllabes) et de tailles comparable pour tous les segments considérés. Si l'on considère par contre des expressions prosodiques qui vont modifier un énoncé complet, l'empan du contour intonatif peut montrer des différences de longueur importantes. Par contre, la forme globale de ce contour peut rester comparable ; c'est ce qu'exprime la notion de *Prosodic movement expansion* proposée par Morlec *et al.* (2001) pour la prosodie attitudinale. Les expressions prosodiques attitudinales se situent au niveau de la phrase ou de l'énoncé et vont permettre au locuteur d'exprimer par exemple du doute, de l'ironie ou de la politesse. La figure 2.2 montre un exemple de ce phénomène d'expansion du mouvement prosodique, obtenu en demandant à un locuteur de répéter les mêmes six expressions attitudinales sur des phrases de taille variable (de Moraes et Rilliard, 2014a). On peut remarquer que la forme des contours montre plus de similarités au sein de la même expression attitudinale (pour chaque ligne), plutôt que du fait de la variation de la taille

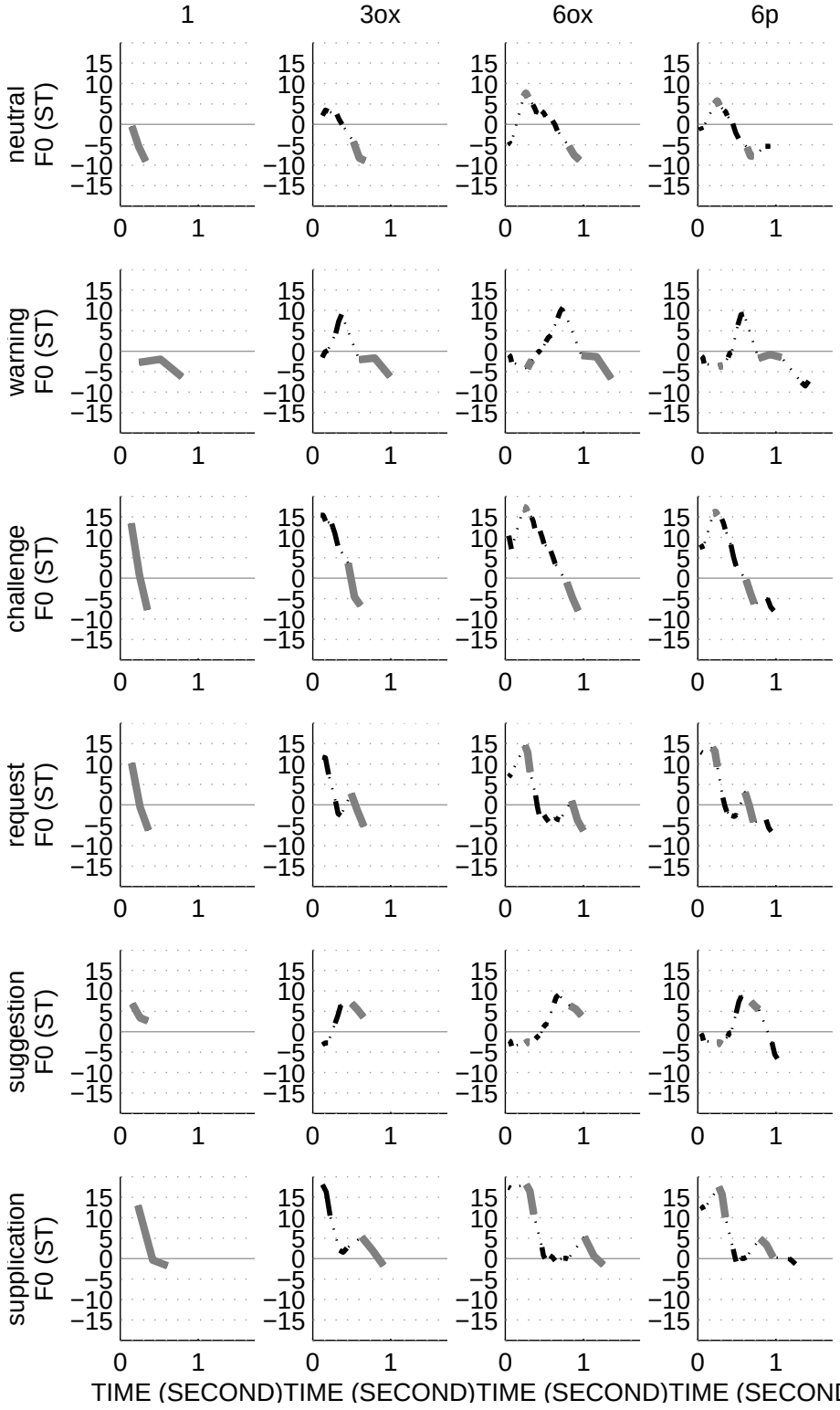


Figure *2.2* – *Tracés de F_0 de six attitudes propositionnelles produites par un locuteur masculin en portugais brésilien, sur des phrases de taille variable (1, 3 et 6 syllabes) et terminées par un oxyton, plus une phrase de 6 syllabes terminée par un paroxyton (voir de Moraes et Rilliard, 2014a).*

ou de la structure morphosyntaxique des phrases (la même phrase est présentée dans chaque colonne). Notons toutefois que ce contour global de phrase respecte la structure accentuelle. Les deux phrases de six syllabes, présentées dans les deux colonnes de droite, se terminent respectivement par un oxyton et un paroxyton (les syllabes accentuées sont représentées sur le graphique par des traits grisés plus larges). On peut remarquer que la position de l'accent final borne le contour attitudinal.

De tels contours intonatifs montrent des différences de durée importantes (différence d'un ordre de grandeur, selon le nombre de syllabes des phrases). Pourtant, cet allongement n'est pas linéaire : certaines parties du contour, qui sont sans doute porteuses de l'information (typiquement les syllabes accentuées dans les exemples de la figure 2.2) montrent des contours ayant des formes proches pour une même attitude, tandis que les autres parties de la phrase sont soumises à un étirement progressif. Cette non-linéarité rend l'usage des stratégies de normalisation de la durée, évoquées précédemment, non souhaitables.

Mais comment procéder pour comparer les variations prosodiques attitudinales de phrases de longueurs potentiellement très différentes, mais transmettant les mêmes indices prosodiques ? Cette question ne connait pas encore, à ma connaissance, de solution satisfaisante. Plusieurs pistes ont été envisagées, mais il s'agit d'abord de mieux comprendre le fonctionnement de ces expressions afin de savoir comment modéliser leur variation qui, de plus, est imbriquée notamment avec le niveau morphosyntaxique de l'expression prosodique. Les chapitres 5 et 6 s'intéressent donc à mieux comprendre et décrire cette variation affective, ses contraintes et ses paramètres, dans le but futur de pouvoir en modéliser la variabilité. Une méthode de mesure non linéaire a cependant été testée afin de tenter de s'abstraire de cette expansion du mouvement prosodique. Cette méthode constitue l'objet de la section suivante.

2.2.3 Alignement non-linéaire de contours prosodiques

La question d'une comparaison objective de la forme de contours prosodiques de différentes durées, mais montrant des similitudes de forme est résumée par les tracés intonatifs de la figure 2.2. Comment décrire des formes pour lesquelles on peut remarquer des similitudes locales, mais entre lesquelles il existe des modifications non-linéaires ? Un problème supplémentaire des métriques décrites auparavant (à l'exception notable des matrices de traits) vient du fait qu'elles ne considèrent généralement que l'influence de la F_0, sans prendre en compte ni l'intensité ni le rythme (notons toutefois que

l'intensité est prise en compte par la métrique proposée par Hermes, 1998b comme un facteur de pondération).

Dans une étude préliminaire (pour des détails, voir Rilliard *et al.*, 2011), une méthode a été testée pour voir si elle permettait une comparaison de ce type de données prosodiques. Pour cela, les vecteurs descripteurs de la prosodie de chaque énoncé sont multiparamétriques et contiennent les valeurs brutes de paramètres de F_0 et d'intensité, des statistiques permettant de décrire l'évolution locale de ces paramètres, ainsi que des informations sur la nature voisée ou non des phonèmes, etc. Il est bien sûr possible d'enrichir encore ces vecteurs prosodiques, avec des données de qualité vocale par exemple. Ensuite, un algorithme de *Dynamic Time Warping* (DTW) a été utilisé pour aligner les formes prosodiques décrites par ces contours multiparamétriques.

Trois mesures objectives ont été testées durant ce travail :

- la mesure de corrélation pondérée proposée par Hermes (1998b), après un alignement linéaire des paires d'énoncés comparés ;
- cette même mesure, mais après l'alignement non linéaire fourni par la DTW (ces deux premières mesures donnent une proximité entre les énoncés) ;
- le coût d'alignement par l'algorithme de DTW des deux énoncés, considéré comme une distance.

Diverses contraintes ont aussi été introduites afin de guider l'algorithme de DTW afin qu'il respecte certaines structures phonétiques ou ne se bloque pas sur un maximum ou un minimum local (Sakoe et Chiba, 1978).

Les résultats comparent ces trois mesures objectives (mesures de similitude ou de divergence, transformées pour exprimer des distances) à la différence perçue entre les attitudes, telle qu'elle a été exprimée lors de tests de reconnaissance de ces stimulus (travail sur la perception abordé notamment à la section 5.3). Pour les deux mesures de corrélation pondérée, les résultats montrent que deux contours prosodiques de la même attitude sont jugés plus proches par ces distances objectives (même s'ils proviennent de phrases de longueurs différentes) que ne le sont deux contours prosodiques provenant de la même phrase, mais porteurs de deux attitudes différentes. Ces mesures vont donc bien dans le sens recherché d'un regroupement des contours prosodiques montrant des formes proches, au-delà des questions de longueur.

Les corrélations pondérées mesurées après le processus d'alignement par DTW sont, bien sûr, plus élevées que celles mesurées après un alignement linéaire ; cependant, l'accroissement de corrélation obtenu est supérieur dans le cas des comparaisons d'attitudes identiques réalisées sur des phrases de tailles différentes, que dans les cas de phrases de tailles identiques véhiculant des attitudes différentes.

Pour observer comment cette mesure objective rassemble les différentes attitudes et voir si cette mesure crée des regroupements comparables à ceux effectués par les auditeurs, les dispersions obtenues à l'aide de ces deux mesures de distance sont soumises à un processus de regroupement hiérarchique. Les dendrogrammes obtenus pour les attitudes du japonais sont présentés à la figure 2.3. Ces arbres montrent des similitudes et des différences entre les regroupements objectifs et subjectifs.

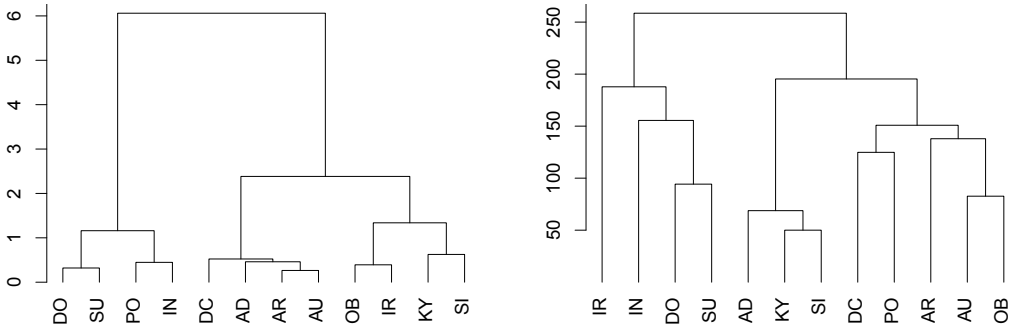


Figure *2.3* – *Dendrogrammes représentant les dispersions entre attitudes obtenues grâce à une mesure objective de corrélation pondérée effectuée après un alongement non linéaire (gauche) et grâce à une mesure perceptive de similarité (voir Rilliard* et al.*, 2011).*

Les trois groupes principaux obtenus grâce à la mesure objective de corrélation après l'alignement par DTW sont (de gauche à droite dans l'arbre) :

1. un groupe d'expressions « dubitatives », plus l'expression de politesse ;
2. un groupe d'expression assertives ;
3. un groupe d'expressions marquées par une qualité vocale particulière.

Ces trois groupes font sens, même s'ils différent des regroupements plus fins effectués par des auditeurs. Ainsi le groupe « dubitatif » (cf. Brandt, 2008, et la section 5.3 pour des détails) regroupe principalement des attitudes caractérisées par un montée finale de F_0. On y retrouve aussi l'attitude de politesse, ce qui peut s'expliquer par sa tendance à être réalisée avec une F_0 moyenne plutôt élevée (cf. Ohala, 1984, et section 5.3). Le troisième groupe rassemble les expressions marquées par des qualités vocales non modales, pour lesquelles on observe une tendance au dévoisement. Comme les indices liés à la qualité vocale contenus dans les vecteurs prosodiques sont limités à ces informations de dévoisement, il est logique que la mesure n'ait pas été capable de différencier plus dans le détail ces expressions. Enfin, le groupe assertif regroupe l'ensemble des autres expressions, qui montrent une intonation descendante.

Les regroupements effectués par cette mesure objective sont intéressants et leur interprétation à la lumière des résultats perceptifs détaillés (résultats présentés aux chapitres 5 et 6) permet de comprendre les limitations de cette approche. Nous y reviendrons dans nos perspectives.

Nous allons maintenant aborder la description de différentes sources de variation de la prosodie ; ces travaux sont destinés d'abord à mieux décrire, pour chacune de ces fonctions, les raisons de la variation prosodique ainsi que les dimensions de cette variation.

3 | Fonction de démarcation prosodique

> La mesure se rapporte au geste, le chant ou la mélodie à l'arrangement des mots, les sons ou la cadence aux inflexions de la voix, qui varient à l'infini dans le discours.
>
> Quintilien, *Institution Oratoire*, Livre I, chapitre XI (trad. C. V. Ouizille, 1829)

Il s'agira donc de s'intéresser à la mesure des différences perceptives induites dans le matériau prosodique par des variations de la structure morphosyntaxique d'un énoncé. Autrement dit, dans quelle mesure la prosodie, seule, permet de percevoir la localisation syntagmatique d'une frontière ? Permet-elle de même de percevoir une différence d'ordre paradigmatique ? Dans le but de répondre à ces questions, une étude basée sur de la parole délexicalisée a testé la capacité d'auditeurs à juger de l'adéquation de variations prosodiques à des structures syntaxiques (Rilliard et Aubergé, 2003).

3.1 Méthodologie expérimentale

Ce travail vise à évaluer l'apport de la prosodie, seule. Pour cela, différentes techniques de délexicalisation ont été proposées dans la littérature (pour plus de détails, voir Rilliard, 2000). Certaines utilisent des outils de synthèse pour créer de la parole inintelligible, mais respectant les structures phonotactiques des phrases originales (voir le paradigme de *saltanaj* proposé par Ramus et Mehler, 1999). D'autres (e.g. Sonntag et Portele, 1998) utilisent un filtrage des énoncés de parole originaux pour enlever tout accès aux informations non prosodiques.

Parce que ce travail s'intéresse à la parole naturelle, une technique de délexicalisation basée sur un processus de réitération de parole a été préférée. Ce paradigme assure une transmission efficace des informations prosodiques (cf. Larkey, 1983), en normalisant d'éventuels effets microprosodiques et phonotactiques (Kelso *et al.*, 1985) qui ne nous intéressent pas ici (voir Rossi *et al.*,

1981, p. 20 sqq.). Par ailleurs la réitération conserve une structure syllabique claire et donc toutes les informations rythmiques qui lui sont liées.

Un corpus de phrases lues a donc été enregistré par deux locuteurs (une femme et un homme) qui répétaient une phrase qui leur était présentée à travers un casque, avec comme consigne de conserver la même prosodie que l'original[1]. Les phrases sont dans un premier temps répétées normalement, puis reproduites en remplaçant chaque syllabe par un /ma/, obtenant ainsi les énoncés réitérés délexicalisés. Les phrases constituant ce corpus sont présentées dans le tableau 3.1. Elles comportent de 5 à 11 syllabes. À l'intérieur de chaque groupe de phrases du même nombre de syllabes, on observe des oppositions systématiques de certaines caractéristiques de leurs structures syntaxiques. Ces oppositions concernent principalement des déplacements syntagmatiques de la position de la frontière syntaxique principale, mais aussi par exemple des inversions nom/adjectif. Certains groupes présentent aussi des différences syntaxiques sur le plan paradigmatique. Pour un groupe de longueur de phrase donné, chaque paire de phrases présente donc des oppositions syntaxiques particulières. C'est l'importance perceptive des corrélats prosodiques associés à ces oppositions dont on va chercher à étudier l'importance.

Pour cela, toutes les paires de phrases possibles pour chaque groupe de longueur sont constituées et donnent un total de 72 oppositions syntaxiques. Ces oppositions peuvent être regroupées selon six catégories différentes :

(a) paires de phrases de structures syntaxiques identiques (paires homogènes) ;

(b) paires de phrases présentant une frontière syntaxique principale à la même position et composées de groupes de même niveau et de même nature (paires de même niveau / même nature) ;

(c) paires de phrases présentant une frontière syntaxique principale à la même position et composées de groupes de même niveau mais de natures différentes (paires de même niveau / natures différentes) ;

(d) paires de phrases présentant une frontière syntaxique principale à la même position et composées de groupes de différents niveaux syntaxiques (paires de différents niveaux) ;

(e) paires de phrases présentant des frontières syntaxiques décalées et composées de groupes de même niveau syntaxique (paires de même niveau avec décalage) ;

[1] le corpus original est tiré des travaux d'Aubergé (1992)

#	**Phrase**	**Structure syntaxique**
	5 syllabes	
1	Ce passant chantait.	$GN_3 \mid V_2$
2	Ce pas sera fait.	$GN_2 \mid V_3$
	6 syllabes	
3	Ce beau passant chantait.	$GN_4 \mid V_2$
4	Ce passant fou chantait.	inversion A_1/N_2
5	On entendait des pas.	$S1\ V_3 \mid GN_2$
	7 syllabes	
6	Ce petit passant chantait.	$GN_5 \mid V_2$
7	Ce passant tout fou chantait.	inversion A_2/N_2
8	Son pas doux retentissait.	$GN_3 \mid V_4$
	8 syllabes	
9	Tu dis que ce passant chantait.	$P_2 \mid P_6$
10	Ce passant chantait l'opéra.	$GN_3 \mid V_2\ GN_3$
11	Je verrai si les enfants jouent.	$P_3 \mid P_5$
	9 syllabes	
12	Ce passant chantait tous les six mois.	$GN_3\ V_2 \mid O_4$
13	Ce passant chantait, Toto dansait.	$P_5 \mid P_4$
14	On entendait plus ou moins les pas.	$S_1\ V_3 \mid A_3\ GN_2$
	10 syllabes	
15	L'enfant pleurait quand il était malade.	$P_4 \mid P_6$
16	Quand l'enfant pleurait, il était malade.	$P_5 \mid P_5$
17	Quand il pleurait, l'enfant était malade.	$P_4 \mid P_6$
18	Ce passant chantait quand Toto dansait.	$P_5 \mid P_5$
	11 syllabes	
19	Je vois le marchand de poissons de Paris.	$S_1\ V_1 \mid GN_9$
20	Ce passant chantait parce qu'il était content.	$P_5 \mid P_6$
21	Je mangeais du vin, du Boursin, et du pain.	$E_{3+2+3+3}$
22	Même si les enfants jouent, je verrai le chat.	$P_6 \mid P_5$

Table *3.1* – *Les 22 phrases composant le corpus de parole réitérée, et leur structure syntaxique avec la taille des groupes, qui permet de dériver les oppositions syntagmatiques. Les abréviations suivantes sont utilisées : A pour Adjectif, E pour Énumération, GN pour Groupe Nominal, N pour Nom, O pour Objet, P pour Proposition, S pour Sujet, V pour Verbe. Les numéros en indice indiquent la taille, en nombre de syllabes, de ces structures syntaxiques. Les frontières syntaxiques principales sont marquées par le symbole |.*

5 syll.	*1*	*2*
1	a	e
2	e	a

6 syll.	*3*	*4*	*5*
3	a	b	c
4	b	a	c
5	c	c	a

7 syll	*6*	*7*	*8*
6	a	b	e
7	b	a	e
8	e	e	a

8 syll.	*9*	*10*	*11*
9	a	f	e
10	f	a	d
11	e	d	a

9 syll.	*12*	*13*	*14*
12	a	d	e
13	d	a	f
14	e	f	a

10 syll.	*15*	*16*	*17*	*18*
15	a	e	c	e
16	e	a	e	c
17	c	e	a	e
18	e	c	e	a

11 syll.	*19*	*20*	*21*	*22*
19	a	f	e	f
20	f	a	f	e
21	e	f	a	f
22	f	e	f	a

Table *3.2* – *Les tables présentent le type d'opposition syntaxique de chaque paire de phrases, pour chaque groupe de longueur de phrase (indiqué en haut à gauche de chaque table). Les lettres dans les tables renvoient au type d'opposition (voir texte), les numéros indiquent les phrases telles que référencées à la table 3.1.*

(f) paires de phrases présentant des frontières syntaxiques décalées et composées de groupes de différents niveaux syntaxiques (paires de différents niveaux avec décalage).

On postulera pour la suite que chacune de ces six catégories regroupe des paires de phrases qui présentent des différences syntaxiques comparables d'un point de vue prosodique. Ces différences sont décrites ci-dessus par ordre croissant de divergence syntaxique. Les divergences perçues entre les paires de ces catégories permettront donc d'avoir une évaluation de la pertinence de la prosodie pour transmettre ce type de différences entre structures syntaxiques. La table 3.2 résume pour chaque paire de phrases les oppositions syntaxiques qui sont présentées aux auditeurs.

Les paires de phrases comportant ces oppositions syntaxiques sont présentées aux auditeurs de la manière suivante. La première partie de la paire est présentée à l'écrit, ou à l'écrit et dans sa version lexicalisée ; la seconde partie est présentée dans sa version délexicalisée seule. Les auditeurs jugent donc si la phrase en /mamama/ pourrait être une réitération valide de la phrase qui leur est présentée sous sa forme lexicalisée. Toutes les paires sont présentées dans les deux sens. Les deux conditions (C1 et C2) de présentation (comportant une première partie présentée soit à l'écrit seul – C1 – soit à l'écrit et à l'oral – C2) permettent de tester si une comparaison auditive directe des deux stimulus (délexicalisé et lexicalisé) rend la tâche de comparaison plus performante. En effet, la tâche demandée aux sujets nécessite une analyse métalinguistique de la structure des énoncés qui leur sont fournis

et elle pourrait s'avérer irréalisable sur la base des seules indications transmises par les stimulus réitérés. La seconde condition permet donc de valider la validité des réponses obtenues lors de la première condition.

3.2 Analyse des résultats

La tâche demandée aux sujets revient à un jugement d'association binaire entre une structure syntaxique et une structure prosodique. Les scores d'association obtenus pour chaque paire sont regroupés en fonction de la catégorie d'opposition syntaxique représentée par la paire considérée (voir 3.2). Les scores moyens obtenus pour chaque catégorie de comparaison prosodie / syntaxe sont résumés à la figure 3.1. On observe un effet significatif (d'après une régression logistique menée sur le score d'association en fonction de la catégorie des paires) de la catégorie d'opposition prosodie / syntaxe, avec un accroissement significatif des jugements de dissociation entre les deux parties des paires, en fonction de l'accroissement des divergences présentées par les deux structures.

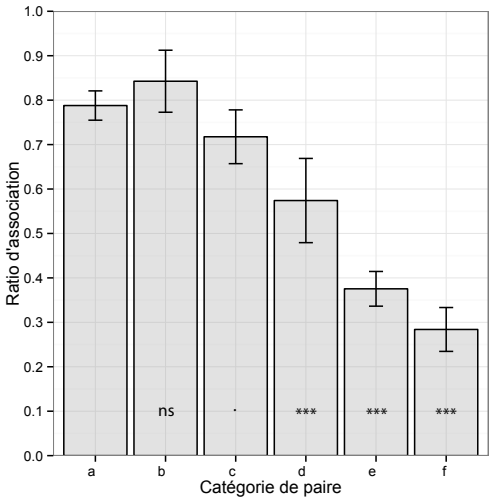


Figure ***3.1*** *– Ratio d'association entre les deux stimulus de chaque catégorie de paire. Les barres d'erreur indiquent l'intervalle de confiance à 5%. Les symboles « ns », « . » et « *** » indiquent une différence par rapport aux paires de catégorie « a » dont la significativité est, respectivement : non significative, <0.05 et <0.0001.*

Les paires de la catégorie « b », composées de groupes syntaxiques de même nature ayant leur frontière syntaxique principale à la même position, ne sont pas distinguées prosodiquement. Ce résultat valide donc perceptivement

que la fonction de segmentation opérée par la prosodie seule fonctionne bien, indépendamment du contenu lexical des énoncés – résultat attendu au vu de la littérature, mais dont la démonstration expérimentale restait à faire.

En ce qui concerne les autres oppositions syntaxiques testées durant cette expérience, elles montrent un degré d'acceptabilité significativement réduit – allant jusqu'à une claire dissociation entre structures prosodiques et syntaxiques. Ainsi, une différence de nature hiérarchique entre les groupes syntaxiques des deux phrases d'une paire (cas « c ») induit déjà des scores d'association significativement moindres que ceux de la référence – bien qu'il ne s'agisse pas pour autant d'un jugement de dissociation. La prosodie véhicule donc bien des indices perceptibles de la structure paradigmatique de la phrase. Toutefois, ces indices ne permettent pas de rejeter une réalisation prosodique. Les sujets perçoivent toutefois une inadéquation entre prosodie et structure syntaxique.

Par contre, à partir du moment où la position de la frontière syntagmatique ne correspond pas dans les deux membres de la paire proposée aux sujets, on observe un rejet de la structure prosodique proposée par rapport à la syntaxe. Ainsi, dans les cas « d » et « e », la dissociation des structures prosodiques et syntaxiques est claire. L'information primordiale qu'un auditeur est à même d'utiliser dans une structure prosodique concerne la *localisation* des frontières – la fonction de démarcation décrite par exemple par Vaissière (1997) constitue bien une réalité perceptive de premier plan.

3.3 Mesure objective des variations

Ces résultats perceptifs montrent que les variations prosodiques d'un énoncé sont à même de transmettre, seules, une information pertinente sur la structuration morphosyntaxique de cet énoncé. Ce résultat montre aussi que des sujets sont à même de réaliser un jugement perceptif pertinent linguistiquement sur la seule base d'informations prosodiques. Dans l'optique du modèle de travail décrit au chapitre précédent, j'ai tenté de relier ces résultats perceptifs à des critères objectifs. Pour cela, des mesures de distance prosodique objectives ont été appliquées aux paires de stimulus jugées par les auditeurs.

Hermes (1998b) montre que deux mesures, parmi celles qu'il évalue, se rapprochent des jugements perceptifs : une mesure de corrélation (cf. section 2.2.1 pour des détails) et une distance des moindres carrés.

Afin de juger de l'importance relative des différents paramètres de la prosodie, la corrélation et la somme des moindres carrés observés entre chacune

des paires de phrases décrite ci-dessus est calculée pour les paramètres suivants :

- les valeurs de F_0 ;
- la durée syllabique ;
- la durée des groupes inter-centres-perceptifs[1] ;
- l'intensité.

Les résultats montrent que la distance de corrélation est proche des données perceptives que la distance des moindres carrés, ce qui vient confirmer les conclusions de Hermes (1998b). En ce qui concerne les différents paramètres acoustiques pris en compte, le paramètre de F_0 est le paramètre prosodique pour lequel la mesure de distance objective montre la meilleure corrélation avec les distances perceptives. Ce résultat confirme les données de la littérature sur l'importance de la fréquence fondamentale dans la réalisation de cette fonction de segmentation en français (Di Cristo, 2013). Le paramètre de durée reçoit aussi de bonnes associations avec les données de perception, particulièrement lorsque la durée est mesurée à partir des centres perceptifs, ce qui conforte les données de Barbosa (1994). L'intensité par contre semble jouer un rôle mineur dans la réalisation de ces fonctions prosodiques en français.

Ce travail est mené sur de la parole de laboratoire conçue pour contraindre la variabilité prosodique. Les enregistrements sont menés de façon à ce que les productions se rapprochent d'un modèle standard, en demandant aux locuteurs de reproduire une phrase montrée en exemple. Si cela ne retire rien aux conclusions concernant les capacités perceptives des auditeurs à utiliser ces variations prosodiques pour effectuer un traitement linguistique, cette expérience ne nous apprend rien non plus sur la variation possible, acceptable, de ces fonctions de segmentation et de hiérarchisation prosodiques. Dans quelle mesure deux structures prosodiques, variant par exemple du fait de l'origine géographique des locuteurs, seront-t-elles perçues par des auditeurs comme également adéquates pour réaliser une même segmentation morphosyntaxique ?

[1] Sur cette notion de centre perceptif (en anglais *P-Center*, voir Marcus, 1981) comme autre unité possible pour la mesure du rythme, voir la « *V-to-V unit* » de Barbosa (2007).

4 | Variation diatopique

> Tout le plaisir a été pour moi, j'y rétorque dans un français se la jouant entre l'inflexion dauphinoise et l'intonation parisienne.
>
> Dard, *Ceci est bien une pipe*, première partie, chapitre 1 (1999).

4.1 Étude de la prosodie dialectale

4.1.1 Abrégé de méthodologie AMPER

Répondre à cette question (dans quelle mesure la variation diatopique change la perception de la fonction démarcative ?) est loin d'être évident. En effet, comme le constatent Contini et Profili (1989) dès les années 1980, la plupart des travaux menés en prosodie utilisent des données issues des variétés « standard » des langues nationales. Par ailleurs, les rares données disponibles et faisant état d'une variation géolinguistique systématique, ne sont pas strictement comparables pour un grand nombre de raisons, au premier plan desquelles une variation non contrôlée des structures syntaxiques des énoncés enregistrés. Ceci rend délicat de comparer les évolutions prosodiques, dans leur fonction de segmentation et de hiérarchisation des énoncés. En réponse à ces problèmes va naître, au centre de dialectologie de Grenoble, le projet d'Atlas Multimédia Prosodique de l'Espace Roman (AMPER) (Contini, 1991; Contini *et al.*, 2002), qui se propose la collecte de données de prosodie dialectale comparables sur l'ensemble de l'espace roman (européen et latino-américain).

Pour cela, le principe de base du projet, tel que décrit par Contini (1991), consiste à se servir de phrases ayant la même structure syntaxique, comme support de l'observation de ces variations prosodiques. Le choix s'est porté sur des phrases ayant une structure sujet—verbe—objet, avec le sujet et

Code	**Groupe nominal**		
	Article	*Nom*	*Extension*
k	1	oxyton	-
t	1	paroxyton	-
p	1	préparoxyton	-
g	1	oxyton	oxyton
d	1	paroxyton	oxyton
b	1	préparoxyton	oxyton
x	1	oxyton	paroxyton
s	1	paroxyton	paroxyton
f	1	préparoxyton	paroxyton
j	1	oxyton	préparoxyton
z	1	paroxyton	préparoxyton
v	1	préparoxyton	préparoxyton

Table *4.1* – *Structure accentuelle des groupes nominaux formant les phrases sujet—verbe—objet du corpus AMPER. Chaque groupe (sujet ou objet) est basé sur des substantifs trisyllabiques accentués sur la dernière, la pénultième ou l'antépénultième syllabe. Un article monosyllabique initie le groupe. Les phrases sont codées selon le type des structures accentuelles des groupes nominaux qui les composent.*

l'objet composés de groupes nominaux comportant ou non une extension adjectivale. Les substantifs et les adjectifs qui composent ces groupes nominaux comportent trois syllabes, dans le but d'étudier l'effet d'une variation de la position accentuelle sur la réalisation prosodique de la phrase[1]. La table 4.1 présente la variation systématique des positions accentuelles des mots pleins utilisés dans chaque groupe, ainsi que les codes utilisés pour chaque combinaison. Ce codage, mis en place dans le cadre du projet AMPER, sert à retrouver les phrases comparables d'un point de vue morphosyntaxique : deux phrases portant le même code ne différent prosodiquement que par les stratégies dialectales des locuteurs (en théorie). Par exemple, les phrases « *(La pàpera)*$_p$ *tocca (la patata càrica)*$_z$. » et « *(O pássaro)*$_p$ *toca (no Toneca cómico)*$_z$. »[2] sont extraites respectivement d'enquêtes menées en Italie[3] et au Portugal[4].

[1] Les enquêteurs peuvent bien entendu étendre à volonté les structures étudiées, l'essentiel étant de conserver un noyau commun.

[2] Les lettres en indices indiquent le code des groupes nominaux, tels que décrits au tableau 4.1.

[3] Point d'enquête *062*, parler du Salento, dialecte de la variété « Meridional estremo » de l'aire italo-romane (Romano, 1999).

[4] Point d'enquête *013*, parler d'Aveiro, dialecte de la région du Beira-Litoral – variété de portugais européen continental, dans l'aire ibéro-romane (de Castro Moutinho *et al.*, 2005).

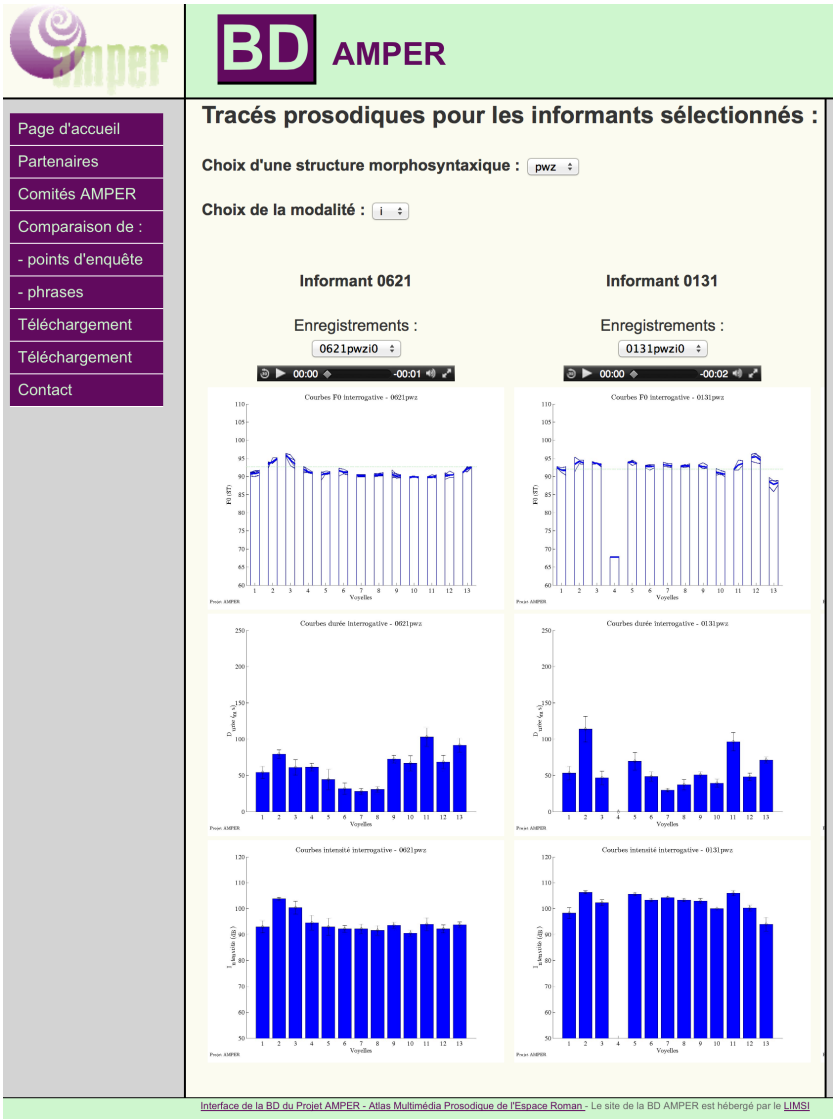


Figure 4.1 – *Exemple de comparaison des tracés prosodiques tels qu'ils sont obtenus grâce à l'interface web de la base de données AMPER. Les tracés prosodiques (de haut en bas pour la F_0, la durée et l'intensité) d'une même structure morphosyntaxique sont affichés pour un point d'enquête dans chaque colonne. Il est possible d'écouter le son et de modifier le mode de la phrase ou la structure affichée.*

Si leur contenu lexical et sémantique varie, ces phrases sont toutes les deux basées sur la même structure et les différences prosodiques observées dans ces deux cas (voir figure 4.1) sont donc liées uniquement à l'origine dialectale des locuteurs – au-delà de leurs caractéristiques idiosyncratiques. Nous verrons

que cette structure syntaxique de base, théoriquement possible au travers de tout l'espace roman, pose tout de même un certain nombre de problèmes, auxquels nous tenterons de proposer des solutions.

Chaque phrase est enregistrée selon les deux modes[1] assertif et interrogatif (question totale pouvant entrainer une réponse en oui/non). Plusieurs répétitions (trois au minimum) de chaque phrase sont demandées aux locuteurs afin d'assurer la prototypicité des variations prosodiques observées. Deux locuteurs au minimum (une femme et un homme) sont enregistrés par point d'enquête. Ces locuteurs sont sélectionnés par les enquêteurs pour la typicité de leur parler et – si possible – l'absence d'influences externes sur ce parler.

4.1.2 Rôle dans le projet

Ces phrases sont enregistrées sur le terrain, auprès d'informants de chaque variété dialectale étudiée, réparties dans toutes les aires dialectales de l'espace roman : gallo-roman, ibéro-roman, italo-roman, rhéto-roman et roumain. Ces collectes de corpus sont donc réalisées par des scientifiques de nombreuses universités[2], qui les analysent à l'aide d'outils communs et suivant une procédure de stylisation prosodique se focalisant sur les voyelles. Les données sont ensuite transmises à la coordination du projet, qui les recueille et met à disposition des outils pour leur étude et leur comparaison.

Mon rôle dans ce projet est multiple. Il consiste d'une part à mettre à disposition des membres du projet des outils leur permettant d'analyser, de recueillir et de partager ou d'interroger les données. Ces outils doivent notamment permettre la création de données qui suivent exactement les principes d'analyse préconisés par le projet. J'ai ainsi créé une interface[3] permettant de générer des représentations graphiques à partir des analyses prosodiques d'un ensemble de phrases. Ces graphiques permettent une superposition des tracés des différents paramètres acoustiques de la prosodie (F_0, durée, intensité) observés pour deux structures morphosyntaxiques, ou pour deux modes donnés. Afin d'extraire ces paramètres acoustiques, une routine[4] Praat per-

[1] Certains travaux présentés ici utilisent les modalités auditive et visuelle ; afin d'éviter les confusions entre les termes désignant ces modalités audiovisuelles et les modes illocutoires de la phrase (assertif, interrogatif, etc.), le terme de « modalité » sera réservé au premier sens, tandis que le terme « mode » sera utilisé pour désigner le mode illocutoire des phrases.

[2] Pour des détails, voir `http://amper.limsi.fr/index.php/partenaires`

[3] `http://groupeaa.limsi.fr/membres:rilliard:outils_amper#interface_de_calcul_des_courbes_prosodiques_du_projet_amper`

[4] `http://groupeaa.limsi.fr/membres:rilliard:outils_amper#script_praat_pour_amper`

met aux enquêteurs d'effectuer des analyses de leurs enregistrements et de réaliser des corrections manuelles sur les valeurs de pitch détectées et stylisées automatiquement sur la base d'une segmentation en voyelles. En effet, ces valeurs sont l'objet d'erreurs notamment du fait des conditions d'enregistrement de terrain, rarement optimales. Ces données prosodiques issues de l'analyse se présentent sous la forme de données numériques (par exemple des valeurs de F_0) et de graphiques. Elles sont liées aux types de structures phrastiques, à leurs modes et aux origines dialectales des locuteurs. Pour permettre de stocker toutes ces informations et d'avoir un outil performant permettant de les trier dans le but de les comparer efficacement, une base de données a été créée. Cette base, qui regroupe actuellement plus d'une centaine de locuteurs, est décrite par Rilliard (2011). Une interface web à cette base de données a été mise en place[1], qui permet aux membres du projet AMPER d'interroger simplement la base de données et d'obtenir des données résumant les variations qui les intéressent. La figure 4.1 présente ainsi les tracés prosodiques de deux phrases des points d'enquête d'Italie et du Portugal dont il a été question ci-dessus. Les variations prosodiques affichées ici correspondent aux phrases de structure t_z[2] de ces points d'enquête, dans leur mode interrogatif. On peut observer une différence de stratégie dans la réalisation des contours intonatifs de l'interrogation pour ces deux locuteurs : la locutrice italienne réalise une montée sur la syllabe finale de l'énoncé, tandis que la locutrice de cette région du Portugal réalise cette montée finale (montée typique de l'intonation interrogative, voir Ohala, 1983; Gussenhoven, 2004) sur la dernière syllabe accentuée (ici la syllabe pénultième), suivie d'une intonation plus basse sur la voyelle finale atone.

Enfin, j'ai contribué à faire avancer la réflexion du projet AMPER autour de la mesure objective de cette variation prosodique diatopique. Sur la base de travaux antérieurs, des propositions de mesures objectives ont été faites, et appliquées au cas des enquêtes du Portugal continental, pour montrer la faisabilité d'une géoprosodie romane grâce aux enquêtes du projet (de Castro Moutinho *et al.*, 2011).

4.2 Objectivation de la variation prosodique

La problématique du projet AMPER consiste dans un premier temps à récolter des données comparables sur les variations prosodiques des dialectes romans. Ensuite, il s'agit de décrire les phénomènes prosodiques observés (voir par exemple : Lai, 2005; Dorta, 2007; Turculeţ, 2008). Ces descriptions

[1] `http://amper.limsi.fr/`

[2] Codes accentuels des groupes sujet et objet, tels que décrits à la table 4.1

permettent la documentation des particularités prosodiques des différents parlers, tant en terme de réalisation accentuelle (Romano, 1999) que de distinction entre modes (Fernández Rei, 2011). Des tests de perception sont aussi menés afin de mesurer la capacité de ces indices prosodiques à permettre une détermination de l'origine dialectale des locuteurs (Fernández Rei, 2013; Alvarellos *et al.*, 2011). Tous ces travaux sont extrêmement consommateurs en temps. Aussi, la possibilité de réaliser des mesures objectives à partir de ces données est-elle un enjeu de taille pour l'avancement du projet.

Comme on l'a vu dans les deux chapitres précédents, cette mesure objective de la variation prosodique perçue constitue une gageure, différentes variations fonctionnelles de la prosodie pouvant s'accompagner de variations acoustiques similaires. Cela entraîne des difficultés d'interprétation de certaines variations – difficultés qui resteront, dans le cadre d'AMPER, du ressort des experts de chaque dialecte. Pour s'abstraire du recours à un expert, il faudrait apprendre des contraintes contextuelles multiples sur des corpus de grande taille et correspondant à de véritables situations de communication, afin d'extraire de ces mesures de distance prosodique une fonction communicative précise (voir par exemple les travaux de Klabbers et van Santen, 2004, , qui ne traitent cependant pas la possible variabilité fonctionnelle d'un même contour, mais constituent une approche intéressante).

D'autres problèmes se posent et en particulier celui de la variation des phrases cibles du corpus AMPER. La théorie veut que chaque phrase d'une structure morphosyntaxique donnée soit représentée dans tous les dialectes de l'espace roman (à condition bien sûr que la structure accentuelle considérée existe dans le parler) et présente un nombre identique de syllabes. La pratique montre de nombreuses sources de variation du nombre de syllabes pour une structure de phrase donnée. En premier lieu, les chutes de voyelles sont très fréquentes, notamment en portugais, mais aussi dans de nombreux autres dialectes[1]. Ensuite, dans l'aire dialectale roumaine, les articles notamment ont un fonctionnement différent de ce que l'on observe dans le reste des langues romanes (articles enclitiques), ce qui compromet la position des accents dans la structure initialement prévue. D'autres spécificités, comme par exemple l'observation d'assimilations dans des variétés d'italien méridional, viennent modifier la structure des phrases enregistrées – ou tout du moins leur syllabation.

Il a donc fallu réfléchir à des méthodes permettant de contourner au mieux ces problèmes, afin d'arriver à mesurer des différences qui soient effectivement structurées par la position des syllabes accentuées dans les phrases observées

[1]Ce phénomène en lui même mériterait sans doute d'être approfondi. On note par exemple de nombreuses chutes en portugais brésilien, ce à quoi les enquêteurs ne s'attendaient pas.

pour chacun des parlers. Un code spécifiant le nombre de syllabes de chaque mot permet de pallier ces difficultés et d'aligner les phrases comparables en fonction de leur structure accentuelle. Il s'agira ensuite de traiter les « trous » laissés par les syllabes manquantes : voir par exemple la quatrième syllabe de la phrase du dialecte portugais présenté à la figure 4.1, un parti-pris graphique consiste à représenter les voyelles tombées avec une fréquence fondamentale (F_0) à 50Hz. Les solutions proposées à ce problème vont dépendre de la distance objective mise en place.

4.2.1 Mesures de distances

Les propositions pour objectiver la variation prosodique présentées à la section 2.1 proviennent du domaine du traitement automatique de la parole. Les phonéticiens ont ainsi tenté de mettre en place des matrices de traits prosodiques (Martin, 1973, 1987) afin de décrire précisément les variations prosodiques pertinentes. Sur cette idée, Contini et Profili (1989) proposent d'utiliser une série élargie de traits afin de réaliser des comparaisons objectives de variations prosodiques dialectales. On a vu que la mesure de corrélation pondérée proposée par Hermes (1995, 1998b,a) reflète le mieux la perception des divergences entre deux continuums intonatifs, parmi l'ensemble des mesures testées – et ce résultat a été reproduit sur les données présentées dans le chapitre 3. La proximité de cette mesure de corrélation (sous une forme simplifiée) avec les distances perçues subjectivement est notamment validée pour évaluer des variations intonatives générées par synthèse (Hirst *et al.*, 1998).

Cette mesure de corrélation entre courbes prosodiques a été utilisée (de Castro Moutinho *et al.*, 2011) afin de tenter une première cartographie de la variation prosodique au sein d'un espace dialectal (de Castro Moutinho *et al.*, 2011). Ces divergences ont été mesurées entre chaque paire de phrases comparables (ayant la même structure morphosyntaxique et le même mode de phrase) enregistrées par des locuteurs de points d'enquêtes répartis dans l'espace dialectal du portugais européen continental. Deux locuteurs représentent chaque point d'enquête, afin de stabiliser la représentation du parler local. Les divergences prosodiques entre deux points d'enquêtes sont estimées d'après la moyenne de toutes les divergences entre paires de phrases. En prenant comme point de référence un point d'enquête donné, il est possible de construire une représentation cartographique de la variation prosodique sur l'espace dialectal considéré, en s'inspirant des travaux de Goebl (1981, 1996) dans le domaine de la représentation de la variation dialectale, aux niveaux lexical et phonologique, pour l'appliquer à la prosodie. On obtient ainsi une

première représentation géoprosodique, pour l'espace du portugais européen continental (voir la figure 4.2).

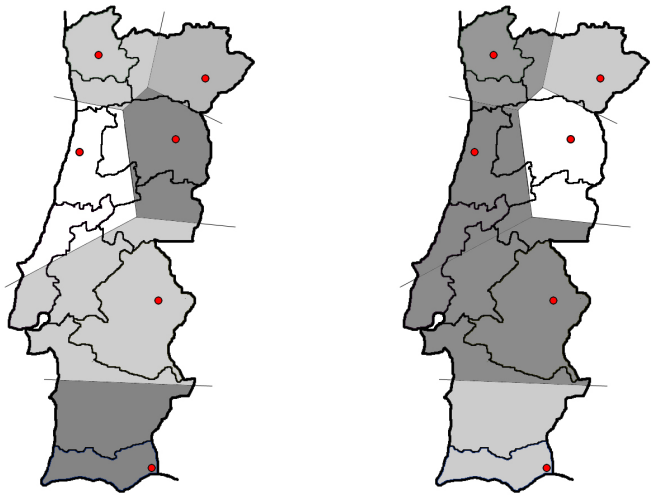


Figure 4.2 – *Représentation géoprosodique de la distance objective mesurée entre le point d'enquête situé au centre de l'aire blanche et chacun des autres points ; le niveau de gris indique la distance prosodique. La figure de gauche représente les distances par rapport au point d'enquête du Beira littoral. La figure de droite représente les distances par rapport au point d'enquête du Beira alta.*

4.2.2 Utilisations possibles de ces mesures objectives

Les distances prosodiques représentées à la figure 4.2 agglomèrent l'ensemble de la variation observée entre les deux locuteurs de chaque point d'enquête considéré. Le point intéressant est que les distances prosodiques obtenues entre dialectes se rapprochent de ce qui avait été observé d'un point de vue perceptif sur ces même données (de Castro Moutinho *et al.*, 2005, 2008). Les dialectes montrant les variations prosodiques perceptivement les plus proches ne sont pas nécessairement les plus proches géographiquement. Ainsi les deux points de référence des cartes de la figure 4.2, correspondant à des régions proches, montrent des contours prosodiques éloignés perceptivement, et aussi selon la mesure de corrélation utilisée ici.

Une telle mesure pourrait être utilisée dans le cadre d'études beaucoup plus fines que celle-ci et en particulier dans le but de mesurer des variations sur certains aspects bien déterminés. On pensera par exemple aux stratégies de réalisation du mode interrogatif. L'un des intérêts du projet AMPER, au-delà de sa fonction descriptive, est d'avoir souligné la variabilité des réali-

sations des contours intonatifs pour les phrases interrogatives (yes/no questions). S'il est généralement postulé que ce mode est marqué par une montée finale (Ohala, 1983), les travaux menés dans le domaine galicien ont notamment montré que ce mode pouvait porter l'essentiel de la variation dialectale observée. Il pourrait donc être souhaitable de focaliser ces mesures de divergence sur des sous-ensembles des corpus enregistrés (dans le cas du galicien, les phrases interrogatives). Ces mesures peuvent bien sûr aussi servir à objectiver des différences plus fines, comme les montées intonatives antérieures à la syllabe accentuée décrites par exemple par Lai (2002); Romano (1999); Romano et Interlandi (2005).

4.2.3 Mesures concurrentes & fiabilité

Une question complexe est celle de la validité de telles mesures objectives en comparaison de la perception qu'en ont les locuteurs de ces parlers. Nous ne pourrons répondre ici plus avant à cette question, sauf à dire comme précédemment que la mesure que nous montrons donne des réponses similaires à une même question posée à des auditeurs. Il est par contre possible de poser la question de la pertinence de cette mesure de corrélation, au regard d'autres mesures objectives possibles. La littérature propose d'autres solutions qui ont été évoquées précédemment. Afin de mieux se rendre compte de la spécificité d'une telle mesure objective, un test a été mené en implémentant une autre mesure objective (présentée aux Xornadas de Dialectoloxía Perceptiva tenues à l'Université de Santiago de Compostela), radicalement différente dans sa philosophie de la mesure de corrélation proposée par Hermes (1995, 1998b). Il s'agit d'estimer des différences entre deux réalisations prosodiques sur la base d'une matrice de traits prosodiques. Cette proposition se base sur les travaux de Martin (1973, 1987) et reprend le détail de la proposition de Contini et Profili (1989), en l'adaptant à une analyse complètement automatique du corpus AMPER.

Matrices de traits

Contini et Profili (1989) proposent 13 traits décrivant la variation prosodique de chaque syllabe. Il peut s'agir de traits locaux décrivant la forme du contour intonatif (par exemple montant, descendant), ou de traits indiquant les variations les plus importantes sur un ensemble de syllabes (par exemple : syllabe la plus haute de la phrase ou du groupe syntaxique). Les indices de forme sont déterminés chez Contini et Profili (1989) sur la base d'une stylisation de la courbe de F_0 en trois points reliés par des segments de droite. Tous les traits prosodiques proposés par les auteurs prennent des valeurs bi-

naires (+/− montant et +/− descendant, par exemple), ce qui amène un codage complexe si l'on cherche à décrire par exemple des formes intonatives circonflexes.

Afin de rationnaliser la création de ces matrices de traits, les onze traits suivants ont été définis :

1. Forme du contour de F_0 de la voyelle ;
2. Vitesse des mouvements du contour de F_0 ;
3. Amplitude des mouvements de F_0 ;
4. Position de la voyelle par rapport au F_0 moyen du locuteur ;
5. Voyelle culminante (F_0) du mot prosodique (WPF) ;
6. Voyelle la plus longue du mot prosodique (WPF) ;
7. Voyelle la plus proéminente du mot prosodique (WPF) ;
8. Voyelle extrême (F_0) de la phrase (SPF) ;
9. Voyelle la plus allongée de la phrase (SPF) ;
10. Voyelle la plus proéminente de la phrase (SPF) ;
11. Pente de la phrase.

Ces traits sont calculés automatiquement à partir d'une stylisation de la courbe intonative réalisée grâce au modèle de perception tonale (d'Alessandro et Mertens, 1995) implémenté dans le logiciel Praat – le Prosogram (Mertens, 2004). Les trois premiers traits définissent la forme, potentiellement complexe, du contour intonatif de la voyelle. Comme le contour peut se composer de plusieurs (et potentiellement plus de 2) segments de droite, la solution des oppositions binaires de traits montants et descendants n'a pas pu être retenue. À sa place, le trait 1 est composé d'une suite de caractères décrivant chaque segment de droite formant la stylisation intonative de la voyelle considérée : segment plat (0), montant (+) ou descendant (−). La longueur de cette suite de caractères indique le nombre de segments composant le contour. Sur la base de cette suite, pour chaque segment dynamique (montant ou descendant), sont estimés les traits 2 & 3, qui indiquent respectivement s'il s'agit d'un mouvement rapide (+) ou non (−)[1], et d'un mouvement ample (+) ou non (−)[2]. Le trait 4 indique quant à lui la position de la voyelle par rapport à la fréquence fondamentale moyenne du locuteur : il mesure donc si la voyelle se trouve au-dessus (+), au-dessous (−), au même niveau ou à cheval sur le registre moyen du locuteur (0).

[1] Un mouvement est décrété rapide si la valeur absolue de la pente de la droite est supérieure à un seuil, fixé à 6 demi-tons pour 100ms.

[2] Un mouvement de F_0 ample aura une amplitude supérieure au seuil de 6 demi-tons décrit par Astésano *et al.* (2007) pour un accent d'insistance.

Les traits 5, 6 et 7 évaluent la voyelle considérée en comparaison des autres voyelles du mot prosodique (le mot prosodique est délimité ici par les trois groupes nominaux ou verbaux qui composent les phrases du projet AMPER, cf. section 4.1.1). Une voyelle peut ainsi montrer la valeur la plus élevée de F_0 pour le mot prosodique (trait 5 : elle sera alors dite « culminante »). Une voyelle peut avoir la durée la plus longue (trait 6). Une voyelle peut enfin être considérée comme la plus « proéminente » du mot prosodique (trait 7) ; la proéminence d'un voyelle étant calculée selon l'équation 4.1.

$$P = \frac{Pg + Pi + Pd}{3} \tag{4.1}$$

Où Pg indique la différence (en demi-tons) entre la dernière valeur de F_0 de la voyelle précédente et la première de la voyelle considérée, Pi indique l'amplitude du mouvement de F_0 interne à la voyelle, et Pd indique la différence entre la dernière valeur de F_0 de la voyelle considérée et la première valeur de la voyelle suivante.

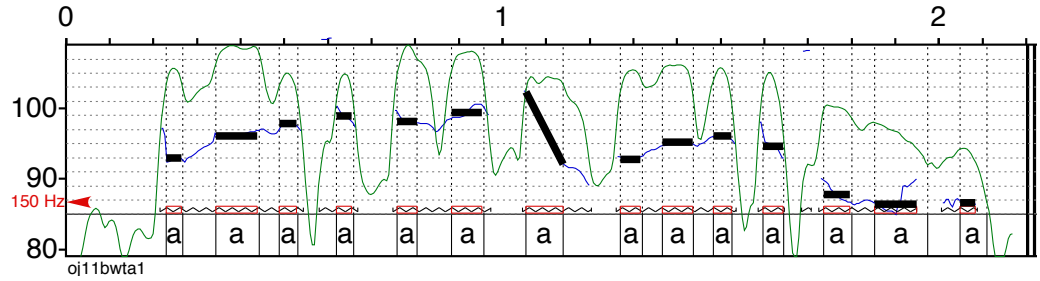
Les traits 8, 9, et 10 évaluent la voyelle considérée par rapport à l'ensemble des autres voyelles de la phrase, de manière similaire à ce qui est fait pour le mot prosodique. Une voyelle peut ainsi montrer la valeur de F_0 la plus élevée de la phrase (trait 8 : voyelle dite « extrême »). Une voyelle peut avoir la durée la plus longue de la phrase (trait 9). Une voyelle peut enfin être considérée comme la plus « proéminente » de la phrase (trait 10), toujours selon l'équation 4.1.

Le dernier trait (trait 11) permet de mesurer la pente de la courbe de déclinaison sur la phrase. Ce trait est noté sur la première et la dernière voyelle seulement (les autres voyelles portant la valeur 0) : la première voyelle par la valeur + et la dernière la valeur − si la pente est descendante et inversement pour les phrases avec une intonation montante.

Un exemple de la stylisation automatique obtenue sur une phrase AMPER est donné à la figure 4.3 (phrase « *oj11bwt* » - cf. Fernández Rei *et al.*, 2007a). On y observe que la seule voyelle qui montre un mouvement dynamique de sa courbe intonative est bien repérée dans la matrice de trait comme celle qui se distingue le plus des autres voyelles de la phrase (voir la notion de « point chaud » chez Contini et Profili, 1989).

Distances objectives entre matrices de traits

Comment passer de telles matrices de traits prosodiques à des distances objectives ? Pour cela, la distance d'édition décrite par Navarro (2001, p. 37) (distance de Levenshtein) est utilisée, qui permet de pénaliser les insertions, délétions et substitutions de symboles (et donc de traits prosodiques) entre



	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
Trait 1	0	0	0	0	0	0	-	0	0	0	0	0	0	0
Trait 2	0	0	0	0	0	0	+	0	0	0	0	0	0	0
Trait 3	0	0	0	0	0	0	+	0	0	0	0	0	0	0
Trait 4	0	0	0	0	0	0	+-	0	0	0	0	-	-	-
Trait 5	0	0	0	0	0	0	+	-	0	+	+	0	-	0
Trait 6	-	+	0	-	0	0	0	0	+	-	-	0	+	0
Trait 7	0	+	0	0	0	-	0	0	+	-	0	-	0	+
Trait 8	0	0	0	0	0	0	+	0	0	0	0	0	-	0
Trait 9	-	0	0	-	0	0	0	0	0	0	0	0	+	0
Trait 10	0	+	0	0	0	0	0	0	0	0	0	-	0	0
Trait 11	+	0	0	0	0	0	0	0	0	0	0	0	0	-

Figure 4.3 – *Stylisation par le* Prosogram *d'une phrase de structure « bwt » pour une locutrice du point d'enquête oj1 de Galice (graphique du haut). La table du bas montre le résultat de l'extraction des 11 traits prosodiques faite à partir de cette stylisation pour chacune des 14 voyelles.*

les deux matrices (voir aussi Oakes, 1998, p. 127). Elle est basée sur l'algorithme de programmation dynamique 4.2 qui calcule une distance $ed(x, y)$ entre deux suites de caractères x et y. Pour cela, la matrice $C_{0..x,0..y}$ contient le nombre minimum d'opérations nécessaires pour transformer $x_{1..i}$ en $y_{1..j}$ (d'après Navarro, 2001).

$$\begin{aligned} C_{i,0} = \quad & i \\ C_{0,j} = \quad & j \\ C_{i,j} = \quad & if(x_i = y_j) \\ & then \quad C_{i-1,j-1} \\ & else \quad 1 + min(C_{i-1,j}, C_{i,j-1}, C_{i-1,j-1}) \end{aligned} \tag{4.2}$$

Avec, à la fin de l'algorithme : $C_{x,y} = ed(x, y)$. La notation $C_{i,0}$ (respectivement $C_{0,j}$) représente la distance d'édition entre une suite de caractères de longueur i (resp. j) et une chaine vide, cas dans lequel il faut effectuer i (respectivement j) suppressions pour approximer la chaine cible.

Cette distance d'édition est déjà utilisée en dialectologie pour mesurer des variations phonétiques entre variétés dialectales (Heeringa, 2004). Elle est calculée ici pour chacun des onze traits de la matrice, puis sommée afin d'obtenir une distance globale entre les deux phrases comparées. Un élément

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
Trait 1	0	0	0	0	0	0	-	0	0	0	0	0	0	0
Trait 2	0	0	0	0	0	0	+	0	0	0	0	0	0	0
Trait 3	0	0	0	0	0	0	+	0	0	0	0	0	0	0
Trait 4	0	0	+	0	0	+	+-	0	0	0	0	-	-	-
Trait 5	0	0	+	0	0	0	-	-	+	0	+	0	-	0
Trait 6	+	0	-	-	0	0	0	-	+	+	-	-	0	+
Trait 7	0	+	0	0	0	-	0	+	0	-	0	-	0	+
Trait 8	0	0	+	0	0	0	0	0	0	0	0	0	-	0
Trait 9	0	0	-	-	0	0	0	0	0	0	0	0	0	+
Trait 10	0	+	0	0	0	0	0	0	0	0	0	-	0	0
Trait 11	+	0	0	0	0	0	0	0	0	0	0	0	0	-

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
Trait 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 4	0	0	0	+	0	0	0	0	0	0	0	-	-	-
Trait 5	-	0	0	+	0	0	0	+	0	-	+	0	0	-
Trait 6	0	+	0	-	0	0	0	-	+	+	-	0	+	0
Trait 7	0	0	+	0	-	0	0	-	0	+	0	0	-	+
Trait 8	0	0	0	+	0	0	0	0	0	0	0	0	0	-
Trait 9	0	0	0	-	0	0	0	0	0	0	0	0	+	0
Trait 10	0	0	+	0	-	0	0	0	0	0	0	0	0	0
Trait 11	+	0	0	0	0	0	0	0	0	0	0	0	0	-

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
Trait 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 4	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Trait 5	+	-	0	0	0	0	0	+	-	0	+	0	0	-
Trait 6	0	+	0	0	-	0	0	0	+	-	-	0	+	0
Trait 7	0	-	0	0	0	+	0	-	+	0	0	-	0	+
Trait 8	+	0	0	0	0	0	0	0	0	0	0	0	0	-
Trait 9	0	0	0	0	-	0	0	0	0	0	0	0	+	0
Trait 10	0	-	0	0	0	+	0	0	0	0	0	0	0	0
Trait 11	+	0	0	0	0	0	0	0	0	0	0	0	0	-

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
Trait 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trait 4	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Trait 5	-	0	0	0	0	0	+	-	0	+	+	0	0	-
Trait 6	0	0	0	-	+	0	0	-	+	0	-	0	+	0
Trait 7	+	0	-	0	0	0	0	-	+	0	+	0	-	0
Trait 8	0	0	0	0	0	0	0	0	0	+	0	0	0	-
Trait 9	0	0	0	-	0	0	0	0	0	0	0	0	+	0
Trait 10	+	0	0	0	0	0	0	0	0	0	0	0	-	0
Trait 11	+	0	0	0	0	0	0	0	0	0	0	0	0	-

Figure 4.4 – *Matrices de traits obtenues pour des phrases assertives de la même structure « bwt » qu'à la figure 4.3, phrases produites par des locutrices de quatre points d'enquêtes de la Galice (de haut en bas, les points oj1, ob1, ok1, om1).*

crucial du calcul de cette distance est l'unité sur laquelle les chaines vont être comparées. Il est en effet possible de comparer deux phrases syllabe par syllabe, ou par unités syntaxiques de plus grande taille : soit pour chaque mot prosodique ou en considérant la phrase entière. Dans le cas de comparaisons par syllabe ou par mot prosodique, les distances obtenues sur chaque sous-ensemble d'une phrase sont sommées pour obtenir la distance finale.

	phrase				**mot prosodique**				**voyelle**			
	ob1	oe1	ok1	om1	ob1	oe1	ok1	om1	ob1	oe1	ok1	om1
Trait 1	0	1	1	1	0	1	1	1	0	1	1	1
Trait 2	0	1	1	1	0	1	1	1	0	1	1	1
Trait 3	0	1	1	1	0	1	1	1	0	1	1	1
Trait 4	2	3	4	4	2	3	4	4	2	3	4	4
Trait 5	4	6	4	3	5	9	7	3	5	9	9	3
Trait 6	5	4	3	5	7	4	3	5	10	4	3	5
Trait 7	2	5	6	6	2	7	6	7	2	10	6	10
Trait 8	2	3	3	3	2	4	4	4	2	4	4	4
Trait 9	4	1	3	1	4	1	3	1	4	1	3	1
Trait 1	0	3	4	3	0	4	4	4	0	4	4	4
Trait 11	0	0	0	0	0	0	0	0	0	0	0	0
Somme	19	28	30	28	22	35	34	31	25	38	36	34

Table *4.2* – *Distances entre les matrices de trait de phrases de structures « bwt », pour 4 points d'enquête comparés au point de référence oj1.*

Ces différentes versions fournissent des résultats de distances objectives variables. Par ailleurs, toutes ces solutions ne prennent pas en compte de la même manière le problème des délétions de voyelles qui peuvent se produire différemment d'un dialecte à l'autre.

Imposer à cette mesure une contrainte syllabique oblige à insérer des syllabes factices (remplies de traits neutres 0) aux endroits correspondants aux voyelles manquantes ; mais c'est aussi la manière la plus stricte de faire en sorte de comparer les syllabes une à une, et donc d'aligner parfaitement les positions accentuelles attendues. Une contrainte au niveau du mot prosodique permet d'une part de s'abstraire de la tâche de reconstruction des voyelles manquantes – qui sera prise en charge par la distance d'édition elle-même, tout en conservant un contrôle assez strict sur l'alignement des syllabes accentuées, puisqu'un même nombre d'accents potentiels (1 ou 2 dans le cas du projet AMPER) sont attendus pour un mot prosodique. La contrainte la plus lâche, qui cherche à aligner les vecteurs de trait au niveau de la phrase, n'impose aucun apriori sur les structures des deux phrases comparées, mais pourrait donner lieu à des alignements « fautifs » de syllabes accentuées dans des cas extrêmes de délétion de syllabes.

Afin de donner une idée des résultats obtenus grâce à cette mesure de distance entre matrices de traits, la figure 4.4 propose quatre matrices, obtenues pour des phrases de structure identique à la phrase d'exemple du point *oj1* (figure 4.3), mais provenant de quatre points d'enquête différents de Galice : *ob1*, *oe1*, *ok1* et *om1* (voir les travaux de Escourido Pernas, 2008; Fernández Rei *et al.*, 2007b,a; Fernández Rei et Escourido Pernas, 2008, pour une

description de ces enquêtes). Ces quatre matrices sont comparées à celle du point *oj1*, trait par trait, puis sommées afin d'obtenir une distance moyenne. Les trois contraintes, au niveau de la voyelle, du mot prosodique et de la phrase, sont utilisées. Les distances obtenues sont résumées à la table 4.2. On observe des différences dans les mesures de distance obtenues grâce à cette technique.

Tout d'abord, des différences sont observables au niveau du poids des différents traits sur la distance finale. Les traits déterminants la forme du contour de chaque voyelle semblent ainsi apporter peu d'information - cela notamment du fait des variations prosodiques limitées qui sont observées sur les corpus AMPER, pour lesquels des phrases très contrôlées sont produites. On se rapproche ici de la prosodie de lecture et peu de tons dynamiques sont observés. Les trois traits liés aux mots prosodiques montrent à l'inverse (sur cet exemple) les valeurs les plus importantes, ce qui pourrait s'expliquer encore une fois par la nature des phrases AMPER, qui sont conçues afin d'observer la variation accentuelle entre dialectes spécifiquement sur cet aspect de la variation prosodique. On notera ainsi que des travaux comme ceux de Romano (1999) ou Lai (2002) montrent une variation diatopique de la montée intonative accentuelle - et donc de la position des voyelles portant les maximums de F_0 pour chaque mot prosodique. Le fait d'observer plusieurs mots prosodiques sur une seule phrase explique aussi le plus grand poids de ce niveau d'analyse, en comparaison des traits calculés au niveau de la phrase. Enfin, le trait 11 n'est pas productif ici, car aucune opposition de mode n'est effectuée. Ce trait pourrait s'avérer intéressant pour la mesure des inversions de pente observées en Galice (Fernández Rei et Escourido Pernas, 2008) ou en Corse (Boula de Mareüil *et al.*, 2014).

Si l'on considère les valeurs globales de distance, et comme la description des contraintes donnée plus haut nous le laissait prévoir, la contrainte la plus lâche (celle agissant au niveau de la phrase), donne les distances les moins grandes, tandis que la contrainte la plus stricte (au niveau de la voyelle) maximise les distances. On observe que la phrase du point *ob1* (puis celle du point *om1*) est jugée globalement plus proche de la prosodie de référence du point *oj1*. Pour les phrases des points *oe1* et *ok1*, l'ordre des distances par rapport à la référence peut par contre changer en fonction de la contrainte considérée. La contrainte la plus permissive (contrainte de phrase) est favorable à la phrase du point *oe1*, car elle sous-estime (par rapport aux autres contraintes) les variations des traits déterminés au niveau du mot prosodique. Il semble donc que les contraintes qui permettent de respecter l'esprit des traits prosodiques[1] qui composent les matrices pourraient être plus intéressantes.

[1] Au sens de Contini et Profili (1989)

Afin de s'en assurer, une mesure sur l'ensemble des données de ces cinq points d'enquête est réalisée. Les trois contraintes (de phrase, de mot prosodique et vocalique) sont utilisées. À des fins de comparaison, la distance de corrélation entre les courbes intonatives (cf. section 4.2) est aussi calculée sur ces données. Les résultats de ces quatre mesures de distance prosodique entre les parlers de cinq points d'enquête de Galice sont résumés aux figures 4.5 et 4.6. Pour cela, les moyennes des distances calculées entre toutes les paires de phrases de même structure AMPER, produites par tous les locuteurs de ces cinq points d'enquête, sont regroupées afin d'obtenir une matrice de distances. Cette matrice de distances permet d'obtenir, grâce à un algorithme de regroupement hiérarchique (Murtagh, 1985), les dendrogrammes de la figure 4.5. Une autre représentation de ces distances est rendue possible grâce à la librairie du logiciel R « RuG/L04 » (Kleiweg, 2010). Cette librairie est utilisée pour réaliser les cartes choroplèthe (Goebl *et al.*, 1982) de la figure 4.6, qui représentent soit directement les distances issues d'un positionnement multidimensionnel par des échelles de couleurs (cartes de gauche), soit le résultat d'un partitionnement réalisé sur ces distances (cartes de droite). Les quatre cartes de chaque partie montrent les résultats obtenus à l'aide des différentes distances testées ici : corrélation pondérée par l'intensité et distances entre matrices de traits prosodiques calculées à l'aides des trois contraintes décrites ci-dessus (contrainte de phrases, de mot prosodique ou de voyelle).

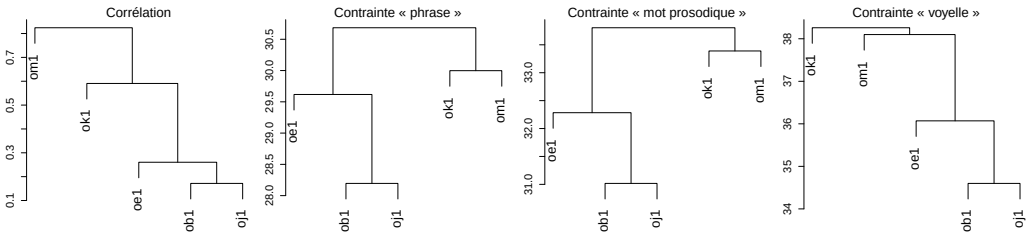


Figure *4.5* – *Dendrogrammes montrant la hiérarchisation des distances prosodiques moyennes mesurées entre cinq points d'enquête AMPER de Galice : ob1, oe1, oj1, ok1, om1 (cf. Fernández Rei* et al.*, 2007b; Fernández Rei et Escourido Pernas, 2008).*

Ce qu'on observe sur les graphiques 4.5 et 4.6 montre la variabilité des différentes distances prosodiques calculées. La variabilité la plus importante oppose la mesure de corrélation aux mesures de matrice de traits : la distance basée sur une mesure de corrélation rapproche beaucoup plus le point *oe1* des deux points *ob1* et *oj1* que ne le font les autres mesures de distance. Cependant, cette différence est principalement qualitative, car on n'observe aucune différence d'ordonnancement entre ces cinq points, quelle que soit la distance

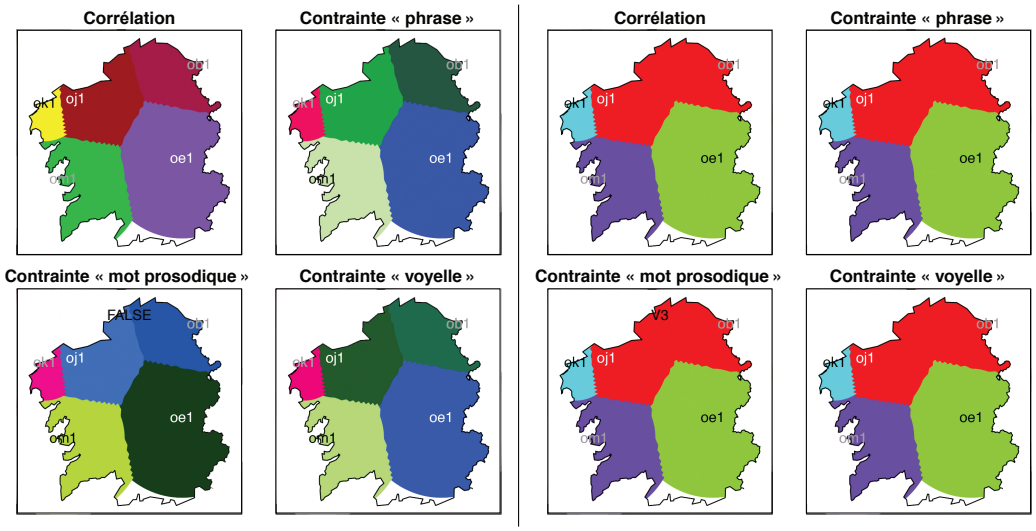


Figure 4.6 – *Cartes choroplèthes pour cinq points d'enquête de Galice (ob1, oe1, oj1, ok1, om1 - cf. Fernández Rei* et al.*, 2007b; Fernández Rei et Escourido Pernas, 2008) montrant, à gauche les distances issues d'un positionnement multidimensionnel entre les points, à droite les 4 principales partitions obtenues sur la base de ces distances.*

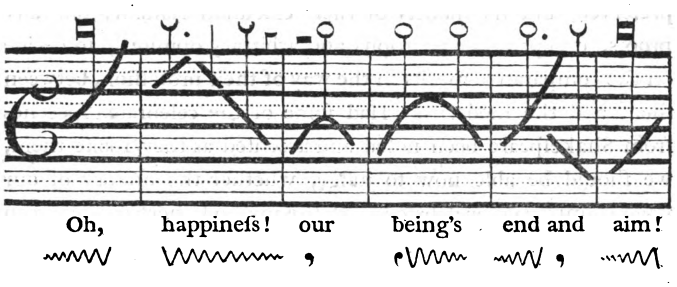
considérée. Cela se reflète le mieux dans les cartes montrant le regroupement en zones dialectales (figure 4.6 droite) : les quatre cartes sont identiques. Par ailleurs, le résultat obtenu est intéressant, car il reflète les descriptions de la variation prosodique dialectale du domaine galicien, en particulier autour de l'intonation des questions totales de chacun des trois points *oj1*, *ok1* et *om1*, proches géographiquement, mais montrant des contours prototypiques distincts, alors que le contour interrogatif du point *oj1* se rapproche de celui de *ob1* (Fernández Rei et Escourido Pernas, 2008). De même, une distinction prosodique claire entre les parlers des points *ok1* et *oe1* est décrite par Fernández Rei *et al.* (2007b), distinction tout à fait capturée par les différentes mesures de distances objectives présentées ici. Par ailleurs, les variations des réalisations des contours intonatifs interrogatifs observés ne seraient perceptivement pas reliés à des variations de typologie de question, mais bien à des distinctions dialectales distinguées par les auditeurs de ces parlers (Fernández Rei, 2011, 2013). La partition objective des variations prosodiques dialectales de l'espace galicien reflète donc assez précisément les descriptions des dialectologues.

Ces résultats sont, certes, préliminaires et il faudrait travailler sur les distances obtenues pour chaque structure morphosyntaxique, dans chacun

des modes – ce travail devant être réalisé par les dialectologues travaillant sur ces données. Pourtant il semble que ces mesures de distances objectives soient un outil prometteur pour une mesure et une représentation efficace des variations prosodiques au sein d'espaces dialectaux comparables. De plus, la relative stabilité de mesures objectives réalisées grâce à des approches radicalement différentes permet d'espérer une certaine robustesse de ces résultats. Un problème complexe reste cependant l'évaluation systématique des liens entre mesures subjectives et objectives. En effet, la distinction fonctionnelle des variations prosodiques mesurées par ces modèles ne peut être guidée que par des évaluations perceptives.

On verra par la suite l'application de cette approche duale entre évaluation subjective d'une fonction prosodique et modélisation objective de la variation grâce à des analyses acoustiques de variations paramétriques que l'on peut (et sait) mesurer à partir du signal de parole.

5 | Fonction expressive



Steele, *Prosodia rationalis* (1779, p. 13)

Les deux premiers chapitres ont traité de la réalisation et de la perception de la fonction démarcative de la prosodie, ainsi que de la variation diatopique de cette fonction. Les « fonctions expressives » qui seront abordées ici englobent très largement tous les aspects de la variation prosodique qui ne participent pas directement à la segmentation et la hiérarchisation des énoncés, tout en prenant part à l'interaction parlée. Nous ne prendrons donc pas en compte les variations prosodiques liées par exemple à l'origine régionale (en tant que variante d'un parler de référence), au sexe ou à l'âge des locuteurs. Il sera ici question de caractériser ce que Uldall (1960) ou Wichmann (2000, 2002) appellent des *attitudes* prosodiques et d'étudier les différences et les similitudes qu'elles entretiennent avec d'autres expressions prosodiques comme l'expression d'*émotions* (Scherer, 1989a, 2003), ou les variations *illocutoires* et *modales* (Di Cristo, 2013, p. 244 sqq.). Il s'agit de comprendre comment des variations prosodiques peuvent exprimer un sens utile à l'interaction parlée et en quoi ces variations prosodiques sont le reflet de contraintes linguistiques ou des autres facteurs constituant le milieu d'un locuteur – son origine culturelle, son état affectif, ses buts communicatifs, etc.

Dans un premier temps, nous verrons comment on peut distinguer les expressions attitudinales des expressions émotionnelles, puis de l'expression

des modes de la phrase. La communication parlée est multimodale[1] ; nous tâcherons donc ensuite de montrer comment des informations provenant des mouvements du visage et des informations prosodiques concourent à la transmission de ce type d'information vers l'interlocuteur.

5.1 Hiérarchisation des expressions affectives

5.1.1 L'*émotion* au centre de la communication

Les travaux en neuroscience et en psychologie proposent de nombreuses définitions de ces processus, liées aux différents modèles de ces « émotions ». Dans un article prospectif, Damasio (1998) définit une émotion comme une commande cérébrale induisant des changements physiologiques dans le corps de celui qui la ressent – commande qui induit donc des états émotionnels. Dans son introduction, Damasio insiste sur le fait que ces processus émotionnels se sont développés au cours d'une phylogénie complexe ; comprendre les émotions nécessite donc la compréhension de phénomènes profondément ancrés au sein du corps et ces processus ont une importance vitale en termes darwiniens de survie (voir aussi Scherer, 1989b). Cette importance du développement phylogénétique de l'espèce humaine dans des processus observés de manière beaucoup plus anodines (les « émotions » de tous les jours) rappelle la proposition d'Ohala (1983, 1984) qui postule l'existence d'un code fréquentiel (dorénavant le « *frequency code* »), universel, partagé par les mammifères comme par les oiseaux et qui constituerait l'un des fondements de l'*expression* des émotions, du fait d'un processus de symbolisme sonore (Ohala, 1994).

Nous sommes passés ici de l'*émotion* comme déclencheur d'états somatiques permettant une réponse adéquate à un stimulus, à l'*expression d'émotions*, qui constitue donc la communication externe (intentionnelle ou non) de ces états, éventuellement pour des buts communicatifs (prévenir d'un danger ou marquer un territoire). Pour relier ces deux aspects des émotions, il nous faut faire appel aux modèles psychologiques proposés, notamment par Scherer (par exemple Scherer, 1984b,a, 1986, 1999, 2001, 2003). Je ne ferai pas ici un résumé du modèle des processus-composantes, mais tenterai d'en extraire des principes aidant à la compréhension de ce qui suit (pour une revue des théories psychologiques des émotions, voir aussi Brosch *et al.*, 2010). Scherer décrit l'émotion comme un processus dynamique, qui évolue

[1] Rappelons que nous utilisons dans ce document le terme de *mode* pour désigner les modes illocutoires de la phrase (déclaratif, interrogatif, impératif), tandis que le terme *modalité* sera réservé aux différentes modalités d'expression (vocale, verbale, faciale, gestuelle, etc.).

en fonction de l'interaction entre un individu et son environnement, et de l'évaluation faite par cet individu des évènements qui surviennent dans cet environnement. Cette évaluation donnera naissance, dans le modèle proposé par Damasio (1998), aux commandes cérébrales induisant un état émotionnel et donc à des changements physiologiques. Ces changements physiologiques vont induire en particulier, pour ce qui touche à la parole, des modifications des paramètres de l'ensemble de l'appareil vocal (pression d'air dans les poumons, tension des plis vocaux, etc.) ; ces changements auront une conséquence acoustique audible, et potentiellement perçue, puis reconnue, par des auditeurs. Scherer (1981, 1989b) prédit ainsi des changements acoustiques liés à différents ressentis émotionnels, sur la base de ce modèle. Ces hypothèses seront en partie validées par Banse et Scherer (1996).

Si l'on s'arrête là, l'*expression émotionnelle* est un sous-produit d'une modification somatique qui est à l'origine de signes perceptibles. Ce serait ignorer l'existence de phénomènes de contrôle de ces manifestations. Les individus soumis à un ressenti émotionnel peuvent, dans une certaine mesure, exercer un contrôle, conscient ou automatisé, sur les signes de leur ressenti émotionnel (voir la distinction entre effets *push* et *pull* chez Scherer, 1989a; Scherer *et al.*, 2003, 2013). Ces processus de contrôle s'exercent essentiellement sous la forme d'une inhibition de la réaction comportementale qui, sans eux, afficherait un signal non conforme aux codes du contexte social ou culturel dans lequel se trouve le *locuteur* (nous nous intéressons ici à la communication *parlée*). Cependant, les travaux de Scherer et Wallbott (1994) concluent à une importance restreinte des facteurs sociaux-culturels dans la variation de la réaction comportementale par rapport à l'ampleur des similitudes observées.

Il se trouve que l'on observe cependant de notables variations dans les expressions émotionnelles. Un processus détaillé par Damasio (1998) peut expliquer cela. En effet, il décrit d'abord le processus « normal » du phénomène émotionnel, qui implique des changements somatiques ; ce processus fait partie de l'expérience émotionnelle de chaque sujet et résulte en une représentation mentale des changements provoqués par l'état émotionnel. Cette représentation mentale peut par la suite être utilisée afin de *simuler* une réponse émotionnelle, cette fois sans que le corps subisse les changements physiologiques associés. Ce processus d'une boucle faisant « comme si » le corps participe (« *as-if-body-loop* », cf. Damasio, 1998) à une émotion a aussi été observé expérimentalement par l'équipe de Damasio (Damasio *et al.*, 2000) sur des sujets se remémorant des émotions passées. C'est cette boucle hors du corps qui permettrait à un sujet de simuler des émotions, en faisant appel à sa mémoire d'émotions ressenties précédemment, et ce à diverses fins. L'une de ces fins permet à un locuteur de reproduire les com-

portements liés à un ressenti émotionnel. Cette capacité des sujets à *jouer* les expressions émotionnelles (Decety et Grèzes, 2006) est particulièrement intéressante dans le contexte qui nous concerne ici, car elle pourrait être à la base d'un grand nombre des variations expressives de la parole. Mes travaux cherchent à mieux comprendre certaines sources de la variation prosodique, et donc s'intéressent à la prosodie de la parole comme outil au service de la communication – je ne m'intéresse donc pas directement au phénomène émotionnel en tant que tel. Cependant, ce phénomène émotionnel, comme un des moteurs important du comportement humain (notamment en interaction) et en tant que processus phylogénétiquement construit, est particulièrement productif pour expliquer des variations prosodiques.

5.1.2 Variation du contrôle du locuteur sur ses affects

L'expressivité émotionnelle pourrait, on l'a vu, n'être qu'un sous-produit des processus physiologiques associés à des états ayant pour but d'améliorer nos chances de survie. Cependant, ces symptômes sont le symbole d'un ressenti (Scherer, 1992) et peuvent être utilisés d'un point de vue communicatif à diverses fins : par un locuteur en exprimant de la peur pour servir d'alarme en présence de danger, par un auditeur percevant la peur de son interlocuteur comme un signal de sa position de dominance. Cette importance communicative majeure des expressions émotionnelles (et pas seulement chez l'homme) donne naissance à des codes symboliques utilisés pour la communication entre individu (Hinton *et al.*, 1994). Ces codes sont aussi, évidemment, détournés pour obtenir un avantage – c'est typiquement l'une des fonctions du « *frequency code* », tel qu'il est décrit par Ohala (1994). Les émotions, et les codes qui leur sont associés, sont aussi complexifiés progressivement dans le cours du développement de l'enfant ; Widen et Russell (2003) montrent ainsi que le nombre de termes désignant des émotions augmente régulièrement (et de manière cohérente) jusqu'à l'âge de quatre ans pour un nombre restreint de concepts émotionnels. Allant plus loin, Zinck et Newen (2008) décrivent ce qu'ils nomment des « attitudes cognitives » – telles que le respect – états mentaux cognitifs faisant référence à des propositions ou à des objets.

Cette dernière catégorie d'affects introduit le terme d'*attitude*. L'expression par la prosodie d'une « attitude cognitive » (au sens de Zinck et Newen, 2008) correspond, je pense, au concept « d'attitude prosodique » décrit par Wichmann (2000, 2002). Il s'agit d'affects ayant pour but l'expression d'un concept, dirigé soit vers un contenu propositionnel, soit vers un objet. Cette distinction entre attitudes propositionnelles et attitudes sociales est utilisée notamment par de Moraes (2008) et on verra les implications qu'elle peut avoir sur la mise en place de l'expression de ces affects. Ces expressions

attitudinales ne sont pas des expressions d'émotions ; elles sont des codes, construits dans une langue et une culture, pour exprimer un concept (voir la notion de cliché mélodique chez Fónagy *et al.*, 1983). En ce sens, il s'agit d'objets linguistiques puisqu'ils permettent la communication et l'on pourra s'attendre à les voir varier d'une langue à l'autre.

Dernier étage de l'expressivité prosodique que nous aborderons ici, les actes illocutoires implémentent prosodiquement les différentes modalités de la phrase (et notamment les phrases assertives, les questions, et les ordres). Le *frequency code* propose que la distinction assertion / question soit implémentée prosodiquement par une hausse de la fréquence fondamentale du locuteur dans le cas des questions. On renverra le lecteur vers les travaux de Gussenhoven (2004) pour une description plus détaillée de différents codes prosodiques productifs de distinctions fonctionnelles en parole. La réalisation des distinctions illocutoires seraient donc bien encore le sous-produit d'un code symbolique hérité de l'expression des émotions.

Cette hiérarchisation de l'expressivité prosodique en trois grands groupes (expression d'émotions, d'attitudes, d'actes illocutoires) fait aussi référence au concept d'*engagement* du locuteur dans le langage, proposé par Daneš (1994). Ce concept d'engagement regroupe l'ensemble de ces trois types d'expressions, que l'on ne doit pas voir comme des catégories distinctes les unes des autres, mais plus comme le produit d'un raffinement conceptuel progressif, effectués par des groupes sociaux dans le cadre de la construction de leur langue commune. Cette notion d'engagement est aussi liée à la force illocutoire des expressions affectives. Un locuteur réalisera des actes de parole véhiculant des forces illocutoires plus grandes dans les cas pour lesquels le contenu de ce qu'il exprime contient un enjeu important et on peut supposer que la prosodie participe à la réalisation de cette force illocutoire dans le signal de parole (via notamment la force de voix, Liénard et Barras, 2013). À des fins d'expressivité, les locuteurs ont réutilisé les symboles issus de l'expression des émotions pour en dériver des codes parfois complexes à décrire (par exemple une expression ironique), même s'il se pourrait que ces codes prosodiques soient réductibles à un nombre restreint de dimensions sémantiques, décodées en contexte.

5.2 Expressions émotionnelles : spontané *vs.* acté

En commençant à m'intéresser aux expressions émotionnelles, j'ai cherché à observer une potentielle distinction entre expressions d'émotions spontanées et actées. Ce travail a été réalisé dans le cadre de la thèse de Nicolas Audibert

(Audibert, 2008). La suite de cette partie est tirée de Audibert *et al.* (2010) ; les détails du processus expérimental s'y trouvent.

5.2.1 Capture contrôlée d'expressions

Les énoncés utilisés pour cette étude sont extraits du corpus expressif Sound Teacher/E-Wiz (Aubergé *et al.*, 2006). Ce corpus a été enregistré en chambre sourde grâce à une technique de Magicien d'Oz (Kelley (1983) – voir aussi Ito et Speer, 2006). Les locuteurs sont supposés interagir avec un logiciel novateur d'aide à l'apprentissage des langues étrangères basé sur un système de reconnaissance vocale, qui s'appuierait sur les principes de la théorie neurologique de la perception-action et la plasticité cérébrale pour permettre à ses utilisateurs d'acquérir rapidement et sans efforts la maitrise de la prononciation de voyelles de langues étrangères. L'interaction des locuteurs avec le système est contrainte par un langage de commande composé des mots monosyllabiques français : « brique » ([bʁik]), « jaune » ([ʒon]), « rouge » ([ʁuʒ]), « sable » ([sabl]) et « vert » ([vɛʁ]), ainsi que de la commande « page suivante » ([paʒsɥivɑ̃t]).

L'interaction locuteur/système était manipulée en jouant sur les performances attribuées aux locuteurs, afin d'induire chez les sujets (sélectionnés pour leur intérêt pour l'apprentissage des langues étrangères) des émotions positives puis négatives. Pour cela, des performances excellentes leur sont attribuées dans les premières phases du test, suivies de performances très médiocres lors d'une phase plus complexe, doublées d'un avertissement stipulant que leurs « *capacités perceptives pourraient avoir été dégradées par les effets du système sur la plasticité cérébrale* ». Afin de rendre crédible une telle dégradation, les voyelles qui leur sont présentées ont été modifiées de façon à être artificiellement proches.

Chaque énoncé du langage de commande a été répété au minimum vingt fois par chaque locuteur, de façon répartie sur les différentes phases du scénario. On obtient ainsi des énoncés phonétiquement identiques, mais porteurs de valeurs affectives variées. Les locuteurs sont filmés et enregistrés durant leur performance ; ces films sont ensuite étiquetés par les locuteurs eux-mêmes, à qui il est demandé d'annoter les états affectifs dans lesquels ils se trouvaient. Les locuteurs étant également acteurs[1], il leur a été demandé, immédiatement après l'enregistrement, de reproduire les affects ressentis lors de l'expérience, sur les mêmes énoncés (rouge, jaune...).

Parmi ces sujets, quatre locuteurs[2] ont été sélectionnés, jugés parmi les plus expressifs pour leurs productions spontanées (Laukka *et al.*, 2007) –

[1]Acteurs ayant une pratique de théâtre de rue et/ou d'improvisation.
[2]Deux femmes (F1 et F2) et deux hommes (M1 et M2).

	Modalité			Moyenne
Locuteur	*A*	*V*	*AV*	
F1	0,11	0,14	0,33	0,20
F2	0,20	0,07	0,15	0,14
M1	0,08	0,10	0,33	0,17
M2	0,33	0,19	0,44	0,32
Émotion	*A*	*V*	*AV*	
Anxiété	0,12	0,14	0,36	0,21
Irritation	0,15	0,23	0,31	0,23
Satisfaction	0,27	0,00	0,27	0,18
Moyenne	0,18	0,12	0,31	0,21

Table *5.1* – *Valeurs moyennes des scores de discrimination, par locuteur, puis par classe d'émotion, pour chaque modalité de présentation des stimulus.*

donc ceux qui sont le mieux entrés dans le protocole d'induction émotionnelle. Trois classes émotionnelles ont été retenues (satisfaction, irritation et anxiété), pour lesquelles deux types d'énoncés pour chaque locuteur (soit un mot monosyllabique de couleur, soit « page suivante ») ont été retenus. Ces 24 stimulus sont appariés dans leur version spontanée et actée.

5.2.2 Discrimination perceptive

Les sujets avaient pour tâche d'indiquer pour chacune de ces paires lequel des stimulus correspondait à une expression spontanée. Les réponses étaient données à l'aide d'une échelle (de type *visual analog scale*, cf. Rietveld et Chen, 2006) allant de « Certainement le premier » à « Certainement le deuxième » (le résultat est codé entre -1 et 1). Chaque paire de stimulus est présentée selon trois modalités : audio seul (A), visuel seul (V) et audiovisuel (AV) ; chaque paire est aussi présentée dans les deux sens (acté/spontané ou spontané/acté). Les présentations des paires se font par modalité et en ordre aléatoire au sein d'une modalité. Les 33 sujets francophones qui ont passé le test devaient donc évaluer 144 paires de stimulus : 4 locuteurs x 3 classes d'émotions x 2 types d'énoncés x 2 ordres dans la paire x 3 modalités. Les réponses sont codées négativement si le curseur tend vers le choix correspondant à un stimulus acté, positivement dans le cas inverse. Les valeurs moyennes, pour chaque condition de présentation, chaque locuteur et chaque classe d'émotion, sont rassemblées dans la table 5.1.

Les sujets tendent à discriminer les stimulus spontanés des stimulus actés : une majorité des paires présentées reçoit un score de discrimination significativement supérieur au hasard. Le résultat le plus remarquable de cette

expérience reste cependant la variabilité des capacités des juges à discriminer entre stimulus actés ou spontanés. On aurait pu s'attendre à ce que cette variabilité inter-sujets soit liée au genre des juges, la littérature indiquant que les femmes perçoivent plus efficacement et plus rapidement les indices émotionnels que les hommes (voir Hall *et al.*, 2000, pour un état de l'art). Cependant, si les femmes montrent des capacités de discrimination en moyenne supérieures à celles des hommes, la différence n'est pas significative (notons tout de même que la faible puissance du test n'autorise pas de conclusion – cf. Rietveld et Van Hout, 2005). Par ailleurs, Émond (2013) observe elle aussi des résultats contraires à ceux de la littérature (de meilleures performances de la part des hommes) pour la perception de la parole souriante, sur des stimulus de parole spontanée. Il se pourrait donc que le contenu affectif de stimulus spontanés soit traité indifféremment du genre de l'auditeur – ou en tout cas que ces expressions affectives spontanées induisent des résultats suffisamment différents de ceux obtenus sur la base de stimulus actés pour insister sur l'importance de plus nombreuses études basées sur ce genre de données.

Autre facteur ayant un effet important sur les scores de discrimination, la modalité de présentation des stimulus. La modalité audiovisuelle reçoit des scores significativement plus élevés que chaque modalité présentée seule (et qui ne montrent pas de différence significative entre elles). S'il ne semble pas y avoir de préséance d'une modalité sur l'autre pour une tâche de discrimination du caractère acté ou non d'expressions affectives, chaque modalité apporte des informations complémentaires et la présentation bimodale permet aux auditeurs des scores significativement meilleurs. Nous reviendrons par la suite sur la multimodalité des expressions affectives, car il s'agit d'un facteur déterminant pour nombre de leurs caractéristiques.

D'autres travaux ont montré que les stimulus actés présentent des affects réalisés avec une force perçue plus importante que celle de stimulus spontanés (Wilting *et al.*, 2006; Shahid *et al.*, 2008; Laukka *et al.*, 2007). Cette différence peut ressortir au caractère caricatural des expressions émotionnelles actées, et pourrait être l'un des facteurs explicatifs de la capacité des sujets à discriminer entre les deux types de stimulus. Une seconde expérience a donc été menée pour évaluer l'intensité perçue relative des paires de stimulus évaluées ci-dessus. Les résultats confirment une corrélation significative (et négative) entre le score de discrimination et l'intensité perçue : plus l'intensité perçue d'un affect est forte (comparativement à l'autre membre de la paire) et plus les auditeurs auront tendance à le juger acté. Ici encore, la modalité de présentation joue un rôle important dans le jugement de la différence d'intensité – la modalité audio-visuelle montrant des différences d'intensités perçues plus

importantes que les deux conditions basées sur une seule modalité.

Les résultats montrent globalement la capacité des auditeurs à discriminer la nature actée ou spontanée d'une majorité des expressions émotionnelles. On observe aussi l'importance cruciale de la multimodalité dans cette tâche complexe. Par ailleurs, cette capacité ne semble pas dépendre de la catégorie émotionnelle présentée (sur les trois représentées dans ce test). Les différences d'intensité émotionnelle perçues dans les paires de stimulus présentées suggèrent que ce paramètre pourrait être responsable d'une part non négligeable de cette capacité à discriminer. Quels indices prosodiques et visuels permettent aux sujets de discriminer entre stimulus actés et spontanés ? L'intensité affective perçue, comme la force illocutoire, pourrait être encodée par exemple dans la valeur moyenne des empans mélodiques et / ou de la force de voix (Martin, 2014). D'autres indices pourraient aussi concourir à ce résultat ; notamment, pourrait-il y avoir des différences dans l'ancrage temporels des variations prosodiques, plus contraintes par la structure linguistique dans le cadre de stimulus actés que pour des stimulus spontanés ? Les indices de l'expression émotionnelle dans la structure de la prosodie peuvent-ils avoir une temporalité indépendante de celle de la langue, qui réponde au temps des émotions ? Pour répondre à ces questions, et sans préjuger de la capacité d'autres acteurs à jouer des émotions de manière plus proche du spontané, l'utilisation de la parole actée comme référence pour la modélisation des expressions émotionnelles devrait être reconsidérée.

Inversement, et bien que les acteurs enregistrés ne soient pas nécessairement les plus performants, tous ont été en mesure de piéger plus de la moitié des juges sur au moins l'une des paires présentées. Cela montre la subtilité des différences observées et ce sont certainement des détails fins qui permettent d'effectuer un jugement de ce type. Cette capacité à simuler un affect est d'une utilité reconnue (Decety et Grèzes, 2006; Brassens, 1972) ; nous verrons dans les parties suivantes que de nombreuses expressions d'affects reposent sur des conventions et sont produites hors de tout ressenti émotionnel – si la genèse de ces conventions a sans doute à voir avec les émotions (cf. section 5.1.1).

5.3 Inventaires d'attitudes

Le deuxième aspect de l'expressivité prosodique abordé dans mes travaux concerne les expressions attitudinales et commence avec les travaux d'Aubergé *et al.* (1997). Comme on l'a vu, les attitudes prosodiques de Wichmann (2000, 2002) constituent des expressions affectives qui suivent des

conventions sociales. Ces variations prosodiques fournissent aux locuteurs un outil expressif leur permettant une plus grande efficacité dans leurs actes de parole, car ces attitudes permettent d'exprimer du sens parallèlement à celui véhiculé notamment par le contenu lexical des énoncés. Il serait intéressant à cet égard d'intégrer la prosodie dans les mesures de densité d'information linguistique (cf. Pellegrino *et al.*, 2011). En tant que signes, les attitudes prosodiques transmettent un sens qui peut avoir pour objet soit le contenu propositionnel de l'énoncé, soit un objet externe à l'énoncé (sur cette distinction, voir de Moraes, 2008, mais aussi Zinck et Newen, 2008, Wichmann, 2000, Gu *et al.*, 2011). Les codes utilisés pour exprimer ces variations de sens peuvent dériver de contraintes ayant évolué phylogénétiquement – typiquement le *frequency code* (Ohala, 1983) – ou dépendant d'autres formes de symbolismes sonores, telles que celles proposées par Léon (1993) ou Gussenhoven (2004). Il est cependant fréquemment proposé que ces expressions prosodiques constituent des codes spécifiques à chaque langue Wichmann (2000), des « clichés mélodiques » conventionnels (Fónagy *et al.*, 1983). Ainsi, ces expressions prosodiques sont l'objet de manuels destinés aux enseignants de langue étrangère, car comme tout code linguistique, elles doivent aussi être enseignées (Martins-Baltar, 1977).

5.3.1 Des études dans différentes langues et cultures

Notre approche des attitudes prosodiques s'est donc d'abord intéressée à la constitution d'inventaires de ces codes conventionnels, dans différentes langues. Le travail le plus emblématique reste pour cela la thèse de Takaaki Shochi (Shochi, 2008) , qui portait sur la comparaison d'expressions attitudinales en trois langues : japonais, anglais britannique et français hexagonal. Les travaux doctoraux de Dang-Khoa Mac (Mac, 2012, sur les attitudes du vietnamien) et de Yan Lu (thèse en cours sur les attitudes du mandarin) ont suivi.

Je donnerai ici un bref aperçu des inventaires d'attitudes obtenus lors de ces travaux, pour ces cinq langues : anglais britannique, français hexagonal, japonais (Shochi *et al.*, 2009c), mandarin (Lu *et al.*, 2012) et vietnamien (Mac *et al.*, 2009). Pour chaque langue, la variété standard est choisie – celle qui est le plus généralement enseignée en cours de langue étrangère.

La table 5.2 fourni l'ensemble des étiquettes utilisées pour décrire les attitudes étudiées dans ces cinq langues, dans leur traduction française. Nous reviendrons par la suite sur les problèmes liés à la traduction. On voit que le nombre d'attitudes varie en fonction de la langue considérée. Dans les premières études, portant sur l'anglais, le français et le japonais, les inventaires obtenus ont bénéficié des explorations de cette dimension expressive par les

Étiquette	**anglais**	**français**	**japonais**	**vietnamien**	**mandarin**
Declaration (DECL)	x	x	x	x	x
Interrogation (INTE)	x	x	x	x	x
Evidence (EVID)	x	x	x	x	x
Ironie (IRON)	x	x	-	x	x
Autorité (AUTO)	x	x	x	x	x
Doute (DOUT)	x	x	x	x	x
Surprise (SURP)	x	x	x	x	x
Surprise positive (SPOS)	-	-	-	x	x
Surprise négative (SNEG)	-	-	-	x	x
Irritation (IRRI)	x	x	x	x	x
Arrogance (ARRO)	-	-	x	-	-
Mépris (MEPR)	x	x	-	x	x
Admiration (ADMI)	-	x	x	x	x
Politesse (POLI)	x	x	x	x	x
Sincérité-politesse (SINC)	-	-	x	-	-
kyoshuku (KYOS)	-	-	x	-	-
Séduction (SEDU)	x	x	-	x	x
Familier (FAMI)	-	-	-	x	-
Infant-Direct-Speech (IDS)	-	-	-	x	x
Intimité (INTI)	-	-	-	-	x
Confiance (CONF)	-	-	-	-	x
Décéption (DECE)	-	-	-	-	x
Résignation (RESI)	-	-	-	-	x
Total	11	12	12	16	19

Table *5.2* – *Étiquettes attitudinales (et leurs abréviations) utilisées pour les inventaires établis dans cinq langues. Une croix indique la présence de l'attitude dans l'étude, pour la langue considérée. La dernière ligne indique le nombre d'attitudes considérées pour la langue.*

enseignants de langue étrangère. Pour le vietnamien, langue peu dotée, et pour le mandarin, les travaux n'ont pas pu bénéficier d'un pareil travail de constitution empirique d'une liste d'expressions reconnues comme prototypiques par des enseignants. Les choix se sont donc basés sur les listes préexistantes dans les autres langues et montrent aussi une tendance certaine à l'inflation du nombre d'étiquettes.

Dans la table 5.2, les deux premières entrées ne sont pas à proprement parler des attitudes, mais des modes de phrases. Les modes assertif et interrogatif sont ajoutés à ces ensembles d'attitudes afin de fournir une référence d'une réalisation des phrases enregistrées « sans » attitude (ou avec une attitude « neutre », si cela veut dire quelque chose). Certaines étiquettes données en japonais ou en anglais sont en italique : l'expression japonaise de *kyoshuku* n'a pas d'équivalent lexical en français ; le terme *infant-directed-speech* n'a

pas non plus de bonne traduction en français. L'expression de *kyoshuku* est décrite comme « *corresponding to a mixture of suffering ashamedness and embarrassment, com[ing] from the speaker's consciousness of the fact his/her utterance of request imposes a burden to the hearer* » (Sadanobu, 2004, p. 34). Sa réalisation prosodique utilise une qualité de voix inattendue pour des auditeurs occidentaux et pour une expression classée parmi les expressions de politesse. Il s'agit de l'expression attitudinale emblématique des travaux de Shochi (2008).

Pour tous ces travaux, les attitudes sont enregistrées par un (ou une) locuteur (locutrice) dans une chambre sourde, sur des phrases construites pour la circonstance. Les mêmes phrases (autant que possible sémantiquement neutres du point de vue des attitudes) sont utilisées pour enregistrer toutes les attitudes, afin d'avoir un matériau prosodique strictement comparable. Le premier travail réalisé lors de ces études consiste à valider la qualité des expressions prosodiques produites par les locuteurs. Pour cela, des tests perceptifs de reconnaissance à choix forcé sont menés, avec des sujets natifs de la langue étudiée. Ces tests donnent deux sortes de résultats : (1) un score de reconnaissance brute de chaque attitude, qui indique la qualité prototypique des variations prosodiques enregistrées et (2) des matrices de confusion (contenant un comptage de toutes les réponses à chaque attitude présentée, pour l'ensemble des choix possibles), qui permettent d'évaluer les liens entre les différentes expressions. En effet, ces « erreurs » de reconnaissance sont plutôt des indicateurs de similarité perçue et permettent une meilleure compréhension de la structuration de l'espace expressif dans lequel se situent ces expressions attitudinales (cf. Banse et Scherer, 1996, p. 632). Nous ne rapporterons pas ici les analyses de taux d'erreurs, qui indiquent diverses qualités de la performance des locuteurs, mais montrent surtout, à mon sens, que de larges inventaires d'attitudes ne peuvent pas être l'objet de telles procédures d'évaluation perceptive : trop d'expressions peu distinctes et une liste de réponses possibles trop longue bruitent les résultats.

5.3.2 Principales dimensions perceptives

Nous allons voir plutôt quelles sont les principales dimensions qui regroupent ces attitudes en grandes catégories (voir les travaux précurseur sur les dimensions du sens de Osgood *et al.*, 1957). Pour cela, des analyses de correspondances (Benzécri, 1973) ont été menées sur les matrices de confusion. Les principales dimensions opposant ces différentes attitudes permettent de mieux comprendre la structuration de ces expressions les unes par rapport aux autres, dans ces cinq langues et pour des auditeurs natifs de chacune de ces langues. Un regroupement hiérarchique est ensuite appliqué aux données

ainsi obtenues (Husson *et al.*, 2013) afin d'extraire les principaux groupes d'attitudes que les auditeurs ont perçus. Il est frappant de constater que, pour des langues de familles et de typologies très différentes (indo-européennes, sino-tibétaine, tonales ou non, etc.), pour des inventaires d'attitudes qui ne se recoupent pas tout à fait, on observe des regroupements expressifs proches dans leur principales caractéristiques.

Analyses des dimensions

Les figures 5.1 et 5.2 présentent les graphiques obtenus pour le japonais, le français, l'anglais, le vietnamien et le mandarin. Les axes retenus pour montrer la dispersion des données sont : soit les deux premières dimensions (japonais, français, vietnamien), soit la première et la troisième dimension (anglais, mandarin) des analyses de correspondances. Ce choix est dû au fait que certaines attitudes ne montrant que peu de confusions avec les autres, elles sont liées, seules, à une dimension de l'analyse. Cette dimension perd alors de son intérêt pour comprendre la dispersion des autres attitudes[1]. C'est le cas des attitudes de séduction réalisées par le locuteur anglais et la locutrice chinoise, qui se distinguent complètement des autres attitudes. Les graphes de la figure 5.1 représentent le résultat de l'analyse des correspondances : les points noirs (dont les noms sont en majuscules) représentent la position des stimulus dans la distribution globale, tandis que les triangles verts (avec des noms en minuscules) représentent la position des étiquettes utilisées par les auditeurs pour juger de ces stimulus. On observe sur ces graphes les deux dimensions expressives[2] qui répartissent le mieux la distribution de l'ensemble des attitudes :

- la première dimension correspond pour ses extrêmes à des actes illocutoires assertifs *vs.* interrogatifs ;
- la seconde dimension oppose des expressions de dominance ou de soumission.

Pour quatre langues sur cinq, cette première dimension expressive correspond aussi à la première dimension de l'analyse de correspondances – et donc à la plus grande part de variance observée ; pour le vietnamien, cette opposition assertive / interrogative correspond à la seconde dimension de l'analyse et est moins claire que pour les autres langues (nous reviendrons sur le cas de cette langue). Cette première dimension peut être décrite, dans les termes

[1]Notons qu'un ensemble d'attitudes faisant l'objet d'une reconnaissance parfaite de la part des sujets donnerait une matrice de confusion diagonale et donc une analyse de correspondances pour laquelle chaque attitude serait liée à une dimension particulière.

[2]Il s'agit de dimensions interprétées à partir de la répartition observée sur les graphes, et non pas des dimensions de l'analyse de correspondances.

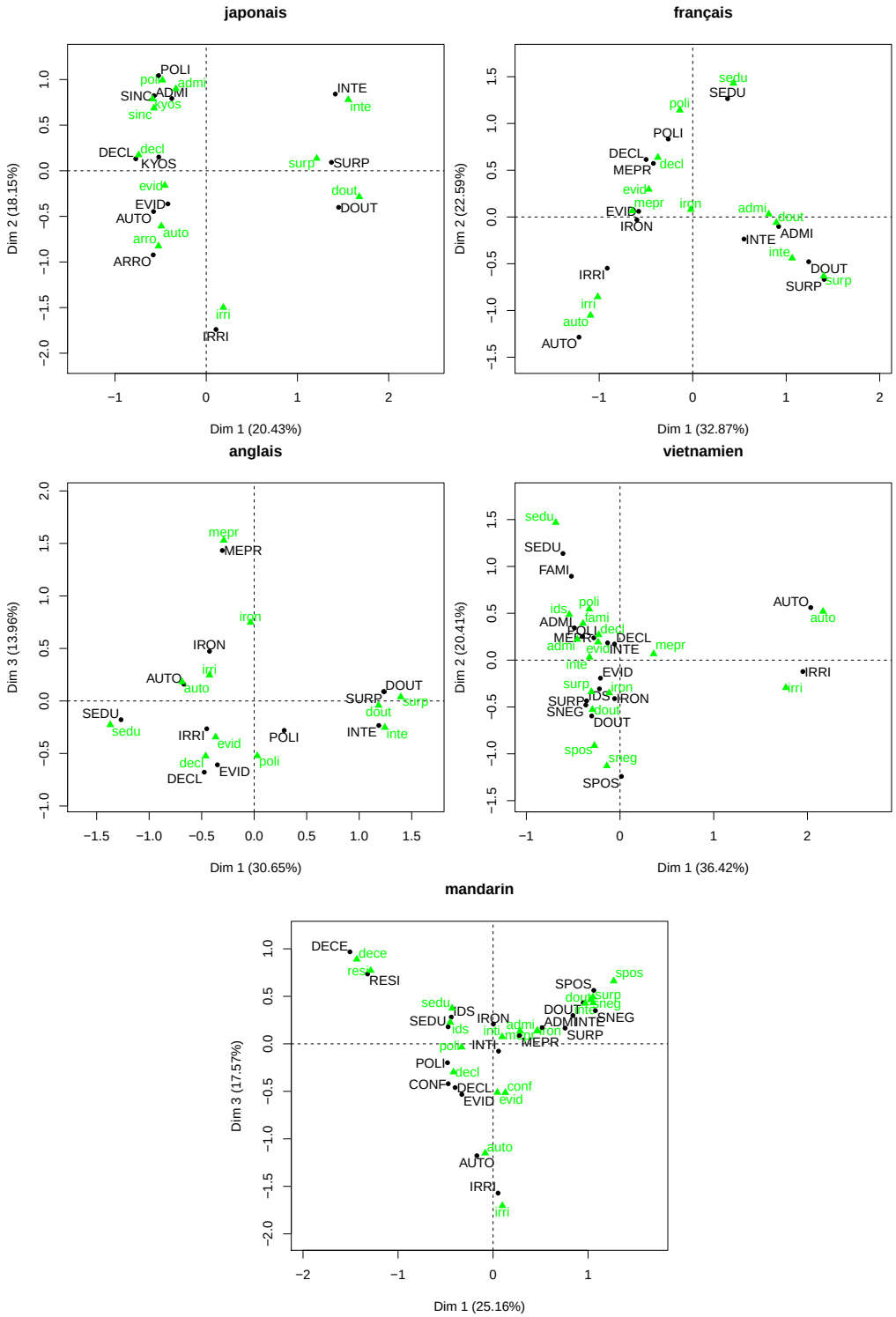


FIGURE 5.1 – *Graphiques plans des analyses de correspondances menées sur les matrices de confusion obtenues pour chaque langue. Les points noirs indiquent la position des stimulus ; les triangles verts indiquent la position des étiquettes d'attitudes utilisées par les auditeurs pour répondre. Les dimensions (1 et 2 ou 3) utilisées sont précisées sur les axes.*

de Brandt (2008), comme une opposition entre actes illocutoires assertifs et dubitatifs ; cette dimension rappelle ainsi une opposition fonctionnelle majeure en linguistique (particulièrement pour les aspects prosodiques) avec la distinction entre actes illocutoires assertifs et interrogatifs. Nous sommes ici au cœur d'une sémantique attitudinale qui adresse le contenu propositionnel de la phrase (au sens de de Moraes, 2008), entre l'assertion d'un contenu ou sa mise en question.

La seconde dimension qui structure ces attitudes concerne la relation entre le locuteur et son interlocuteur et donc organise les attitudes sociales (de Moraes, 2008), en permettant l'expression d'un rapport hiérarchique entre eux. Il peut s'agir de l'expression directe d'une dominance, exprimant par exemple l'autorité du locuteur, ou à l'inverse d'une stratégie de *politesse positive*, au sens de Brown et Levinson (1987), exprimant le désir d'être reconnu, accepté par l'autre. Cette seconde dimension expressive fait écho à des propositions comme le *frequency code* (comme marqueur de dominance, avec la prédiction d'une fréquence fondamentale plus élevée pour la politesse, Ohala, 1994), ou comme une dimension de *potency* ou de *dominance* proposée notamment (mais pas seulement) pour les expressions émotionnelles (Osgood *et al.*, 1957; Russell, 1980; Russell *et al.*, 1989; Russell et Barrett, 1999). On retrouverait donc ici des expressions attitudinales qui seraient le résultat de l'encodage culturel d'expressions émotionnelles (les attitudes cognitives de Zinck et Newen, 2008), conventionnalisées par une société pour gérer les relations et les interactions efficacement.

Analyses des regroupements

Les graphiques de la figure 5.2 montrent les mêmes données que celles de la figure 5.1, sans les points correspondant aux étiquettes utilisées pour les réponses, mais cette fois-ci les stimulus sont regroupés selon leur similarité perceptive, selon les 6 premiers groupes du regroupement hiérarchique effectué sur les résultats de l'analyse de correspondances. Ces groupes (*cluster* sur les graphiques) portent les mêmes noms (et couleurs) sur les graphes de chaque langue, mais ces numéros ne présentent *aucun* caractère de correspondance d'une langue à l'autre : les analyses sont menées de manière indépendante pour chaque langue. Afin de comparer ces groupes entre les différentes langues, la procédure suivante a été suivie.

- Les deux premiers groupes considérés sont ceux opposant les actes illocutoires assertifs et interrogatifs (principale dimension expressive observée auparavant) – et donc le premier groupe contient l'expres-

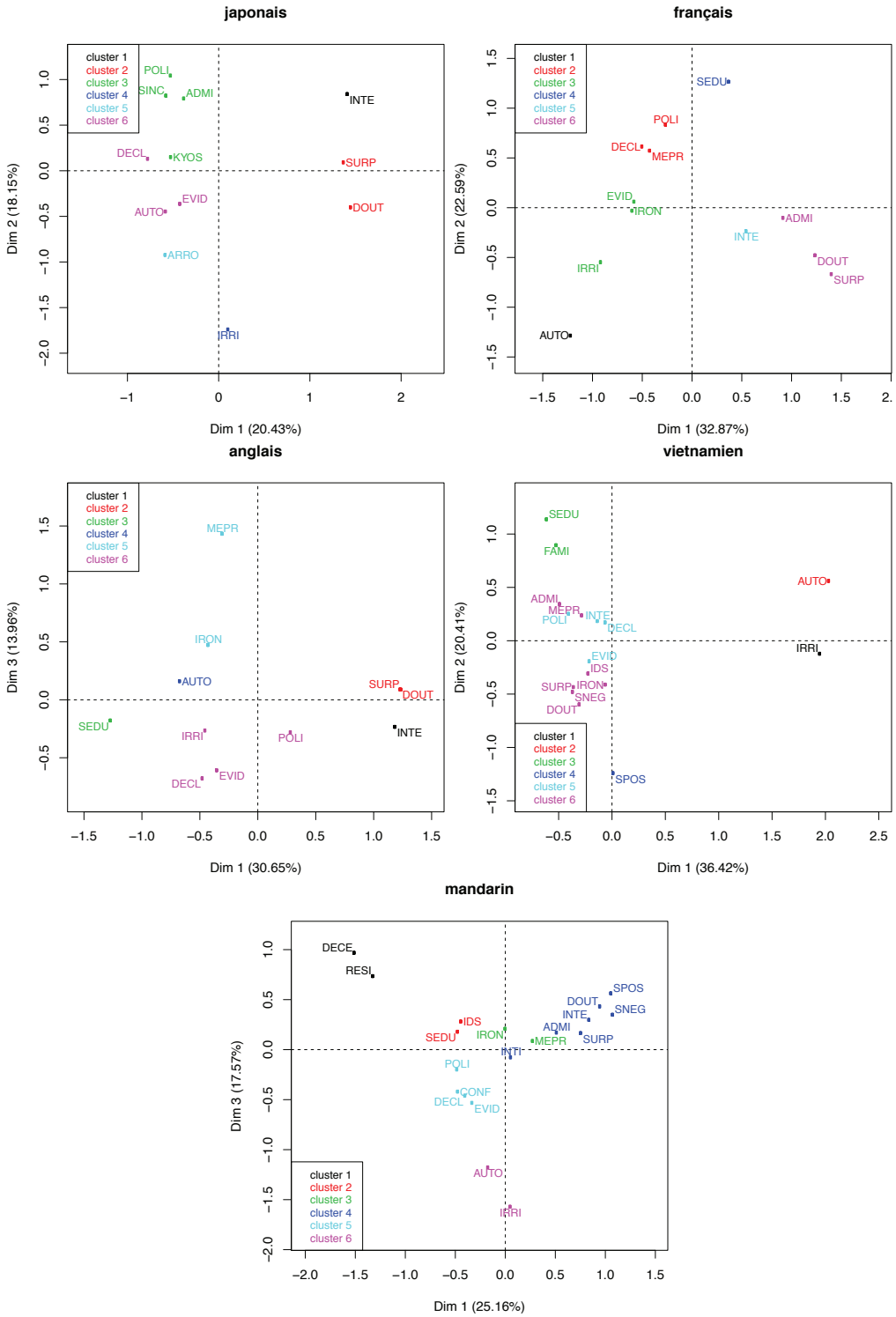


Figure 5.2 – *Graphiques plans des analyses de correspondances menées sur les matrices de confusion obtenues pour les attitudes des cinq langues étudiées. Les couleurs correspondent aux groupes d'attitudes obtenus à l'aide d'une agglomération hiérarchique menée sur la distribution des analyses de correspondances.*

sion de *déclaration* et le second celle d'*interrogation* (quand cela est possible).

- Les deux groupes suivants reprennent la dimension de dominance : le troisième groupe contenant l'expression d'*autorité* et le quatrième une ou plusieurs expression(s) de proximité ou de soumission – l'expression de *politesse* étant le plus souvent déjà regroupée avec la *déclaration*.
- Les caractéristiques des deux derniers groupes ont été déterminées à partir de l'observation des données restantes à la suite de la classification des quatre premiers groupes : il reste essentiellement des expressions à valence négative. Le cinquième groupe contient une ou des expression(s) de dominance (à valence négative, à la différence de l'autorité), tandis que le dernier regroupe des expressions dubitatives (au sens de Brandt, 2008), souvent à valence négative (à la différence de l'interrogation).

Pour chaque langue, ce processus d'identification des groupes d'attitudes permet de construire la table 5.3, qui montre une remarquable cohérence de la répartition des attitudes dans les six principaux groupes obtenus par ce processus de classification des données perceptives. On observe, bien sûr, des différences entre les langues dans cette classification (les inventaires eux-mêmes diffèrent) ; nous tâcherons de voir lesquelles et d'en proposer une analyse.

Le premier groupe est construit sur l'attitude « neutre » de déclaration. Ce groupe est aussi celui qui contient le plus grand nombre d'attitudes, sans doute parce que les auditeurs tendent à rapprocher les expressions peu marquées prosodiquement, ou celles qu'il ne comprennent pas, à une simple déclaration, qui fonctionnera alors comme une classe par défaut. On observe le rapprochement de certaines attitudes cohérentes (d'un point de vue pragmatique) avec la déclaration, en ce que ces attitudes peuvent être décrites comme étant essentiellement porteuses d'un acte assertif et différenciées par un (ou des) trait(s) supplémentaire(s). Les attitudes suivantes sont dans ce cas :

- La *politesse*, conçue comme une marque de courtoisie lors d'un acte assertif ; le but illocutoire est bien celui de l'assertion. La politesse appartient au groupe #1 pour quatre langues sur cinq – japonais exclu.
- L'expression d'*évidence* est une assertion, à laquelle le locuteur ajoute sa perception du caractère trivial du contenu propositionnel. L'évidence appartient au groupe #1 pour quatre langues sur cinq – français exclu.

	japonais	**français**	**anglais**	**vietnamien**	**mandarin**
#1	decl, evid, *auto*	decl, poli, *mepr*	decl, poli, evid, *irri*	decl, poli, evid, *inte*	decl, poli, evid, conf
#2	inte	inte	inte	spos	inte, surp, sneg, spos, dout, *inti*
#3	irri	auto	auto	auto	auto, irri
#4	poli, sinc, kyos, admi	sedu	sedu	sedu, fami	sedu, ids
#5	arro	iron, irri, *evid*	iron, mepr	irri	iron mepr
#6	dout, surp	dout, surp, *admi*	dout, surp	dout, sneg , surp, *ids*, *iron*, *mepr*	*dece*, *resi*

Table ***5.3*** *– Regroupement des attitudes de chaque langue en six groupes d'attitudes. Les groupes d'attitudes sont obtenus pour une langue grâce à leur proximité perçue par des auditeurs natifs. Les règles d'association des groupes sont expliquées dans le texte. Les attitudes qui ne correspondent pas à l'interprétation d'un groupe sont indiquées en caractères italiques. Les abréviations sont détaillées à la table 5.2*

- L'expression de *confiance*, qui n'est réalisée ici qu'en mandarin, peut être aussi analysée comme une assertion pour laquelle le locuteur marque la confiance qu'il a dans la véracité du contenu propositionnel.

Pour certaines langues, d'autres attitudes sont regroupées avec la déclaration, sans que cela puisse être interprété comme une tendance. Les expressions suivantes sont dans ce cas :

- pour le japonais, l'expression d'autorité ;
- pour le français, l'expression de mépris ;
- pour l'anglais, l'expression d'irritation ;
- pour le vietnamien, l'expression d'interrogation.

Plusieurs cas se présentent. L'attitude française de mépris (respectivement l'attitude d'irritation de l'anglais ; d'interrogation du vietnamien) est confondue avec la déclaration ou de la séduction (respectivement la déclaration et l'évidence ; la déclaration). Il s'agit ici d'erreurs de classification, qui ne sont bien sûr pas imputables aux auditeurs, mais plutôt à un manque d'informations prosodiques pertinentes – soit du fait d'une contreperformance des locuteurs, soit du fait que le trait sémantique est normalement exprimé via une autre modalité (visuelle ou verbale). Le cas de l'interrogation vietnamienne est particulièrement intéressant, car le mode interrogatif dans cette langue est généralement marqué lexicalement (Đô *et al.*, 1998). Cette utilisa-

tion de la prosodie pour marquer l'interrogation en vietnamien n'est donc pas habituelle et par ailleurs (ou justement de ce fait) certains tons (descendant, typiquement) sur la syllabe finale, peuvent venir contraindre la réalisation et la perception d'une montée intonative finale (Mac *et al.*, 2010, 2012). À l'inverse des cas précédents, l'attitude japonaise d'autorité est reconnue sans confusion majeure ; il serait alors possible d'avancer une interprétation de l'autorité dans le contexte social japonais comme étant essentiellement un acte assertif, pour lequel une personne effectivement en position d'autorité marque son rôle, socialement reconnu, par cette prosodie. Ce qui prime alors pour l'auditeur, c'est le contenu propositionnel, pas l'affirmation de la position sociale.

Le second groupe se construit autour de la modalité interrogative. La réalisation prosodique de ce mode constitue l'une des fonctions les plus souvent citées pour la prosodie dans son rôle linguistique et est l'objet d'une vaste littérature autour de la caractérisation des contours intonatifs qui en permettent la dénotation, dont le plus classique est un marquage par une montée intonative finale (mais voir Boula de Mareüil *et al.*, 2014). En terme de classification, trois langues sur les cinq (japonais, français, anglais) ont un groupe #2 dans lequel cette expression interrogative est seule. Pour le vietnamien, l'interrogation est confondue avec la déclaration. On a vu qu'en vietnamien, la dimension dubitative a une moindre importance que pour les autres langues (en terme de marquage prosodique, bien sûr). On voit cependant sur le graphique de la figure 5.1 que cette dimension existe, et ce en premier lieu du fait de l'expression de surprise positive. Nous avons donc choisi cette attitude (qui, elle aussi, forme un groupe à elle seule) pour constituer le groupe #2 en langue vietnamienne. Enfin, en mandarin, l'expression d'interrogation est regroupée avec d'autres expressions dubitatives : surprise, surprise négative, surprise positive et doute ; nous reviendrons sur ce regroupement lors de l'analyse du groupe #6, mais il pourrait être lié notamment au grand nombre d'attitudes dans cette langue (19), qui pourrait nécessiter un regroupement plus détaillé. L'expression d'intimité est aussi regroupée avec l'interrogation, mais il s'agit ici d'une erreur liée à une mauvaise reconnaissance de l'attitude prosodique.

Le troisième groupe caractérise l'expression de dominance de la seconde dimension expressive observée sur les analyses de correspondances. L'expression typique de cette dominance, pour quatre langues sur cinq, est représentée par l'attitude d'autorité. En japonais, c'est l'expression d'irritation qui a été choisie (l'autorité étant regroupée avec la déclaration), car elle exprime bien la volonté du locuteur d'imposer par sa prosodie son point de vue à son interlocuteur. Par ailleurs, le mandarin regroupe autorité et irritation. Il serait donc possible d'interpréter ce groupe prosodique comme un ensemble d'ex-

pressions dont le but communicatif est d'imposer son point de vue. Il est intéressant de noter dans cette optique que l'attitude prosodique d'autorité en japonais ne contient pas (ou moins) ce trait d'imposition, car l'autorité est donnée par le rôle social – ce qui n'est pas le cas de l'irritation, pour laquelle le locuteur doit exprimer une dominance qui n'est pas liée à son rôle sociétal.

À l'inverse du précédent, le quatrième groupe contient des expressions de proximité et / ou de déférence. Pour ce groupe encore, les différences entre les langues sont assez claires. Alors que le japonais rassemble trois expressions de politesse plus l'attitude d'admiration, les quatre autres langues montrent des groupes contenant l'expression de séduction, plus une expression de familiarité pour le vietnamien et l'expression d'*infant-directed-speech* pour le mandarin. Il faut toutefois noter qu'aucune de ces expressions n'est présente dans le corpus de japonais. Il est par contre plus frappant de voir que les japonais distinguent les expressions de politesse de la déclaration, alors que toutes les expressions de politesse des autres langues se retrouvent dans le groupe #1. Des travaux supplémentaires sur cette question seront nécessaires afin de déterminer s'il s'agit là d'un trait particulier des expressions de politesse japonaise, pour lesquelles l'expression de politesse en elle-même pourrait constituer le contenu déterminant de l'acte illocutoire : il est certain que le respect de normes sociales clairement définies est important au Japon (Hill *et al.*, 1986; Ide, 2002). Il est toutefois plus difficile d'avoir une opinion tranchée dans le cas des autres langues et de savoir dans quelle mesure ce résultat pourrait venir d'un artéfact des différentes attitudes représentées dans les corpus, ou du principe même du regroupement hiérarchique.

Les cinquième et sixième groupes sont marqués par une valence négative qui n'est pas présente – pas ou peu présente, suivant les diverses interprétations de l'irritation que l'on peut proposer – dans les groupes précédents. La dimension de valence est l'une des plus importante et des plus systématiquement mise en valeur dans les études qui s'intéressent aux dimensions sémantiques (Osgood *et al.*, 1957; Russell, 1980; Russell *et al.*, 1989; Russell et Barrett, 1999). On peut aussi rappeler la présence des deux primitives sémantiques *good* et *bad*, proposées comme universelles dans le *Natural Semantic Metalanguage* de Wierzbicka (1985, 1986b, 1996b, 1999, 2005). Il n'est donc pas étonnant de retrouver la valence des attitudes comme l'un des axes structurant de leur regroupement. Ce qui est aussi intéressant dans ces deux derniers groupes est la répartition de ces attitudes négatives entre (1) un groupe d'expressions dirigées vers l'interlocuteur (groupe #5), exprimant une dominance négative (il s'agit donc d'un groupe d'expressions sociales au sens de de Moraes, 2008), et (2) un groupe (#6) d'expressions interroné-

gatives, typiquement l'expression de doute (il s'agit pour ce second groupe d'attitudes propositionnelles au sens de de Moraes, 2008).

Le groupe #5 comprend des attitudes aux étiquettes variées. Il faudrait réaliser un travail d'analyse plus complet pour comparer les traits sémantiques de ces attitudes dans chacune de ces langues ; on peut toutefois faire les remarques suivantes. Le japonais apporte une expression d'arrogance, expression d'impolitesse délibérée clairement négative et perçue comme telle (Rilliard *et al.*, 2014b). L'anglais et le mandarin apportent les expressions d'ironie (définie comme une ironie sarcastique) et de mépris, là aussi clairement négatives. Le vietnamien apporte une expression d'irritation. Le français apporte les expressions négatives d'ironie et d'irritation (rappelons que l'expression de mépris a mal été reconnue dans cette langue). Cependant, on retrouve aussi dans le groupe #5 en langue française l'expression d'évidence ; une observation détaillée des résultats montre que ce regroupement est essentiellement lié à des confusions de l'expression d'ironie avec l'évidence.

Le groupe #6 est essentiellement composé d'expressions dubitatives et d'expressions à valence négative. L'expression la plus commune (présente dans les groupes de quatre langues sur cinq) étant la remise en question d'une assertion de son interlocuteur, une expression de doute qui combine ces deux traits dans une expression interronégative. Avec cette expression se retrouve une autre expression dubitative négative, la surprise négative, en vietnamien. On retrouve aussi en japonais, français, anglais et vietnamien la surprise, expression qui a une valence neutre dans la définition qui en est faite pour cette langue. Le cas du mandarin est particulier, puisque toutes les expressions dubitatives se trouvent dans le groupe #2 ; ce qui est intéressant, c'est qu'on trouve dans le groupe #6 du mandarin des expressions à valence négative, mais opposées à celles du groupe #5 (au sens où elles n'expriment pas une imposition de la part du locuteur, au contraire) ; il s'agit des expressions de déception et de résignation.

5.3.3 Limite des inventaires pour les études interculturelles

Cette méta-analyse des inventaires d'attitudes effectués sur cinq langues permet de confirmer l'importance d'un certain nombre de dimensions communes à différentes langues et perçues au travers de ces réalisations prosodiques :

- la distinction propositionnelle entre actes illocutoires assertifs et interrogatifs (cf. de Moraes, 2008) ;
- l'acte social d'imposition de sa volonté *vs.* une marque prosodique de déférence (une dimension de dominance, cf. Russell, 1980) ;

- la valence des actes illocutoires, valable tant pour des attitudes propositionnelles que sociales (Osgood *et al.*, 1957).

Cependant, cette approche par inventaires, basée sur des concepts tirés d'une culture particulière et donc imprégnés par cette culture (voir, sur la notion de « folk-concept », Wierzbicka, 2005), ne permet pas de faire des comparaisons directes de ces concepts, traduits d'une langue à l'autre, sur la base d'une proximité supposée des traductions.

En effet, l'un des principaux problèmes lié à l'utilisation d'étiquettes pour des études interculturelles est celui de leur traduction, ou autrement dit celui de la précision des concepts utilisés. Il s'agit d'un point largement argumenté, notamment pour les études interculturelles sur les émotions, par Wierzbicka (1986a, 1992, 1999, 2004, 2005, 2010). Elle remarque justement que, s'il peut exister des universaux (et tout son travail de recherche est axé vers la recherche d'universaux sémantiques), baser des travaux scientifiques sur des termes directement issus d'une langue (l'anglais, en général) induit un biais car ces entrées lexicales n'ont pas nécessairement d'équivalents (ou leurs traductions ne recouvrent pas exactement les mêmes concepts) dans d'autres langues et d'autres cultures et ceci constitue un biais méthodologique important :

> *"Emotion" is an English word. Evidence suggests that all languages have a word for "feel" (as in "I felt something good/bad"), (...) but not for "emotion." Thus, the English word "emotion" imposes a certain language - and culture - specific perspective on human feelings. If Anglophone scholars find it convenient to use this word in their discussions of human feelings (and their bodily correlates), there is of course nothing wrong with that, as long as they recognize that it is a complex cultural construct of modern English (...), and do not fall into the trap of taking it for a neutral analytical tool which can carve nature at its joints* (Wierzbicka (2010), p. 379).

On peut observer dans les résultats exposés précédemment des variations potentiellement liées à ce phénomène de non-recouvrement de concepts. Prenons le cas de la politesse, certainement celui sur lequel nous avons le plus spécifiquement travaillé et pour lequel nous pouvons avancer des analyses (Shochi *et al.*, 2009a,b; Rilliard *et al.*, 2012, 2014b). Parmi les attitudes du japonais, celles dites « de politesse » occupent un place importante (Shochi, 2008). L'analyse de correspondances de la figure 5.1 montre que, parmi les étiquettes utilisées par les auditeurs pour donner leurs réponses (les triangles verts), celles des trois expressions de politesse (la politesse de courtoisie, la sincérité-politesse et l'expression de *kysohuku*) sont regroupées (avec l'expres-

sion d'admiration). Les réalisations prosodiques de ces concepts (les points noirs) sont proches de leurs étiquettes respectives (et donc bien reconnues) pour toutes les attitudes – sauf l'expression de *kyoshuku* qui se trouve un peu éloignée de ce groupe de politesse. Cette différence demande une analyse plus approfondie : si l'on regarde les confusions effectuées par les auditeurs natifs qui perçoivent une expression prosodique de *kyoshuku*, on observe que la plupart des réponses se partagent entre les étiquettes *kyoshuku* et *sincérité-politesse* - deux concepts proches ; cependant quinze pour cent des réponses vont vers l'expression d'autorité. C'est ce biais vers l'étiquette d'autorité qui tire l'expression de *kyoshuku* plus près des expressions de dominance, selon la seconde dimension expressive du graphe 5.1.

Reprenons la définition du *kyoshuku* : c'est une expression « *corresponding to a mixture of suffering ashamedness and embarrassment, com[ing] from the speaker's consciousness of the fact his/her utterance of request* imposes *a burden to the hearer* » (Sadanobu, 2004, p. 34). Cette expression contient donc bien un trait sémantique d'*imposition* de la volonté du locuteur à son interlocuteur. Il se trouve que cette stratégie de *kyoshuku* correspond typiquement à une stratégie prosodique de politesse négative selon la définition de Brown et Levinson (1987, p. 188). Il est donc intéressant de constater l'écart entre l'étiquette, que les natifs associent uniquement à de la politesse, et l'expression – dans laquelle ils perçoivent et notent cette imposition.

D'autres résultats confortent cette analyse. Dans une étude sur l'acquisition des expressions de politesse et d'impolitesse chez les enfants japonais (Shochi *et al.*, 2009a,b), on observe que l'expression prosodique du *kyoshuku* est placée par les auditeurs adultes et les enfants natifs a peu près au milieu de l'échelle de politesse sur laquelle ils devaient juger les stimulus (figure 5.3, graphiques du haut). Dans une autre étude s'intéressant à l'apprentissage de ces mêmes expressions par des adultes francophones apprenants de japonais (Shochi *et al.*, 2014), on observe des jugements moyens pour l'expression de *kyoshuku* parmi les plus élevés sur l'échelle des réponses – en modalité audiovisuelle (figure 5.3, graphiques du bas).

Ces résultats nous apprennent plusieurs choses sur la perception de l'expression prosodique du *kyoshuku* :

- son acquisition par des enfants natifs progresse jusqu'au-delà de la dixième année ; auparavant, l'expression est jugée plutôt neutre (et moins polie qu'une déclaration) – sans effet de la modalité de présentation.
- des auditeurs ne connaissant pas le japonais (FR0) perçoivent eux aussi la réalisation acoustique (seule) comme une expression neutre (sur une échelle de politesse / impolitesse).

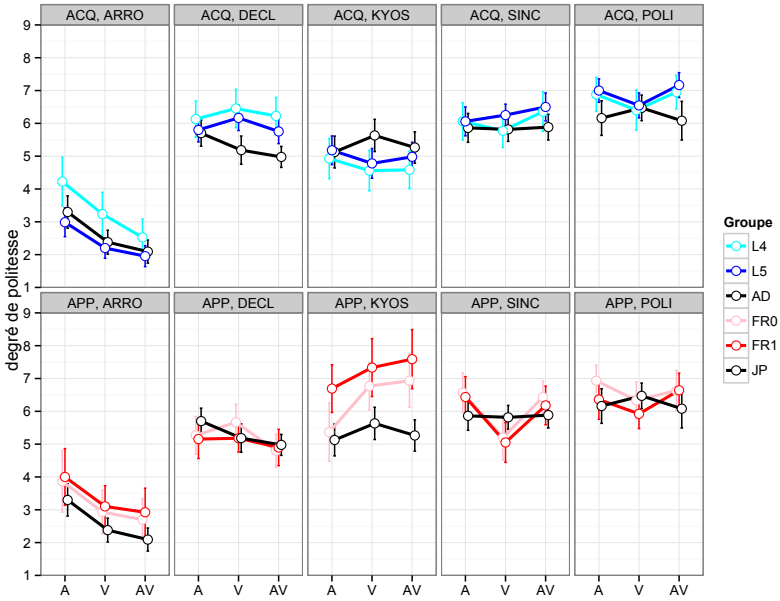


Figure 5.3 – *Jugements moyens du degré de politesse d'attitudes prosodiques japonaises (arrogance, déclaration, kyoshuku, sincérité-politesse, politesse). Graphiques du haut : étude de l'acquisition (ACQ) chez des enfants japonais de niveaux scolaire 4 et 5 (L4, L5), et par des adultes (AD). Graphiques du bas : étude de l'apprentissage (APP) par des adultes francophones ne connaissant pas (FR0) ou peu (FR1) le japonais, comparé au groupe adulte natif (JP). Les stimulus sont présentés selon trois modalités : audio (A), visuel (V) et audio-visuel (AV).*

- les francophones ayant déjà rencontré ce concept, ou bénéficiant d'une présentation multimodale de l'expression de *kyoshuku*, jugent la performance comme étant la plus polie de l'ensemble des stimulus qui leur sont présentés.

On peut conclure de ces observations que les enfants japonais ont déjà vers 10 ans une représentation du concept de *kyoshuku* assez proche de celles des adultes et savent reconnaitre la qualité vocale des stimulus ; pour autant ils ont encore du mal à en maitriser l'usage social (voir Shochi *et al.*, 2009a,b). Les locuteurs francophones apprenants de japonais appliquent une échelle de politesse « française » à cette expression de *kyoshuku* en lui attribuant une caractéristique de politesse extrême. Enfin, les sujets français ne connaissant pas le japonais effectuent la même attribution d'un expression de politesse extrême lorsqu'ils perçoivent le stimulus visuel ; mais ne comprennent pas l'expression en modalité audio seule et ils ne peuvent pas la situer sur une échelle de politesse de la même manière que les sujets japonais.

Nous sommes donc bien ici en présence de deux visions différentes de la politesse. Une vision provenant de la culture japonaise, culture pour laquelle la norme sociale est très présente (Hill *et al.*, 1986; Ide, 2002), et qui a conventionnalisé une forme prosodique particulière pour servir d'expression dans le cas très précis décrit par Sadanobu (2004). Ce concept de *kyoshuku* correspond bien à un comportement poli (et donc dans la norme sociale de cette culture), mais n'exprime pas uniquement de la politesse – son expression véhicule aussi un trait d'imposition perçu et noté de manière constante par les sujets natifs. On est donc en présence d'une culture qui accepte et conventionnalise plusieurs formes prosodiques de politesse, plus ou moins complexes, et destinées à être utilisées dans des contextes d'interaction précis, en fonction des buts illocutoires du locuteur. L'autre concept de politesse, français, est plus simple, au sens où la politesse a pour but dans cette culture de maintenir l'harmonie sociale durant une interaction verbale. Dans la description qu'en fait Kerbrat-Orecchioni (2005), il s'agit de protéger la « face »[1] de l'interlocuteur, soit en adoucissant une « menace » potentielle, soit en ajoutant une onction supplémentaire. Il s'agit bien d'une vision unidimensionnelle de la politesse, dont le contraire est l'impolitesse. Les sujets francophones – lorsqu'ils doivent juger la *politesse* d'une expression pour laquelle un locuteur japonais[2] effectue une inclination de la tête (voir figure 5.4) en face de son interlocuteur, jugent donc cette attitude « extrêmement polie » (cela serait interprété dans le modèle de Kerbrat-Orecchioni, 2005, comme une onction).

On voit donc, à la lumière de ces exemples que l'utilisation de *folk-concepts*, comme base d'inventaires prosodiques *destinés à comparer les attitudes prosodiques dans différentes langues* soulève des problèmes complexes : les expressions prosodiques actualisant les concepts dans une culture donnée sont liées à l'ensemble de la structure conceptuelle de cette culture. On ne peut donc pas considérer les réalisations prosodiques de deux concepts similaires du point de vue de leurs traductions comme équivalentes (par exemple en français, anglais et portugais : ironie, irony, ironia), et il n'est pas possible, sur la base de ces données, de savoir si une variation prosodique observée sera liée à des divergences conceptuelles ou à des différences de conventionnalisation de formes prosodiques associées à des concepts équivalents. Notons toutefois que ces inventaires sont tout à fait valides dans une optique d'enseignement des langues étrangères (Shochi, 2008), s'il conviennent mal à la comparaison interculturelle.

[1] « Face » au sens de Brown et Levinson (1987)

[2] Le fait qu'il s'agisse d'un japonais peut aussi jouer un rôle, les japonais ayant une solide réputation de politesse auprès des français

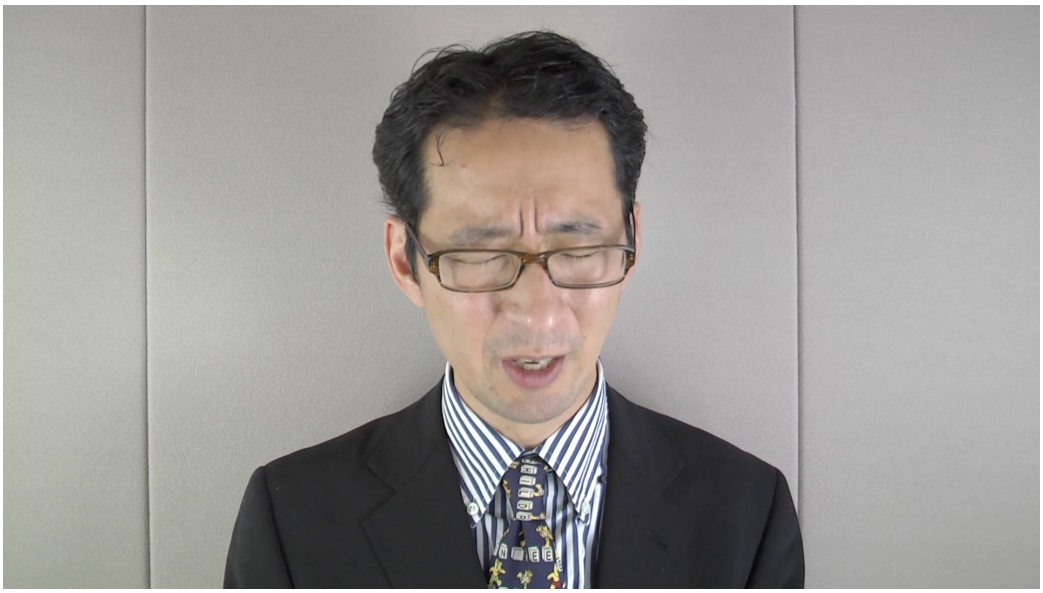


Figure 5.4 – *Image du locuteur japonais réalisant une expression de kyoshuku, prise au milieu de la phrase.*

5.4 Illocution & variations expressives

Afin de poursuivre cette étude des fonctions expressives de la prosodie, intéressons-nous maintenant à la distinction entre actes illocutoires et autres fonctions expressives. Les fonctions prosodiques illocutoires sont liées à la structure linguistique du message, au travers des distinctions modales, notamment déclarative, interrogative et impérative. L'approche de ces fonctions se fait dans le cadre d'une collaboration avec J. A. de Moraes et pose la question des relations entre le mode de la phrase et la réalisation d'autres fonctions expressives, que ce soit des attitudes prosodiques ou l'expression simulée d'émotions. La distinction entre attitudes propositionnelles et sociales (de Moraes, 2008), comme entre attitude / expression émotionnelle et mode de phrase sont centrales dans cette analyse. Ces études portent sur le portugais brésilien ; on en trouvera tous les détails dans de Moraes et Rilliard (2014a) pour les aspects attitudinaux et de Moraes et Rilliard (2014b) pour les expressions émotionnelles.

5.4.1 À propos de multimodalité

Avant d'aller plus loin, notons que cette expressivité prosodique a sa place dans la communication parlée en face-à-face, dont le fonctionnement est fondamentalement multimodal. Les travaux de Swerts et Krahmer (2005)

décrivent une « prosodie audiovisuelle », pour laquelle les deux modalités concourent à la transmission d'informations parallèles au message lexical. Les auteurs définissent les informations « prosodiques » comme « *the whole gamut of features that do not determine* what *speakers say, but rather* how *they say it* » (Swerts et Krahmer, 2005, p. 81). Notons que cette vision de la « prosodie » exclut toutes les fonctions prosodiques participant à la constitution du message linguistique. Dans une optique différente, Nadeu et Prieto (2011) utilisent des informations visuelles comme un contexte permettant à l'auditeur d'interpréter les variations prosodiques perçues. Pour ces auteurs, la modalité visuelle est parallèle à la modalité prosodique ; les informations visuelles contextualisent les informations prosodiques qui, dans leur approche expérimentale basée sur de courtes phrases modifiées par resynthèse, peuvent être délicates à interpréter pour les sujets. Malgré les différences importantes entre ces deux approches, ces travaux montrent l'importance de la modalité visuelle et des expressions faciales en particulier, dans le processus de communication. De même, plusieurs de nos travaux sur les attitudes valident l'importance de la multimodalité pour la compréhension des variations attitudinales (Rilliard *et al.*, 2008; Shochi *et al.*, 2008; Rilliard *et al.*, 2009; Shochi *et al.*, 2014).

Il nous a donc semblé particulièrement important, afin de mieux comprendre ces distinctions expressives, d'enregistrer et d'utiliser pour nos tests perceptifs des stimulus audiovisuels – auxquels il faudrait d'ailleurs rajouter la gestualité. Afin de nous concentrer ici sur les aspects prosodiques (au sens classique du terme), nous ne rapporterons pas ici le détail de la modalité visuelle, mais seulement les résultats qui permettent de mieux comprendre le fonctionnement des variations prosodiques.

5.4.2 Corpus

Le tableau 5.4 donne la liste des étiquettes des attitudes utilisées pour l'étude des attitudes, en fonction de leur caractère propositionnel ou social (de Moraes, 2008) et des trois modes étudiés (déclaratif, interrogatif, impératif). On notera que l'inventaire des attitudes propositionnelles varie en fonction du mode[1], ce qui n'est pas le cas des attitudes sociales. Toutes ces attitudes ont été enregistrées sur des phrases attitudinalement neutre, de longueur variable (1, 3 et 6 syllabes), et pour lesquelles le dernier mot de la phrase est soit oxyton, soit paroxyton (pour les phrases de plus d'une syllabe...). Les mêmes phrases sont utilisées pour les modes déclaratif et in-

[1]Il serait incohérent d'exprimer par exemple de l'évidence sur une question.

terrogatif, tandis que d'autres phrases sont nécessaires pour le mode impératif – phrases qui ont toutefois la même structure accentuelle. Toutes ces phrases ont été enregistrées dans les modalités audio et visuelles, par deux locuteurs (1 femme et 1 homme) ayant le portugais brésilien comme L1. Trois répétitions de chaque attitude et pour chaque phrase ont ainsi été enregistrées.

De manière similaire, une phrase de six syllabes terminée par un paroxyton est utilisée par les mêmes locuteurs afin d'enregistrer l'expression prototypique des émotions de colère, peur, joie et tristesse. Ces expressions émotionnelles sont enregistrées sur des phrases à modalités déclarative, interrogative et impérative (la même phrase est utilisée pour la déclaration et l'interrogation, une autre pour la phrase impérative).

5.4.3 Analyse prosodique

Les paramètres prosodiques de fréquence fondamentale (F_0) et de durée ont été analysés sur l'ensemble de ces phrases. L'analyse est faite grâce au logiciel STRAIGHT (Kawahara, 2008). Les valeurs de F_0 sont exprimées en demi-tons par rapport à la référence de 1 Hz, puis normalisées pour chaque

	Propositionnelles **portugais** *français [abréviation]*	**Sociales** **portugais** *français [abréviation]*
Déclaratif	dúvida *doute (DOU)* ironia *ironie (IRO)* incredulidade *incrédulité (INC)* evidência *évidence (OBV)* surpresa *surprise (SUR)*	arrogância *arrogance (ARR)* autoridade *autorité (AUT)* desprezo *mépris (CONT)* irritação *irritation (IRR)* polidez *politesse (POL)* charme *séduction (SED)*
Interrogatif	confirmação *confirmation (CONF)* estranheza *incrédulité (INC)* retoricidade *rhétoricité (RET)* surpresa *surprise (SUR)*	arrogância *arrogance (ARR)* autoridade *autorité (AUT)* desprezo *mépris (CONT)* irritação *irritation (IRR* polidez *politesse (POL)* charme *séduction (SED)*
Impératif	desafio *défi (CHAL)* pedido *requête (REQ)* sugestão *suggestion (SUG)* súplica *supplique (SUP)* conselho *défi (WAR)*	arrogância *arrogance (ARR)* autoridade *autorité (AUT)* desprezo *mépris (CONT)* irritação *irritation (IRR)* polidez *politesse (POL)* charme *séduction (SED)*

TABLE *5.4* *– Étiquettes des attitudes du portugais brésilien, en fonction du mode et du type d'attitude.*

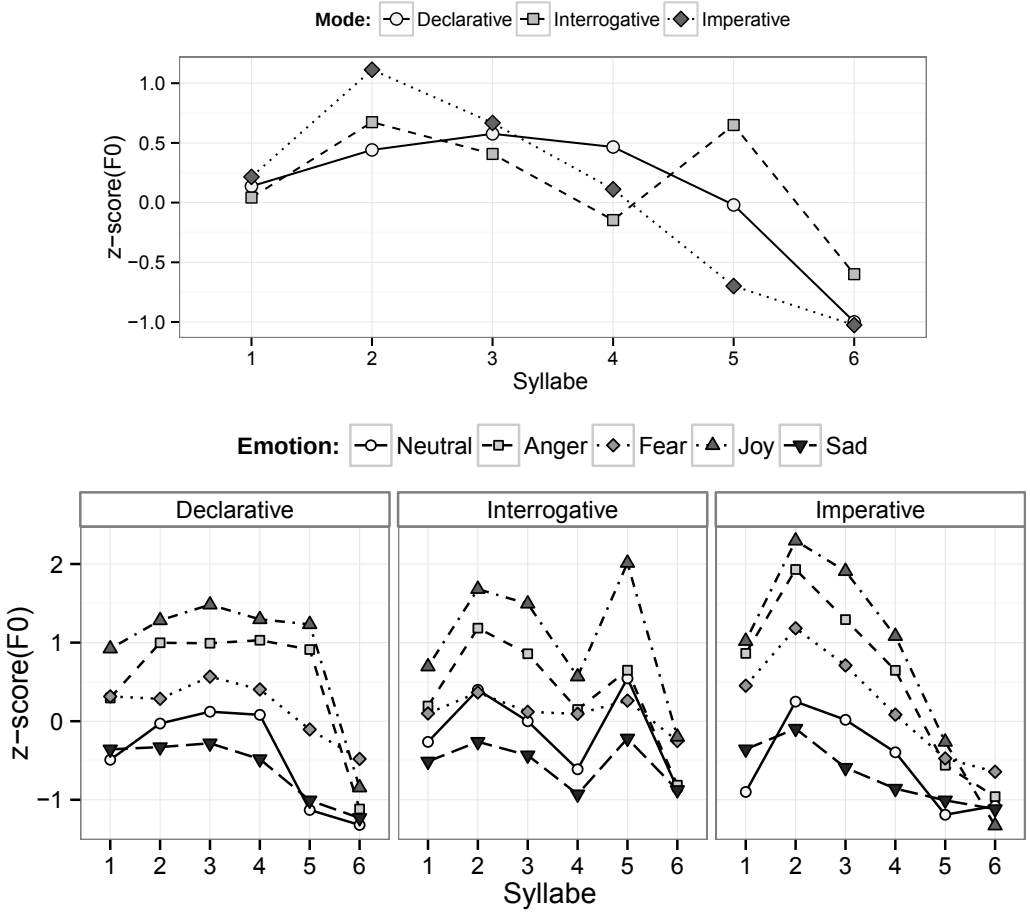


Figure 5.5 – *Tracés des contours intonatifs de chaque mode (en haut) et de chaque expression émotionnelle dans chaque mode de phrase (en bas), moyennes établies pour les deux locuteurs, sur une phrase paroxytonique de six syllabes. Chaque point représente la moyenne des valeurs de F_0 observées sur une voyelle particulière pour les trois répétitions d'une même attitude ; la F_0 est exprimée en terme de z-score des valeurs en demi-tons observées.*

locuteur, en en prenant le z-score. La figure 5.5 montre les contours obtenus pour le corpus d'expressions émotionnelles (et donc pour une phrase paroxytonique de 6 syllabes). Afin de rendre les données plus facilement comparables, les figures 5.6 et 5.7 représentent les variations de F_0 pour les expressions attitudinales, sur des phrases de six syllabes (oxytonique et paroxytonique).

La figure 5.5 montre sur le graphique du haut les formes moyennes des contours des trois modes de phrase – déclaratif, interrogatif, impératif. Ces contours de mode sont tout à fait compatibles avec les descriptions de la

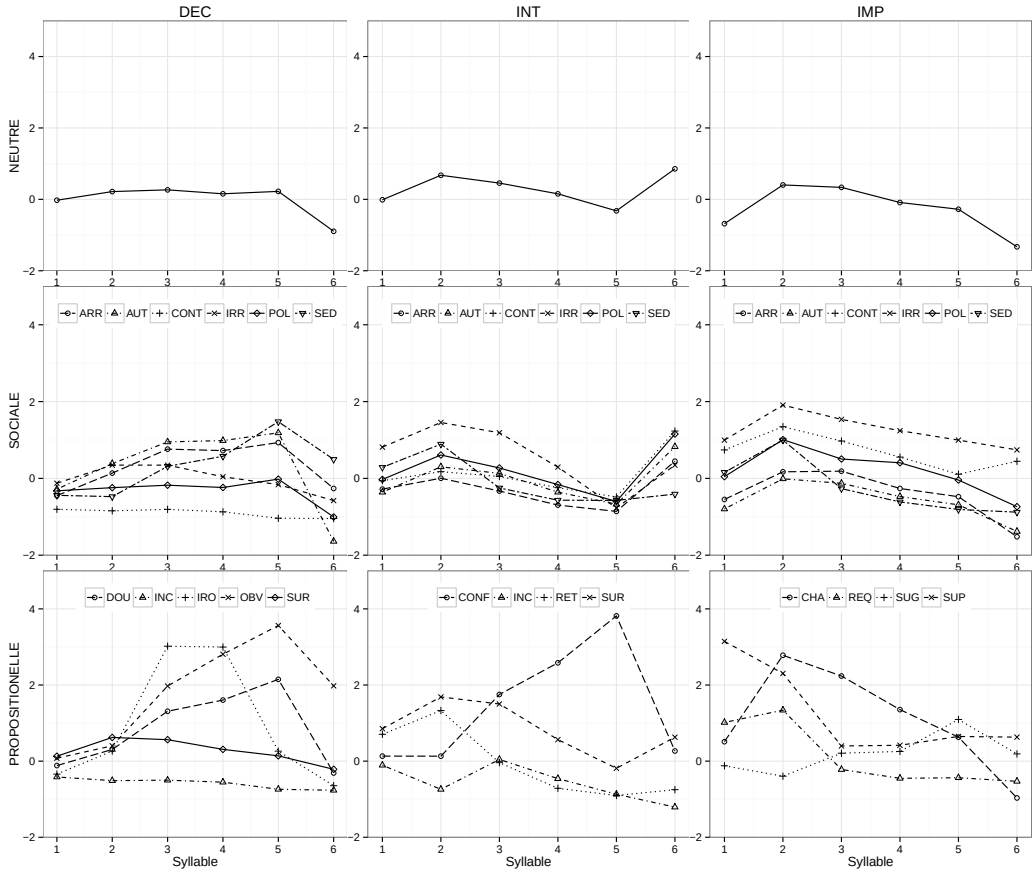


Figure *5.6* – *Tracés des contours intonatifs de chaque attitude produite par la locutrice, sur la phrase de six syllabes terminée par un oxyton. Chaque point représente la moyenne des valeurs de F_0 observées sur une voyelle particulière pour les trois répétitions d'une même attitude ; la F_0 est exprimée en terme de z-score des valeurs en demi-tons observées pour la locutrice. Les tracés des contours sont regroupés, par colonne selon le mode de la phrase (déclaratif, interrogatif, impératif), par ligne selon le type des attitudes (neutre, sociales, propositionnelles).*

littérature (de Moraes, 2008) ; on y observe aussi l'influence de la position de la dernière syllabe accentuée sur la forme du contour, et particulièrement pour la phrase interrogative (la montée finale se faisant sur la pénultième). Le graphique du bas montre, pour chacun de ces modes, la forme des contours intonatifs des quatre expressions émotionnelles. L'effet le plus important (voir les détails statistiques dans de Moraes et Rilliard, 2014b) est bien le respect de la forme du mode, quelle que soit l'émotion exprimée : on observe essentiellement une dilatation ou une contraction de l'amplitude de la variation

de F_0 (voir aussi Martin, 2014). Le facteur principal influant sur cette dilatation des contours semble être l'activation des expressions émotionnelles jouées par les deux locuteurs : la joie, puis la colère et la peur étant exprimées de manière plus intense, tandis que la tristesse aplatit le contour de F_0.

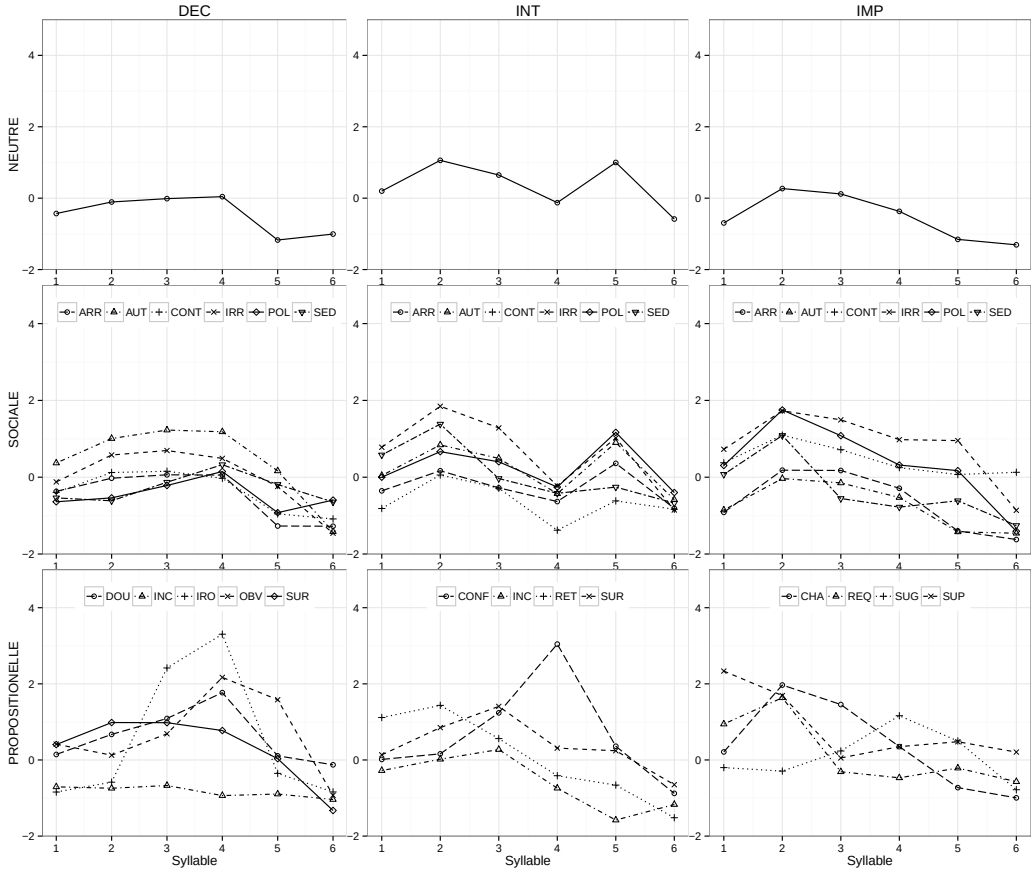


Figure 5.7 – *Tracés des contours intonatifs de chaque attitude produite par la locutrice, sur la phrase de six syllabes terminée par un paroxyton. Chaque point représente la moyenne des valeurs de F_0 observées sur une voyelle particulière pour les trois répétitions d'une même attitude ; la F_0 est exprimée en terme de z-score des valeurs en demi-tons observées pour la locutrice. Les tracés des contours sont regroupés, par colonne selon le mode de la phrase (déclaratif, interrogatif, impératif), par ligne selon le type des attitudes (neutre, sociales, propositionnelles).*

Si l'on considère les contours de F_0 qui sont proposés aux figures 5.6 et 5.7, cette fois pour les expressions attitudinales, on peut remarquer :

- que les contours des modes sont similaires à ceux observés pour le corpus d'expressions émotionnelles, avec toutefois des changements

liés à la position de l'accent : dans le cas des phrases interrogatives, la montée finale coïncide avec la position accentuelle ;
- que les attitudes sociales modifient peu la forme du contour du mode, particulièrement en comparaison de l'effet flagrant des attitudes propositionnelles sur la forme de ce contour.

En effet, les attitudes propositionnelles montrent des contours intonatifs propres à chacune d'elles, indépendamment de la modalité de la phrase. Cette spécificité des contours attitudinaux propositionnels respecte toutefois (comme les contours de mode) la structure accentuelle de la phrase : ainsi les contours des attitudes propositionnelles sont similaires tant pour la phrase oxytonique que la paroxytonique, compte tenu de la position de l'accent.

Cette observation renforce l'analyse d'un ancrage linguistique des informations modales, en opposition aux informations des expressions émotionnelles, qui n'affectent pas directement le contenu propositionnel, mais sont à l'origine de changements plus globaux, liés notamment aux différences d'activation (ou d'engagement) du locuteur. De manière similaire, les attitudes sociales affectent peu le contour de F_0 et leurs différences sont même moins marquées que ce que l'on observe pour les expressions émotionnelles – ceci sans doute du fait de moindres changements d'activation ou d'engagement (Daneš, 1994) pour ces expressions. Les attitudes propositionnelles, par contre, transmettant un contenu informatif affectant le sens de la phrase, modifient la forme du contour de modalité – tout en respectant la structure morpho-syntaxique.

5.4.4 Analyse perceptive

Interactions expressions émotionnelle *vs.* mode

On a vu que les modes montrent des contours intonatifs spécifiques, dont les expressions émotionnelles viennent perturber la dynamique. Deux tests de perception ont donc été menés afin de savoir dans quelle mesure ces variations expressives perturbent la reconnaissance mutuelle des expressions émotionnelles et des modes (de Moraes et Rilliard, 2014b).

Pour le test A, les sujets avaient pour tâche de reconnaitre l'émotion exprimée, quel que soit le mode ; pour le test B, il leur fallait reconnaitre le mode (entre déclaratif et interrogatif, l'impératif utilisant une phrase distincte empêchant son utilisation), quelle que soit l'émotion exprimée. Les figures 5.8 et 5.9 détaillent les résultats obtenus pour les tests A et B, respectivement.

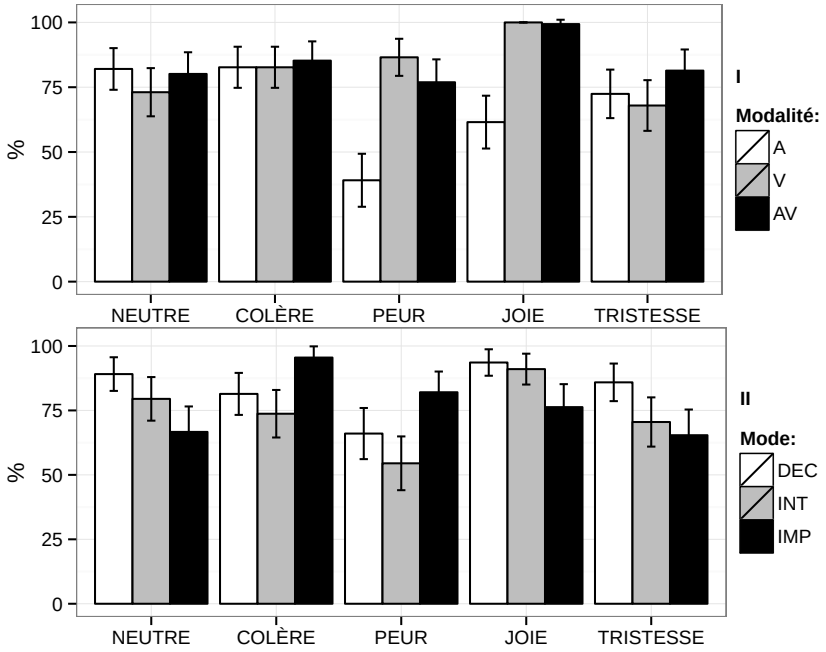


Figure 5.8 – *Pourcentage de reconnaissance de chaque expression émotionnelle dans le test A, en fonction de la modalité (Audio, Visuel, Audio-Visuel) de présentation (graphique du haut) et du mode – déclaratif, interrogatif ou impératif (graphique du bas). Les barres d'erreur indiquent l'intervalle de confiance à 95%.*

L'analyse des résultats du test A montre que la modalité de présentation des expressions émotionnelles (audio seul, visuel seul ou audiovisuel) ainsi que les interactions entre (1) la modalité de présentation et l'émotion exprimée, (2) l'expression émotionnelle et le locuteur et (3) l'expression émotionnelle et le mode de la phrase, sont les principaux facteurs explicatifs de la variabilité des réponses. La modalité est donc un facteur important dans le choix expressif du locuteur. Comme on peut l'observer sur le graphique du haut de la figure 5.8, la modalité visuelle domine la reconnaissance des expressions émotionnelles et ceci principalement du fait des expressions de peur et de joie, pour lesquelles l'audio prend une part bien moins déterminante que le visuel.

En ce qui concerne l'impact du mode sur la perception des expressions émotionnelles, il est moins important que celui de la modalité, mais tout de même significatif. Les phrases déclaratives reçoivent ainsi les meilleurs scores de reconnaissance des expressions émotionnelles (83%), alors que les interrogatives dégradent le plus les performances (74% ; et 77% pour les impératives). Mais cet effet du mode dépend surtout de l'expression émotionnelle

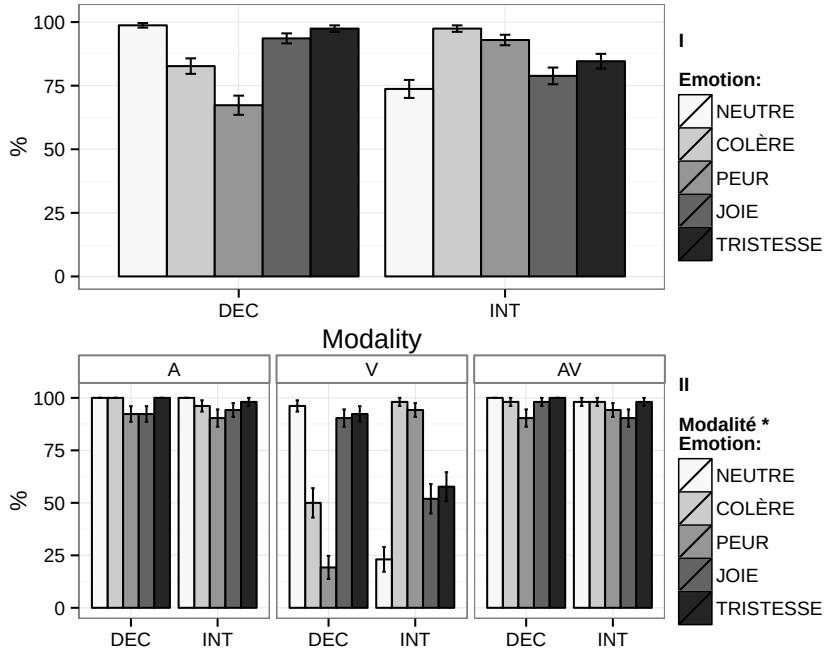


Figure 5.9 – *Pourcentage de reconnaissance de chaque mode (déclaratif ou interrogatif) dans le test B, en interaction avec l'expression émotionnelle (graphique du haut) et avec l'expression émotionnelle et la modalité (Audio, Visuel, Audio-Visuel) de présentation (graphique du bas). Les barres d'erreur indiquent l'intervalle de confiance à 95%.*

(voir le détail sur le graphique du bas de la figure 5.8). Ainsi, les phrases impératives permettent une meilleure reconnaissance des expressions de colère et de peur, tandis que les phrases interrogatives montrent un effet inhibiteur pour ces expressions. Inversement, les phrases impératives reçoivent des scores de reconnaissance plus bas que les phrases déclaratives pour les trois autres expressions (joie, tristesse et neutre). Le mode interrogatif montre aussi un effet inhibiteur sur la reconnaissance des expressions neutre et triste, mais un effet positif pour la reconnaissance de la joie.

Les effets observés ressortissent à une logique assez simple. Ainsi, une phrase sera moins *neutre* dès que sa prosodie véhicule un contenu supplémentaire à celui du niveau lexical ; les modes interrogatifs et impératifs ne sont donc pas *neutres*, au moins d'un point de vue illocutoire. Une phrase impérative véhicule un ordre ; cela convient donc bien à la colère, mais aussi à l'expression d'avertissements liés à la peur. Inversement, ordre et manifestations de joie ou de tristesse sont plus antagonistes. Le mode interrogatif ne montre aucun effet facilitateur et cela peut être dû à l'absence d'expressions émotionnelles dubitatives telle que la surprise (cf. section 5.3.2) : toutes les

expressions étudiées sont assertives, aussi au mieux le mode interrogatif ne dégrade pas les performances de reconnaissance. Par contre l'interrogation transmet une demande, qui semble antagoniste aux expressions de colère et de peur.

Pour le test B, on observe un effet principal de la modalité de présentation des stimulus sur la capacité des auditeurs à reconnaitre le mode. La modalité visuelle est ici la moins performante pour indiquer ce type d'information : en modalité visuelle seule, les sujets obtiennent 67% de bonnes réponses, contre 96% voir 97% pour les modalités audio et audiovisuelle. Le second facteur (par ordre d'importance de l'effet) est l'interaction entre le mode et l'expression émotionnelle (cf. le graphe du haut de la figure 5.9) : les expressions de peur et de colère induisent un biais vers la perception de l'interrogation, tandis que les autres expressions émotionnelles biaisent les réponses vers le mode déclaratif (avec toutefois des scores de reconnaissance supérieurs au hasard). On observe un effet de l'interaction triple entre le mode, l'expression émotionnelle et la modalité de présentation : il est clair que les mauvais scores de reconnaissance de certains modes pour certaines expressions émotionnelles sont liés à la présentation visuelle seule. En modalité visuelle, les expressions de colère et de peur empêchent la reconnaissance du mode déclaratif, tandis qu'à l'inverse, les expressions émotionnelles neutres, de joie ou de tristesse rendent aléatoire la perception de l'interrogation. Par contre, à partir du moment où les sujets ont accès à la modalité audio, ils reconnaissent parfaitement le mode de la phrase.

Ces tests de reconnaissance supportent donc l'idée que la modalité audio est avant tout utilisée pour transmettre le mode de la phrase, puis des informations d'ordre émotionnel, tandis que la modalité visuelle véhicule essentiellement des informations du niveau de l'expression émotionnelle.

Les expressions d'attitudes constituent des exemples de l'engagement du locuteur dans son discours, selon le concept discuté par Daneš (1994) ; il s'agit donc d'expressions mettant en œuvre le corps dans son entier. Ces attitudes sont clairement multimodales et les différentes modalités transmettent une certaine redondance, qui participe de la robustesse nécessaire à la communication parlée. Pour autant, ces différentes modalités semblent aussi se répartir le type d'information préférentiellement transmis. Les aspects les plus fortement encodés dans le langage sont essentiellement transmis par le canal audio, tandis que plus les informations concernent la gestion de la relation interpersonnelle, plus la modalité visuelle est recrutée.

Les inventaires attitudinaux étudiés dans différentes langues ne sont pas strictement les mêmes ; cela peut être interprété comme un indice du caractère culturellement spécifique de ces expressions. Cependant, les principales dimensions conceptuelles qui organisent la répartition de ces concepts attitudinaux montrent des ressemblances frappantes.

Les études décrites dans ce chapitre ne peuvent toutefois pas répondre à la question de savoir ce qui est similaire ou différent d'une langue/culture à l'autre, du symbolisme (ou du code) transmis par la prosodie, ou des concepts sous-jacents à ces expressions. Afin de tenter de répondre à ces questions, il est nécessaire de séparer les deux niveaux – l'expressivité prosodique et la description conceptuelle de ces attitudes communicatives. C'est à cela que les travaux présentés dans le prochain chapitre vont s'atteler, cette tâche constituant bien entendu un travail en cours.

6 | Variation interculturelle

> - Eh bien, puisque tu te sacrifies à mon service, Carlotta, il faut, pour mon service toujours, que tu sois très éprise du roi de Navarre, et très jalouse surtout, jalouse comme une Italienne.
> – Mais, madame, demanda Charlotte, de quelle façon une Italienne est-elle jalouse ?
>
> Dumas, *La reine Margot*,
> tome 1, chapitre XV (1847)

Le chapitre précédent a montré un ensemble d'expressions transmises par les variations prosodiques, allant d'actes illocutoires jusqu'à l'expression d'émotions simulées ou spontanées. Toutes ces expressions n'influent pas de la même manière sur les variations intonatives. En particulier, les expressions qui modifient le sens du message exprimé verbalement par le locuteur, changent considérablement la réalisation prosodique. Ces changements sont étroitement liés à la structure morphosyntaxique de l'énoncé ; partant, on peut s'attendre à ce que ces variations puissent avoir des aspects spécifiques aux différentes langues dans lesquelles elles sont réalisées. On a aussi vu (cf. partie 5.3.2) que les inventaires d'attitudes varient en fonction des langues, mais que ces expressions se regroupent en quelques grandes dimensions sémantiques qui sont communes aux différentes langues étudiées (dimensions perçues par des auditeurs natifs de chacune des langues étudiées).

Pour aller plus loin dans la compréhension de cette variation expressive interculturelle, de nombreux problèmes sont à surmonter, qui ont déjà été évoqués (cf. partie 5.3.2) : la traduction d'un concept laisse en général la place à une variation des représentations de ces concepts dans les différentes langues considérées, comme l'ont bien montré les travaux de Wierzbicka (1985, 1986b, 1996b, 1999, 2005). Mais comment étudier la perception d'expressions attitudinales par des locuteurs étrangers à une langue donnée, sans recourir à la traduction ?

Un paradigme a été développé pour répondre à ce problème, qui est décrit de manière exhaustive dans Rilliard *et al.* (2014b) (voir aussi Rilliard *et al.*, 2014a). Ce travail se donne pour but d'analyser la perception, par des sujets de différentes cultures, d'un ensemble d'expressions prosodiques japonaises exprimant diverses formes de politesse et d'impolitesse, et de comparer cette perception des variations prosodiques à la compréhension des concepts associés à ces réalisations dans la langue japonaise.

Afin de dissocier les traitements perceptifs des performances prosodiques et des concepts, plusieurs modalités de stimulus sont utilisées en concurrence :

- la performance comportementale du sujet exprimant l'attitude considérée, c'est-à-dire la présentation audiovisuelle du locuteur ;
- la présentation audio seule du même enregistrement ;
- la présentation visuelle seule du même enregistrement ;
- la présentation du concept, sans la performance.

Pour comparer les différentes attitudes dans chacune de ces modalités, un test de comparaison par paire a été choisi, les auditeurs ayant la tâche de juger de la différence perçue entre deux stimulus présentés dans la même modalité. Ce test, et les traitements statistiques des données qui en sont issues, sont basés sur la méthode présentée par Romney *et al.* (1997) ; plus généralement, les travaux de Romney *et al.* (1996); Romney et Moore (1998); Romney *et al.* (2000) ont beaucoup inspiré la méthode d'analyse. On pourra aussi se référer aux travaux pionniers d'Osgood *et al.* (1957, 1975) dans ce domaine de la mesure des variations sémantiques et au concept de « *semantic differential* » qui y est développé.

6.1 Scripts culturels pour les attitudes

Comment présenter à des sujets de langues maternelles différentes les concepts de ces cinq attitudes en limitant au maximum le biais de traduction ? La solution adoptée consiste à utiliser le principe des *scripts culturels* (Wierzbicka, 1996a; Goddard, 1997), décrits dans le « *Natural Semantic Metalanguage* » (NSM) proposé et développé par Wierzbicka (1985, 1986b); Goddard (2002); Wierzbicka (2005); Goddard (2007). Le NSM consiste en un ensemble restreint de primitives renvoyant à des concepts indéfinissables, observés dans toutes [1] les langues ; ces primitives universelles forment un (méta)langage permettant de définir des concepts plus complexes (et non

[1] Les défenseurs de cette théorie (Goddard, 2002; Wierzbicka, 2005) soutiennent l'universalité de ces primitives ; voir les travaux de Bohnemeyer (2004); Koptjevskaja-Tamm et Ahlgren (2003); Riemer (2006) ou Wawrzyniak (2010) pour des critiques de cette approche, au-delà de leur postulat d'universalité.

universaux) de manière équivalente dans différentes langues, en utilisant le *Natural semantic métalanguage* – l'ensemble des primitives – de chacune de ces langues.

Nous avons donc défini les scripts culturels correspondant aux cinq attitudes de politesse et d'impolitesse retenues pour notre étude. Ces cinq attitudes prosodiques japonaises sont les suivantes :

- une expression de politesse de courtoisie (POLI – 丁寧 – *teinei*), qui est utilisée par le locuteur dans un contexte social neutre, pour s'adresser à un interlocuteur avec lequel il n'a pas de différence hiérarchique ;
- une expression de sincérité-politesse (SINC – 誠意 – *seii*), qui est utilisée lorsque le locuteur se trouve socialement inférieur à son interlocuteur, dans le but de l'assurer de ses intentions sérieuses et sincères ;
- une expression de *kyoshuku* (KYOS – 恐縮 – *kyoshuku*), qui est utilisée lorsque le locuteur doit exprimer une opinion contradictoire, ou rechercher une faveur auprès d'un interlocuteur socialement supérieur ; alors, l'expression de *kyoshuku* est définie comme « *corresponding to a mixture of suffering ashamedness and embarrassment, comes from the speaker's consciousness of the fact his/her utterance of request imposes a burden to the hearer* » (Sadanobu, 2004, p. 34) ;
- une déclaration neutre (DECL – 平叙 – *heijo*), qui est utilisée pour donner une information, sans exprimer de point de vue (cette expression n'est en soit ni polie ni impolie) ;
- une expression d'arrogance (ARRO – そんさい – *zonzai*), qui est utilisée pour exprimer le sentiment impoli de supériorité que ressent le locuteur envers son interlocuteur.

La table 6.1 donne les scripts NSM qui définissent les traits sémantiques de ces expressions (nous donnons ici la version anglaise seulement ; pour les autres langues se référer à Rilliard *et al.*, 2014b). Ces scripts ont été établis en s'inspirant des travaux sur la définition des émotions de Wierzbicka (1992, 1996a); Harkins et Wierzbicka (2001); Goddard (2007). Plus spécifiquement, les travaux de Aznárez-Mauleón et González-Ruiz (2006) ont été importants pour exprimer la notion de sincérité contenue dans l'expression prosodique de sincérité-politesse ; de même, le travail de Bartens et Sandström (2006) sur les fonctions sémantiques des diminutifs en espagnol, exprimant euphémisme et intensification, a été utile à la création de ces définitions. Les scripts ont été écrits sur la base des exposants de l'anglais tels que présentés par Goddard et Wierzbicka (2002). Les traductions des scripts NSM se sont basées sur les travaux de Goddard (2012) pour le japonais et de Peeters (2006) pour le portugais et le français.

Chacune des attitudes prosodique correspondant à ces concepts sont produites par un locuteur japonais (il s'agit des mêmes stimulus que ceux utilisés

Declaration	**Arrogance**
X is someone like me I want to say something to X I think like this : when I say it to X, I want X to know this Because of this, I want to say this X knows it because of this I say it like this	X is someone below me I want to say something to X I think like this : when I say it to X, I want X to feel something bad Because of this, I want to say something bad X knows what I think because of this I say it like this
Courtesy Politeness	**Sincerity Politeness**
X is someone like me I want to say something to X I think like this : when I say it to X, I want X to feel something good Because of this, I want to say something good X knows what I think because of this I say it like this	X is someone above me I want to say something to X I think like this : when I say it to X, I want X to know it is true Because of this, I want to say something more X knows what I think because of this I say it like this
Kyoshuku	
X is someone above me X is someone near me I want something I want to say this to X because I want X to know what I want I think like this : when I say it to X, X will feel something very bad I don't want not to say it because of this Because of this, I want to say I feel something very bad X knows what I feel because of this I say it like this	

Table *6.1* – *Scripts NSM écrits sur la base des primitives de l'anglais, décrivant les concepts transmis par les cinq attitudes étudiés.*

dans les expériences sur le japonais décrites auparavant) et enregistrées dans les modalités audio et visuelle. Les réalisations sont produites sur une même phrase ayant un contenu lexical neutre du point de vue des expressions étudiées (la phrase japonaise « *Nagoyade nomimas[u]* », « il boit à Nagoya »). Les réalisations utilisées ont été choisies pour la qualité des performances, évaluée lors d'autres travaux (Rilliard *et al.*, 2009). Ces performances audiovisuelles actualisent dans la parole les concepts discutés ici, dans le but de les communiquer à un interlocuteur.

6.2 Evaluation comparative des attitudes

La question est donc maintenant d'évaluer dans quelle mesure ces concepts attitudinaux de la langue japonaise sont compris par des locuteurs de dif-

férentes langues et dans chacune de ces modalités – les concepts purs, ou leur actualisation dans les différentes modalités. La question de cette mesure du sens est abordée dans une optique provenant des travaux d'Osgood *et al.* (1957, 1975), en suivant la méthodologie expérimentale proposée par Romney *et al.* (1996, 1997, 2000). Au lieu de demander explicitement aux sujets de juger telle ou telle qualité d'un stimulus, leur tâche consiste ici à juger de la similarité entre deux stimulus. La consigne ne fait donc en aucun cas explicitement référence aux différentes attitudes qui sont exprimées – il s'agit seulement de savoir dans quelle mesure le locuteur exprime la même intention, pour deux stimulus.

Un intérêt particulier de cette consigne, outre celui d'études interculturelles, est de se passer de toute référence conceptuelle pour les tests de perception. Il est donc possible d'interroger la perception d'enfants ne maitrisant pas nécessairement l'écriture, ni certaines lexies complexes. Ces travaux sur le développement de la perception des attitudes prosodiques par des enfants japonais ne seront pas rapportés ici (voir Rilliard *et al.*, 2012).

Les dix paires formées sur la base de ces cinq attitudes sont donc présentées à des auditeurs originaires du Japon, des États-Unis, du Brésil et de France. Ces paires sont présentées selon les quatre modalités suivantes : audio seule, visuelle seule, audiovisuelle et conceptuelle (par présentation des scripts NSM de la table 6.1). Les réponses sont données sur une échelle de similarité allant de 1 à 9. Une matrice des distances perçues par chaque sujet entre chaque paire de stimulus, dans chaque modalité est construite sur la base de ces données et une analyse de correspondances est menée sur cette matrice (voir le détail des traitements et des normalisations effectués dans Rilliard *et al.*, 2014b). En regroupant la position des points obtenus par attitude, modalité et groupe linguistique, il est possible d'obtenir une représentation de la structure de cet ensemble de données (voir la figure 6.1). Il ressort clairement de ces représentations que les « structures sémantiques » (Romney *et al.*, 2000) perçues par les sujets sont essentiellement similaires, quelles que soit la modalité de présentation ou l'origine culturelle. La première dimension des analyses de correspondances oppose l'expression de *kyoshuku* aux quatre autres attitudes. La seconde dimension oppose l'expression d'arrogance aux attitudes de déclaration, de politesse de courtoisie et de sincérité-politesse. La troisième dimension (non représentée sur les graphiques, et qui explique 13% de la variance) oppose la déclaration aux deux politesses.

Ces grandes oppositions conceptuelles ne doivent pas masquer l'existence de différences entre modalités, comme entre groupes linguistiques. La présentation en modalité audio seule est particulièrement intéressante à cet égard. Cette modalité occupe, comparée aux trois autres, l'espace le plus restreint (les points sont plus concentrés vers le centre du graphique plan, notamment

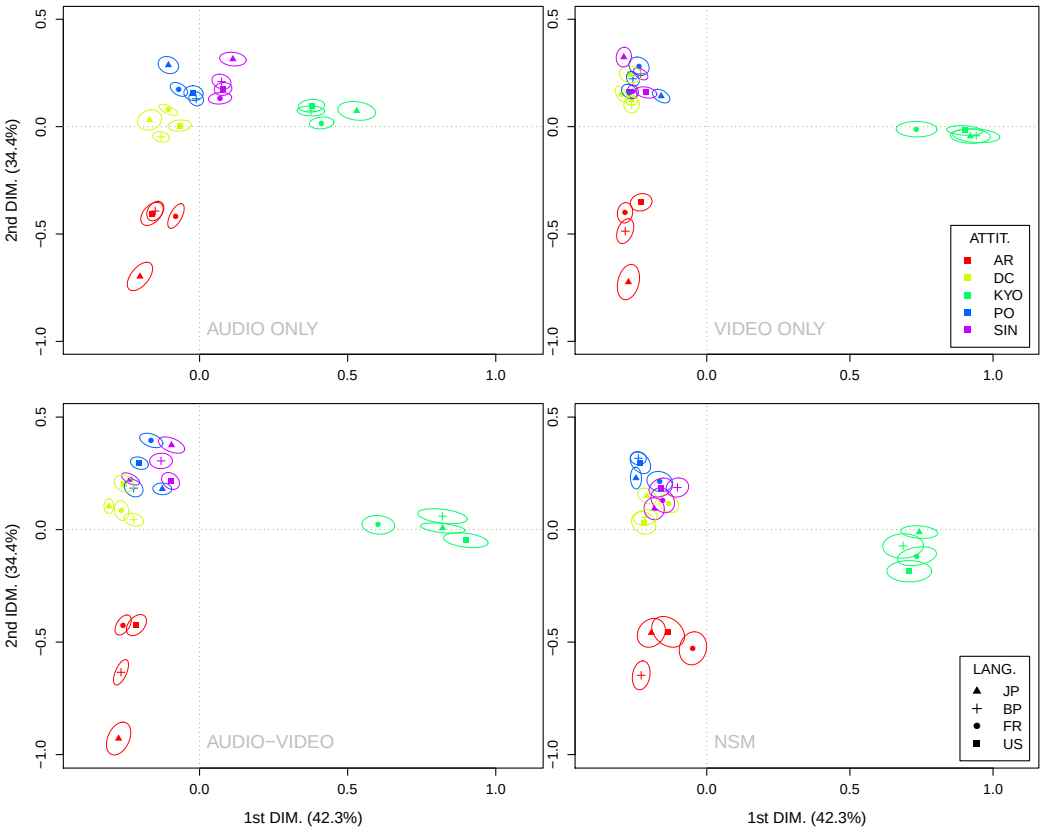


Figure 6.1 – *Dispersion des attitudes sur le graphique plan des deux principales composantes de l'analyse de correspondances menée sur le test de jugement par paires. Les données sont présentées par modalité, pour les quatre groupes linguistiques.*

du fait de la position de l'expression de *kyoshuku*). On peut en induire que les différences transmises par la prosodie sont moins fortes que celles transmises par les autres modalités. Cependant, c'est dans cette modalité (avec la modalité audio-visuelle) que les distinctions entre les cinq attitudes sont les plus claires : les points des différents groupes linguistiques forment cinq groupes distincts, un pour chaque attitude, alors que seuls trois groupes peuvent être observés pour la modalité visuelle. Les indices prosodiques semblent donc être moins clairs, mais plus subtils, plus complets, que les indices visuels, pour ces attitudes. De même, on observe des différences entre groupes linguistiques. Toujours pour la modalité audio, les points des attitudes perçues par des sujets japonais forment le groupe le plus large : il semble donc (et cela était attendu) que des auditeurs natifs soient à même de plus clairement distinguer les différents affects sociaux sur la seule base des indices prosodiques.

Ces analyses qualitatives de la structure sémantique de ces cinq attitudes doivent cependant être quantifiées : dans quelle mesure les jugements liés aux groupes de modalités et de langues diffèrent-ils ? Pour répondre à cette question, on calcule la similitude des formes des dispersions dans l'espace de l'analyse de correspondances, observées pour les cinq attitudes de chaque groupe de modalité et d'origine linguistique (la méthode de comparaison employée vient des travaux de Rao et Suryawanshi (1996), adaptés par Romney *et al.* (2000) – voir Rilliard *et al.* (2014b) pour plus de détails sur l'application de la méthode à ces données). En observant ainsi la part de variance commune ou spécifique à ces groupes, on obtient la répartition de la figure 6.2. Ce calcul montre qu'une part majoritaire (60%) de la structure sémantique est partagée par l'ensemble des sujets, tandis que 8% de cette structure sont spécifiquement liés à la modalité de présentation, et seulement 3% à l'origine linguistique. Les 29% restants sont liés aux différences entre sujets et au bruit inhérent à tout type d'évaluation perceptive.

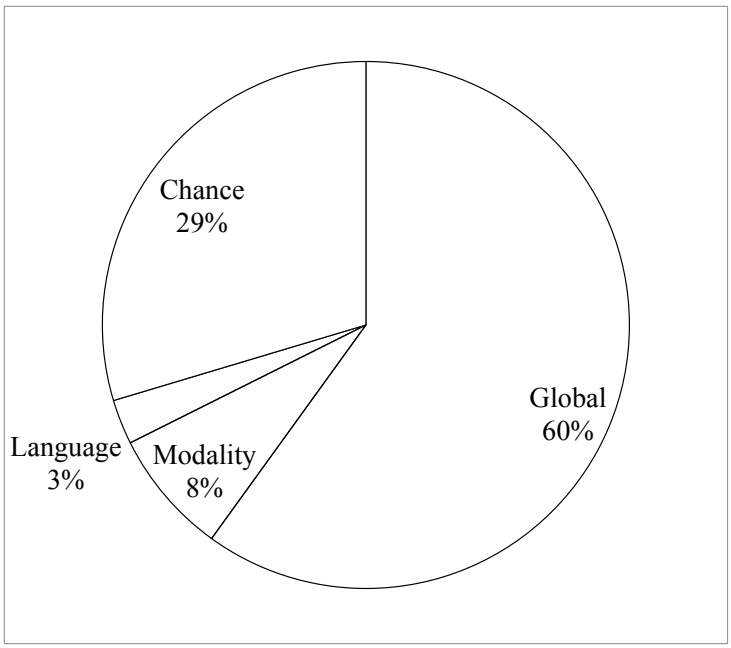


Figure *6.2* – *Contribution relative des différents facteurs à la connaissance partagée de la structure sémantique des cinq attitudes japonaises de (im)politesse (d'après Rilliard* et al.*, 2014b).*

Si l'on compare les réponses dans les différentes modalités de présentation, on remarque que la présentation des attitudes dans les modalités NSM et audio seule induisent le moins de cohésion entre sujets, tandis que les modalités visuelle et audiovisuelle montrent la plus grande cohérence entre les

sujets. Par ailleurs, les deux modalités audio seule et NSM sont les modalités montrant la plus grande divergence entre elles. La plus grande complexité des distinctions faites dans ces modalités peut en partie expliquer la dispersion des jugements. Cependant, le résultat intéressant est que la présentation conceptuelle (NSM) diverge au maximum de la présentation audio – on peut conclure de cette observation que les informations prosodiques véhiculent certains aspects de la sémantique attitudinale, mais moins certains autres, qui sont plus prégnants dans la modalité visuelle. Si l'on observe les répartitions de la figure 6.1, on peut voir que la principale spécificité de la modalité audio vient de la position de l'attitude de *kyoshuku*, bien moins à droite de la première dimension qu'elle ne l'est pour les autres modalités. La perception de la performance vocale de l'expression de *kyoshuku* constituerait donc une part importante des différences perçues entre modalités. Il est intéressant de noter que cette différence est similaire, quelle que soit l'origine des auditeurs, alors que l'expression de *kyoshuku* est typiquement japonaise et n'a pas été conceptualisée dans les autres cultures observées durant ce test.

6.3 Comparer la production d'attitudes

6.3.1 Des situations pour comparer les expressions

Les résultats de l'étude précédente montrent que l'organisation perceptive de quelques affects sociaux véhiculant des concepts typiques d'une culture donnée (ici, la culture japonaise) dépend peu de la culture de l'auditeur. Cette capacité à décoder des variations expressives n'ayant pas fait l'objet d'une convention dans sa propre culture soulève des questions, et en premier lieu celle de la variation des productions attitudinales au travers des cultures. Si des sujets perçoivent grosso-modo la même information, produisent-ils aussi la même information pour encoder des variations expressives ? Et comment étudier ces variations des stratégies d'encodage prosodique dans différentes cultures, étant donnés les problèmes conceptuels soulevés par Wierzbicka (1992) ? Pour répondre à ce défi, et dans l'esprit des « scripts culturels » prônés par Wierzbicka (1996a), nous avons choisi de mettre en place des enregistrements qui ne demandent pas aux locuteurs la production d'expressions attitudinales liées à une étiquette (étiquette qui renverrait aux « concepts populaires » liés aux termes utilisés dans les différentes langues), mais plutôt des enregistrements qui mettent les locuteurs dans des situations d'interaction décrites par des scripts et aboutissant à la production de phrases cibles. Ces scripts ne sont pas décrits en utilisant le NSM, mais la langue d'interaction (ils diffèrent en cela des préconisations de Wierzbicka, 2005), qui peut

être le japonais, l'anglais des États-Unis, le français hexagonal (Rilliard *et al.*, 2013), le portugais brésilien (travail en cours) ou l'allemand (Hönemann *et al.*, 2014)[1]. Ce choix d'utilisation de la langue naturelle est motivé par la plus grande simplicité de son utilisation par des locuteurs non experts (et surtout pour décrire des dialogues ayant un minimum de vraisemblance) : les enregistrements sont déjà complexes et la compréhension de situations exprimées au travers de scripts NSM aurait pu rajouter encore à la charge cognitive des locuteurs. Toutefois la description des situations d'interaction en termes de traits sémantiques simples a bénéficié de la démarche prônée par Wierzbicka (2005) et Goddard (2007).

Les situations d'interaction utilisées pour enregistrer différentes attitudes permettent de proposer à des locuteurs de différentes origines culturelles la même tâche expressive, sans avoir à leur demander une performance prosodique dépendante d'une étiquette. Les locuteurs doivent réaliser un acte de parole répondant à un but communicatif dans un contexte d'interaction bien spécifié ; et ces caractéristiques de l'échange verbal sont les mêmes pour tous les locuteurs, de quelque origine qu'ils soient. La principale variable est alors la langue que les locuteurs vont utiliser durant l'échange verbal, leur maitrise de cette langue, et leur capacité expressive.

Seize contextes d'interaction ont été définis. Ces contextes correspondent à différentes situations de communication attestées dans différentes langues et définies d'après des inventaires d'attitudes étudiées dans différentes langues (Shochi *et al.*, 2009c; de Moraes *et al.*, 2010). Une sélection de ces attitudes a été faite afin d'inclure des attitudes qui se retrouvent dans chacune des langues étudiées, mais aussi des attitudes spécifiques à certaines de ces langues et/ou qui ne font pas partie de l'inventaire spécifique d'une de ces langues. Le choix d'inclure des attitudes spécifiques à une langue vise à comparer le comportement de locuteurs possédant ces attitudes dans les inventaires de leur L1, *vs.* celui d'autres locuteurs ne possédant pas ces attitudes dans l'inventaire de leur L1. Il est parfaitement possible de faire une telle étude sur la base de ce paradigme, car il ne s'agit pas de demander à un locuteur de réaliser une attitude qui corresponde à un concept qu'il ne maitrise pas (et donc qu'il pourrait comprendre différemment), mais plutôt de lui demander de se comporter de manière naturelle dans une situation de communication plausible. Ces seize attitudes correspondent aux situations prototypiques suivantes, dans lesquelles le sujet incarne le locuteur A interagissant avec un interlocuteur B (l'expérimentateur durant les enregistrements) :

[1]Les quatre premières langues étaient prévues dès l'origine du projet PADE, l'allemand s'est rajouté récemment et d'autres langues, notamment l'hébreu, pourraient être l'objet de la même procédure d'enregistrement.

- **déclaration** (DECL) : A & B sont des collègues du même âge ; A donne une information sans y ajouter de perspective personnelle ; la scène se passe au café.
- **évidence** (OBVI) : A & B sont des collègues du même âge ; tout le monde sait que Paul n'aime pas les films indiens, mais A demande à B si Paul aime les films indiens ou non ; la scène se passe au café.
- **ironie** (IRON) : A & B sont des amis du même âge ; A doit se rendre à Marseille pour un match de foot important et B, qui habite à Marseille téléphone à A. Manque de chance, il neige à Marseille et B dit que c'est magnifique ; la scène se passe à la gare.
- **admiration** (ADMI) : A & B sont du même âge et se connaissent bien. Tous deux adorent la cuisine française et parlent du repas délicieux qu'ils ont partagé hier dans un fameux restaurant français. La scène se passe dans un café.
- **politesse** (POLI) : A & B sont du même âge et ne se connaissent pas vraiment, mais travaillent dans la même entreprise. A est assis à côté de B et les deux échangent des mondanités. La scène se déroule à un cocktail organisé par l'entreprise.
- **sincérité** (SINC) : B est le chef de la section dans laquelle A travaille ; B est plus âgé que A. Le chef (B) voudrait que A prenne la responsabilité d'un grand projet ; A est content de se voir confier cette responsabilité et exprime sincèrement son enthousiasme et sa volonté de bien remplir sa tâche. La scène se passe dans le bureau de B.
- « **marcher sur des œufs** » (WOEG) : B est le chef de la section dans laquelle A travaille ; B est plus âgé que A. Le chef (B) voudrait que A prenne en charge une tâche qui demande beaucoup de travail, et cela semble impossible à A. A tente de refuser cette demande, de façon à ce que B ne soit pas fâché de ce refus. La scène se passe dans le bureau de B.
- **arrogance** (ARRO) : A & B étudient dans la même université, mais A est plus âgé et son père est le doyen de l'université et il est très snob. Tous les deux se connaissent, mais ne sont pas amis. A organise une soirée à laquelle B n'est pas invité, mais A apprend la présence de B à la soirée. La scène se déroule à la salle des fêtes et A dit à B qu'il n'est pas invité.
- **autorité** (AUTH) : A est un agent de la police aux frontières ; B est un voyageur. B est en face de A, et demande la permission d'entrer dans le pays ; A doit montrer son autorité ; la scène se déroule au guichet des douanes de l'aéroport.
- **mépris** (CONT) : A & B étudient dans la même université, mais A est plus âgé. Tous les deux se connaissent, mais ne sont pas amis. En

fait, A déteste B. A organise une soirée à laquelle B n'est pas invité, mais A apprend la présence de B à la soirée. La scène se déroule à la salle des fêtes et A dit à B qu'il doit partir.

- **irritation** (IRRI) : A & B sont étudiants et travaillent ensemble à la bibliothèque. A chantonne et gigote sur sa chaise, ce qui empêche A de se concentrer. A veut que B s'arrête et exprime son irritation. La scène se passe à la bibliothèque universitaire.
- **séduction** (SEDU) : A est amoureux de B et ils vivent en couple. A fait un compliment à B de manière explicitement provocante afin d'attirer son attention et son intérêt. La scène se passe dans une discothèque.
- **question** (QUES) : A & B sont des collègues du même âge. A demande une information, sans perspective personnelle, attendant une réponse simple. La scène se passe au café.
- **incertitude** (UNCE) : A & B des collègues du même âge. A pense avoir vu Paul au match de foot la veille, mais n'est pas sûr à 100% qu'il s'agissait bien de Paul ; la scène se déroule au café.
- **doute** (DOUB) : A & B sont des collègues, du même âge. A sait que B n'a pas été voir le match de foot la veille. Pourtant B prétend y être allé, et A ne le croit pas. La scène se passe au café.
- **surprise** (SURP) : A & B sont des amis du même âge. A ne sait pas que B est un excellent chanteur. Un jour, B montre sa belle voix à A. La scène se déroule chez B.

Parmi ces situations, celles de « marcher sur des œufs » et de sincérité sont inspirées des travaux sur le japonais. Il s'agit respectivement de situations correspondant aux expressions de *kyoshuku* (Sadanobu, 2004) et de sincérité-politesse (voir section 5.3.1), typiques de la culture japonaises, mais qui ne sont pas l'objet d'une convention dans les cultures occidentales. Inversement, la situation de séduction n'est pas conventionnalisée en japonais, au moins pour les locuteurs masculins.

Pour chacune de ces seize situations, les paramètres suivants sont contrôlés : s'il s'agit d'une attitude propositionnelle ou sociale (de Moraes, 2008), la distance hiérarchique entre A et B, leur distance sociale (voir Spencer-Oatey, 1996, pour plus de détails sur ces notions de distance), la valence de l'acte de parole et s'il comporte ou non une imposition de A envers B. Ces caractéristiques sont résumées à la table 6.2 pour chaque situation.

Les locuteurs recrutés sont enregistrés dans deux langues[1] : leur langue seconde (L2) puis leur langue première (L1), dans les mêmes contextes, et

[1]Un petit nombre de locuteurs trilingues sont enregistrés dans trois langues, tandis que certains ne sont enregistrés que dans une seule langue.

	Type	**Distance hiérarchique**	**Distance sociale**	**Valence**	**Imposition de A sur B ?**
DECL	Prop.	A = B	2	-	-
QUES	Prop.	A = B	2	-	-
UNCE	Prop.	A = B	2	-	-
SURP	Prop.	A = B	2	-	-
DOUB	Prop./Soc.	A = B	2	Negative	Yes
OBVI	Prop./Soc.	A = B	2	Negative	Yes
ADMI	Prop./Soc.	A = B	1	Positive	No
IRON	Prop./Soc.	A = B	1	Pos./Neg.	Yes
SEDU	Soc.	A = B	1	Pos./Neg.	Yes
AUTH	Soc.	A > B	3	Pos./Neg.	Yes
IRRI	Soc.	A = B	2	Negative	Yes
ARRO	Soc.	A > B	2	Negative	Yes
CONT	Soc.	A > B	2	Negative	Yes
POLI	Soc.	A = B	3	Positive	No
SINC	Soc.	A < B	2	Positive	No
WOEG	Soc.	A < B	2	Positive	Yes

Table *6.2* – *Valeurs des paramètres des seize situations utilisées pour l'enregistrement d'expressions attitudinales interculturelles.*

pour produire approximativement les mêmes phrases cibles. Soixante-neuf locuteurs ont jusqu'à présent été enregistrés selon ce paradigme, dans le cadre du projet PADE. La table 6.3 précise le nombre de locuteurs enregistrés dans chaque langue, avec leur niveau de langue (L1 ou L2). Dix autres locuteurs ont été enregistrés en allemand, mais seulement dans leur L1 (Hönemann *et al.*, 2014).

6.3.2 Analyses prosodiques

Nous avons vu à la section 5.1.1 la proposition d'Ohala (1983, 1984, 1994) d'un *frequency code*, code symbolique lié à la hauteur de la fréquence fondamentale qui permet l'expression d'affects, en exprimant essentiellement une dimension de dominance. Les implications universelles de ce code (pour ce qui concerne les affects sociaux) sont liées notamment à des expressions véhiculant des sèmes de dominance (par exemple l'autorité) ou de soumission (par exemple la politesse ou une interrogation). Gussenhoven (2004) propose une extension de cette théorie à trois codes, ajoutant un « *effort code* » et un « *production code* » avec d'autres implications sur la variation prosodique. Ainsi, l'« *effort code* » est lié selon l'auteur à des variations d'implication du locuteur dans son acte de parole (voir Daneš, 1994, à ce sujet). Notons qu'Ohala (1994) relie le « *frequency code* » à l'« autorité » (p. 329), alors

		Langue d'enregistrement		
L1	**BP**	**FR**	**JP**	**US**
BP	**21**	8	6	6
FR	-	**13**	8	7
JP	-	12	**23**	6
US	-	6	8	**14**

Table *6.3* – *Nombre de locuteurs enregistrés pour le portuguais brésilien (BP), le français hexagonal (FR), le japonais (JP) et l'anglais des États-Unis d'Amérique (US). Les chiffres sur la diagonale correspondent aux enregistrements en L1, et aux enregistrements en L2 pour les autres. Tous les locuteurs enregistrés dans une L2 sont aussi enregistrés dans leur L1.*

que Gussenhoven (2004) cite l'autorité parmi les implications affectives de l'« *effort code* » (p.88). Ces deux attributions ont des implications opposées sur l'usage de la F_0 fait par les locuteurs souhaitant marquer leur *autorité* : un abaissement de F_0 marque la dominance selon le « *frequency code* », tandis qu'une montée de F_0 est postulée par l'« *effort code* ». Notons toutefois que Gussenhoven (2004) fait une différence entre le « *frequency code* » comme marqueur d'un rôle social (p.81), tandis que l'« *effort code* » est relié à un engagement plus momentané. Nous reviendrons sur cette distinction. Le « *production code* » est lié à la notion de déclinaison de la F_0 et de l'énergie au long de la phrase, du fait de la raréfaction progressive de l'air dans les poumons en fin de groupe de souffle (pour une étude détaillée, cf. Vaissière, 1983). Ce code n'est pas relié à des interprétations d'ordre affectif, mais plutôt informatif – signaler les différence thème/rhème, les continuations ou les fins d'énoncés.

Est-il possible de relier les variations prosodiques enregistrées dans ce corpus à ces codes symboliques proposés comme des universaux liés à des contraintes biologiques ? Pour tester cette hypothèse, l'analyse des signaux enregistrés est réalisée. Cette analyse porte pour l'instant sur une petite partie du corpus (la version anglaise de la phrase « banana » : les enregistrements des huit locuteurs ayant l'anglais en L1 et le japonais comme L2, et des six locuteurs ayant le japonais en L1 et l'anglais en L2) ; l'analyse des autres locuteurs est en cours. La phrase cible « banana » a donc été extraite des enregistrements de ces locuteurs, manuellement segmentée au niveau phonémique, puis les paramètres suivants ont été estimés :

- la F_0 moyenne de la voyelle, exprimée en demi-tons par rapport à la fréquence moyenne de chaque locuteur (on normalise ainsi par rapport au registre de chaque locuteur) ;

- l'intensité moyenne de la voyelle (en dB – ces deux mesures sont effectuées grâce au logiciel Praat Boersma et Weenink, 2012) ;
- la durée normalisée en z-score (cf. Campbell, 1993, pour des détails sur la méthode).

Afin de pouvoir comparer les mesures d'intensité d'un locuteur à l'autre, un microphone placé à un mètre du locuteur est utilisé ; le niveau d'enregistrement de chaque locuteur est ensuite calibré.

Une analyse de variance multivariée a été menée sur ces trois paramètres, avec les facteurs attitude, genre du locuteur et niveau de langue (L1 ou L2). Les résultats montrent que seule l'attitude a un effet significatif sur la variation observée de ces trois paramètres prosodiques ($Wilks'\lambda = 0.53, F_{45,2356.6} = 12.3, p < 0.0001$). La table 6.4 présente les valeurs moyennes et les écarts-types de ces paramètres pour les groupes de locuteurs de L1 ou de L2 anglaise. Les attitudes sont regroupées dans la table en fonction de leur type – sociales ou propositionnelles – et les valeurs obtenues pour la déclaration sont répétées en début de chaque groupe comme référence de productions non marquées.

Les expressions sociales de mépris, arrogance et autorité correspondent à des situations décrites (cf. table 6.2) comme des expressions de dominance du fait d'une différence hiérarchique souhaitée (mépris et arrogance) ou de facto (autorité). D'après la distinction faite plus haut, cette différence de rôle social serait donc exprimée en suivant les règles du « *frequency code* », qui postule une baisse de la F_0 produite pour ces expressions. L'irritation est aussi une expression d'imposition, mais qui n'implique pas de différence sociale ; il s'agit plutôt d'une réaction à un comportement non souhaité, et sa réalisation suivrait donc les prédictions de l'« *effort code* » (hausse de F_0, cf. section 5.3.2). La « séduction » (définie ici comme la marque prosodique d'un intérêt d'ordre sexuel envers l'interlocuteur/trice) est une expression plus complexe, qu'il est difficile de prédire à l'aide de ces codes : entre brame du cerf et marqueur de « féminité », les stratégies possibles sont multiples. Enfin, le « *frequency code* » prédit une hausse de F_0 pour les trois expressions de politesse (politesse, sincérité et « marcher sur des œufs »).

Pour les expressions propositionnelles, le « *frequency code* » prédit une hausse de la F_0 dans les cas de productions dubitatives (question, incertitude, doute, cf. Brandt, 2008), et une baisse pour les assertions. L'« *effort code* », quant à lui, prédit une hausse de F_0 pour les marques d'implications, qui pourraient correspondre à la fois à la surprise et à l'admiration.

Pour les locuteurs de L1, ces prédictions sont confirmées. Les mesures de F_0 montrent bien une baisse pour les expressions d'imposition marquant une différence de statut social (mépris, arrogance, autorité). On observe aussi une

Attitude	L1 speakers F_0 (ST)	INT (dB)	DUR (z)	L2 japanese speakers F_0 (ST)	INT (dB)	DUR (z)
			Attitudes sociales			
DECL	-2.02	1.91	-0.478	-2.19	2.19	-0.386
CONT	-2.68	1.18	-0.324	-0.80	3.06	-0.114
ARRO	-2.21	2.52	-0.228	-3.25	3.20	-0.248
AUTH	-2.23	2.90	-0.453	-1.73	2.11	-0.376
IRRI	1.27	7.76	0.318	0.88	7.60	0.422
SEDU	-2.80	-2.74	0.260	-0.69	0.83	0.168
POLI	0.78	0.26	-0.312	-1.02	1.32	-0.360
SINC	-0.88	-0.11	-0.110	-0.08	2.12	-0.103
WOEG	-1.46	-1.01	0.419	-0.93	0.72	0.272
			Attitudes propositionnelles			
DECL	-2.02	1.91	-0.478	-2.19	2.19	-0.386
IRON	-2.87	1.61	0.108	0.37	3.96	0.528
OBVI	-1.68	1.65	0.042	-0.59	4.90	-0.093
ADMI	6.69	9.92	0.669	5.73	8.22	0.822
QUES	1.01	1.17	-0.384	0.95	-0.76	-0.237
UNCE	0.11	-0.53	0.377	-1.34	0.33	0.592
DOUB	0.60	0.84	0.211	1.53	0.46	0.309
SURP	3.10	5.69	-0.052	2.98	5.38	0.202

Table *6.4* – *Valeurs moyennes des paramètres de F_0 (exprimés en demi-tons : ST), d'intensité et durée (voir texte) pour la phrase « a banana » du corpus en anglais des États-Unis produits par les huit locuteurs natifs et 6 locuteurs japonais de L2 anglaise, pour chacune des 16 attitudes (cf. supra pour les abbréviations).*

hausse pour la production de l'irritation – associée à une hausse importante (proche de 5dB en moyenne) de l'intensité. Les trois expressions de politesse sont aussi produites avec une élévation de la F_0 ; cette hausse étant moins marquée pour la situation de « marcher sur des œufs », analogue au concept de *kyoshuku* et qui correspond à une stratégie de politesse négative, selon les termes de Brown et Levinson (1987) (les deux autres étant des stratégies de politesse positive). De même, les expressions dubitatives de question, incertitude et doute montrent une hausse modérée de F_0 et celles d'admiration et de surprise un hausse plus marquée, pouvant être reliée à l'implication plus grande du locuteur qu'elles supposent (« *effort code* »). Les expressions de séduction, d'ironie et d'évidence sont délicates à prédire selon ces codes. On observe une baisse de F_0 et d'intensité et un allongement des phonèmes pour l'ironie, résultat similaire à celui rapporté pour l'allemand par Niebuhr (2014). L'évidence est quant à elle marquée par une hausse de F_0 et un allongement (voir de Moraes et Rilliard, 2014a, pour le portugais brésilien). Enfin

la séduction est réalisée avec une F_0 très basse, mais surtout la plus basse intensité moyenne observée (sommes-nous ici en présence d'une iconicité de proximité ?).

Pour les locuteurs de L2 anglaise d'origine japonaise, les paramètres prosodiques observés montrent des similitudes mais aussi des différences importantes. L'expression de référence – la déclaration – est comparable à celle des locuteurs de L1, mais dans le bas des valeurs moyennes de F_0. En ce qui concerne les expressions d'imposition, les stratégies observées chez les locuteurs de L2 diffèrent généralement de celles des L1 : l'arrogance suit la tendance prédite par le « *frequency code* », mais pas l'autorité ni le mépris. L'autorité est réalisée avec une F_0 plutôt basse, mais supérieure à celle de la déclaration ; le mépris montre une valeur moyenne de F_0 encore plus élevée, mais surtout une énergie plus marquée que celle observée chez les locuteurs de L1. Par contre l'irritation suit les prédictions de l'« *effort code* » ; comme pour les L1. Pour les expressions de politesse, en ce qui concerne les valeurs de F_0, les performances des locuteurs de L2 suivent les prédictions, et sont similaires à celles de L1, mais leur stratégie diffère quant à l'utilisation de l'intensité, bien plus marquée que celle utilisée par les locuteurs de L1, surtout pour les deux situations typiques de la culture japonaise sincérité & « marcher sur des œufs »). Pour les attitudes propositionnelle dubitatives (question, incertitude, doute), on observe aussi la hausse de F_0 prédite. De même, les trois situations relevant d'une implication forte prédite par l'« *effort code* » (irritation, admiration surprise), les valeurs moyennes de F_0 observées sont très hautes. La réalisation de l'évidence diffère de celle des locuteurs de L1 de par l'utilisation d'une intensité plus marquée. Pour les deux expressions ne faisant pas partie du répertoire conventionnel des attitudes prosodiques japonaises (séduction et ironie), on observe des stratégies différentes de celles des locuteurs de L1. La séduction est ainsi marquée par une F_0 assez élevée (de l'ordre de celle utilisée pour la politesse), avec une baisse d'intensité moins marquée que pour les locuteurs de L1 (on n'aurait donc pas ici ce sème de proximité iconique observé chez les locuteurs de L1). Enfin, l'ironie semble être marquée par une exagération (F_0 plutôt haute, intensité forte et allongement très marqué), et diffère tout à fait de la stratégie des locuteurs de L1.

Ces résultats supportent globalement les hypothèses des deux codes symboliques décrits ci-dessus, pour les locuteurs de L1. Toutefois, ces codes sont largement insuffisants pour permettre de comprendre l'ensemble de la variation et leur application est délicate à attribuer : le cas des locuteurs de L2 est intéressant de ce point de vue, puisqu'ils montrent des comportements parfois divergents. Les raisons de ces divergences sont à chercher d'une part dans l'interprétation culturelle de ces contextes : les contextes sont les mêmes,

mais il est probable que leur interprétation diverge – d'où des performances différentes. On a ainsi vu (cf. section 5.3.2) que l'expression d'autorité est regroupée perceptivement avec la déclaration par des auditeurs japonais natifs, à la différence de ce qui est observé pour l'anglais britannique ou le français. Ce résultat est parfaitement en cohérence avec les performances enregistrées ici pour la situation d'autorité : une interprétation possible serait que le locuteur japonais dans cette situation possède un rôle social lui conférant autorité. Il ne lui serait donc pas nécessaire de marquer cette autorité prosodiquement. À l'inverse, dans des sociétés pour lesquelles la hiérarchie sociale est moins prégnante (Hill *et al.*, 1986; Ide, 2002), il serait plus important de marquer son autorité pour qu'elle soit perçue et/ou respectée (cependant, voir l'importance de la notion de « power » dans les situations décrites par Culpeper *et al.*, 2003). Pour le cas de la politesse, on observe bien, pour les deux groupes de locuteurs, une tendance suivant les prédictions du « *frequency code* » (i.e. une hausse de F_0). Pourtant, on a aussi observé (cf. section 5.3.2) un comportement perceptif divergent entre les regroupements d'attitudes effectués par des auditeurs japonais *vs.* des auditeurs occidentaux pour la politesse (les japonais effectuant un regroupement particulier pour la politesse, tandis qu'elle est regroupée avec la déclaration par les autres groupes). Toutefois, les stratégies des deux groupes de locuteurs ne sont pas complètement équivalentes et diffèrent notamment du point de vue de l'intensité plus importante utilisée par les locuteurs japonais dans leur L2. Le résultat de cet usage de l'intensité est un affect plus marqué, ce qui va aussi dans le sens des observations faites lors des regroupements d'attitudes et pourrait relever de l'« *effort code* ».

Il est prématuré d'avancer des explications plus détaillées à ce stade de notre travail, mais cette observation des stratégies expressives en anglais des États-Unis, par des locuteurs de L1 et de L2, fournit d'intéressants parallèles avec les travaux précédents.

6.3.3 Mesures de performance en L1 et L2

Pour mieux comprendre les variations prosodiques observées dans ces productions de locuteurs de L1 ou de L2[1], un test de perception a été mené sur ces stimulus, auprès d'auditeurs natifs de l'anglais des États-Unis (Rilliard *et al.*, 2013; Shochi *et al.*, 2013). Les sujets avaient pour tâche de juger de la qualité de la performance des productions qui leur étaient présentées, sur une échelle allant de 1 à 9, sachant quels étaient la situation et le but com-

[1]Les performances des locuteurs français de L2 anglaise ont aussi été évaluées (Rilliard *et al.*, 2014c), mais ne seront pas présentées ici car leurs performances n'ont pas encore été analysées prosodiquement ; le travail est en cours.

municatif du locuteur. Les stimulus étaient présentés dans leur performance audio-visuelle.

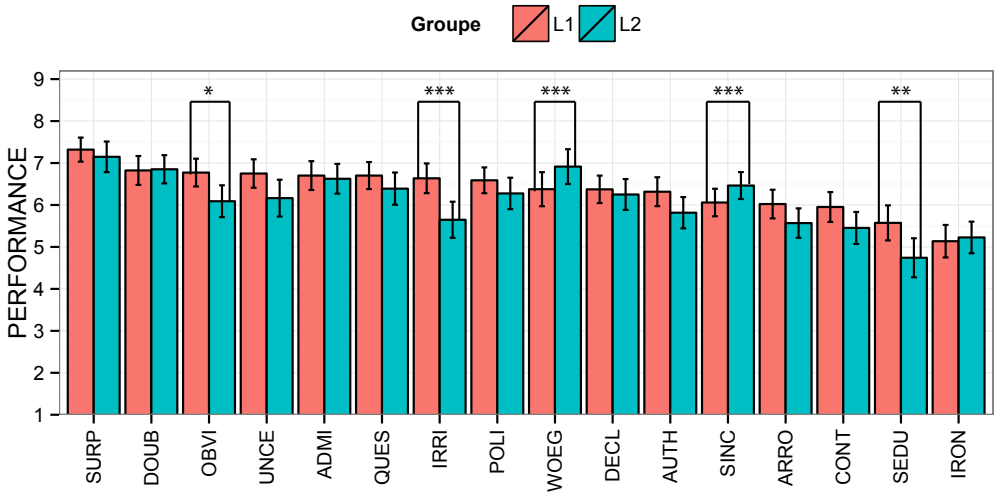


Figure 6.3 – *Scores moyens de performance obtenus par chacun des groupes de locuteurs de L1 ou de L2 japonais, pour chaque attitude. Les différences significatives entre deux groupes pour une attitude sont notées par des astérisques (* : p<0.05 ; ** : p<0.01 ; *** : p<0.001).*

Les résultats de cette mesure de la performance amènent plusieurs éclairages sur les observations faites auparavant. D'abord, les locuteurs de L1 obtiennent des performances globalement supérieures à celle des locuteurs de L2, ce qui est attendu, mais appuie la meilleure qualité des observations faite sur des locuteurs de L1 – il faudra donc regarder ce qu'il se passe en japonais pour les locuteurs de L2. Toutefois, ces performances ne sont significativement différentes entre les deux groupes que pour cinq attitudes sur les seize présentées (cf. figure 6.3 ; notons que les performances considérées ici sont les moyennes de chaque groupe, pas les performances individuelles). Parmi ces cinq attitudes, trois reçoivent des scores de performances meilleurs pour les locuteurs de L1 (évidence, irritation et séduction), par contre les deux autres sont jugées plus performantes lorsque produites par les locuteurs de L2 japonais plutôt que les locuteurs de L1 (sincérité et « marcher sur des œufs »). Ce qui est particulièrement intéressant dans ce résultat est l'attribution de meilleures performances (par des auditeurs de L1) à des locuteurs de L2 – précisément pour les situations conventionnalisées dans la culture de ces lo-

cuteurs[1] – et inversement des performances significativement meilleures aux locuteurs de L1 dans la situation de séduction, non conventionnalisée prosodiquement dans la culture japonaise.

L'interprétation de ces résultats est délicate. On pourrait conclure que le fait qu'une expression soit conventionnalisée dans une culture donnée fournit aux locuteurs ayant cette langue comme L1 un entrainement à se comporter de manière appropriée dans cette situation. Cette interprétation rencontre toutefois le problème que les auditeurs qui ont fourni les jugements de performance ne sont pas japonais – et donc n'ont pas, eux non plus, de prototype pour cette situation. Il se pourrait que les choix comportementaux conventionnalisés dans la culture japonaise puissent s'appliquer à la culture des États-Unis, mais c'est une hypothèse très forte, qui mériterait de plus amples travaux. Une autre hypothèse possible serait que les auditeurs aient des attentes, des préjugés, liés à l'origine culturelle des locuteurs qu'ils jugent. Les japonais sont réputés pour être polis ; une performance du domaine de la politesse réalisée par un japonais pourrait donc bénéficier d'un apriori positif. Le raisonnement inverse pourrait être tenu pour l'expression de séduction. Notons que les locuteurs français de L2 anglaise ont été jugés plus performants que les natifs pour la séduction (Rilliard *et al.*, 2014c), ce qui renforce cette interprétation d'un biais des jugements de performance lié à des aprioris du fait de l'origine culturelle des locuteurs.

Par ailleurs, deux autres attitudes (l'évidence et l'irritation) reçoivent de meilleurs scores de performance pour les locuteurs de L1. Ces deux situations correspondent à des expressions conventionnelles aux Etats-Unis comme au Japon. Dans ce cas-ci, il se pourrait que les locuteurs de L2 aient utilisé des stratégies inspirées de leur L1, et moins adaptées à l'anglais.

Un autre facteur d'importance à considérer à propos de ces scores de performance concerne les importantes (et significatives) différences de performances observées entre les différents locuteurs. La variation des jugements de performance vient en partie de leur origine culturelle, mais aussi de leur individualité. Les variations de performance ne se résument pas à la qualité d'« acteur » des locuteurs (certains sont clairement plus à l'aise que d'autres dans cette tâche), car on observe des performances différentes en fonction des situations. Untel sera jugé particulièrement performant dans une situation d'autorité, mais moins dans celle de sincérité ; une autre configuration sera observée pour un second locuteur.

[1]Rappelons-le, ces situations correspondent respectivement aux attitudes de sincérité-politesse et de *kyoshuku* (Shochi *et al.*, 2009c).

Ces résultats sont en cours d'analyse. Ils soulèvent déjà des questions intéressantes – et difficiles à traiter. De nouveaux tests de perception sont en cours, dont les résultats préliminaires semblent conforter certaines de ces analyses. D'autres paradigmes sont à mettre au point afin d'étudier par exemple l'effet de la personnalité perçue des locuteurs sur la qualité de leur performance. Des tests sont aussi menés sur les autres langues et des analyses prosodiques exhaustives doivent être réalisées. Enfin, d'autres projets ont vu le jour sur la base de ce corpus (notamment le projet « MAVOIX » au LaBRI, dirigé par T. Shochi, des enregistrements dans d'autres langues, etc.) et bénéficient des apports de cette méthodologie pour tester différentes hypothèses. Ce corpus est aussi étudié de manière à en envisager les potentialités en didactique des langues étrangères, une des raisons pour lesquelles il a vu le jour.

7 | Mesures en parole spontanée : diachronie & expressivité

> « [...] Lors le print à la gorge, luy disant : « Tu escorches le latin, par sainct Jehan je te feray escorcher le renard : car je te escorcheray tout vif. » Lors commença le pauvre Limousin à dire : « Vée dicou, gentilastre ! Ho sainct Marsault, adiouda my ! Hau, hau, laissas à quau, au nom de Dious, et ne me touquas grou ! » A quoy dist Pantagruel : « A ceste heure parles tu naturellement. »
>
> Rabelais, *Pantagruel*,
> livre II, chapitre VII (ca 1530)

Les travaux présentés jusqu'ici suivent une approche classique en phonétique (Xu, 2010), qui utilise des corpus de taille relativement petite, très contrôlés afin de ne faire varier si possible que les objets destinés à être observés. Le cout de cette approche est bien sûr le manque de naturel. On a vu que les travaux visant une meilleure compréhension des fonctions expressives se heurtent à ce problème du naturel et cherchent à créer des corpus contrôlés et spontanés (cf. section 5.2, Audibert *et al.*, 2010).

Une autre avenue pour sortir de ces contraintes des corpus de laboratoire consiste à travailler sur des corpus de parole spontanée. La méthode préconisée par Rabelais pour obtenir de la parole spontanée pourrait se heurter aux exigences d'éthique. J'entends par « parole spontanée » une production de parole qui ne soit pas contrainte par le laboratoire (voir Campbell, 2005, pour une discussion). Une parole lue par un journaliste à la radio sera considérée comme spontanée – même si ce style de parole lue possède des caractéristiques différentes de celles d'une conversation. Cette approche est de plus en plus utilisée dans les sciences de la parole, notamment grâce aux avancées des logiciels de traitement automatique de la parole (voir par exemple Campbell, 2000, 2004; Adda-Decker *et al.*, 2005; Roy, 2009; Quené, 2013).

Ma venue au LIMSI a été notamment l'occasion pour moi d'aborder cette approche (tout en continuant mes autres travaux), soumise à d'autres contraintes que celles liées aux travaux présentés jusqu'à maintenant – et permettant d'obtenir d'autres résultats aussi. Notamment, les principes de mesure objective d'indices prosodiques suivent une logique différente – même si on pourrait imaginer appliquer des mesures de distance sur des corpus spontanés (cf. Klabbers et van Santen, 2004). Les travaux résumés ci-dessous constituent un nouveau paradigme dans mes recherches, intéressant pour aller au-delà de certaines des limitations liées aux corpus de laboratoire. On verra comment une convergence entre ces deux approches sera recherchée dans la suite de mes recherches.

7.1 Variation diachronique

Un corpus d'enregistrements radiophoniques d'archives datant des années 40 jusqu'aux années 90 a été l'objet d'une collaboration avec Philippe Boula de Mareüil et Alexandre Allauzen, afin d'étudier une éventuelle évolution du style journalistique depuis la seconde guerre mondiale jusqu'à la fin du vingtième siècle (pour plus de détails, voir Boula de Mareüil *et al.*, 2012b). Le style journalistique étant décrit dans un certain nombre de travaux comme marqué en français par un accent initial (Fónagy et Fónagy, 1976; Lucci, 1983; Fónagy, 1989; Carton, 2000), c'est notamment cet indice dont nous avons cherché à comparer l'importance et l'évolution.

Pour cela, et sur la base des transcriptions de ces enregistrements, des contextes susceptibles de porter un accent initial ont été sélectionnés dans le corpus : des séquences de mots clitique–non clitique telles que l'exemple « un FAbricant de MAtériaux de CONstruction » proposé par Di Cristo (1998)[1]. Sur de telles séquences, la différence de hauteur de F_0 entre la voyelle du clitique et la voyelle initiale du mot non clitique est calculée et sert d'indice de la présence d'un accent sur la séquence considérée ('t Hart *et al.*, 1991; Jankowski *et al.*, 1999). Notons que dans la description de l'accent initial fait par Welby (2006, 2007), celui-ci peut porter sur la première ou la seconde syllabe du mot non clitique. Cependant, Welby (2006, 2007) s'intéresse à la position du pic sur des phrases continument voisées, traitement qu'il est impossible de réaliser sur des archives radiophoniques ; les valeurs moyennes de F_0 sur les voyelles sont considérées ici, car plus robustes aux erreurs de détection. Une discussion détaillée de ce point est proposée dans Boula de Mareüil *et al.* (2012b).

[1]Les accents initiaux sont alors portés par les syllabes en majuscules.

Les données sont ensuite regroupées en quatre périodes, formant des données de taille comparable : 1940-1959, 1960-1969, 1970-1979, 1980-1997. Les résultats de ces analyses montrent, sur ces quatre périodes, une décroissance progressive de cette mesure de la présence d'un accent initial dans les archives radiophoniques. La figure 7.1 présente l'histogramme de cette mesure de différence de F_0. On peut y observer la progressive diminution de la proportion de contextes clitique-nonclitique porteurs d'une montée mélodique. D'autres indices prosodiques (et notamment un allongement de la syllabe pénultième de mots au moins trisyllabiques) ont été mesurés, et soutiennent l'hypothèse d'une évolution de la prosodie des annonceurs radiophoniques sur cette période.

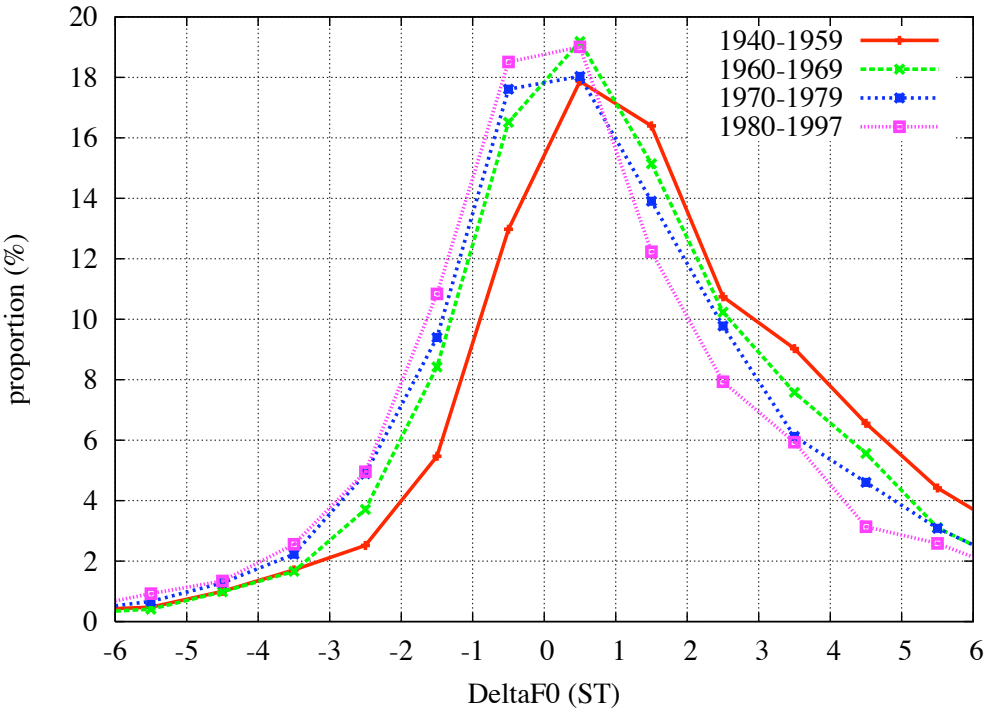


Figure 7.1 – *Histogrammes des différences de F_0 observées entre les voyelles de suites clitique-non clitique, pour les quatre périodes considérées (voir texte).*

La décroissance d'une telle mesure acoustique est un indice robuste d'un changement mélodique. L'étape suivante consiste à vérifier que le changement mesuré correspond à une évolution perceptible. À cette fin, un test de perception a été mené, permettant de mesurer l'importance des différents indices sur la perception d'une évolution. Pour cela, des processus de transplantation prosodique et de resynthèse ont été utilisés, afin de contrôler soit le contenu segmental, soit la prosodie présentée aux sujets, dans le but

de mesurer l'impact relatif des indices lexicaux, segmentaux et prosodiques. Les sujets avaient pour tâche de dater les extraits qui leur étaient proposés. Les résultats montrent que les indices prosodiques, même lorsqu'ils sont le seul indice utilisable, induisent une évolution des dates moyennes attribuées, conformément aux mesures acoustiques observées sur le corpus. Cependant, les autres indices apportent des informations plus prégnantes pour juger de la date des extraits – et en premier lieu les indices lexicaux.

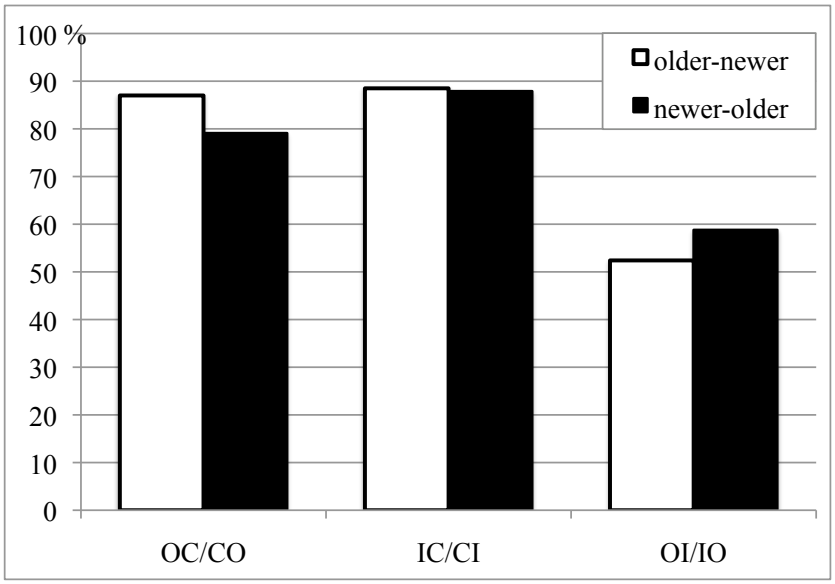


Figure 7.2 – *Pourcentage de jugements « plus ancien » attribués au premier (respectivement au second) membre des paires de stimulus O/C, I/C, O/I (resp. C/O, C/I, I/O).*

Afin de s'abstraire de ces contraintes liées au lexique, une expérience basée sur une imitation d'un style daté et d'un style contemporain a été menée, grâce à l'enregistrement d'un journaliste à qui il a été demandé de produire les deux styles, selon sa propre représentation de ce qu'ils sont. Les résultats de cet enregistrement[1] sont conformes, en terme d'indices prosodiques (présence d'accents initiaux, etc.), à ce qui est observé dans les archives. Il a ensuite été demandé à des sujets de juger de l'ancienneté relative de paires de stimulus, composées uniquement à partir de la prosodie de ces enregistrements originaux (O—tirés du corpus d'archives), imités (I—imitation du style ancien par le journaliste), ou contemporain (C—phrases produites dans un style contemporain). Les résultats de ces comparaisons sont présentés à la figure 7.2. Ils indiquent un jugement « plus ancien » des stimulus origi-

[1]Phrases extraites de la période 1940-1959 du corpus original.

naux ou imités, comparés aux stimulus contemporains, tandis que les sujets n'arrivent pas à départager les deux styles originaux et imité.

Ces résultats des tests de perception supportent les conclusions des mesures acoustiques menées sur les archives. Les évolutions des paramètres prosodiques constituent donc bien des indices perceptifs prégnant d'un changement stylistique. Déterminer les causes de ce changement stylistique est un autre travail, qui ne sera pas abordé ici.

7.2 Lecture expressive de contes

La thèse de David Doukhan (Doukhan, 2013)[1] a été réalisée dans le cadre du projet ANR GVLEX (Gelin *et al.*, 2010), qui avait notamment pour but d'améliorer la synthèse de parole de contes pour enfants. Ce travail de thèse s'est donc intéressé[2] à la caractérisation des variations prosodiques typiques de la lecture de contes pour enfants. Pour mener cette étude à bien, un corpus de contes (écrits) a été réuni et leur structure analysée (Doukhan *et al.*, 2012). Parmi ce matériau, douze contes ont été sélectionnés et donnés à lire à un locuteur professionnel. Ce corpus de douze contes lus, correspondant à près d'une heure de parole, constitue un matériel riche en variations prosodiques (Doukhan *et al.*, 2011). Un des travaux de cette thèse a donc consisté en une objectivation de certaines de ces variations, afin de pouvoir fournir des commandes au système de synthèse à partir du texte utilisé. L'une des principales difficultés rencontrée pour cela venait du fait que le système de synthèse utilisé est basé sur une technologie de concaténation d'unités sélectionnées dans un corpus et ne pouvait donc accepter que des commandes de modification minimum – ce type de synthétiseur s'appuyant sur le contenu du corpus de synthèse pour générer une parole de haute qualité.

7.2.1 Analyses prosodiques

La première tâche a donc consisté à analyser les variations prosodiques afin de mieux comprendre quels aspects de la variation sont les plus pertinents dans le cas de la lecture de contes, afin de fournir des consignes pour enregistrer des corpus expressifs les plus pertinents pour la tâche.

Pour cela, deux stylisations ont été produites, inspirées du modèle de perception tonal de d'Alessandro et Mertens (1995) : la première utilise le Prosogram (Mertens, 2004), la seconde implémente un modèle similaire, en cherchant à maximiser la présentation de différents paramètres prosodiques

[1] Thèse co-encadrée avec C. d'Alessandro.

[2] Je n'aborderais pas ici les aspects spécifiquement liés à l'analyse du texte des contes.

sur une même représentation (F_0, intensité, apériodicité et segmentation phonémique, voir figure 7.3 pour un exemple).

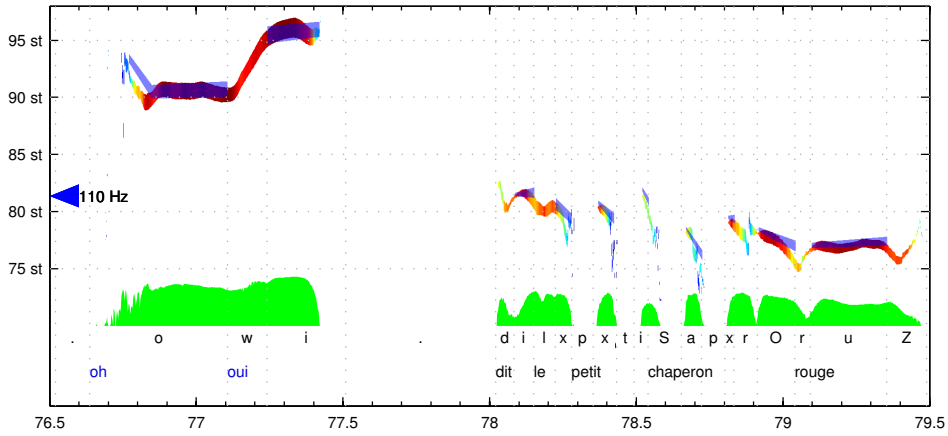


Figure 7.3 – *Représentation de la variation prosodique de l'extrait « oh oui, dit le petit chaperon rouge » du corpus de contes lus du projet GVLEX (Doukhan, 2013).*

La stylisation réalisée grâce au Prosogram permet entre autre une comparaison avec d'autres travaux utilisant le même outil pour extraire des informations prosodiques. Ainsi, 13% des noyaux vocaliques stylisés dans le corpus de contes sont considérés comme des tons dynamiques, alors que Patel *et al.* (2006) rapportent seulement 2% et 4% de tons dynamiques pour des phrases lues respectivement en français et en anglais. Cette proportion importante de tons dynamiques corrobore la nature expressive des variations prosodiques contenues dans le corpus de contes. Les travaux de Roekhaut *et al.* (2010) rapportent des analyses de différents styles de parole (information radio, discours politique et conversation) qui, là encore, montrent que les variations contenues dans le corpus de contes sont importantes. L'étendue de F_0 observée pour le conteur est de 17 demi-tons, alors qu'elle est pour les styles rapportés par Roekhaut *et al.* (2010) de, respectivement 10.5, 10.5 et 7 demi-tons. On observe cependant une similitude entre ce style de lecture de contes et le discours politique en ce qui concerne le pourcentage de temps de pause élevé (autour de 30%). Afin d'avoir un aperçu de la variabilité prosodique induite par l'incarnation des personnages par le conteur (et donc de ses stratégies expressives) dans ces 12 contes, les descripteurs prosodiques suivants ont été extraits :

- **NS** : le nombre de syllabes ;
- **TP** : le taux de parole, mesurée en nombre de syllabes par seconde, hors pauses ;

- **PTP** : le pourcentage de temps de pause ;
- **NSP** : la moyenne du nombre de syllabes entre pauses ;
- **HES** : le pourcentage de syllabes marquées comme des hésitations ;
- **PVR** : le pourcentage de voyelles rejetées par le PROSOGRAM (voir Mertens, 2004, section4.5.5) ;
- **PTD** : le pourcentage de tons dynamiques ;
- **HM** : la hauteur moyenne (exprimée en demi-tons) ;
- **IM** : l'intensité moyenne (pour les voyelles) ;
- **DHIS** : la différence de hauteur inter-syllabique (exprimée en demi-tons) ;
- **DIIS** : la différence d'intensité inter-syllabique (exprimée en dB) ;
- **TES** : la tessiture (intervalle entre le cinquième et le quatre-vingt quinzième centile des valeurs de F_0) ;
- **EI** : l'étendue d'intensité (intervalle entre le cinquième et le quatre-vingt quinzième centile des valeurs d'intensité).

Les 32 personnages pour lesquels au moins 50 syllabes sont observées ont été conservés. L'analyse est menée en séparant les passages de contes correspondant à des personnages féminins (14 personnages), de ceux correspondant à des personnages masculins (15 personnages). Les propriétés prosodiques du narrateur sont conservées dans les deux groupes, afin de servir de référence. Deux analyses en composantes principales (ACP) ont été menées sur ces données (Husson *et al.*, 2010). Le graphique représentant les deux premières dimensions de l'ACP menée sur le groupe féminin est présenté par la figure 7.4.

Une inspection des deux premières composantes principales de l'ACP menée sur le groupe féminin montre que la première composante principale est liée positivement avec les mesures variation locale de F0 et d'intensité (comme les glides et les différences inter-syllabiques) et l'intensité moyenne, et négativement avec le pourcentage de voyelles dévoisées. La seconde composante principale est liée à la fréquence fondamentale moyenne et à la taille des segments de parole entre deux pauses. Un regroupement hiérarchique menée sur la sortie de l'ACP (Husson *et al.*, 2010) permet de créer quatre groupes de personnages féminins, en prenant comme seuil un tiers de la distance maximum. Ces quatre groupes sont représentés sur le dendrogramme de la figure 7.5.

Les quatre groupes de personnages féminins présentent des caractéristiques narratives pertinentes :

- Deux jeunes filles ayant un rôle dirigeant sont incarnées par le conteur avec une voix dans un registre élevée, une mélodie plate et un fort pourcentage de voyelles dévoisées.

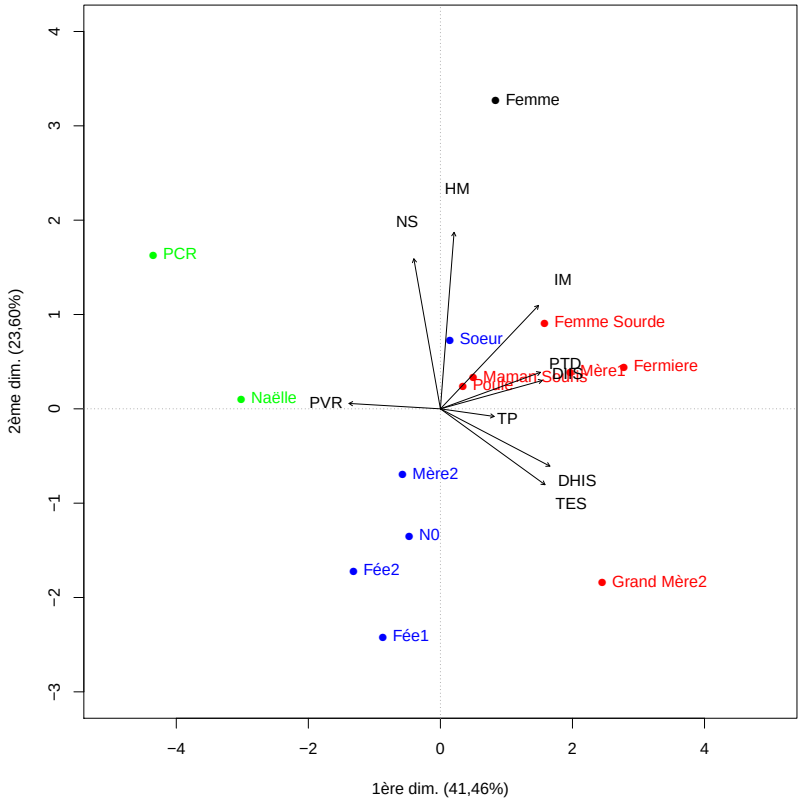


Figure 7.4 – *Dispersion des personnages féminins sur les deux premières composantes de l'analyse en composantes principales montrant leur différences en fonction des paramètres prosodiques extraits de la stylisation par* Prosogram.

- Des femmes adultes ayant un rôle de support proche de celui du narrateur sont incarnées avec une prosodie assez proche de celle du narrateur.
- Des femmes plus âgées ou des personnages ayant la charge d'une importante responsabilité (par exemple « maman souris » à la responsabilité d'élever des enfants) sont incarnées avec une intensité moyenne plus élevée et plus de mouvements mélodiques locaux.
- Un personnage féminin chargé d'aider le héros et incarné avec une voix élevée et intense.

Le graphique représentant les deux premières dimensions de l'ACP menée sur le groupe masculin est présenté par la figure 7.6. L'inspection de la répartition des variables sur ces dimensions montre que la première composante principale est lié négativement à la fréquence fondamentale moyenne, et positive-

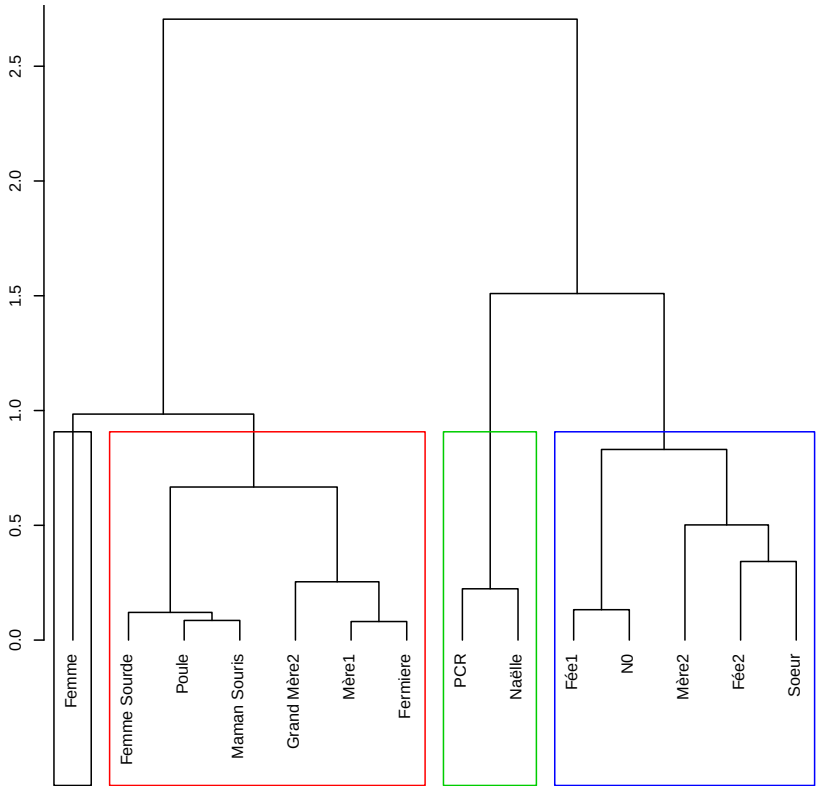


Figure 7.5 – *Regroupement hiérarchique des personnages féminins incarnés par le conteur, obtenu à partir de la dispersion sur l'analyse en composantes principales.*

ment aux mouvement locaux de F0 (glides et différences inter-syllabiques). La seconde composante principale oppose l'empan de F0 au pourcentage de voyelles dévoisées et à la différence d'intensité inter-syllabique.

Le groupe masculin présente une dispersion un peu plus complexe que celle observée pour le groupe féminin. Un regroupement hiérarchique menée sur la sortie de l'ACP permet de créer cinq groupes de personnages masculins (toujours avec le critère d'un seuil au tiers de la distance maximale, cf. figure 7.7), groupes ayant les caractéristiques suivantes :

- Des personnages de héros, jeunes et inexpérimentés, incarnés avec une fréquence fondamentale et une intensité moyenne élevée, et peu de variations locales.
- Des personnages ayant un rôle de support du héros, montrant des caractéristiques prosodiques proches de celles du groupe précédent.
- Des personnages conseillant et aidant le héros, joués dans un registre bas ou moyen et avec une importante dynamique de F0.

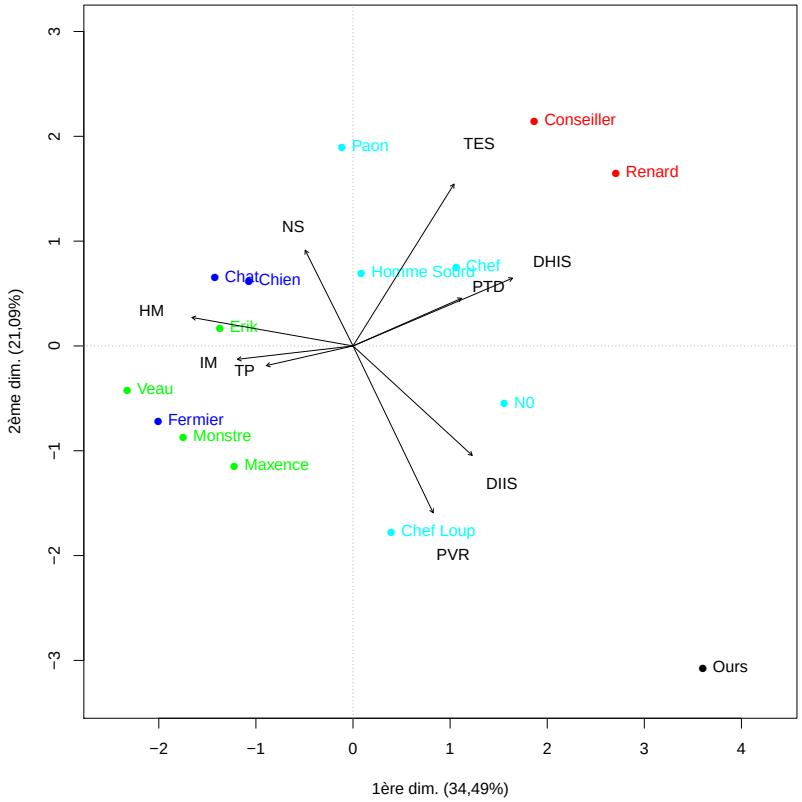


Figure 7.6 – *Dispersion des personnages masculins sur les deux premières composantes de l'analyse en composantes principales montrant leur différences en fonction des paramètres prosodiques extraits de la stylisation du* Prosogram.

- Des personnages plus âgés ou plus puissants (prosodiquement proches du narrateur), ils sont incarnés avec un registre bas ou moyen mais avec plus d'intensité et de notables variations prosodiques locales.
- Un personnage particulier, un ours agressif, ayant une voix très basse, avec d'important changement locaux d'intensité et plus de dévoisement que les autres personnages.

Les groupes de personnages obtenus sur la base de leurs caractéristiques prosodiques sont cohérents, du point de vue de leur rôle dans les récits. On est donc bien en présence de choix stratégiques de la part du conteur pour incarner les différents types de personnages. Ces choix peuvent se résumer par une symbolique liée à la hauteur mélodique, expliquée par le « *frequency code* » (Ohala, 1994) : une utilisation d'un registre plus élevé pour des personnages très jeunes (et donc plus faible et plus petits), et d'un registre bas

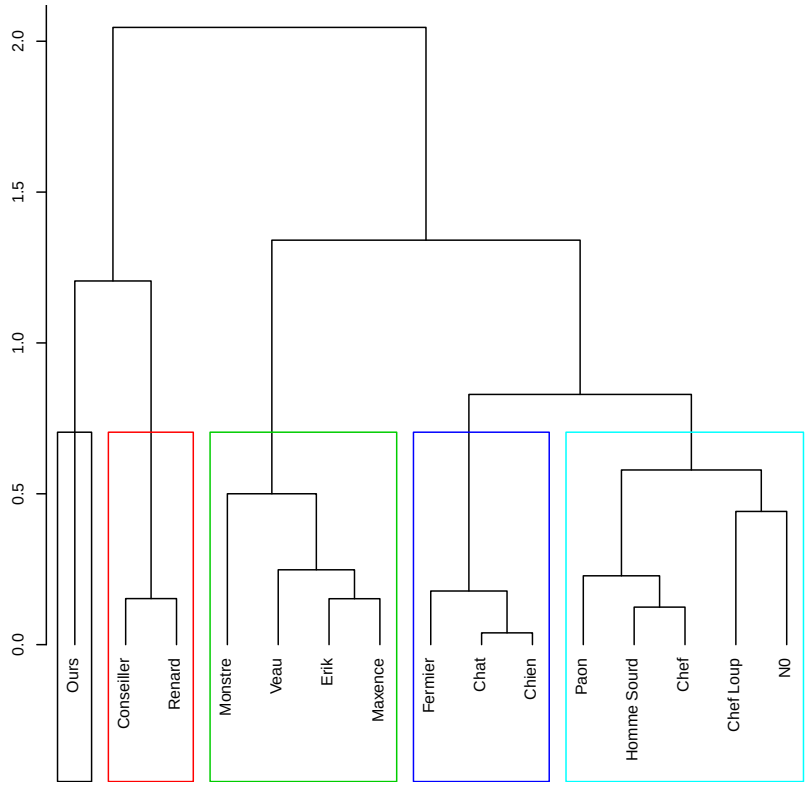


Figure 7.7 – *Regroupement hiérarchique des personnages masculins incarnés par le conteur, obtenu à partir de la dispersion sur l'analyse en composantes principales.*

pour les personnages puissants et/ou méchant (typiquement l'ours). Ce lien entre fréquence et rôle dans le récit serait à l'œuvre pour incarner des personnages ayant un rôle plus ou moins empreint de pouvoir. Il est aussi intéressant de noter que l'intensité joue un rôle important dans cette mise en place des personnages.

Le projet GVLEX (Gelin *et al.*, 2010) devait notamment produire un système de synthèse à même d'ajouter une expressivité adéquate à la lecture de contes. Pour cela, l'enregistrement de quatre bases expressives était possible. Les choix qui ont été fait pour ces quatre bases correspondent en partie à l'observation de cette répartition des personnages – et donc aux stratégies mises en place par le conteur. Au lieu de proposer quatre expressivités correspondant à des catégories perceptives, qui aurait été décrites par des étiquettes par exemple d'émotions, il a été choisi de donner dans une certaine mesure comme consignes d'enregistrement des types de qualités vocales (d'Alessandro, 2006), qui puissent être utilisées pour produire les catégories

perceptives. La dimension d'effort vocal (Liénard et Barras, 2013) nous a semblé particulièrement productive à ce niveau, car corrélée à la dimension affective d'activation (Scherer, 1989b; Russell et Barrett, 1999), et pouvant de plus rendre aussi les aspects symboliques du « *frequency code* ». La seconde dimension affective est liée à la valence. Une claire marque de positivité dans la voix est liée à la production du sourire (Tartter et Braun, 1994; Aubergé et Cathiard, 2003; Émond, 2013) ; une troisième consigne a été donnée pour produire de la parole souriante. La quatrième base, visant une production vocale à valence négative, est moins évidente à décrire en terme acoustico-articulatoires ; la consigne donnée au locuteur pour la quatrième base a donc été de produire de la parole triste. Les quatre consignes suivantes ont donc été définies :

- une voix portée, voix forte destinée à être entendue de loin, mais aussi voix présentant un effort vocal plus marqué ;
- une voix de proximité, avec un taux de voisement bas ;
- une voix souriante (produite en demandant au locuteur de sourire pendant qu'il parle) ;
- une voix triste.

Toute la difficulté étant ensuite de prédire, à partir du texte d'un conte, laquelle de ces bases sera la plus adaptée au contexte, avec éventuellement de simples modifications de registre. Pour des détails à ce propos, cf. Doukhan (2013).

8 | Travaux en cours et perspectives

> Elle déguisait sa voix et se servait d'un vibraphone vocal,
> mais à l'ampli et avec le décomposeur Blochard, on
> perçoit des accents féminins dans les syllabes muettes.
>
> Dard, *Alice au pays des merguez*,
> première partie, chapitre tchlaoff ! (1986)

POUR améliorer la capacité à prédire, décrire et comprendre des changements prosodiques particuliers, liés à un contexte d'occurrence et à un but communicatif particulier, il est important de faire des progrès dans la compréhension des paramètres acoustiques transmettant ces variations prosodiques et de leur décours temporel. Dans ce but, plusieurs pistes de recherches sont actuellement suivies, ou envisagées dans un futur proche.

La première de ces pistes consiste à avoir des outils de mesure de cette variation prosodique, à même d'objectiver, de contrôler et de reproduire des variations perçues.

8.1 Pour une synthèse paramétrique de parole expressive

Les travaux sur la synthèse de parole expressive décrits dans la section 7.2 se continuent actuellement avec la thèse de Marc Évrard, dans le cadre du projet FUI ADN-TR. Ce projet a notamment pour but l'imitation de l'expressivité d'un locuteur.

À la différence de l'approche précédente, ces travaux se basent sur une approche de synthèse par modèles statistiques (Tokuda *et al.*, 2013). Un intérêt particulier de cette approche vient de sa nature paramétrique, qui permet donc de manipuler les paramètres d'entrée d'un vocodeur (paramètres générés par le système d'apprentissage) et donc de tester des hypothèses scientifiques sur la pertinence de certains aspects de la prosodie pour la perception de telle ou telle fonction expressive. Il est possible de contraindre certains paramètres,

ce qui était impossible avec les systèmes de synthèse par concaténation d'unités de tailles variables. En particulier, les aspects de la qualité vocale liés à la source glottique (d'Alessandro *et al.*, 1998; d'Alessandro, 2006; Yegnanarayana et Dhananjaya, 2013), qui sont particulièrement difficiles à mesurer dans la parole du fait des effets supraglottiques, mais qui participent à l'expressivité prosodique, seraient très intéressants à étudier dans cette optique de synthèse de parole paramétrique.

Pour cela, des travaux sont menés à plusieurs niveaux. D'une part sur les corpus : pour tout système basé sur un apprentissage statistique, le corpus d'apprentissage est crucial. Il est important de lui fournir des données en quantité, mais aussi en qualité. Un approfondissement de ces problèmes est mené à l'heure actuelle autour d'une notion d'espace vocal expressif d'un locuteur, qui serait défini à partir des dimensions de la source vocale et des capacités d'un locuteur donné à se déplacer dans son espace (en quelque sorte, une extension du phonétogramme Lamesch *et al.*, 2012). Il s'agirait donc de juger d'une production donnée en fonction de sa position dans l'ensemble des productions *possibles* pour un conduit vocal donné. Les principales dimensions de cet espace sont décrites par d'Alessandro (2006) :

- la fréquence fondamentale (et ses mécanismes, cf. Henrich *et al.*, 2005) ;
- une mesure de l'effort vocal – ou de la force de voix (Liénard et Barras, 2013) ;
- une dimension tendu-relâché, liée à la tension des plis vocaux (Henrich *et al.*, 2004) ;
- une mesure de l'apériodicité (d'Alessandro *et al.*, 1998; Yegnanarayana *et al.*, 2011)

À ces dimensions de la source glottique, il faut rajouter des effets supralaryngés tels que le sourire ou la nasalisation (Laver, 1980).

Les travaux menés pour avancer dans cette direction se concentrent sur la possibilité de décrire des variations affectives à l'aide de telles dimensions d'ordre acoustico-phonétiques, ainsi que sur le choix de paramètres acoustiques adéquats pour les mesurer. Il est aussi primordial de travailler avec des vocodeurs de qualité afin de pouvoir modifier ces paramètres. Des vocodeurs de référence, tels que ceux proposés par Kawahara et Morise (2011), peuvent être avantageusement utilisés pour travailler sur des transformations entre deux prototypes. Il est plus délicat d'arriver à la caractérisation de paramètres articulatoires grâce à ce type d'outils. D'autres approches (Bozkurt *et al.*, 2005; d'Alessandro et Sturmel, 2011), basées sur des modèles d'analyse de la source (par exemple Doval *et al.*, 2006), permettent d'envisager une décomposition du signal. Les travaux en contrôle gestuel (d'Alessandro, 2010, 2011, 2014) de tels modèles semblent une piste prometteuse pour l'étude des variations acoustiques sous-jacentes aux variations expressives (d'Alessandro

et al., 2006; D'Alessandro *et al.*, 2007; d'Alessandro *et al.*, 2011, 2014). Des progrès importants restent à réaliser concernant l'articulation, notamment, mais ausi le contrôle du débit. D'importantes questions sont aussi à travailler en ce qui concerne les espaces de contrôle des dimensions acoustiques et/ou expressives. Mais ces approches sont prometteuses pour fournir de nouvelles façons d'envisager les aspects dynamiques de l'expression.

Cela pourrait avoir des implications intéressantes, en particulier pour les travaux sur la stylisation prosodique au sens large, qui sous-tendent la représentation de la variation prosodique depuis longtemps déjà (cf. figure 8.1 et la section 2.1) afin de mieux représenter les indices pertinents de cette variation. Des progrès pour proposer différentes stylisations sont sans doute encore souhaitables, mais nécessiteront des avancées sur la description de la pertinence percptive des variations au sein d'un espace vocal expressif. Une piste de recherche consisterait par exemple à mieux comprendre et prévoir comment les auditeurs savent extrapoler, sur la base des courts extraits de parole, la position d'un locuteur dans son registre propre (Bishop et Keating, 2012).

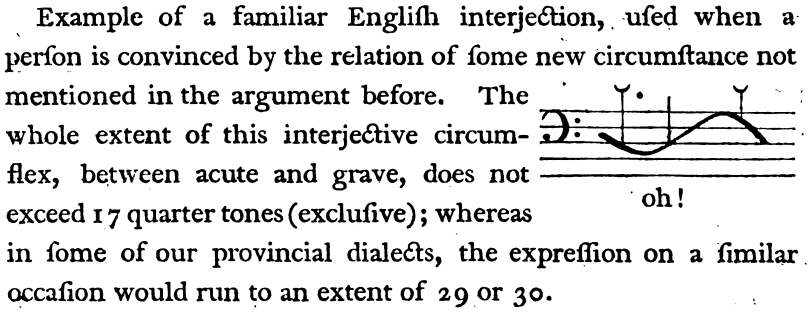


Figure 8.1 – *Exemple de stylisation prosodique d'une interjection de surprise à la cour d'Angleterre (pas dans les provinces), proposée par Steele (1779, p. 86).*

Les perspectives que pourraient ouvrir un système de synthèse expressif paramétrique sont par ailleurs très variées. Les sections suivantes explorent certaines des applications possibles, au regard des travaux présentés dans ce document.

8.2 Rôles sociaux, personnalités & affects

Les derniers travaux présentés sur la production d'attitudes en contexte montrent l'impact du locuteur sur la variation des stratégies expressives.

Cette variation doit être reliée à un très grand nombre de facteurs, notamment physiologiques, sociologiques et individuels (Pierrehumbert, 2006). Parmi ces facteurs, la personnalité du locuteur mais aussi son rôle social et son habitude (aptitude ?) à remplir un tel rôle me semblent jouer un rôle possible à évaluer à partir des données collectées dans le cadre du corpus présenté à la section 6.3. Ces notions de personnalité et de rôle social sont elles-mêmes complexes, et renvoient à une littérature importante.

Les travaux sur la notion de personnalité font état de plusieurs approches de cette notion. On trouve dans la littérature des modèles de description de la « personnalité » sur la base d'ensembles de traits – par exemple un nombre de traits qui s'organisent hiérarchiquement à partir d'un facteur général de personnalité (Rushton et Irwing, 2011). Ces traits mesurent une adaptation des individus à la vie en société, de manière très générale. Ils recoupent donc d'une certaine manière les rôles sociaux abordés ci-dessous, dans une perspective descriptive plus individuelle et d'inspiration darwinienne (un individu « doué » d'une personnalité positivement évaluée sur ces échelles aura plus de chances de se reproduire, etc.). Ces traits de personnalités peuvent être évalués par les sujets eux-mêmes, ou par des personnes externes, avec une certaine cohérence entre les deux sources d'information (Dobewall *et al.*, 2014).

Une première évaluation d'un possible lien entre traits de personnalité (mesurés dans le cadre d'un modèle à cinq traits, cf. Costa et McCrae, 1992) et perception d'expressions attitudinales a été mené sur un ensemble d'attitudes françaises et japonaises (Clavel *et al.*, 2009). On observe peu d'effets significatifs de la personnalité sur les affects les mieux reconnus. Par contre, des effets facilitateurs de certains traits de personnalité pour la reconnaissance de certaines expressions attitudinales peu marquées sont observés. En particulier, les locuteurs les moins performants sont mieux compris, dans leurs expressions les moins efficaces, par les auditeurs dont la personnalité est « compatible » avec les expressions.

De telles observations permettent de dire par exemple qu'une auto-évaluation élevée sur l'échelle d'« openness » est liée à une meilleure perception de la surprise ; inversement un score bas sur cette échelle correspond à des auditeurs percevant mieux l'expression de doute. La différence sémantique entre les attitudes de surprise et de doute peut se caractériser par une tendance à accepter (resp. rejeter) une information nouvelle ; de même, l'échelle d'« openness » est décrite comme une ouverture d'esprit à de nouvelles expériences (resp. des personnes socialement conservatrices). Les personnes ayant un score bas sur cette échelle d'« openness » sont aussi celles qui perçoivent le mieux les différentes expressions de politesse conventionnelles, pour les ex-

pressions japonaises ; ces expressions sont l'objet de conventions fortes dans la société japonaise (cf. section 6.1).

L'échelle de « *conscientiousness* » correspond, pour des scores élevés, à des personnes ayant une forte sensibilité aux règles. Les auditeurs ayant des scores importants sur cette échelle ont obtenu de meilleurs taux de reconnaissance pour les expressions d'imposition de la volonté du locuteur (cf. section 5.3.2), telles que l'attitude d'évidence en français et celles d'arrogance et d'irritation en japonais. Ces résultats soulignent l'importance des facteurs sociaux et individuels tant dans la production que pour la perception des attitudes.

Plus générique que la notion de personnalité - notion mettant l'individu en avant – le rôle social d'une personne est défini par une société en fonction de ses attentes par rapport à des caractéristiques socio-politiques telles que l'âge, le sexe, le métier ou l'origine dialectale. Les travaux de Sadanobu (2012) sur la notion de « *character* » sont intéressants pour traiter de cette notion de rôle. Ils montrent comment des individus peuvent se comporter de différentes manières selon les situations et comment il est attendu qu'ils se comportent en fonction de leur rôle et de leur place dans la société (évidemment, ces attentes varient d'un groupe social à l'autre). Aristote (2007) donne aussi, dans le second livre de sa rhétorique, de nombreux portraits de types de personnes ayant des caractéristiques différentes au regard des passions, du caractère, etc. (vieillards, riches, nobles, etc.) types qu'il est important de connaitre pour mieux les convaincre. Dans un autre registre, Podesva (2007) montre des variations dans l'utilisation de certaines variétés de qualité vocale, attribuées à des différences situationnelles, liées en outre à la construction d'une identité. Les travaux de Sadanobu (2012) montrent que ces rôles (*adult, young lady, well-bred young lady, child, professor, Osaka native,* etc.) sont définis selon des normes culturelles qui peuvent dérouter une personne d'une autre culture. Comme on l'a vu dans le cas de la politesse, certaines cultures (et typiquement la culture japonaise) ont des attentes sociales fortes à l'égard des individus : leur place et leur rôle est important, normé, et répond à l'utilisation de nombreux codes qu'il est difficile d'appréhender car ils ne sont pas nécessairement formalisés. Ces rôles – et bien sûr les spécificités lexicales et multimodales de leurs performances – sont donc des notions qu'il faudrait enseigner. Il s'agit en effet aussi de facteurs qui participent à la variation linguistique (Pierrehumbert, 2001). Pour cela, il est important de mieux comprendre comment ils participent à contraindre l'expressivité, produite et perçue. Est-on capable de déterminer quel est le « personnage » d'un locuteur, hors contexte et sur la base de son expressivité multimodale ? En quoi un contexte donné pourra-t-il induire un comportement proche de tel ou tel aspect de ces rôles sociaux, en fonction du personnage réel du locuteur ?

En quoi les paramètres socio-culturels propres à un locuteur (âge, sexe, etc.) peuvent-ils contraindre la perception de son rôle social et son expressivité ?

Il est possible de travailler sur ces questions sur la base des données du corpus présenté à la section 6.3. Il serait intéressant pour cela d'aller vers une annotation plus précise des variations multiples qui sont capturées lors des interactions ayant mené à la création de ce corpus, afin notamment d'être à même de comparer l'apport des différentes modalités ; un cadre d'annotations tel que celui proposé par Seinturier *et al.* (2012) amènerait certainement de meilleures possibilités d'analyse des variations produites. En ce qui concerne l'analyse de la perception de ces variations, il faut mettre en place des paradigmes expérimentaux qui permettent une évaluation raisonnable de ces différents paramètres, par exemple en imposant des contraintes à l'aide d'outils de (re)synthèse de la parole (voir la section précédente). Ce champ de recherche offre de nombreuses possibilités pour une meilleure compréhension de la variation interculturelle – pourquoi pas afin de pouvoir comprendre ce que pourrait signifier « traduire de la prosodie » ? Notons qu'à ce titre, des travaux très intéressants sont menés au LAEL (Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo) sur les techniques utilisées par les interprètes de conférence pour transmettre les informations prosodiques dans une autre langue.

8.3 Variation prosodique & interrogatives

On a vu (cf. section 4.2) que la variation prosodique dialectale prenait un intérêt et une complexité particulière avec l'étude des phrases interrogatives. Il semble en effet qu'il y ait plus de variations pour ce mode illocutoire que dans le cas des assertives : on observe de manière récurrente des patrons se démarquant de la forme prototypique, qui prône une montée mélodique finale (Delattre, 1966).

Pour les principales langues romanes (espagnol, français, italien, portugais, roumain), la réalisation des questions oui/non est décrite comme suivant la norme d'une montée finale (Hirst et Di Cristo, 1998)[1] (voir aussi del Mar Vanrell *et al.*, 2012, pour le catalan). Cependant, de nombreuses variantes sont relevées dès qu'on s'intéresse à des travaux sur des dialectes ; par exemple pour le galicien, Fernández Rei *et al.* (2007a); Fernández Rei et Escourido Pernas (2008); Fernández Rei (2011, 2013) montrent l'existence de différents contours intonatifs pour l'interrogation, contours qui sont reliés à la variation diatopique, avec une indexicalité de l'origine dialectale des locuteurs. D'autres exemples de telles variations intonatives sont décrits dans

[1]Il peut s'agir de montées sur la dernière syllabe ou sur la dernière syllabe accentuée.

la littérature ; pour le corse : Boula de Mareüil *et al.* (2012c,a, 2014) ; pour d'autres variétés de l'aire italo-romane : Savino (2012) ; pour le portugais brésilien : Seara *et al.* (2014) ; mais aussi en dehors du domaine roman – pour certaines langues africaines, voir Rialland (2007).

Les travaux de Hedberg et Sosa (2002), menés sur des données de parole spontanée, partent des attendus théoriques de la littérature et cherchent à les observer - sans forcément arriver aux mêmes conclusions. L'étude très détaillée de Fontaney (1991), qui porte sur l'intonation des questions en français, montre différentes typologies d'actes interrogatifs, avec des tendances à les réaliser suivant une forme intonative particulière, mais avec beaucoup de variations possibles.

Ces variations intonatives du mode interrogatif soulèvent donc des questions complexes et en particulier celle des fonctions communicatives et culturelles que véhicule la prosodie (voir par exemple Rebollo Couto *et al.*, 2010; Edlund *et al.*, 2012). La question plus générale de la langue dans son usage sociologique va aussi transmettre des indices de prestige et/ou identitaires (Mufwene, 2007; Vigouroux, 2008) ; la capacité de marquage diatopique de ces réalisations intonatives devient alors particulièrement imbriquée dans l'identité du locuteur et dans ce qu'il veut en afficher. Pour rester sur l'hypothèse d'un possible marquage diatopique, et/ou identitaire, de l'intonation des questions, les travaux de (Boula de Mareüil *et al.*, 2013, 2014) montrent de possibles transferts prosodiques du substrat dialectal vers la langue nationale, ici dans le cas du corse vers le français.

La poursuite d'un travail d'abord descriptif des formes intonatives de la modalité interrogative me semble donc important pour mieux cerner et comprendre ces phénomènes. Pour cela, la méthodologie AMPER (Contini, 1991; Contini *et al.*, 2002) fournit un cadre qui s'est déjà révélé productif pour observer des variations d'origine dialectale (cf. section 4), et qui pourrait l'être dans le cadre d'études intéressées par d'autres sources de variations (sociale par exemple). Les zones de contacts linguistiques sont ainsi particulièrement pertinentes pour observer des variations (Rebollo Couto *et al.*, 2008) et pour les tester expérimentalement (Boula de Mareüil *et al.*, 2014). En effet, les fonctions prosodiques sont subtiles et les distinctions proposées – par exemple entre question oui/non ou question de confirmation – peuvent être délicates à évaluer lors de tests perceptifs. De même, la perception d'une variation d'origine diatopique se testera mieux à proximité d'une frontière, là où les auditeurs auront une conscience plus affinée des schémas prototypiques des deux variétés.

Du point de vue de la description des variations prosodiques liées à ces variantes, des outils de stylisation sont importants pour n'observer que des différences pertinentes (par exemple la stylisation par Prosogram utilisée

par Cunha *et al.*, 2008). Il est aussi primordial d'effectuer des analyses automatiques afin de pouvoir traiter des données en quantités importantes, car la taille des corpus permet d'abord pour avoir une représentativité des variations et éviter l'écueil d'étudier un idiolecte, mais aussi de pouvoir proposer des hypothèses fortes, soutenues par des récurrences statistiques suffisantes (voir par exemple les travaux de Ferragne et Pellegrino, 2007, pour la variation dialectale de l'anglais britannique).

L'application des mesures objectives de distance prosodique qui sous-tendent ces travaux reste ainsi à effectuer sur ces données. Des avancées sont cependant encore nécessaires pour comparer de telles réalisations prosodiques de manière objective. En effet, si l'on regarde le détail des résultats obtenus grâce à la méthode d'alignement par DTW (cf. section 2.2.3 et Rilliard *et al.*, 2011), on peut remarquer des problèmes liés aux effet des réalisations de la fonction morphosyntaxique, que cette approche ne sait pas dissocier des variations liées à la réalisation d'attitudes. Ainsi, les phrases présentées à la figure 8.2 présentent toutes la même attitude prosodique de doute, caractérisée notamment par une montée finale. Les deux phrases des graphiques du haut, comme les deux phrases des graphiques du bas, partagent le même nombre d'accents lexicaux – et elles sont très bien alignées par DTW. Par contre, cette méthode ne sait pas aligner les contours prosodiques des phrases du haut, montrant un seul accent lexical, avec ceux du bas, montrant deux accents lexicaux. Il faudrait pour cela être capable d'extraire le contour attitudinal de la phrase et celui des accents.

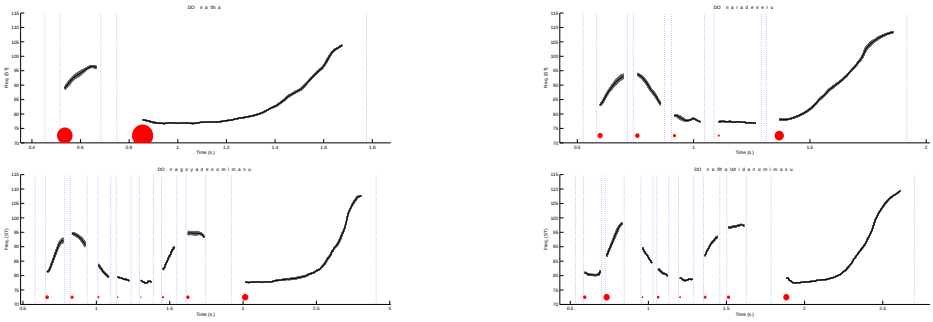


Figure *8.2* – *Tracés de F_0 de l'attitude de doute produite par un locuteur masculin en japonais, sur des phrases de taille variable (2, 5syllabes en haut, 8 syllabes en bas). Les deux phrases du bas ont le même nombre de syllabe mais des accents lexicaux placés sur des syllabes différentes (voir Shochi* et al.*, 2009c).*

Ce type de séparation des différents niveaux hiérarchiques de la prosodie est typiquement ce qui est proposé par l'implémentation du modèle de

Fujisaki (1983) faite par Mixdorff (2000). Il faudrait toutefois tester cette implémentation pour voir comment elle est capable de capturer la diversité des contours attitudinaux réalisés sur un même empan. Les travaux sur la réalisation des tons dans le cadre de ce modèle devrait pouvoir fournir des pistes de réponse à ce problème (Mixdorff *et al.*, 2002).

8.3.1 Prosodie & performance

L'approche de la prosodie par son contrôle gestuel, présentée à la section 2.1 (d'Alessandro *et al.*, 2006), offre d'intéressantes possibilités d'investigation des fonctions prosodiques. D'une part, cette méthode, comme toute méthode d'analyse par la synthèse (Öhman et Lindqvist, 1965), permet de s'abstraire des programmes d'estimation des divers paramètres acoustiques, qui introduisent souvent des incertitudes. Cela est d'autant plus vrai quand on commence à s'intéresser aux aspects de qualité vocale, pour lesquels les procédures d'inversion destinées à retrouver les paramètres d'un modèle source-filtre (Doval *et al.*, 2006) hypothétique sont notamment sensibles au contenu segmental et demandent des normalisations (Hanson, 1997).

À l'inverse, l'approche de la variation prosodique par la synthèse donne un accès direct aux paramètres du modèle – avec le défaut inverse, que ce qui est observé est borné par le modèle et les capacités expressives du synthétiseur. L'un des avantages du contrôle gestuel de la synthèse consiste justement dans la possibilité offerte à un humain de produire lui-même une variation prosodique qu'il jugera adéquate à une tâche donnée. À charge par la suite aux expérimentateurs d'analyser les variations des contrôles et des paramètres liés à ces performances expressives (voir d'Alessandro *et al.*, 2014, pour le chant).

Un second avantage de l'approche par contrôle gestuel, déjà évoqué à la section 2.1, vient de la normalisation des variations microprosodiques introduite par le processus chironomique : le geste effectué ne reproduit pas ces variations, et l'étude des tracés n'est donc aucunement biaisé par ce niveau de variation. Ce modèle constitue donc un paradigme propice à l'étude de la dynamique des paramètres – mélodiques en particulier.

Il reste bien sûr à trouver une solution adéquate au problème majeur posé par cette approche : l'articulation. S'il parait improbable – et sans doute non souhaitable – de pouvoir contrôler par les geste l'articulation de synthèse, on peut envisager d'utiliser la sortie d'un système de synthèse de la parole comme entrée d'une série de phonèmes, avec les paramètres adéquats pour les faire produire par un vocodeur. Le geste viendrait alors contrôler les

paramètres *prosodiques* à proprement parler : débit, intonation, force de voix et autres variations de qualité vocale.

De manière plus prospective, il me semble important de développer et d'utiliser des méthodes de mesure de paramètres acoustiques robustes à extraire de la parole. Ceci est particulièrement vrai dans le cas des mesures de qualité de voix, qui souffrent par ailleurs d'un manque de consensus sur les dimensions à prendre en compte. Mais cela est vrai aussi par exemple pour la fréquence fondamentale, qui reste l'indice le plus important des études en prosodie, mais souffre d'une nature impropre à nombre de traitements statistiques, notamment du fait de sa discontinuité (voir Garner *et al.*, 2013). Il reste donc encore du travail à effectuer concernant des outils d'analyse du signal, plus standards et fonctionnels (à cet égard PRAAT et les scripts du PROSOGRAM ou de MOMEL constituent d'importantes avancées). Ces paramètres robustes pourront aussi servir à mieux comprendre la variation temporelle de la prosodie.

Pour finir sur une note brésilienne, la cooccurrence d'évènements externes porteurs d'une charge émotionnelle forte peut perturber localement le cours de la production de parole et fournir des indices forts (même s'il sont aussi rares), sur l'importance de l'évènement ; c'est typiquement ce qui se produit dans le discours des commentateurs sportifs, à l'occasion d'un but.

9 | Bibliographie

Adda-Decker, M., Boula de Mareüil, P., Adda, G. et Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous french. *Speech Communication*, 46(2):119–139.

d'Alessandro, C. (2001). 33 ans de synthèse de la parole à partir du texte : une promenade sonore (1968-2001). *Traitement Automatique des Langues*, 42(1):1–29.

d'Alessandro, C. (2006). Voice source parameters and prosodic analysis. *In* Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. et Schliesser, J., éditeurs : *Methods in empirical prosody research*, pages 63–87. Berlin : Walter de Gruyter.

d'Alessandro, C., Darsinos, V. et Yegnanarayana, B. (1998). Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources. *IEEE Transactions on Speech and Audio Processing*, 6(1):12–23.

d'Alessandro, C., Feugère, L., Le Beux, S., Perrotin, O. et Rilliard, A. (2014). Drawing melodies : Evaluation of chironomic singing synthesis. *The Journal of the Acoustical Society of America*, 135(6):3601–3612.

d'Alessandro, C. et Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech & Language*, 9(3):257–288.

d'Alessandro, C., Rilliard, A. et Le Beux, S. (2011). Chironomic stylization of intonation. *Journal of the Acoustical Society of America*, 129(3):1594–1604.

d'Alessandro, C. (2010). La parole comme mouvement : glossolalies chironomiques. *In Conférence invitée aux XXVIIIèmes Journées d'Étude sur la Parole JEP 2010 - 25-28 mai 2010, Mons, Belgique.*

d'Alessandro, C. (2011). Computerized chironomy : Five years of gesture-controlled voice and speech synthesis at LIMSI. *In Keynote at the 1st International Workshop on Performative Speech and Singing Synthesis (p3s), Vancouver, 14-15 mars 2011.*

d'Alessandro, C. (2014). Glossolalies électroniques. *In CUTE - Materclass series on culture and technology, march 12-15, 2014, University of Mons, Belgique.*

d'Alessandro, C., D'Alessandro, N., Le Beux, S. et Doval, B. (2006). Comparing time-domain and spectral-domain voice source models for gesture controlled vocal instruments. *In Proc. of the 5th International Conference on Voice Physiology and Biomechanics, Tokyo*, volume 34, page 37.

d'Alessandro, C. et Sturmel, N. (2011). Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude. *Sadhana*, 36(5):601–622.

D'Alessandro, N., Woodruff, P., Fabre, Y., Dutoit, T., Le Beux, S., Doval, B. et d'Alessandro, C. (2007). Realtime and accurate musical control of expression in singing synthesis. *Journal on Multimodal User Interfaces*, 1(1):31–39.

Alvarellos, M., Muñiz, C., Díaz, L. et González, R. (2011). La entonación en las variedades lingüísticas de asturias : estudio contrastivo. *Revista internacional de lingüística iberoamericana*, pages 111–120.

Aristote (trad. 2007). *Rhétorique*. GF Flammarion, Paris.

Aston, J. A., Chiou, J.-M. et Evans, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 59(2):297–317.

Astésano, C., Bard, E. G. et Turk, A. (2007). Structural influences on initial accent placement in french. *Language and Speech*, 50(3):423–446.

Aubergé, V., Audibert, N. et Rilliard, A. (2006). De E-Wiz à C-Clone : recueil, modélisation et synthèse d'expressions authentiques. *Revue d'intelligence artificielle*, 20(4-5):499–527.

Aubergé, V. et Cathiard, M. (2003). Can we hear the prosody of smile ? *Speech Communication*, 40(1):87–97.

Aubergé, V. (1992). Developing a structured lexicon for synthesis of prosody. *In* Bailly, G., Benoît, C. et Sawallis, T. R., éditeurs : *Talking machines : theories, models, and designs*, pages 307–321. Amsterdam : Elsevier Science.

Aubergé, V., Grépillat, T. et Rilliard, A. (1997). Can we perceive attitudes before the end of sentences ? the gating paradigm for prosodic contours. *In EuroSpeech'97, Rhodes, Grèce*, pages 871–877.

Audibert, N. (2008). *Prosodie de la parole expressive : dimensionnalité d'énoncés méthodologiquement contrôlés authentiques et actés*. Thèse de doctorat, Institut National Polytechnique de Grenoble-INPG.

Audibert, N., Aubergé, V. et Rilliard, A. (2010). Discrimination perceptive d'expressions émotionnelles actées vs. spontanées : Variabilité interindividuelle et influence de l'intensité de l'émotion. *TSI. Technique et science informatiques*, 29(7):833–857.

Aznárez-Mauleón, M. et González-Ruiz, R. (2006). Francamente, el rojo te sienta fatal – semantics and pragmatics of some expressions of sincerity in present-day spanish. *In* Peeters, B., éditeur : *Semantic Primes and Universal Grammar : Empirical Findings from the Romance Languages*, pages 307–330. Amsterdam : John Benjamins.

Banse, R. et Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70:614–636.

Barbosa, P. (1994). *Caractérisation et génération automatique de la structuration rythmique du français*. Thèse de doctorat, Institut National Polytechnique, Grenoble, France.

Barbosa, P. (2007). From syntax to acoustic duration : A dynamical model of speech rhythm production. *Speech Communication*, 49(9):725–742.

Bartens, A. et Sandström, N. (2006). Towards a description of spanish and italian diminutives within the NSM framework. *In* Peeters, B., éditeur : *Semantic Primes and Universal Grammar : Empirical Findings from the Romance Languages*, pages 331–360. Amsterdam : John Benjamins.

Bell, A. M. (1849). *A new elucidation of the principles of speech and elocution*. W.P. Kennedy, Edinburgh.

Benzécri, J.-P. E. (1973). *L'analyse des données - II L'analyse des correspondances*, volume 2. Dunod, Paris.

Bishop, J. et Keating, P. (2012). Perception of pitch location within a speaker's range :

Fundamental frequency, voice quality and speaker sex. *The Journal of the Acoustical Society of America*, 132(2):1100–1112.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the institute of phonetic sciences*, 17:97–110.

Boersma, P. et Weenink, D. (2012). Praat : doing phonetics by computer (version 5.3.32)[computer program]. retrieved october 17, 2012.

Bohnemeyer, J. (2004). NSM without the strong lexicalization hypothesis. *Theoretical linguistics*, 29(3):211–222.

Boula de Mareüil, P., Mairano, P., Rilliard, A. et Lai, J. (2012a). Corsican French questions : is there a prosodic transfer from Corsican to French and how to highlight it ? *In 6th International Conference on Speech Prosody, Shanghai, China*, pages 418–421, Shanghai, China.

Boula de Mareüil, P., Rilliard, A. et Allauzen, A. (2012b). A diachronic study of initial stress and other prosodic features in the French news announcer style : corpus-based measurements and perceptual experiments. *Language and Speech*, 55:263–293.

Boula de Mareüil, P., Rilliard, A., Lehka-Lemarchand, I. et Ivent, F. (2013). Regional accents and languages in France : a contrastive prosodic analysis of Romance varieties. *In Workshop on Phonetics, Phonology and Languages in Contact (PPLC 2013)*, pages 72–75.

Boula de Mareüil, P., Rilliard, A., Lehka-Lemarchand, I., Mairano, P. et Lai, J.-P. (2014). Falling yes/no questions in Corsican French and Corsican : evidence for a prosodic transfer. *In* Delais-Roussarie, E., Avanzi, M. et Herment, S., éditeurs : *Languages in contact*. Springer Verlag, Berlin.

Boula de Mareüil, P., Rilliard, A., Mairano, P. et Lai, J. (2012c). Questions corses : peut-on mettre en évidence un transfert prosodique du corse vers le français ? *In Journées d'Etude sur la Parole (JEP 2012)*, Actes de la conférence conjointe JEP-TALN-RECITAL, pages 609–616, Grenoble, France.

Bozkurt, B., Doval, B., d'Alessandro, C. et Dutoit, T. (2005). Zeros of z-transform representation with application to source-filter separation in speech. *Signal Processing Letters, IEEE*, 12(4):344–347.

Brandt, P. A. (2008). Thinking and language. A view from cognitive semio-linguistics. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 649–654.

Brassens, G. (1972). Quatre-vingt-quinze pourcent. *In Fernande*. Philips.

Brosch, T., Pourtois, G. et Sander, D. (2010). The perception and categorisation of emotional stimuli : A review. *Cognition & Emotion*, 24(3):377–400.

Brown, P. et Levinson, S. C. (1987). *Politeness : Some Universals in Language Usage*. Studies in Interactional Sociolinguistics. Cambridge University Press.

Campbell, N. (1993). Automatic detection of prosodic boundaries in speech. *Speech communication*, 13(3):343–354.

Campbell, N. (2000). Databases of emotional speech. *In ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, Newcastle, North Ireland*, pages 34–38.

Campbell, N. (2004). Speech & expression ; the value of a longitudinal corpus. *In 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal*, pages 183–186.

Campbell, N. (2005). Getting to the heart of the matter : Speech as the expression of

affect ; rather than just text or language. *Language Resources and Evaluation*, 39(1):109–118.

Carton, F. (2000). La prononciation. *In* Antoine, G. et Cerquiglini, B., éditeurs : *Histoire de la langue française 1945–2000*, pages 25–60. CNRS Éditions, Paris, France.

Chambers, J. K. et Trudgill, P. (1998). *Dialectology (2nd edn)*. Cambridge University Press, Cambridge.

Clavel, C., Rilliard, A., Shochi, T. et Martin, J.-C. (2009). Personality differences in the multimodal perception and expression of cultural attitudes and emotions. *In IEEE International Workshop on Social Signal Processing, Amsterdam, The Netherlands*, page 6p.

Contini, M. (1991). Vers une géoprosodie. *In Nazioarteko Dialektologia Biltzarra. Agiriak*, pages 83–109.

Contini, M., Lai, J.-P., Romano, A., Roullet, S., Moutinho, L. d. C., Coimbra, R. L., Bendiha, U. P. et Ruivo, S. S. (2002). Un projet d'atlas multimédia prosodique de l'espace roman. *In 1st International Conference on Speech Prosody, Aix-en-Provence, France*.

Contini, M. et Profili, O. (1989). L'intonation de l'italien régional - un modèle de description par traits. *In Mélanges de phonétique générale et expérimentale offerts à Péla Simon*, pages 855–870.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. et Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, 24(6):597–606.

Cooper, F. S., Liberman, A. M. et Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences of the United States of America*, 37(5):318.

Costa, P. T. et McCrae, R. R. (1992). Normal personality assessment in clinical practice : The NEO personality inventory. *Psychological Assessment*, 4(1):5 – 13.

Culpeper, J., Bousfield, D. et Wichmann, A. (2003). Impoliteness revisited : with special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, 35:1545–1579.

Cunha, C., Felismino, G., Rebollo, L. et Silva, M. (2008). Focus and intonational marking in boundaries dialects : Brazilian Portuguese and Uruguayan Spanish in yes/no questions. *4th International Conference on Speech Prosody, Campinas, Brésil*, pages 551–554.

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Research Reviews*, 26(2–3):83–86.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J. et Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3(10):1049–1056.

Daneš, F. (1994). Involvement with language and in language. *Journal of pragmatics*, 22(3-4):251–264.

Dard, F. (1986). *Alice au pays des merguez*. Fleuve noir, Paris.

Dard, F. (1999). *Ceci est bien une pipe*. Fleuve noir, Paris.

Decety, J. et Grèzes, J. (2006). The power of simulation : imagining one's own and other's behavior. *Brain research*, 1079(1):4–14.

Delattre, P. (1966). Les dix intonations de base du français. *The French review*, 40(1):1–14.

Di Cristo, A. (1998). Intonation in French. *In* Hirst, D. et Di Cristo, A., éditeurs : *Intonation systems : A survey of twenty languages*, pages 195–218. Cambridge University Press, Cambridge.

Di Cristo, A. (2013). *La prosodie de la parole*. de Boeck, Bruxelles.

Dobewall, H., Aavik, T., Konstabel, K., Schwartz, S. H. et Realo, A. (2014). A comparison of self-other agreement in personal values versus the Big Five personality traits. *Journal of research in personality*, 50:1–10.

Dorta, J., éditeur (2007). *La prosodia en el ámbito lingüístico románico*. La Página Ediciones, Santa Cruz de Tenerife.

Doukhan, D. (2013). *Synthèse de parole expressive au delà du niveau de la phrase : le cas du conte pour enfant*. Thèse de doctorat, Université Paris Sud.

Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M. et d'Alessandro, C. (2011). Prosodic analysis of a corpus of tales. *In Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, pages 3129–3132, Firenze, Italy.

Doukhan, D., Rosset, S., Rilliard, A., d'Alessandro, C. et Adda-Decker, M. (2012). Designing French tale corpora for entertaining text to speech synthesis. *In International Conference on Language Resources and Evaluation (LREC 2012)*, pages 1003–1010, Istanbul, Turkey.

Doval, B., d'Alessandro, C. et Henrich, N. (2006). The spectrum of glottal flow models. *Acta Acustica united with Acustica*, 92(6):1226–1246.

Đô, T. D., Tran, T. H. et Boulakia, G. (1998). Intonation in Vietnamese. *In* Hirst, D. et Di Cristo, A., éditeurs : *Intonation systems : A survey of twenty languages*, pages 398–420. Cambridge University Press, Cambridge.

Dumas, A. (1847). *La Reine Margot*. Pétion, Paris.

Edlund, J., House, D. et Strömbergsson, S. (2012). Question types and some prosodic correlates in 600 questions in the spontal database of swedish dialogues. *In 6th International Conference on Speech Prosody, Shanghai, Chine*, pages 737–740.

Émond, C. (2013). *Les corrélats prosodiques et fonctionnels de la parole perçue souriante en français québécois spontané*. Thèse de doctorat, Université du Québec à Montréal.

Escourido Pernas, A. B. (2008). A prosodia de Camelle : descrición acústica dunha fala. *Cadernos de lingua*, 30/31:75–122.

Evans, J., Chu, M.-n., Aston, J. A. D. et Su, C.-y. (2010). Linguistic and human effects on f0 in a tonal dialect of qiang. *Phonetica*, 67(1-2):82–99.

Fernández Rei, E. (2011). La declinación en las interrogativas totales del gallego : estudio perceptivo. *Revista Internacional de Lingüística Iberoamericana*, pages 121–132.

Fernández Rei, E. (2013). Estudo perceptivo das distancias prosódicas dialectais. *In Présentation aux « Xornadas de Dialectoloxía Perceptiva », Santiago de Compostela, 17-18 janvier 2013*.

Fernández Rei, E. et Escourido Pernas, A. (2008). La entonación de las interrogativas totales a lo largo de la costa gallega. *In* Turculeţ, A., éditeur : *La variation diatopique de l'intonation dans le domaine roumain et roman*, pages 151–166. Editura universităţii «Alexandru Ioan Cuza», Iaşi.

Fernández Rei, E., Escourido Pernas, A. et Caamaño Varela, M. (2007a). Caracterización prosódica das interrogativas da Costa da Morte e do Morrazo. *In Actas do III Congreso Internacional de Fonética Experimental, Santiago de Compostela : Xunta*

de Galicia, pages 305–315.

Fernández Rei, E., Escourido Pernas, A. et Gómez Castro, S. (2007b). Acento e entoación nas frases con extensión en dúas variedades do galego (O Incio e Camelle). *In* Dorta, J., éditeur : *La prosodia en el ámbito lingüístico románico.*, pages 55–71. Tenerife : La Página.

Ferragne, E. et Pellegrino, F. (2007). Automatic dialect identification : A study of British English. *In Speaker classification II*, pages 243–257. Springer, Berlin.

Ferraty, F. et Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer, Berlin.

Fónagy, I. (1989). Le français change de visage ? *Revue Romane*, 24:225–254.

Fónagy, I., Bérard, E. et Fónagy, J. (1983). Clichés mélodiques. *Folia linguistica*, 17(1-4):153–186.

Fónagy, I. et Fónagy, J. (1976). Prosodie professionnelle et changements prosodiques. *Le français moderne*, 3:193–227.

Fontaney, L. (1991). A la lumière de l'intonation. *In* Kerbrat-Orecchioni, C., éditeur : *La question*, pages 113–161. Lyon : Presses Universitaires de Lyon.

Fourer, D., Guerry, M., Shochi1, T., Rouas, J.-L., Aucouturier, J.-J. et Rilliard, A. (2014). Analyse prosodique des affects sociaux dans l'interaction face à face en japonais. *In Journées d'Etude sur la Parole (JEP 2014)*, Le Mans, France.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. *In* MacNeilage, P., éditeur : *The production of speech*, pages 39–55. Springer, New York, NY.

Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. *In* Fujimura, O., éditeur : *Vocal fold physiology : voice production, mechanisms and functions*, pages 347–355. Raven, New York, NY.

Garcin de Tassy, J. H. (1873). *Rhétorique et prosodie des langues de l'orient musulman - à l'usage des élèves de l'école spéciale des langues orientales vivantes.* Maisonneuve et Cie, Paris.

Garner, P. N., Cernak, M. et Motlicek, P. (2013). A simple continuous pitch estimation algorithm. *Signal Processing Letters, IEEE*, 20(1):102–105.

Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., Pelachaud, C., Rilliard, A. et Rosset, S. (2010). Towards a storytelling humanoid robot. *In AAAI 2010 Fall Symposia, "Dialog with Robots", Arlington (VA)*, pages 2263–2266.

Goddard, C. (1997). Cultural values and 'cultural scripts' of Malay (Bahasa Melayu). *Journal of pragmatics*, 27(2):183–201.

Goddard, C. (2002). The search for the shared semantic core of all languages. *In* Goddard, C. et Wierzbicka, A., éditeurs : *Meaning and Universal Grammar – Theory and Empirical Findings. Volume I*, pages 5–40. Amsterdam : John Benjamins.

Goddard, C. (2007). A culture-neutral metalanguage for mental state concepts. *In* Schalley, A. C. et Khlentzos, D., éditeurs : *Mental States. Volume 2 : Language and Cognitive Structure*, pages 11–35. Amsterdam : John Benjamins.

Goddard, C. (2012). Semantic primes, semantic molecules, semantic templates : Key concepts in the nsm approach to lexical typology. *Linguistics*, 50(3):711–743.

Goddard, C. et Wierzbicka, A., éditeurs (2002). *Meaning and universal grammar :*

Theory and empirical findings, volume 1. John Benjamins Publishing.

Goebl, H. (1981). Eléments d'analyse dialectométrique (avec application à l'AIS). *Revue de linguistique romane*, 45:349–420.

Goebl, H. (1996). La convergence entre les fragmentations géo-linguistiques et géo-génétiques de l'italie du nord. *Revue de linguistique romane*, 60:25–49.

Goebl, H., Selberherr, S., Rase, W.-D. et Pudlatz, H. (1982). Atlas, matrices et similarités : Petit aperçu dialectométrique. *Computers and the Humanities*, 16(2):69–84.

Grabe, E., Kochanski, G. et Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and speech*, 50(3):281–310.

Gu, W., Zhang, T. et Fujisaki, H. (2011). Prosodic analysis and perception of mandarin utterances conveying attitudes. *In Interspeech 2011 - 12th Annual Conference of the International Speech Communication Association, Florence, Italy*, pages 1069–1072.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Hadjipantelis, P. Z., Aston, J. A. et Evans, J. P. (2012). Characterizing fundamental frequency in mandarin : A functional principal component approach utilizing mixed effect models. *The Journal of the Acoustical Society of America*, 131(6):4651–4664.

Hall, J. A., Carter, J. D. et Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. *In* Fischer, A., éditeur : *Gender and emotion : Social psychological perspectives*, pages 97–117. Cambridge University Press, Cambridge, UK.

Hanson, H. M. (1997). Glottal characteristics of female speakers : Acoustic correlates. *The Journal of the Acoustical Society of America*, 101(1):466–481.

Harkins, J. et Wierzbicka, A. (2001). *Emotions in crosslinguistic perspective*, volume 17. Mouton de Gruyter.

't Hart, J. (1991). F0 stylization in speech : straight lines versus parabolas. *The Journal of the Acoustical Society of America*, 90(6):3368–3372.

't Hart, J., Collier, R. et Cohen, A. (1991). *A perceptual study of intonation : an experimental-phonetic approach to speech melody*. Cambridge University Press, Cambridge, UK.

Hedberg, N. et Sosa, J. M. (2002). The prosody of questions in natural discourse. *In 1st International Conference on Speech Prosody, Aix-en-Provence, France*.

Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. Thèse de doctorat, University of Groningen.

Henrich, N., d'Alessandro, C., Doval, B. et Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America*, 115(3):1321–1332.

Henrich, N., d'Alessandro, C., Doval, B. et Castellengo, M. (2005). Glottal open quotient in singing : Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *The Journal of the Acoustical Society of America*, 117(3):1417–1430.

Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *The journal of the acoustical society of America*, 83(1):257–264.

Hermes, D. J. (1995). Measuring the perceptual similarity of pitch contours. *In 4th European Conference on Speech Communication and Technology Eurospeech'95*, pages

2051–2054.

Hermes, D. J. (1998a). Auditory and visual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41(1):63–72.

Hermes, D. J. (1998b). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41(1):73–82.

Hill, B., Ide, S., Ikuta, S., Kawasaki, A. et Ogino, T. (1986). Universals of linguistic politeness : Quantitative evidence from Japanese and American English. *Journal of Pragmatics*, 10:347–371.

Hinton, L., Nichols, J. et Ohala, J. J. (1994). *Sound symbolism*. Cambridge University Press.

Hirst, D. (2005). Form and function in the representation of speech prosody. *Speech Communication*, 46(3/4):334–347.

Hirst, D. et Di Cristo, A. (1998). *Intonation systems : a survey of twenty languages*. Cambridge University Press, Cambridge.

Hirst, D. et Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15:75–85.

Hirst, D., Rilliard, A. et Aubergé, V. (1998). Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. *In The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*.

Husson, F., Josse, J., Le, S. et Mazet, J. (2013). *FactoMineR : Multivariate Exploratory Data Analysis and Data Mining with R*. R package version 1.25.

Husson, F., Lê, S. et Pagès, J. (2010). *Exploratory Multivariate Analysis by Example Using R*. Chapman and Hall.

Hönemann, A., Mixdorff, H. et Rilliard, A. (2014). Social attitudes - recordings and evaluation of an audio-visual corpus in German. *In 7th Forum Acusticum 2014, Krakow*.

Ide, S. (2002). The speaker's viewpoint and indexicality in a high context culture. *In* Ide, S. et Kataoka, K., éditeurs : *Bunka, Intaakushon, Gengo [Culture, Interaction, and Language]*, pages 3–20. Hituzi Syobor, Tokyo, Japon.

Ito, K. et Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. *In* Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. et Schliesser, J., éditeurs : *Methods in empirical prosody research*, pages 229–257. Walter de Gruyter, Berlin, Germany.

Jankowski, L., Astésano, C. et Di Cristo, A. (1999). The initial rhythmic accent in French : Acoustical and perceptual prosodic cues. *In 14th International Congress of Phonetic Sciences (ICPhS), San Francisco, CA*, pages 257–260.

Kawahara, H. (2008). Tandem-straight, a research tool for L2 study enabling flexible manipulations of prosodic information. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 619–628.

Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T. et Irino, T. (2005). Nearly defect-free f0 trajectory extraction for expressive speech modifications based on straight. *In Interspeech 2005 - Annual Conference of the International Speech Communication Association, Lisbon, Portugal*, pages 537–540.

Kawahara, H. et Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana*, 36(5):713–727.

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. *In Proceedings of the SIGCHI Conference on Human Factors in*

Computing Systems, CHI '83, pages 193–196, New York, NY, USA. ACM.

Kelso, J. S., Vatikiotis-Bateson, E., Saltzman, E. L. et Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production : Phase portraits, kinematics, and dynamic modeling. *The Journal of the Acoustical Society of America*, 77:266.

Kerbrat-Orecchioni, C. (2005). Politeness in France : How to buy bread politely. *In* Hickey, L. et Stewart, M., éditeurs : *Politeness in Europe*, pages 29–44. Multilingual Matters, Clevedon, UK.

Klabbers, E. et van Santen, J. P. (2004). Clustering of foot-based pitch contours in expressive speech. *In Fifth ISCA Workshop on Speech Synthesis*, pages 73–78.

Kleiweg, P. (2010). *iL04 : An interface to RuG/L04, software for dialectometrics and cartography*. R package version 1.15.

Koptjevskaja-Tamm, M. et Ahlgren, I. (2003). NSM : Theoretical, methodological and applicational problems. *Theoretical linguistics*, 29(3):247–262.

Koster, W. J. W. (1936). *Traité de métrique grecque ; suivi d'un précis de métrique latine*. Sijthoff, Leyde.

Lai, C. (2014). Interpreting final rises : task and role factors. *In 7th International Conference on Speech Prosody, Dublin, Ireland*, pages 520–524.

Lai, J.-P. (2002). *L'intonation du parler de Nuoro (Sardaigne)*. Thèse de doctorat, Université Stendhal Grenoble 3.

Lai, J.-P. (2005). Projet AMPER. Atlas Multimédia Prosodique de l'Espace Roman. *In Géolinguistique, hors-série 3*. Éditions littéraires et linguistiques de l'Université de Grenoble.

Lamesch, S., Doval, B. et Castellengo, M. (2012). Toward a more informative voice range profile : The role of laryngeal vibratory mechanisms on vowels dynamic range. *Journal of Voice*, 26(5):672–e9.

Larkey, L. S. (1983). Reiterant speech : An acoustic and perceptual validation. *The Journal of the Acoustical Society of America*, 73(4):1337–1345.

Laukka, P., Audibert, N. et Aubergé, V. (2007). Graded structure in vocal expression of emotion : What is meant by "prototypical expressions". *In 1st International Workshop on Paralinguistic and Speech–Between Models and Data*.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge, UK.

Leipp, E., Castellengo, M. et Liénard, J.-S. (1968). La synthèse de la parole à partir de diagrammes phonétiques. *In 6th International Congress on Acoustic, Tokyo, Japan*.

Levitt, H. et Rabiner, L. R. (1971). Analysis of fundamental frequency contours in speech. *The Journal of the Acoustical Society of America*, 49:569.

Liénard, J.-S. et Barras, C. (2013). Fine-grain voice strength estimation from vowel spectral cues. *In Interspeech 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*, pages 128–132.

Liénard, J.-S., Signol, F. et Barras, C. (2007). Speech fundamental frequency estimation using the alternate comb. *In Interspeech 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgique*, pages 2773–2776.

Liénard, J.-S. et Teil, D. (1970). Les éléments phonétiques et la traduction automatique du message écrit en message parlé. *Automatisme, nr10, octobre*.

Lu, Y., Aubergé, V. et Rilliard, A. (2012). Do you hear my attitude ? Prosodic perception of social affects in Mandarin. *In 6th International Conference on Speech Prosody,*

Shangai, China, pages 685–688.

Lucci, V. (1983). *Étude phonétique du français contemporain à travers la variation situationnelle*. Publications de l'Université des Langues et Lettres, Grenoble, France.

Léon, P. (1993). *Précis de phonostylistique. Parole et expressivité*. Nathan Université, Paris.

Mac, D.-K. (2012). *Génération de parole expressive dans le cas des langues à tons*. Thèse de doctorat, Université de Grenoble.

Mac, D.-K., Aubergé, V., Castelli, E. et Rilliard, A. (2012). Local vs. global prosodic cues : effect of tones on attitudinal prosody in cross-perception of Vietnamese by French. *In 6th International Conference on Speech Prosody, Shangai, China*, pages 222–225.

Mac, D.-K., Aubergé, V., Rilliard, A. et Castelli, E. (2009). Audio-Visual prosody of social attitudes in Vietnamese : building and evaluating a tones balanced corpus. *In Interspeech 2009 - 10th Annual Conference of the International Speech Communication Association, Brighton, UK*, pages 2263–2266.

Mac, D.-K., Aubergé, V., Rilliard, A. et Castelli, E. (2010). Vietnamese multimodal social affects : How prosodic attitudes can be recognized and confused. *In Spoken Languages Technologies for Under-Resourced Languages*, pages 24–28.

Marcus, S. M. (1981). Acoustic determinants of perceptual center (p-center) location. *Perception & psychophysics*, 30(3):247–256.

del Mar Vanrell, M., Mascaró, I., Torres-Tamarit, F. et Prieto, P. (2012). Intonation as an encoder of speaker certainty : information and confirmation yes-no questions in Catalan. *Language and speech*, 56(2):163–190.

Martin, J.-C., d'Alessandro, C., Jacquemin, C., Katz, B., Max, A., Pointal, L. et Rilliard, A. (2007). 3D audiovisual rendering and real-time interactive control of expressivity in a Talking Head. *In International Conference on Intelligent Virtual Agents (IVA 2007)*, page 8p, Paris, France.

Martin, P. (1973). Les problèmes de l'intonation : recherches et applications. *Langue française*, 19(1):4–32.

Martin, P. (1982). Comparison of pitch detection by cepstrum and spectral comb analysis. *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'82)*, volume 7, pages 180–183. IEEE.

Martin, P. (1987). Prosodic and rhythmic structures in French. *Linguistics*, 25(5):925–950.

Martin, P. (2012). Automatic detection of voice creak. *6th International Conference on Speech Prosody, Shanghai, Chine*, pages 26–28.

Martin, P. (2014). Emotions and prosodic structure - Who is in charge ? *In* Baider, F. et Cislaru, G., éditeurs : *Linguistic Approaches to Emotions in Context*, pages 215–229. John Benjamins Publishing Company, Amsterdam.

Martins-Baltar, M. (1977). *De l'énoncé à l'énonciation : une approche des fonctions intonatives*. Didier, Paris.

Mertens, P. (2004). Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues*, 45(2):109–130.

Mertens, P. (2006). A predictive approach to the analysis of intonation in discourse in french. *In* Kawaguchi, Y., Fonagy, I. et Moriguchi, T., éditeurs : *Prosody and Syntax*, pages 64–101. John Benjamins, Amsterdam.

Mixdorff, H. (2000). A novel approach to the fully automatic extraction of fujisaki model

parameters. *In Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 3, pages 1281–1284. IEEE.

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H. et Charnvivit, P. (2002). Perception of tone and vowel quantity in thai. *In Interspeech 2002 - Annual Conference of the International Speech Communication Association, Denver, Colorado*.

de Moraes, J., Rilliard, A., Alberto, B. et Shochi, T. (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. *In 5th International Conference on Speech Prosody, Chicago, USA*, page 4p, Chicago, USA.

de Moraes, J. A. (2008). The pitch accents in Brazilian Portuguese : analysis by synthesis. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 389–397.

de Moraes, J. A. et Rilliard, A. (2014a). Illocution, Attitudes and Prosody. *In* Raso, T., éditeur : *Spoken Corpora and Linguistic Studies*. John Benjamins Publisher, Amsterdam.

de Moraes, J. A. et Rilliard, A. (2014b). Prosody and Emotion. *In* Armstrong, M., Henriks, N. et del Mar, M., éditeurs : *Interdisciplinary approaches to intonational grammar in Ibero-Romance*. John Benjamins Publisher, Amsterdam.

Morlec, Y., Bailly, G. et Aubergé, V. (2001). Generating prosodic attitudes in french : data, model and evaluation. *Speech Communication*, 33(4):357–371.

Moulines, E. et Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication*, 9(5):453–467.

de Castro Moutinho, L., Coimbra, R. L., Rilliard, A. et Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. *Estudios de fonética experimental*, 20:33–55.

de Castro Moutinho, L., Coimbra, R. L., Teixeira, A. et Pereira, M. (2005). Variação entoacional em três áreas dialectais de portugal continental. *Géolinguistique*, Hors série 3:19–37.

de Castro Moutinho, L., Vaz, A. M. et Coimbra, R. L. (2008). Variantes prosódicas do português europeu : O barlavento e o sotavento algarvio. *In* Turculeţ, A., éditeur : *La variation diatopique de l'intonation dans le domaine roumain et roman*, pages 91–104. Editura universităţii «Alexandru Ioan Cuza», Iaşi.

Mufwene, S. S. (2007). Population movements and contacts in language evolution. *Journal of language contact*, 1(1):63–92.

Murtagh, F. (1985). Multidimensional clustering algorithms. *Compstat Lectures, Vienna : Physika Verlag*, 1.

Nadeu, M. et Prieto, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of pragmatics*, 43(3):841–854.

Nakatani, L. H. et Schaffer, J. A. (1978). Hearing"words"without words : Prosodic cues for word perception. *The Journal of the Acoustical Society of America*, 63(1):234–245.

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88.

Nguyen, T. T. T., Rilliard, A., Tran, D. D. et d'Alessandro, C. (2014). Prosodic phrasing modeling for Vietnamese TTS using syntactic information. *In Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, Singapore.

Niebuhr, O. (2014). "A little more ironic" – Voice quality and segmental reduction differences between sarcastic and neutral utterances. *In 7th International Conference*

on Speech Prosody, Dublin, Ireland, pages 608–612.

Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh University Press, Edinburgh.

Ohala, J. J. (1983). Cross-language use of pitch : an ethological view. *Phonetica*, 40(1):1–18.

Ohala, J. J. (1984). An ethological perspective on common cross - language utilization of f0 of voice. *Phonetica*, 41(1):1–16.

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. *In* Hinton, L., Nichols, J. et Ohala, J. J., éditeurs : *Sound symbolism*, pages 325–347. Cambridge University Press, Cambridge, UK.

Öhman, S. et Lindqvist, J. (1965). Analysis-by-synthesis of prosodic pitch contours. *Quarterly Progress and Status Report (STL-QPSR)*, 6(4):1–6.

Osgood, C. E., May, W. H. et Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. University of Illinois Press.

Osgood, C. E., Suci, G. J. et Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana : University of Illinois Press.

Patel, A. D., Iversen, J. R. et Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music : The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5):3034–3047.

Peeters, B., éditeur (2006). *Semantic primes and universal grammar : Empirical evidence from the Romance languages*, volume 81. John Benjamins Publishing.

Pellegrino, F., Coupé, C. et Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3):539–558.

Pierrehumbert, J. B. (2001). Exemplar dynamics : Word frequency, lenition, and contrast. *In* Bybee, J. L. et Hopper, P., éditeurs : *Frequency effects and the emergence of lexical structure*, pages 137–157. Amsterdam : John Benjamins.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, 34:516–530.

Podesva, R. J. (2007). Phonation type as a stylistic variable : The use of falsetto in constructing a persona1. *Journal of sociolinguistics*, 11(4):478–504.

Pánini (1897). *Ashtádhyáyí*. Sindhu Charan Bose, Benares.

Quené, H. (2013). Longitudinal trends in speech tempo : The case of queen Beatrix. *The Journal of the Acoustical Society of America*, 133(6):EL452–EL457.

Quintilien (trad. 1829). *Institution oratoire*. C. L. F. Panckoucke, Paris.

Rabelais, F. (ca 1530). *Les Horribles et Espoventables Faictz et Prouesses du très renommé Pantagruel, roy des Dipsodes, filz du grand géant Gargantua, composez nouvellement par Maistre Alcofrybas Nasier*. C. Nourry, Lyon.

Ramsay, J. O. et Silverman, B. W. (2005). *Functional data analysis. second edition*. Springer, New York.

Ramus, F. et Mehler, J. (1999). Language identification with suprasegmental cues : A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, 105:512.

Rao, C. R. et Suryawanshi, S. (1996). Statistical analysis of shape of objects based on landmark data. *Proceedings of the National Academy of Sciences*, 93:12132–12136.

Rastier, F. (2001). *Arts et sciences du texte*. Presses Universitaires de France, Paris.

Rebollo Couto, L., Cunha, C., Pinto, M. d. S. et Santos, G. F. (2008). Marcas ento-

nacionais em dialetos de fronteira : o continuum entre o português brasileiro e o espanhol uruguaio em enunciados interrogativos totais. *In* Roncarati, C. et Abraçado, J., éditeurs : *Português Brasileiro II - contato lingüístico, heterogeneidade e história*, pages 20–38. Editora UFF, Niterói, Brésil.

Rebollo Couto, L., dos Santos Figueiredo, N., da Silva Pinto, M. et Sosa, J. M. (2010). Pragmática intercultural e entoaação : os enunciados interrogativos (perguntas) em português e em espanhol. *In Congresso Internacional de Professores de Línguas Oficiais do MERCOSUL (CIPLOM)*, pages 599–607.

Reichel, U. D., Markó, A. et Mády, K. (2014). Parametrization and automatic labeling of Hungarian intonation. *In 7th International Conference on Speech Prosody, Dublin, Ireland*, pages 738–742.

Rialland, A. (2007). Question prosody : an african perspective. *In* Riad, T. et Gussenhoven, C., éditeurs : *Tones and tunes : Typological studies in word and sentence prosody*, pages 35–62. Mouton de Gruyter, Berlin, Germany.

Riemer, N. (2006). Reductive paraphrase and meaning : A critique of wierzbickian semantics. *Linguistics and philosophy*, 29(3):347–379.

Rietveld, T. et Chen, A. (2006). How to obtain and process perceptual judgements of intonational meaning. *In* Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. et Schliesser, J., éditeurs : *Methods in empirical prosody research*, pages 283–319. Walter de Gruyter, Berlin, Germany.

Rietveld, T. et Van Hout, R. (2005). *Statistics in language research : Analysis of variance*. Walter de Gruyter.

Rilliard, A. (2000). *Vers une mesure de l'intelligibilité linguistique de la prosodie – Évaluation diagnostique des prosodies synthétique et naturelle*. Thèse de doctorat, Institut National Polytechnique de Grenoble.

Rilliard, A. (2011). La base de données AMPER. *Géolinguistique*, hors-série 4.

Rilliard, A., Allauzen, A. et Boula de Mareüil, P. (2011). Using dynamic time warping to compute prosodic similarity measures. *In Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, pages 2021–2024, Florence, Italy.

Rilliard, A. et Aubergé, V. (2003). Prosody evaluation as a diagnostic process : subjective vs. objective measurements. *International Journal of Speech Technology*, 6(4):409–418.

Rilliard, A., de Moraes, J. A., Erickson, D. et Shochi, T. (2014a). Social affect production and perception across languages and cultures – the role of prosody. *Leitura*, 52:16–41.

Rilliard, A., Erickson, D., de Moraes, J. A. et Shochi, T. (2014b). Cross-cultural perception of some Japanese politeness and impoliteness expressions. *In* Baider, F. et Cislaru, G., éditeurs : *Linguistic Approaches to Emotions in Context*. John Benjamins Publishing Company, Amsterdam.

Rilliard, A., Erickson, D., Shochi, T. et de Moraes, J. A. (2013). Social face to face communication – American English attitudinal prosody. *In Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)*, pages 1648–1652, Lyon, France.

Rilliard, A., Erickson, D., Shochi, T. et de Moraes, J. A. (2014c). US English attitudinal prosody performances in L1 and L2 speakers. *In 7th International Conference on Speech Prosody*, pages 895–899, Dublin, Ireland.

Rilliard, A., Martin, J.-C., Aubergé, V. et Shochi, T. (2008). Perception of French audio-visual prosodic attitudes. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 685–688, Campinas, Brazil.

Rilliard, A., Shochi, T., Erickson, D. et de Moraes, J. (2012). Developmental perception of polite & impolite non-verbal behaviours in Japanese. *In* Mello, H., Pettorino, M. et Raso, T., éditeurs : *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, pages 167–171. Firenze University Press.

Rilliard, A., Shochi, T., Martin, J.-C., Erikson, D. et Aubergé, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech*, 52(2-3):223–243.

Roekhaut, S., Goldman, J.-P. et Simon, A. C. (2010). A model for varying speaking style in TTS systems. *In 5th International Conference on Speech Prosody, Chicago, USA*.

Romano, A. (1999). *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento (Italie) : approche linguistique et instrumentale*. Thèse de doctorat, Université Stendhal Grenoble 3.

Romano, A. et Interlandi, G. M. (2005). Variabilità geo-socio-prosodica - dati linguistici e statistici. *Géolinguistique*, hors-série 3:259–280.

Romney, A. K., Boyd, J. P., Moore, C. C., Batchelder, W. H. et Brazill, T. J. (1996). Culture as shared cognitive representations. *Proceedings of the National Academy of Sciences*, 93(10):4699–4705.

Romney, A. K. et Moore, C. C. (1998). Toward a theory of culture as shared cognitive structures. *Ethos*, 26(3):314–337.

Romney, A. K., Moore, C. C., Batchelder, W. H. et Hsia, T.-L. (2000). Statistical methods for characterizing similarities and differences between semantic structures. *Proceedings of the National Academy of Sciences*, 97(1):518–523.

Romney, A. K., Moore, C. C. et Rusch, C. D. (1997). Cultural universals : Measuring the semantic structure of emotion terms in english and japanese. *Proceedings of the National Academy of Sciences*, 94(10):5489–5494.

Rossi, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica*, 23(1):1–33.

Rossi, M. (1978). Interactions of intensity glides and frequency glissandos. *Language and speech*, 21(4):384–396.

Rossi, M., Di Cristo, A., Hirst, D., Martin, P. et Nishinuma, Y. (1981). L'intonation : de l'acoustique à la sémantique. *Paris, Klincksieck*.

Roy, D. (2009). New horizons in the study of child language acquisition. *In 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom*, pages 13–20.

Rushton, J. P. et Irwing, P. (2011). The general factor of personality : Normal and abnormal. *In* Chamorro-Premuzic, T., von Stumm, S. et Furnham, A., éditeurs : *The Wiley–Blackwell handbook of individual differences*. Blackwell Publishing.

Russell, J. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Russell, J. et Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called *emotion* : Dissecting the elephant. *Journal of personality and social psychology*, 76(5):805.

Russell, J., Lewicka, M. et Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of personality and social psychology*, 57(5):848.

Sadanobu, T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, 8(1):29–44.

Sadanobu, T. (2012). An unofficial guide for Japanese characters. http ://dictionary.sanseido-publ.co.jp/wp/author/sadanobu-e/. Vu le 2014-06-16.

Sakoe, H. et Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1): 43–49.

van Santen, J., Shih, C. et Möbius, B. (1998). Intonation. *In* Sproat, R., éditeur : *Multilingual Text-to-Speech synthesis : the Bell Labs approach*, pages 141–189. Kluwer academic publishers, Dordrecht.

Savino, M. (2012). The intonation of polar questions in Italian : Where is the rise ? *Journal of the International Phonetic Association*, 42:23–48.

Scherer, K. R. (1981). Speech and emotional states. *In* Darby, J., éditeur : *Speech evaluation in psychiatry*, pages 189–220. Grune & Stratton, New York, USA.

Scherer, K. R. (1984a). Emotion as a multicomponent process : A model and some cross-cultural data. *In* Shaver, P., éditeur : *Review of Personality and Social Psychology*, volume 5, pages 37–63. CA : Sage, Beverly Hills, USA.

Scherer, K. R. (1984b). On the nature and function of emotion : A component process approach. *In* Scherer, K. R. et Ekman, P., éditeurs : *Approaches to emotion*, chapitre 14, pages 293–317. NJ : Erlbaum, Hillsdale.

Scherer, K. R. (1986). Vocal affect expression : A review and a model for future research. *Psychological Bulletin*, 99:143–165.

Scherer, K. R. (1989a). Vocal correlates of emotional arousal and affective disturbance. *In* Wagner, H. et Manstead, A., éditeurs : *Handbook of Psychophysiology : Emotion and social behavior*, chapitre 17, pages 165–197. Wiley, London, UK.

Scherer, K. R. (1989b). Vocal measurement of emotion. *In* Plutchik, R. et Kellerman, H., éditeurs : *Emotion : Theory, research, and experience.*, volume 4. The measurement of emotion, chapitre 9, pages 233–260. Academic Press, New York, USA.

Scherer, K. R. (1992). Vocal affect expression as symptom, symbol, and appeal. *In* Papousek, H., Jürgens, U. et Papousek, M., éditeurs : *Nonverbal vocal communication : Comparative and developmental approaches*, pages 43–60. Cambridge University Press, Cambridge and New York.

Scherer, K. R. (1999). On the sequential nature of appraisal processes : Indirect evidence from a recognition task. *Cognition and Emotion*, 13:763–793.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. *In* Scherer, K. R., Schorr, A. et Johnstone, T., éditeurs : *Appraisal processes in emotion : Theory, Methods, Research*, pages 92–120. Oxford University Press, New York and Oxford.

Scherer, K. R. (2003). Vocal communication of emotion : A review of research paradigms. *Speech Communication*, 40:227–256.

Scherer, K. R., Bänziger, T. et Grandjean, D. (2003). L'étude de l'expression vocale des émotions : Mise en évidence de la dynamique des processus affectifs. *In* Coletta, J.-M. et Tcherkassof, A., éditeurs : *Les émotions, cognition, langage et développement*, pages 39–58. Mardaga, Liège, Belgique.

Scherer, K. R., Mortillaro, M. et Mehu, M. (2013). Understanding the mechanisms underlying the production of facial expression of emotion : A componential perspective. *Emotion Review*, 5:47–53.

Scherer, K. R. et Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66:310–328.

Seara, I., Sosa, J. M. et Nunes, V. (2014). Sentence type and prenuclear contours in Brazilian Portuguese : production and perception. *In 7th International Conference on Speech Prosody, Dublin, Ireland*, pages 448–452.

Seinturier, J., Murisasco, E., Bruno, E. et Blache, P. (2012). An ontological approach to model and query multimodal concurrent linguistic annotations. *In International Conference on Language Resources and Evaluation, Istanbul, Turkey*, pages 2602–2605.

Shahid, S., Krahmer, E. et Swerts, M. (2008). Real vs. acted emotional speech : Comparing south-asian and caucasian speakers and observers. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 669–772.

Shochi, T. (2008). *Prosodie des affects socioculturels en japonais, et anglais : à la recherche des vrais et faux-amis pour le parcours de l'apprenant.* Thèse de doctorat, Université Stendhal, Grenoble III.

Shochi, T., Erickson, D., Rilliard, A., Aubergé, V. et Martin, J.-C. (2008). Recognition of Japanese attitudes in audio-visual speech. *In 4th International Conference on Speech Prosody, Campinas, Brésil*, pages 689–692, Campinas, Brazil.

Shochi, T., Erickson, D., Rilliard, A. et de Moraes, J. A. (2013). Prosodic differences between L1 and L2 performance in social face to face American English context. *In Workshop on Affective Social Speech Signals (WASSS 2013)*, page 5p, Grenoble, France.

Shochi, T., Erikson, D., Sekiyama, K., Rilliard, A. et Aubergé, V. (2009a). Comparison between Japanese children and adults perception of prosodic politeness expressions. *In Meeting of Acoustical Society of America (Acoustics 2009)*, Proceedings of Meetings on Acoustics, Vol 6 (1), page 7p, Portland, USA.

Shochi, T., Erikson, D., Sekiyama, K., Rilliard, A. et Aubergé, V. (2009b). Japanese children's acquisition of prosodic politeness expressions. *In Annual Conference of the International Speech Communication Association (INTERSPEECH 2009)*, pages 1743–1746, Brighton, UK.

Shochi, T., Kamiyama, T., Rilliard, A. et Aubergé, V. (2014). Effet d'apprentissage des expressions prosodiques et gestuelles de politesse en japonais chez des apprenants français. *In* Butel, J., éditeur : *Japon Pluriel, 9.* Philippe Picquier, Paris.

Shochi, T., Rilliard, A., Aubergé, V. et Erickson, D. (2009c). Intercultural perception of English, French and Japanese social affective prosody. *In* Hancil, S., éditeur : *The role of prosody in affective speech*, pages 189–220. Linguistic Insights, 97, Bern, Germany.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. et Hirschberg, J. (1992). Tobi : a standard for labeling english prosody. *In Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP'92), Banff, Alberta, Canada*, pages 867–870.

Sonntag, G. P. et Portele, T. (1998). PURR—a method for prosody evaluation and investigation. *Computer Speech & Language*, 12(4):437–451.

Spencer-Oatey, H. (1996). Reconsidering power and distance. *Journal of Pragmatics*,

26:1–24.

Steele, J. (1779). *Prosodia Rationalis : or, an essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols (the second edition amended and enlarged)*. J. Nichols, London.

Swerts, M. et Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1):81–94.

Tartter, V. C. et Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *The Journal of the Acoustical Society of America*, 96(4):2101–2107.

Tokuda, K., Nankaku, Y., Toda, T., Zen, H., Yamagishi, J. et Oura, K. (2013). Speech synthesis based on Hidden Markov Models. *Proceedings of the IEEE*, 101(5): 1234–1252.

Turculeţ, A., éditeur (2008). *La variation diatopique de l'intonation dans le domaine roumain et roman*. Editura universităţii «Alexandru Ioan Cuza», Iaşi.

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech*, 3(4):223–234.

Vaissière, J. (1983). Language-independent prosodic features. *In* Cutler, A. et Ladd, D. R., éditeurs : *Prosody : Models and Measurements*, pages 53–66. Springer Verlag, Berlin.

Vaissière, J. (1997). Langues, prosodies et syntaxe : Prosodie et syntaxe. *Traitement automatique des langues*, 38(1):53–82.

Vigouroux, C. B. (2008). "the smuggling of la francophonie" : Francophone africans in anglophone Cape Town (South Africa). *Language in Society*, 37(03):415–434.

Wawrzyniak, J. K. (2010). Native speakers, mother tongues and natural semantic metalanguages. *Language Sciences*, 32(6):648–670.

Welby, P. (2006). French intonational structure : Evidence from tonal alignment. *Journal of Phonetics*, 34(3):343–371.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, 49(1):28–48.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *In ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 143–148.

Wichmann, A. (2002). Attitudinal intonation and the inferential process. *In 1st International Conference on Speech Prosody, Aix-en-Provence, France*, pages 11–16.

Widen, S. C. et Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental psychology*, 39(1):114.

Wierzbicka, A. (1985). A semantic metalanguage for a crosscultural comparison of speech acts and speech genres. *Language in society*, 14(4):491–513.

Wierzbicka, A. (1986a). Human emotions : universal or culture-specific ? *American Anthropologist*, 88(3):584–594.

Wierzbicka, A. (1986b). A semantic metalanguage for the description and comparison of illocutionary meanings. *Journal of Pragmatics*, 10(1):67–107.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science*, 16(4):539–581.

Wierzbicka, A. (1996a). Japanese cultural scripts : Cultural psychology and "cultural grammar". *Ethos*, 24(3):527–555.

Wierzbicka, A. (1996b). *Semantics : Primes and universals*. Oxford University Press,

USA.

Wierzbicka, A. (1999). *Emotions across languages and cultures : Diversity and universals*. Cambridge University Press.

Wierzbicka, A. (2004). Emotion and culture : arguing with Martha Nussbaum. *Ethos*, 31(4):577–600.

Wierzbicka, A. (2005). Empirical universals of language as a basis for the study of other human universals and as a tool for exploring cross-cultural differences. *Ethos*, 33(2):265–291.

Wierzbicka, A. (2010). On emotions and on definitions : A response to Izard. *Emotion Review*, 2(4):379.

Wilting, J., Krahmer, E. et Swerts, M. (2006). Real vs. acted emotional speech. *In Interspeech 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA*, pages paper 1093–Tue1A3O.4.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, 38(3):329–336.

Yegnanarayana, B., Anand Joseph, M., Suryakanth, V. et Dhananjaya, N. (2011). Decomposition of speech signals for analysis of aperiodic components of excitation. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5396–5399. IEEE.

Yegnanarayana, B. et Dhananjaya, G. N. (2013). Spectro-temporal analysis of speech signals using zero-time windowing and group delay function. *Speech Communication*, 55(6):782–795.

Zinck, A. et Newen, A. (2008). Classifying emotion : a developmental account. *Synthese*, 161(1):1–25.

10 | Curriculum Vitæ

10.1 Parcours

État civil

Nom : Rilliard
Prénom : Albert
Date de naissance : 2 Janvier 1973
Statut : Chargé de Recherche (CR1) CNRS, section 34
Laboratoire : LIMSI-CNRS
Téléphone : (+33) 1 69 85 81 52
Email : albert.rilliard@limsi.fr
Page web : http ://groupeaa.limsi.fr/membres :rilliard
Adresse postale : LIMSI-CNRS – Rue John von Neumann
Campus Universitaire d'Orsay – Bât. 508
F-91405 Orsay cedex

Carrière

2007 : CR1 CNRS au LIMSI (UPR 3251), Orsay
2002 : CR2 CNRS à l'ICP (UMR 5009), Grenoble
2001 : ATER en Informatique à l'Université Stendhal (Grenoble III)
2000 : Service national

Titres Universitaires

2000 : Doctorat de Sciences Cognitives, Institut National Polytechnique de Grenoble. Mention très honorable avec félicitations du Jury. Jury : V. Aubergé, N. Campbell, A. Di Cristo, D. Hirst, H. Méloni, J.-L. Schwartz
2000 : DEA de Sciences du Langage, Université Stendhal (Grenoble III), Grenoble
1996 : DEA de Sciences Cognitives, Institut National Polytechnique de Grenoble
1995 : - Bachelor of Sciences (computing), University of Northumbria, Newcastle
- Maitrise d'Informatique, Université de Bourgogne, Dijon
1994 : Licence d'Informatique, Université de Bourgogne, Dijon
1993 : DEUG A Sciences et Structure des Matériaux, Université de Bourgogne, Dijon
1991 : Baccalauréat série C

10.2 Production scientifique

Logiciels et bases de données

- Création de la base de données et responsable des méthodologies d'extraction des données : `http://amper.limsi.fr/`
- Création et mise à disposition d'outils de stylisation prosodique et de mesure de distances prosodiques pour le projet AMPER : `http://groupeaa.limsi.fr/membres:rilliard:outils_amper`
- Corpus de contes écrits et lus (projet GV-LEx) en cours de dépôt à l'ELRA

Invitations

Conférences invitées dans des congrès

- « On the measurement of the perceptive distance of prosodic social affects through its acoustic correlates ». Fall meeting of the Technical Committee of Psychological and Physiological Acoustics, The Acoustical Society of Japan, octobre 2009, Wakayama, Japon.
- « Stratégies expressives et variations interculturelles – perception et mesure d'expressions d'(im)politesse », Journées Perception Sonore, Société Française d'Acoustique, 10-11 décembre 2012, Marseille, France.
- « Metodoloxía cuantitativa para a medida das distancias prosódicas », Xornadas de Dialectoloxía Perceptiva, Universidade de Santiago de Compostela, 17-18 janvier 2013.
- « Measures of distances between prosodies », Workshop on Cross Cultural research on Speech Communication & Second Language learning processing, LaBRI, Bordeaux, 15 mars 2013
- « Cross-Cultural Perception of Attitudes and Concepts », IV Colóquio Brasileiro De Prosódia Da Fala, Universidade Federal de Alagoas, Maceio, Brésil, 16-18 octobre 2013.

Séminaires invités

- « Les affects sociaux : multimodalité et interculturalité. Perception, analyse et contrôle - enjeux didactiques », Master *Industrie de la Langue*, Université Stendhal-Grenoble 3, 2008.
- « The expressive function of prosody : a multimodal and cross-cultural approach ». Séminaire donné à Kumamoto University, Japon, Octobre 2009.
- « Cross-Cultural Perception of Attitudes and their concepts », Séminaire donné à la Pontifícia Universidade Católica de São Paulo, 30 octobre 2013.
- « Comparaison interculturelle de la perception et de la production d'expressions attitudinales », Séminaire donné à l'Université du Québec à Montréal, 12 Décembre 2013.

Programmes d'échanges, collaborations, réseaux internationaux, projets nationaux et européens

- Membre du projet AMPER (Atlas Multimédia Prosodique de l'Espace Roman, réseau international d'étude des variations prosodiques au sein de l'espace linguistique roman, regroupant une trentaine d'universités)
- Collaborations (externes au LIMSI et ayant donné lieu à publication) avec des chercheurs des universités de Bordeaux III & Grenoble III en France, Showa Music University & Kumamoto University au Japon, Universidade Federal do Rio de Janeiro au Brésil, Universidade de Aveiro au Portugal, Università di Torino en Italie, Budapest University of Technology and Economics en Hongrie.
- Participation au network of excellence SIMILAR (FP6)

10.3 Liste complète des publications scientifiques

La liste complète de mes publications scientifiques parues ou sous presse est présentée ci-dessous. Une sélection des publications les plus significatives est proposée dans un document annexe (« Sélection de publications »), avec une présentation de l'apport particulier de chacune d'elle, au vu du présent document.

- articles dans des revues avec comité de lecture : 22
- chapitres de livres : 8
- articles dans des actes de congrès avec comité de lecture : 93
- nombre total de publications : 123
- conférences invitées ou séminaires : 9

10.3.1 Articles de revues & chapitres d'ouvrages

1. de Moraes, J. A. et Rilliard, A. (sous presse). Prosody and Emotion. In Armstrong, M., Henriks, N. et del Mar, M. (Eds.), *Interdisciplinary approaches to intonational grammar in Ibero-Romance*, Amsterdam : John Benjamins Publisher.
2. de Moraes, J. A. et Rilliard, A. (sous presse). Illocution, Attitudes and Prosody. In Raso, T., et al. (Eds.), *Spoken Corpora and Linguistic Studies*, Amsterdam : John Benjamins Publisher.
3. Boula de Mareüil, P., Rilliard, A., Lehka-Lemarchand, I., Mairano, P. et Lai, J.-P. (sous presse). Falling yes/no questions in Corsican French and Corsican : evidence for a prosodic transfer. In Delais-Roussarie, E., Avanzi, M., Herment, S. (Eds.), *Languages in contact*, Berlin : Springer Verlag.
4. d'Alessandro, C., Feugère, L., Le Beux, S., Perrotin, O. et Rilliard, A. (2014). Drawing melodies : Evaluation of chironomic singing synthesis. *The Journal of the Acoustical Society of America*, 135 (6) :3601–3612.
5. Rilliard, A., de Moraes, J.A., Erickson, D. et Shochi, T. (2014). Social affect production and perception across languages and cultures – the role of prosody. *Leitura*, 52 : 16-41.
6. Shochi, T., Kamiyama, T., Rilliard, A. et Aubergé, V. (2014). Effet d'apprentissage des expressions prosodiques et gestuelles de politesse en japonais chez des apprenants français. In J.M. Butel (Ed.) *Japon Pluriel*, 9, Paris : Philippe Picquier.

7. Rilliard, A., Erickson, D., de Moraes, J.A. et Shochi, T. (2014). Cross-Cultural Perception of some Japanese Expressions of Politeness and Impoliteness. In F. Baider et G. Cislariu (eds.) *Linguistic approaches to emotions in context*. Amsterdam : John Benjamins, 251-276.
8. Lu, Y., Aubergé, V., Rilliard, A. et Gu, W. (2013). 普通度音的感知研究, *Journal of the School of Chinese Language and Culture*, Nanjing Normal University, 3 : 169-174
9. Boula de Mareüil, P., Rilliard, A. et Allauzen, A. (2012). Variation diachronique dans la prosodie du style journalistique : le cas de l'accent initial. *Revue Française de Linguistique Appliquée*, 17(1) : 97-111.
10. Boula de Mareüil, P., Rilliard, A. et Allauzen, A. (2011). A diachronic study of initial stress and other prosodic features in the French news announcer style : corpus-based measurements and perceptual experiments. *Language and Speech*, 55(2) : 263–293.
11. d'Alessandro, C., Rilliard, A. et Le Beux, S. (2011). Chironomic stylization of intonation. *Journal of the Acoustical Society of America*, 129(3), 1594-1604
12. de Castro Moutinho, L., Coimbra, R.-L., Rilliard, A., Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental*, XX : 35-55.
13. Romano, A., Contini, M., Lai, J.-P. et Rilliard, A. (2011). Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. *Revista Internacional de Lingüística Iberoamericana*, IX(1) : 13-25.
14. Rilliard, A. (2011). La base de données AMPER. *Géolinguistique*, hors-série n° 4.
15. Audibert, N., Aubergé, V. et Rilliard, A. (2010). Discrimination perceptive d'expressions émotionnelles actées vs. spontanées. Variabilité interindividuelle et influence de l'intensité de l'émotion. *Technique et Science Informatiques*, numéro spécial sur les Agents Conversationnels Animés, 29(7) : 833-857.
16. Rilliard, A., Shochi, T., Martin J.C., Erickson D. et Aubergé, V. (2009). Multimodal Indices To Japanese And French Prosodically Expressed Social Affects. *Language and Speech* 52 (2&3) : 223-243.
17. Rilliard, A. (2009). On the measurement of the perceptive distance of prosodic social affectrs through its acoustic correlates. *Transaction on Technical Committee of Psychological and Physiological Acoustics*, The Acoustical Society of Japan, 39 (6) : 471-476.
18. Shochi, T., Rilliard, A., Aubergé, V. et Erickson, D. (2009). Intercultural Perception of English, French and Japanese Social Affective Prosody. in S. Hancil (Ed.), *The role of prosody in Affective Speech*, Linguistic Insights 97, Bern : Peter Lang AG, 31-59.
19. Shochi, T., Gagnié, G., Rilliard, A., Erickson, D. et Aubergé, V. (2009). 日本人仏語学習者によるフランス語社会的情動情報の知覚(Perception of prosodic French social affects for Japanese learners of French language). *Transaction on Technical Committee of Psychological and Physiological Acoustics*, The Acoustical Society of Japan, 39 : 555-560.
20. Shochi, T., Aubergé, V. et Rilliard, A. (2009). 「日本語母語話者は発話の一 部を聞いて態度を 知覚で きるか ？–Gating パラダイムによる実験を通して」(Peut-on percevoir les attitudes japonaises avant la fin de phrase ? - test perceptif selon le paradigme de gating). 『フランス日本語教育』 n° 4 : 65-75.

21. Mokhtari, A., Tabuchi, S., Shochi, T. et Rilliard, A. (2009). A Cross-Linguistic and Cultural Comparison of Perception for Japanese Communication - A Contrastive Study between Japanese and French - (in Japanese). *Enseignement du Japonais en France*, 4 : 86-94.
22. Tabuchi, S., Shochi, T., Mokhtari, A. et Rilliard, A. (2009) Nihongo shigenkaiwa niokeru hatuwataido no ninshiki : Nihongobogowasha to kankokujin shokyugakushusha wo taishoni. *Grammer and Speech* VI, Kurosio publisher, Osaka, Japan.
23. Erickson, D., Shochi, T., Menezes, C., Kawahara, H., Sakakibara, K. et Rilliard, A. (2009) Nichi eigobogowasha niyoru kanjouonseichikaku no hikaku - kihonshuhasu igai no onkyotekitokusho kara erareru jouhou ni chumokushite. *Grammar and Speech* VI, Kurosio publisher, Osaka, Japan.
24. Contini, M., Lai ; J.P., Rilliard, A., Romano, A., Coimbra, R.L. et de Castro Moutinho, L. (2008). La collaboration scientifique franco-portugaise : une constance dans l'orientation scientifique du Centre de Dialectologie de Grenoble. *Bollettino dell'Atlante Linguistico Italiano*, III Serie, 32 : 205-214.
25. Shochi, T., Aubergé, V. et Rilliard, A. (2007). Hatsuwataido no bunkatekitokusei to niseno tomodachi. In T. Sadanobu et M. Nakagawa (Eds.). *Onsei bunpo no taisho*, Tokyo : Kurosio publisher, 55-78.
26. Lai, J.P. et Rilliard, A. (2007). L'intonation du parler occitan de la Viadène et présentation de la base de données AMPER. In J. Dorta (Ed.) *La prosodia en el ámbito lingüístico románico*. Santa Cruz de Tenerife : La Página Ediciones, S.L., 73-100.
27. Shochi, T., Aubergé, V. et Rilliard, A. (2007). Nihongo inritsu taido no chikaku ni ataeru nihongo syutoku no eikyo. *Enseignement du Japonais en France*, 3, 47-56.
28. Aubergé, V., Audibert, N. et Rilliard, A. (2006). De E-Wiz à C-Clone : recueil, modélisation et synthèse d'expressions authentiques. *Revue d'Intelligence Artificielle*, 20(4-5) : 499-527.
29. Shochi, T., Aubergé, V. et Rilliard, A. (2005). Furansugo bogowasha wa donoyouni nihongo wo kikunoka – inritu taido ni chumokushite. *Japanese Language Education in Europe*, 9 : 105-110.
30. Rilliard, A. et Aubergé, V. (2003). Prosody evaluation as a diagnostic process : subjective *vs.* objective measurements. *International Journal of Speech Technology*, 6(4) :409–418.

10.3.2 Actes de conférences à comité de lecture

2014

1. Nguyen, T. T. T., Rilliard, A., Tran, D. D. et d'Alessandro, C. (2014). Prosodic phrasing modeling for Vietnamese TTS using syntactic information. In *Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, Singapore.
2. Do, C.-T., Evrard, M., Leman, A., d'Alessandro, C., Rilliard, A. and Crebouw, J.-L. (2014). Objective Evaluation of HMM-based Speech Synthesis System Using Kullback-Leibler Divergence. In *Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)*, Singapore.

3. Nguyen, T.T.T., Tran, D.D., Rilliard A., d'Alessandro, C., Pham, T.N.Y. (2014). Intonation issues in HMM-based speech synthesis for vietnamese. In *4th International Workshop On Spoken Language Technologies For Under-resourced Languages*, St. Petersburg, 98-104.
4. Hönemann, A., Mixdorff, H. et Rilliard, A. (2014). Social attitudes - recordings and evaluation of an audio-visual corpus in German. In *7th Forum Acusticum 2014*, Krakow.
5. Fourer, D., Guerry, M., Shochi1, T., Rouas, J.-L., Aucouturier, J.-J. et Rilliard, A. (2014). Analyse prosodique des affects sociaux dans l'interaction face à face en japonais. In *Journées d'Etude sur la Parole (JEP 2014)*, Le Mans, France.
6. Rilliard, A., Erickson, D., Shochi, T. et de Moraes, J.A. (2014). US English attitudinal prosody performances in L1 and L2 speakers. In *7th International Conference on Speech Prosody (SP 2014)*, Dublin, Ireland, 20/05 au 23/05, 895-899.
7. Lu, Y., Aubergé, V. et Rilliard, A. (2014). Prosodic profiles of social affects in mandarin chinese. In *7th International Conference on Speech Prosody (SP 2014)*, Dublin, Ireland, 20/05 au 23/05, 125-129.
8. Lu, Y., Aubergé, V., Audibert, N. et Rilliard, A. (2014). Audiovisual Perception of Expressions of Mandarin Chinese social affects by French L2 Learners. In *7th international conference on Speech Prosody (SP7 2014)*, Dublin, Ireland, 20/05 au 23/05, 2014, 169-173.

2013

9. Nguyen, T.T.T., d'Alessandro, C., Rilliard, A. et Tran, D.D. (2013). HMM-based TTS for Hanoi Vietnamese : issues in design and evaluation. In *14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)*, Lyon, France, 25/08 au 29/08, 2311-2315.
10. Rilliard, A., Erickson, D., Shochi, T. et de Moraes, J.A. (2013). Social face to face communication – American English attitudinal prosody. In *14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)*, Lyon, France, 1648-1652.
11. Rilliard, A., de Moraes, J.A., Shochi, T. et Erickson, D. (2013). Perception of emotions across sentence's mode in Brazilian Portuguese. In *Workshop on Affective Social Speech Signals (WASSS 2013)*, Grenoble, France, ISCA, 2013.
12. Lu, Y., Aubergé, V. et Rilliard, A. (2013). Cognitive distance of attitudes in Chinese and French, In *Workshop on Affective Social Speech Signals (WASSS 2013)*, Grenoble, France, ISCA, 2013
13. Shochi, T., Erickson, D., Rilliard, A. et de Moraes, J.A. (2013). Prosodic differences between L1 and L2 performance in social face to face American English context, In *Workshop on Affective Social Speech Signals (WASSS 2013)*, Grenoble, France.
14. Boula De Mareüil, P., Rilliard, A., Lehka-Lemarchand, I. et Ivent, F. (2013). Regional accents and languages in France : a contrastive prosodic analysis of Romance varieties, In *Workshop on Phonetics, Phonology and Languages in Contact (PPLC 2013)*, Paris, France.

2012

15. Doukhan, D., Rilliard, A., Rosset, S. et d'Alessandro, C. (2012). Modelling pause duration as a function of contextual length. In *Proceedings of Interspeech 2012*, Portland, Oregon.
16. Boula de Mareüil, P., Mairano, P., Rilliard, A. et Lai, J.-P. (2012). Corsican French questions : is there a prosodic transfer from Corsican to French and how to highlight it ?, In *6th International Conference on Speech Prosody (SP 2012)*, Shanghai, China, 22/05 au 25/05, 418-421.
17. Mac, D.K., Aubergé, V., Castelli, E. et Rilliard, A. (2012). Local vs. global prosodic cues : effect of tones on attitudinal prosody in cross-perception of Vietnamese by French, In *6th International Conference on Speech Prosody (SP 2012)*, Shanghai, China, 22/05 au 25/05, 222-225
18. Lu, Y., Aubergé, V. et Rilliard, A. (2012). Do you hear my attitude ? Prosodic perception of social affects in Mandarin, In *6th International Conference on Speech Prosody (SP 2012)*, Shanghai, China, 22/05 au 25/05, 685-688
19. Rilliard, A., de Moraes, J.A., Erickson, D. et Shochi, T. (2012). Prosodic analysis of Brazilian Portuguese attitudes, In *6th International Conference on Speech Prosody (SP 2012)*, Shanghai, China, 22/05 au 25/05, 677-680
20. Boula de Mareüil, P., Rilliard, A., Mairano, P. et Lai, J.-P. (2012). Questions corses : peut-on mettre en évidence un transfert prosodique du corse vers le français ?, In *29e Journées d'Etude sur la Parole (JEP 2012)*, Grenoble, France, 04/06 au 08/06, 609-616.
21. Lu, Y., Aubergé, V. et Rilliard, A. (2012). Entends-tu mes attitudes ? Perception de la prosodie des affects sociaux en chinois Mandarin. In *29e Journées d'Etude sur la Parole (JEP 2012)*, Grenoble, France, 04/06 au 08/06, 25-32.
22. Doukhan, D., Rosset, S., Rilliard, A., d'Alessandro, C. et Adda-Decker, M. (2012). Designing French Tale Corpora for Entertaining Text To Speech Synthesis. In *Proceedings of LREC*, Istanbul, Turkey.
23. Lu, Y., Aubergé, V., Rilliard, A. et Gu, W. (2012). Prosodic cross-linguistic perception of social affects in Mandarin Chinese by native, French and Vietnamese listeners, In H. Mello, M. Pettorino, T. Raso (Eds.) *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, Firenze University Press, 141-145.
24. Mac, D.K., Aubergé, V., Rilliard, A. et Castelli, E. (2012). Can the tones influence the acoustic perception of the Vietnamese attitudes by French listeners ? In H. Mello, M. Pettorino, T. Raso (Eds.) *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, Firenze University Press, 146-150.
25. de Moraes, J.A., Rilliard, A., Erickson, D. et Shochi, T. (2012). Acoustic analysis of a corpus of Brazilian Portuguese attitudes, In H. Mello, M. Pettorino, T. Raso (Eds.) *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, Firenze University Press, 162-166.
26. de Moraes, J.A., Miranda, L. et Rilliard, A. (2012). Facial gestures in the expression of prosodic attitudes of Brazilian Portuguese, In H. Mello, M. Pettorino, T. Raso (Eds.) *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, Firenze University Press, 157-161.

27. Rilliard, A., Shochi, T., Erickson, D. et de Moraes, J.A. (2012). Developmental perception of polite et impolite non-verbal behaviours in Japanese, In H. Mello, M. Pettorino, T. Raso, *Proceedings of the VIIth GSCP International Conference : Speech and Corpora*, Firenze University Press, 167-171.
28. Lu, Y., Aubergé, A. et Rilliard, A. (2012). Tonal Influences on the prosodic Cross-linguistic Perception of Mandarin Social Affects by French and Vietnamese listeners, In *the Third International Symposium of Tonal Aspect of Language Proc.*, Nanjing, China.

2011

29. Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M. et d'Alessandro, C. (2011). Prosodic analysis of a corpus of tales. In *12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, Firenze, Italy, 27/08 au 31/08, 3129-3132.
30. Rilliard, A. Allauzen, P. Boula de Mareüil (2011). Using dynamic time warping to compute prosodic similarity measures. In *12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, Florence, Italy, 27/08 au 31/08, 2021-2024.
31. de Moraes, J.A., Rilliard, A. Erickson, D. et Shochi, T. (2011). Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China.

2010

32. d'Alessandro, C., Quok, A., Deroo, O., Doukhan, D., Gelin, R., Martin, J.C., Pelachaud, C., Rilliard, A. et Rosset, S. (2010). Towards a storytelling humanoid robot, In *AAAI 2010 Fall Symposium on Dialog With Robots*, Arlington, USA.
33. Le Beux, S., d'Alessandro, C. et Rilliard, A. (2010). Calliphony : a tool for real-time gestural modification and analysis of intonation and rythm, In *International Conference on Speech Prosody (SP 2010)*, Chicago, USA, May 11-14, 2010, 4p.
34. d'Alessandro, C., Le Beux, S. et Rilliard, A. (2010). Contrôle gestuel du modèle source / filtre de production de la voix, In *10ème Congrès Français d'Acoustique (CFA 2010)*, Lyon, France, 12-16 avril 2010, 4p.
35. Mac, D.K., Aubergé, V., Rilliard, A. et Castelli, E. (2010). Cross-cultural perception of Vietnamese audio-visual prosodic attitudes, In *International Conference on Speech Prosody (SP 2010)*, Chicago, USA, May 11-14, 4p.
36. Mac, D.K., Aubergé, V., Rilliard, A. et Castelli, E. (2010). Perception interculturelle des attitudes audio-visuelles vietnamiennes, In *28e Journées d'Etude sur la Parole (JEP 2010)*, Mons, Belgique, 25-28 mai, 4p.
37. de Moraes, J.A., Rilliard, A., Alberto, B. et Shochi, T. (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese, In *International Conference on Speech Prosody (SP 2010)*, Chicago, USA, May 11-14, 4p.

38. Shochi, T., Gagnié, G., Rilliard, A., Erickson, D. et Aubergé, V. (2010). Learning effect of French prosodic social affects for Japanese learners of French language, In *International Conference on Speech Prosody (SP 2010)*, Chicago, USA, May 11-14, 4p.
39. Audibert, N., Aubergé, V. et Rilliard, A. (2010). Prosodic correlates for the discrimination of acted vs spontaneous expressive speech : a pilot study, In *International Conference on Speech Prosody (SP 2010)*, Chicago, USA, May 11-14, 4p.

2009

40. Clavel, C., Rilliard, A., Martin, J.C. et Shochi, T. (2009). Personality Differences in the Multimodal Perception and Expression of Cultural Attitudes and Emotions. In *1st IEEE International Workshop on Social Signal Processing*, Amsterdam, September 13th.
41. Mac, D.K., Aubergé, V., Rilliard, A. et Castelli, E. (2009). Audio-Visual prosody of social attitudes in Vietnamese : building and evaluating a tones balanced corpus. In *Proceedings of Interspeech 2009*, 6-10 September 2009, Brighton, UK.
42. Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A. et Aubergé, V. (2009). Japanese children's acquisition of prosodic politeness expressions. In *Proceedings of Interspeech 2009*, 6-10 September 2009, Brighton, UK.
43. Boula de Mareüil, P. Rilliard, A. et Allauzen, A. (2009). Perception of the evolution of prosody in the French broadcast news style. In *Proceedings of Interspeech 2009*, 6-10 September 2009, Brighton, UK.
44. Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A. et Aubergé, V. (2009). Comparison between Japanese children and adults perception of prosodic politeness expressions. In *Proceedings of Meetings on Acoustics, ASA*, Vol 6(1), 062001, url : http ://link.aip.org/link/ ?PMA/6/062001/1 , DOI :10.1121/1.3171002

2008

45. Rilliard, A. et Lai, J.-P. (2008). Outils pour le calcul et la comparaison prosodique dans le cadre du projet AMPER - l'exemple des variétés Occitane et Sarde. In *Actes du symposium international sur la Variation diatopique de l'intonation dans le domaine roumain et roman.* Iasi, Octobre 2008.
46. Erickson, D., Chun-Fang, H., Shochi, T., Rilliard, A., Dang, J., Iwata, R., Lu, X. (2008). Acoustic and articulatory cues for Taiwanese, Japanese and American listeners' perception of Chinese happy and sad speech. In *Proceeding of 2008 Autumn meeting*, The Acoustic Society of Japan.
47. Erickson, D., Shochi, T., Kawahara, H., Rilliard, A. et Menezes, C. (2008). Formant lowering in spontaneous crying speech. In *Proceedings of the 156th Meeting of the Acoustical Society of America*, Miami, Florida, 10–14 November 2008.
48. Erickson, D., Rilliard, A., Shochi, T., Han, F., Kawahara, H., Sakakibara, K.I. (2008). A cross-linguistic comparison of perception to formant frequency cues in emotional speech. In *Proceedings of Oriental COCOSDA 2008*, Kyoto, Japan, Nov. 25-27 2008.

49. Shochi, T., Rilliard, A., Erickson, D., Martin, J.C. et Aubergé, V. (2008). Perception interculturelle des affects sociaux japonais et français. In *Actes du 3ème Workshop sur les Agents Conversationnels Animés*, Paris, France, 28 novembre 2008.
50. Audibert, N., Aubergé, V. et Rilliard, A. (2008). Emotions actées vs. spontanées : variabilité des compétences perceptives. In *Journées d'Etude de la Parole*. Avignon, France.
51. Boula de Mareüil, P., Rilliard A. et Allauzen, A. (2008). Étude diachronique de l'accent initial au travers d'archives audio. In *Journées d'Etude de la Parole*, Avignon, France.
52. Audibert, N., Aubergé, V. et Rilliard, A. (2008). How we are not all competent the same for discriminating acted from spontaneous expressive speech. In *Speech Prosody 2008*, Campinas, Brésil, 693-696.
53. Boula de Mareüil, P., Rilliard A. et Allauzen, A. (2008). A diachronic study of prosody through French audio archives. In *Speech Prosody 2008*, Campinas, Brésil, 531-534.
54. Rilliard, A., Martin, J.C., Aubergé, V. et Shochi, T. (2008). Perception of French Audio-Visual Prosodic Attitudes. In *Speech Prosody 2008*, Campinas, Brésil, 685-688.
55. Shochi, T., Erickson, D., Rilliard, A., Aubergé, V. et Martin, J.C. (2008). Recognition of Japanese attitudes in Audio-Visual speech. In *Speech Prosody 2008*, Campinas, Brésil, 689-692.
56. Fék, M., Audibert, N., Szabó, J., Rilliard, A., Németh, G. et Aubergé, V. (2008). Multimodal Spontaneous Expressive Speech Corpus for Hungarian. In *International Conference on Language Resources and Evaluation*, Marrakech, Maroc.
57. Rilliard, A. et Lai, J.-P. (2008). La Base de Données AMPER et ses interfaces : structure et formats de données, exemple d'utilisation pour une analyse comparative de la prosodie de différents parlers romans. In L. de Castro Moutinho et R.L. Coimbra, *Actas I Jornadas Cientificas AMPER-POR*, Universidade de Aveiro, Portugal, 29-30 Outubro 2007, 127-139.

2007

58. d'Alessandro, C., Rilliard, A. et Le Beux, S. (2007). Computerized chironomy : evaluation of hand-controlled Intonation reiteration. In *Proceedings of Interspeech 2007*, Antwerpen, Belgium, 1270-1273
59. Audibert, N., Aubergé, V. et Rilliard, A. (2007). When is the emotional information ? A gating expreriment for gradient and contours cues. In *Proceedings of ICPhS 2007*, Saarbrucken, Germany, 2137-2140.
60. Le Beux, S., Rilliard, A. et d'Alessandro, C. (2007). Calliphony : A real-time intonation controller for expressive speech synthesis. In *Proceedings of the 6th ISCA Speech Synthesis Research Workshop*, Bonn, Germany.
61. Martin, J.-C., d'Alessandro, C., Katz, B., Jacquemin, C., Max, A., Pointal, L. et Rilliard, A. (2007). 3D Audiovisual Rendering and Real-Time Interactive Control of Expressivity in a Talking Head. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents*, Paris, France.

62. Shochi, T., Aubergé, V. et Rilliard, A. (2007). Cross-Listening of Japanese, English and French social affect : about universals, false friends and unknown attitudes. In *Proceedings of ICPhS 2007*, Saarbrucken, Germany, 2097-2100.

2006

63. Aubergé, V. et Rilliard, A. (2006). More than pointing with the prosodic focus : The Valence-Intensity-Domain (VID) model. In *Proceedings of the Speech Prosody 2006 Conference*, Dresden, Germany.
64. Shochi, T., Aubergé, V. et Rilliard, A. (2006). How prosodic attitudes can be false friends : Japanese vs. French social affects. In *Proceedings of the Speech Prosody 2006 Conference*, Dresden, Germany.
65. Aubergé V., Rilliard A. et Audibert N. (2006). Auto-annotation : an alternative method to label expressive corpora. In *Proceedings of the Workshop « Corpora for research on emotion and affect »*, Gènes, Italie
66. Aubergé V. et Rilliard A. (2006). Le focus prosodique n'est pas que déictique : Le modèle VID (Valence-Intensité-Domaine). In *Actes des Journées d'Étude sur la Parole*, Dinard, France.
67. Audibert, N., Vincent, D., Aubergé, V., Rilliard, A. et Rosec, O. (2006). Dimensions acoustiques de la parole expressive : Poids relatifs des paramètres resynthétisés par Praat vs. LF-ARX. In *Actes des Journées d'Étude sur la Parole*, Dinard, France.
68. Shochi, T., Aubergé, V. et Rilliard, A. (2006). Comment les attitudes prosodiques sont parfois de « faux-amis » : Les affects sociaux du japonais vs. français. In *Actes des Journées d'Étude sur la Parole*, Dinard, France.

2005

69. Aubergé, V. et Rilliard, A. (2005). The focus prosody : more than a simple binary function. In *Proceedings of Interspeech 2005*, Lisbon, Portugal, 1373-1376.
70. Aubergé, V., Rilliard, A. et Audibert, N. (2005). De E-Wiz à E-Clone : méthodologie expérimentale pour la modélisation des émotions et affects authentiques. In *Actes du premier Workshop francophone sur les Agents Conversationnels Animés*, Grenoble, France, 125-134.
71. Audibert, N., Rilliard, A. et Aubergé, V. (2005). La plateforme E-Wiz (Expressive-Wizard of Oz) : capture d'expressions authentiques en Interaction Homme-Machine. In *Actes du premier Workshop francophone sur les Agents Conversationnels Animés*, Grenoble, France, 161-164.
72. Audibert, N., Aubergé, V. et Rilliard, A. (2005). The Prosodic Dimensions of Emotion in Speech : the Relative Weights of Parameters. In *Proceedings of Interspeech 2005*, Lisbon, Portugal, 525-528.
73. Audibert, N., Aubergé, V. et Rilliard, A. (2005). The Relative Weights of Prosodic Parameters for the Expression of Emotion in Speech : a Resynthesis Study. In *Proceedings of the first international conference on Affective Computing et Intelligent Interaction*, Beijing, China, 527-534.

74. Shochi, T., Aubergé, V. et Rilliard, A. (2005). Because attitudes are social affects, they can be false friends... In *Proceedings of the first international conference on Affective Computing et Intelligent Interaction*, Beijing, China, 482-489.

2004

75. Aubergé V., Audibert N., Rilliard A (2004). Acoustic Morphology of Expressive Speech : What about Contours ? In *Speech Prosody*, Nara, Japon. Mars 2004.
76. Aubergé V., Audibert N., Rilliard A (2004). E-Wiz : A Trapper Protocol for Hunting the Expressive Speech Corpora. In *Lab. 4th International Conference on Language Resources and Evaluation*, Lisbonne, Portugal. Mai 2004.
77. Audibert N., Aubergé V., Rilliard A (2004). EWiz : contrôle d'émotions authentiques. In *XXVe Journées d'Etude de la Parole*, Fès, Maroc. Avril 2004.
78. Rilliard A., Aubergé V., Audibert N. (2004). Evaluating an Authentic Audio-Visual Expressive Speech Corpus. In *4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisbonne, Portugal, 175-178.

2003

79. Aubergé, V., Audibert, N. et Rilliard, A. (2003). Why and how to control the authentic emotional speech corpora ? In *Eurospeech 2003*, 185-188. Genève.
80. Rilliard, A. (2003). Objective vs. Subjective Evaluation of Synthetic Prosody. In *15th International Congress of Phonetic Sciences*, 2953-2956. Barcelone.

2002

81. Rilliard, A. et Aubergé, V., (2002). Towards a Linguistic Validation of a Prosodic Generation Model. In *Speech Prosody*, 607-610, Aix-en-Provence.

2001

82. Rilliard, A. et Aubergé, V. (2001). Mesure de l'intelligibilité de la démarcation prosodique. In *Actes des Journées Prosodie*, Grenoble, France.
83. Rilliard, A. et Aubergé, V. (2001). Prosody evaluation as a diagnostic process : subjective vs. objective measurements, In *4th ISCA Workshop on Speech Synthesis*, Atholl, Scotland.

2000

84. Rilliard, A. et Aubergé, V. (2000). Perception and Analysis of a Reiterant Speech Paradigm : a Functional Diagnostic of Synthetic Prosody, In *Actes de la 2nde International Conference on Linguistic Ressources and Evaluation*, Athènes, Grèce, 661-664.
85. Aubergé, V et Rilliard, A. (2000). Prosody evaluation : quality measurement or diagnostic ?, In *Workshop COST 258*, Stockholm, février.

1999

86. Rilliard, A. et Aubergé, V. (1999). Prosody Diagnostic using Reiterant Speech. In *Proceedings of the 14th International Congress on Phonetic Sciences*, Berkeley, USA, 37-40.

1998

87. Rilliard, A. et Aubergé, V. (1998). Mesure perceptive de la fonction linguistique de la prosodie pure à l'aide d'énoncés réitérés, In *Annales des XXII Journées d'Études de la Parole*, 123-126, Martigny, Suisse.
88. Morlec Y., Rilliard A., Bailly G. et Aubergé V. (1998). Evaluating the Adequacy of Synthetic Prosody in Signalling Syntactic Boundaries : Methodology and First Results. In *Actes de la 1ère International Conference on Linguistic Ressources and Evaluation*, Grenade, Espagne, 647-650.
89. Rilliard, A. et Aubergé, V. (1998). A perceptive measure of pure prosody linguistic functions with reiterant sentences. In *Actes de la 5ème International Conference on Spoken Language Processing*, Sydney, Australie, 675-678.
90. Rilliard, A. et Aubergé, V. (1998). Reiterant Speech for the Evaluation of Natural vs. Synthetic Prosody. In *Actes du 3ème Workshop on Speech Synthesis*, Jenolan Caves, Australie, 87-91.
91. Hirst, D. J., Rilliard, A. et Aubergé, V. (1998). Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. In *Actes du 3ème Workshop on Speech Synthesis*, Jenolan Caves, Jenolan Caves, Australie, 1-4.

1997

92. Rilliard, A., Aubergé, V., Bailly, G. et Morlec, Y. (1997). Vers une Mesure de l'Information Linguistique Véhiculée par la Prosodie. In *Journées FRANCIL'97*, Avignon, 481-487.
93. Aubergé, V., Grépillat , T. et Rilliard, A. (1997). Can we Perceive Attitudes before the End of Sentences ? The Gating Paradigm for Prosodic Contours. In *EuroSpeech'97*, Rhodes, Grèce, 871 877.

10.4 Enseignement

Direction de thèses et de stages

- Stages de Master : 9
- Thèses : 3
- Stages post-doctoral : 3

Jurys, commissions

- Participation à 4 jurys de thèses
- Présidence d'un jury de recrutement d'un IE CNRS en informatique

Cours

- « Prosody & Emotion » Cour magistral, Summer School, Universidade Federal de Alagoas, Brésil, 2013.
- « Production et perception de la parole » Cour magistral, Licence 2 MIASS, Université Grenoble 2 de 2002 à 2006.
- « Corpus oraux » Cours magistral, Master 1 de Sciences du Langage, Université Grenoble 3 de 2004 à 2006
- « Initiation à l'Informatique » Travaux dirigés, Licence 1 de Sciences du langage, Université Grenoble 3 en 2004

10.5 Valorisation

Contrats

Participation à des projets de recherche financés :

- Projet du Pôle Rhône-Alpes de Sciences Cognitives « émotion » (2002-2003)
- Projet « Expressive Speech Project » – CREST Japan Science Technology (2003-2005)
- PPF Pegasus « Communication Expressive » (2003-2006)
- Contrat de Recherche Externalisée avec France Telecom R&D (2005-2006)
- ATIP CNRS « La prosodie des affects : des attitudes aux émotions » (2006-2008)
- Projet ANR appel CONTINT « GV-LEx » (2009-2012)
- Projet ANR blanc appel JCJC « PADE » (2010-2014)
- Projet FUI « ADN T-R » (2012-2014)
- Projet ANR appel CONTINT « CHANTER » (2014-2016)

Administration liée à la recherche

- Responsable de la Base de Données du projet AMPER
- Responsabilité scientifique de projets de recherche :
  - ◇ ATIP CNRS jeunes chercheurs « La prosodie des affects : des attitudes aux émotions » 2006-2008
  - ◇ ANR blanc JCJC « PADE » 2011-2014

10.6 Organisation de la recherche

Participation à des comités, Editorial boards, organisation de colloques...

- Membre de l'Editorial Board du Journal of Speech Sciences.

- Membre du comité d'organisation du WACA'05, Grenoble.
- Membre du comité d'organisation d'eNTERFACE'08, LIMSI, Orsay.
- Membre du comité de programme du WASSS'13, Grenoble.

Sociétés savantes

- Membre de l'Association Francophone de la Communication Parlée (AFCP)
- Membre de l'International Speech Communication association (ISCA)

Activités liées à l'administration

- Responsable du thème « Prosodie Expressive » du groupe Audio & Acoustique au LIMSI
- Membre élu au CA de l'Association Francophone de la Communication Parlée (2006-2010)
- Membre élu de la CCSU de l'Université Paris Sud (en informatique – depuis 2010)