



Prosodie des affects socioculturels en japonais, et anglais: à la recherche des vrais et faux-amis pour le parcours de l'apprenant

Takaaki Shochi

► To cite this version:

Takaaki Shochi. Prosodie des affects socioculturels en japonais, et anglais: à la recherche des vrais et faux-amis pour le parcours de l'apprenant. Sciences de l'Homme et Société. Université Stendhal - Grenoble III, 2008. Français. NNT: . tel-00366612

HAL Id: tel-00366612

<https://theses.hal.science/tel-00366612>

Submitted on 9 Mar 2009

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

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

UNIVERSITÉ GRENOBLE III – STENDHAL

UFR des Sciences du Langage

Numéro attribué par la bibliothèque

/ _ / _ / _ / _ / _ / _ / _ / _ / _ /

THÈSE

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE III

(Discipline : Science du Langage)

Présentée et soutenue publiquement le 21 avril 2008 par

Takaaki SHOCHI

Prosodie des affects socioculturels en japonais, français
et anglais : à la recherche des vrais et faux-amis
pour le parcours de l'apprenant

Sous la direction de Véronique AUBERGÉ

Membres du jury :

Daniel HIRST	Directeur de Recherche	Rapporteur
Toshiyuki SADANOBU	Professeur	Rapporteur
Christian BOITET	Professeur	Examineur
Jean-Marc COLLETTA	Maître de Conférences	Examineur
Donna ERICKSON	Professeur	Examineur
Tomoko HIGASHI	Maître de Conférences	Examineur
Véronique AUBERGÉ	Chargé de Recherche	Directeur de thèse
Albert RILLIARD	Chargé de Recherche	Co-directeur

Remerciements

Tout d'abord, je voudrais remercier Professeur Véronique Aubergé qui m'a encadré durant toutes ces années de thèse. J'admire toujours sa qualité scientifique et sa personnalité. Elle restera toujours ma seconde mère paternelle.

Je voudrais remercier également Dr. Albert Rilliard qui m'a encadré comme co-encadrant depuis en DEA. J'apprécie particulièrement sa patience pour comprendre ma logique japonaise parfois difficile à saisir. Sans lui, je ne pouvais pas réaliser cette thèse. Il restera toujours mon second père maternel.

Je remercie Professeur Toshiyuki Sadanobu pour ses conseils et ses soutiens tout au long de ma thèse.

Je remercie également Professeur Donna Erickson pour ses conseils et son aide pour les recrutements des sujets américains au Etats-Unis.

Je remercie Dr. Boris Doval pour ses conseils sur l'évaluation de la qualité de voix pour chacune des attitudes japonaises.

Je remercie Pro. Tomoko Higashi, Pro. Yasuko Takeuchi et Pro. Kazuko Ushiyama pour leur contribution pour les recrutements des apprenants japonais aux tests de perception.

Je remercie Professeur Noriko Kobayashi pour nous avoir donné la permission d'utiliser le test d'aptitude de la compétence de la langue japonaise.

Je remercie tous les participants à tous mes tests perceptifs réalisés dans les trois pays (Japon, France et Etats-Unis).

Un grand merci à mes collègues, Nicolas Audibert, Anne Vanpé, Fanny Loyau, Mathilde Fort pour la correction de cette thèse.

Et, en dernier lieu à ma famille qui m'a donné tant de courage même au moment difficile pour terminer cette thèse.

Table des matières

1. AVANT PROPOS : L'EXPRESSIVE SPEECH PROJECT.....	25
1.1. PRÉSENTATION GÉNÉRALE	25
1.2. LA PLACE DE NOS TRAVAUX AU SEIN DU PROJET ESP	25
2. INTRODUCTION	27
2.1. COMMUNIQUER.....	27
2.2. L'ENSEIGNEMENT DE LA COMMUNICATION VERBALE	28
3. ANALYSE DE L'ÉTAT DE L'ART	30
3.1. LES ÉMOTIONS ET LES ATTITUDES	30
3.1.1. <i>Les émotions</i>	30
3.1.1.1. La perspective évolutionniste.....	30
3.1.1.2. L'approche physiologique des émotions	31
3.1.1.3. L'approche cognitive des émotions.....	31
3.1.1.4. La perspective du constructivisme social.	31
3.1.2. <i>Les attitudes</i>	32
3.1.2.1. La perspective unidimensionnelle	32
3.1.2.2. La théorie multidimensionnelle	33
3.1.2.3. L'approche linguistique	34
3.1.3. <i>Définition des attitudes et des émotions</i>	35
3.1.4. <i>Problème des étiquettes données aux attitudes</i>	35
3.2. LA PROSODIE.....	36
3.2.1. <i>Tentative de définition de la prosodie</i>	36
3.2.2. <i>La Prosodie et les attitudes</i>	38
3.2.3. <i>L'approche didactique</i>	38
3.2.4. <i>L'approche « biologique »</i>	40
3.2.5. <i>La prosodie dans le système de communication parlée</i>	40
3.3. UN MODÈLE DE SYSTÈME DE COMMUNICATION	41
3.3.1. <i>Les trois niveaux prosodiques du système de communication parlée</i>	41
3.3.1.1. Les fonctions linguistiques	42
3.3.1.2. Les fonctions émotionnelles.....	42
3.3.1.3. Les fonctions attitudinales	42
4. LES ATTITUDES FRANÇAISES ET ANGLAISES À L'I.C.P.	45
4.1. ÉTUDES DES ATTITUDES PROSODIQUES DU FRANÇAIS	45
4.2. VALIDATION DES ATTITUDES FRANÇAISES	46
4.2.1. <i>Objectif du test perceptif</i>	47

4.2.2. <i>Choix des attitudes du français</i>	47
4.2.3. <i>Définition des attitudes</i>	47
4.2.4. <i>Composition du Corpus</i>	48
4.2.5. <i>Enregistrement</i>	49
4.2.6. <i>Validation perceptive</i>	49
4.2.6.1. <i>Méthodologie expérimentale</i>	49
4.2.7. <i>Résultats du test perceptif</i>	50
4.2.7.1. <i>Distribution des réponses par rapport au hasard</i>	50
4.2.7.2. <i>Effet de la longueur des énoncés</i>	51
4.2.7.3. <i>Taux d'identification des attitudes</i>	51
4.2.7.4. <i>Note sur l'analyse statistique des données perceptives</i>	53
4.2.7.5. <i>Analyse des résultats grâce à l'A.C.</i>	54
4.2.7.6. <i>Classification hiérarchique</i>	56
4.2.8. <i>Conclusion</i> :.....	57
4.3. VALIDATION DES ATTITUDES ANGLAISES	58
4.3.1. <i>Objectif du test perceptif</i>	58
4.3.2. <i>Choix des attitudes de l'anglais</i>	58
4.3.3. <i>Définition des attitudes</i>	58
4.3.4. <i>Composition du Corpus</i>	59
4.3.5. <i>Enregistrement</i>	60
4.3.5.1. <i>Locuteur</i>	60
4.3.5.2. <i>Dispositif et procédure expérimentale</i>	60
4.3.6. <i>Validation perceptive</i>	61
4.3.6.1. <i>Méthodologie expérimentale</i>	61
1.1.1.1. <i>Stimuli</i>	61
4.3.6.2. <i>Matériel</i>	61
4.3.6.3. <i>Procédure</i>	61
4.3.6.4. <i>Sujets</i>	62
4.3.7. <i>Résultats du test perceptif</i>	62
4.3.7.1. <i>Distribution des réponses par rapport au hasard</i>	62
4.3.7.2. <i>Taux d'identification des attitudes</i>	64
4.3.7.3. <i>Analyse des résultats grâce à l'A.C.</i>	64
4.3.7.4. <i>Classification hiérarchique</i>	66
4.3.8. <i>Conclusion</i>	67
5. ÉTUDE DES ATTITUDES JAPONAISES	68
5.1. INTRODUCTION	68
5.2. DESCRIPTION DE LA STRUCTURE PHONOLOGIQUE DU JAPONAIS. ...	69
5.2.1. <i>Présentation générale</i>	69
5.2.2. <i>La structure phonologique du dialecte de Tokyo</i>	69
5.2.2.1. <i>La notion de « more »</i>	69
5.2.2.2. <i>Le système accentuel de Tokyo</i>	69

5.3. LES EXPRESSIONS ATTITUDINALES DU JAPONAIS	72
5.3.1. <i>Six « attitudes » dans l'enseignement du japonais langue étrangère.....</i>	72
5.3.2. <i>Caractéristiques phonétiques des six attitudes.....</i>	72
5.3.3. <i>L'approche fonctionnelle à partir du contour final.....</i>	73
5.3.4. <i>Le mouvement du contour intonatif de l'interrogation japonaise</i>	75
5.3.5. <i>Le rôle de la fréquence fondamentale pour 4 types d'interrogations</i>	75
5.3.6. <i>Etude comparative de la phrase interrogative (anglais vs. japonais)</i>	76
5.3.7. <i>Discussion</i>	77
5.3.8. <i>La politesse japonaise</i>	77
5.3.8.1. <i>Présentation générale</i>	77
5.3.8.2. <i>« Code de fréquence » (J. J. Ohala).....</i>	77
5.3.8.3. <i>Les attitudes de courtoisie et l'impolitesse/l'arrogance</i>	78
5.3.8.4. <i>Les travaux des sociolinguistes pour le rôle de F0.....</i>	79
5.3.8.5. <i>La stratégie de politesse chez les hommes japonais</i>	83
5.3.8.6. <i>Stratégie de politesse des hommes japonais en parole spontanée</i>	85
5.3.8.7. <i>La stratégie de l'aspiration ?</i>	86
5.3.8.8. <i>Stratégie de « Kyoshuku » et « sincérité/sérieux »</i>	87
5.3.8.9. <i>Les particules honorifiques vs. la prosodie.....</i>	88
5.4. VALIDATION DES ATTITUDES JAPONAISES.....	89
5.4.1. <i>Choix des attitudes du japonais</i>	90
5.4.2. <i>Définition des attitudes.....</i>	90
5.4.3. <i>Composition du Corpus.....</i>	92
5.4.4. <i>Enregistrement.....</i>	93
5.4.4.1. <i>Locuteur.....</i>	93
5.4.4.2. <i>Dispositif et procédure expérimentale</i>	93
5.4.5. <i>Validation perceptive.....</i>	94
5.4.5.1. <i>Problématique</i>	94
5.4.5.2. <i>Méthodologie expérimentale</i>	94
5.4.5.3. <i>Sujets</i>	95
5.4.6. <i>Analyse des résultats</i>	95
5.4.6.1. <i>Distribution des réponses par rapport au hasard</i>	95
5.4.6.2. <i>Effet de la longueur des énoncés</i>	96
5.4.6.3. <i>Effet de l'accent lexical</i>	96
5.4.6.4. <i>Taux d'identification des attitudes.....</i>	97
5.4.6.5. <i>Analyse des résultats grâce à l'analyse des correspondances</i>	99
5.4.6.6. <i>Classification hiérarchique.....</i>	101
5.4.7. <i>Analyse acoustiques</i>	102
5.4.7.1. <i>Paramètre prosodiques et analyse</i>	102
5.4.7.2. <i>Extraction des paramètres prosodiques</i>	103
5.4.7.3. <i>Analyse en composantes principales & classification hiérarchique.....</i>	105
5.4.7.4. <i>Résultats de l'ACP sur les paramètres prosodiques</i>	106
5.4.7.5. <i>Classification hiérarchique des paramètres prosodiques</i>	107

5.4.7.6.	Analyse de la qualité de voix	109
5.4.7.7.	Classification hiérarchique de la qualité de voix	110
5.4.7.8.	Analyse prosodie et qualité de voix.....	111
5.4.8.	<i>Conclusion</i>	112
5.5.	RECONNAISSANCE AUDIO-VISUELLE DES ATTITUDES	115
5.5.1.	<i>Introduction</i>	115
5.5.2.	<i>Enregistrement du corpus audio-visuel</i>	115
5.5.3.	<i>Perception des attitudes audio-visuelles</i>	116
5.5.3.1.	Méthodologie expérimentale	117
5.5.3.2.	Résultats du test perceptif	118
5.5.3.3.	Paramètres objectifs – acoustiques et visuels	130
5.5.3.4.	Conclusion	136
5.6.	TEST DE GATING SUR LES ATTITUDES JAPONAISES.....	137
5.6.1.	<i>Introduction</i>	137
5.6.2.	<i>Premier test de gating</i>	138
5.6.2.1.	Corpus.....	138
5.6.2.2.	Protocole expérimental	138
5.6.2.3.	Résultats.....	139
5.6.2.4.	Conclusion	144
5.6.3.	<i>Deuxième test de gating</i>	145
5.6.3.1.	Corpus.....	145
5.6.3.2.	Protocole expérimental	145
5.6.3.3.	Résultats.....	146
5.6.3.4.	Conclusion	152
5.6.4.	<i>Troisième test de gating</i>	153
5.6.4.1.	Introduction.....	153
5.6.4.2.	Méthodes	153
5.6.4.3.	Résultats.....	155
5.6.4.4.	Paramètres prosodiques et analyse	163
5.6.4.5.	Extraction des paramètres prosodiques	163
5.6.4.6.	ACP et classification hiérarchique.....	163
5.6.4.7.	ACP sur les paramètres prosodiques	163
5.6.4.8.	Classification hiérarchique des paramètres prosodiques	165
5.6.4.9.	ACP sur les attitudes et la qualité de voix	165
5.6.4.10.	Classification hiérarchique de la qualité de voix.....	167
5.6.4.11.	Analyse de la prosodie et de la qualité de voix	167
5.6.4.12.	Relations entre prosodie et perception.....	168
5.6.4.13.	Conclusion.....	170
6.	PERCEPTION INTERCULTURELLE DES ATTITUDES	173
6.1.	INTRODUCTION	173
6.2.	ÉTUDES INTERCULTURELLES DE PERCEPTION	173

6.3. ATTITUDES JAPONAISES PERÇUES PAR DES FRANÇAIS ET DES AMÉRICAINS	179
6.3.1. <i>Objectif du test perceptif</i>	179
6.3.2. <i>Composition du Corpus</i>	179
6.3.3. <i>Validation perceptive</i>	180
6.3.3.1. <i>Méthodologie expérimentale</i>	180
6.3.3.2. <i>Sujets</i>	181
6.3.4. <i>Résultats du test perceptif</i>	181
6.3.4.1. <i>Les sujets français</i>	181
6.3.4.2. <i>Les sujets américains</i>	189
6.3.5. <i>Conclusion</i>	196
6.4. ATTITUDES ANGLAISES PERÇUES PAR DES FRANÇAIS, DES AMÉRICAINS ET DES JAPONAIS.....	199
6.4.1. <i>Objectif du test perceptif</i>	199
6.4.2. <i>Choix des attitudes de l'anglais</i>	199
6.4.3. <i>Composition du Corpus</i>	200
6.4.4. <i>Méthodologie expérimentale</i>	200
6.4.4.1. <i>Stimuli</i>	200
6.4.4.2. <i>Matériel</i>	200
6.4.4.3. <i>Procédure</i>	201
6.4.4.4. <i>Sujets</i>	201
6.4.5. <i>Résultats du test perceptif</i>	202
6.4.5.1. <i>Les sujets américains</i>	202
6.4.5.2. <i>Les sujets français</i>	207
6.4.5.3. <i>Les sujets japonais</i>	213
6.4.6. <i>Conclusion</i>	219
6.5. ATTITUDES FRANÇAISES PERÇUES PAR DES AMÉRICAINS ET DES JAPONAIS	222
6.5.1. <i>Objectif du test perceptif</i>	222
6.5.2. <i>Choix des attitudes du français</i>	222
6.5.3. <i>Composition du Corpus</i>	222
6.5.4. <i>Méthodologie expérimentale</i>	223
6.5.4.1. <i>Stimuli</i>	223
6.5.4.2. <i>Matériel</i>	223
6.5.4.3. <i>Procédure</i>	223
6.5.4.4. <i>Sujets</i>	223
6.5.5. <i>Résultats du test perceptif</i>	224
6.5.5.1. <i>Les sujets américains</i>	224
6.5.5.2. <i>Les sujets japonais</i>	229
6.5.5.3. <i>Conclusion</i>	236
7. APPRENTISSAGE DES ATTITUDES JAPONAISES	239

7.1. PROBLÉMATIQUE.....	239
7.2. REVUE DE LITTÉRATURE SUR L'APPRENTISSAGE DES ATTITUDES .	239
7.3. OBJECTIF DU TEST PERCEPTIF.....	242
7.4. SUJETS	242
7.5. PROCÉDURE	242
7.5.1. <i>Deux entretiens</i>	<i>242</i>
7.5.2. <i>Catégorisation des sujets selon le niveau de langue</i>	<i>243</i>
7.5.2.1. <i>Objectif.....</i>	<i>243</i>
7.5.2.2. <i>Méthodes d'évaluation de la compétence en langue japonaise.....</i>	<i>243</i>
7.5.2.3. <i>Qu'est-ce que « SPOT » ?.....</i>	<i>244</i>
7.5.2.4. <i>Protocole du test « SPOT ».....</i>	<i>244</i>
7.5.2.5. <i>Résultats du test « SPOT »</i>	<i>244</i>
7.6. LES SUJETS FRANÇAIS DE NIVEAU 1	247
7.6.1. <i>Répartition des attitudes par rapport au hasard</i>	<i>247</i>
7.6.2. <i>Répartition des différentes longueurs et des attitudes par rapport au</i> <i>hasard</i>	<i>247</i>
7.6.3. <i>Effet de l'accent lexical.....</i>	<i>248</i>
7.6.4. <i>Taux d'identification des attitudes</i>	<i>248</i>
7.6.5. <i>Comportement perceptif des sujets français niveau 1 en analyse de</i> <i>correspondance.....</i>	<i>250</i>
7.6.6. <i>Confusion entre attitudes</i>	<i>252</i>
7.7. LES SUJETS FRANÇAIS DE NIVEAU 2	254
7.7.1. <i>Répartition des attitudes par rapport au hasard</i>	<i>254</i>
7.7.2. <i>Effet de la longueur des énoncés</i>	<i>255</i>
7.7.3. <i>Effet de l'accent lexical.....</i>	<i>255</i>
7.7.4. <i>Taux d'identification des attitudes.....</i>	<i>256</i>
7.7.5. <i>Analyse des correspondances pour les sujets français de niveau 2.....</i>	<i>257</i>
7.7.6. <i>Analyse des confusions.....</i>	<i>259</i>
7.8. CONCLUSION.....	261
8. CONCLUSIONS ET PERSPECTIVES	263
9. BIBLIOGRAPHIE.....	272
10. ANNEXES	287
11. RÉSUMÉ D'UNE SÉLECTION D'ARTICLES PUBLIÉS.....	317
12. SÉLECTION D'ARTICLES PUBLIÉS	320

Tables des figures

Figure 1 : Trois composants pour l'attitude selon Rosenberg & Hovland, Attitude Organization and Change. (Ajzen & Fishbein, 1980 p.19)	33
Figure 2 : Architecture cognitive du système de communication (Aubergé, 2004)	41
Figure 3 : Les trois fonctions de la prosodie (Aubergé, 2000)	42
Figure 4 : Paradigme d'études sur les attitudes selon le modèle développé au Gipsa-lab.	43
Figure 5 : Taux d'identification (pourcentage) de chaque attitude par 30 auditeurs français. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).....	52
Figure 6 : distribution des résultats des 30 auditeurs français pour les 12 attitudes françaises obtenus par l'AC (la figure du haut présente les dimensions 1 et 2 et celle du bas les dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitudes. Les cercles bleus indiquent les attitudes les mieux reconnues, tandis que les disques violets soulignent les principales confusions entre attitudes. Les autres couleurs indiquent les principales confusions. .	55
Figure 7 : dendrogramme résultant de la classification hiérarchique des expressions attitudinales du corpus de français, obtenue en utilisant la distance de Ward pour calculer les distances interclasses.....	56
Figure 8 : interface du test de perception en anglais pour les sujets britanniques (Diaferia, 2002).	62
Figure 9 : pourcentage d'identification de chaque attitude par les 10 auditeurs anglais. La ligne pointillée indique le seuil du hasard (9%) (Diaferia, 2002).....	63
Figure 10 : distribution des résultats des 10 auditeurs anglais pour les 11 attitudes anglaises par l'AC (figure du haut : dimensions 1 et 2 ; figure du bas : dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitude. Les cercles bleus indiquent les attitudes les mieux reconnues ; les disques violets soulignent les principales confusions inter-attitudes. Les autres couleurs indiquent les principales confusions.	65
Figure 11 : dendrogramme résultant de la classification hiérarchique des expressions attitudinales du corpus anglais, obtenue en utilisant la distance de Ward pour calculer les distances interclasses.....	66
Figure 12 : La distribution géographique des 4 types principaux de formes accentuelles (Hirata 1960). La couleur qui encadre la région signifie : rouge ; la forme accentuelle de Tokyo, bleue ; la forme accentuelle de Kansai, vert ; forme accentuelle variante de Kansai et gris ; l'absence totale d'accent lexical.....	70
Figure 13 : Réalisation phonétique de l'accent japonais avec la phrase /amearare/ par six locuteurs japonais. L'accent lexical est localisé en 1 ^{ère} more. Le contour de F0 en ligne	71

Figure 14 : pourcentage de reconnaissance de chaque attitude par les 15 auditeurs japonais. La ligne pointillée indique le seuil du hasard (8,4%).....	97
Figure 15 : distribution des résultats de 15 auditeurs japonais pour les 12 attitudes japonaises obtenus par l'AC (la figure du haut présente les dimensions 1 et 2, et celle du bas les dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitudes. Les cercles bleus indiquent les attitudes les mieux reconnues, tandis que les disques violets soulignent les principales confusions entre attitudes.	100
Figure 16 : résultats de l'analyse de classification hiérarchique, obtenue grâce à la distance de Ward sur les réponses moyennes de chaque attitude.....	102
Figure 17 : Deux dimensions principales de l'ACP pour les indices prosodiques. Les étiquettes rouges représentent les projections des vecteurs d'indices sur les deux premières composantes de l'ACP. Les attitudes (abréviations noires) sont projetées sur ces deux principales composantes, qui représentent 76% de la variance.	106
Figure 18 : classification hiérarchique des attitudes à partir de leurs indices prosodiques.....	107
Figure 19 : représentation des deux premières dimensions de l'ACP pour la qualité de voix. Les étiquettes rouges indiquent la projection des paramètres de qualité de voix ; tandis que les attitudes (en noir) sont projetées sur ces deux mêmes dimensions qui expliquent presque 80% de la variance.	108
Figure 20 : catégorisation des attitudes à partir de la similarité des valeurs de qualité de voix, obtenue par une classification hiérarchique.	109
Figure 21 : représentation des deux premières dimensions de l'ACP pour les indices de prosodie et de qualité de voix. Les étiquettes rouges indiquent la projection des paramètres acoustiques ; tandis que les attitudes (en noir) sont projetées sur ces deux mêmes dimensions qui n'expliquent que 56% de la variance.....	110
Figure 22 : catégorisation des attitudes à partir de la similarité des paramètres prosodiques et de qualité de voix par une classification hiérarchique.....	111
Figure 23 : taux de reconnaissance des 12 attitudes selon les groupes d'ordre de présentation des modalités. Les attitudes sont présentées sur l'axe des abscisses, et les taux d'identification en ordonnée.....	120
Figure 24 : effet de la performance des locuteurs sur la perception des attitudes. Les attitudes sont présentées en abscisse et le taux de reconnaissance moyen en ordonnée.	121
Figure 25 : effet du locuteur et de la modalité sur les taux d'identification des attitudes. Les attitudes sont présentées en abscisse et les scores de reconnaissance en ordonnée.	121
Figure 26 : résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « audio seul ». Les triangles rouges correspondent à la perception des stimuli, les ronds noirs indiquent les catégories d'attitude.	123
Figure 27 : résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « audio seul ».	124
Figure 28 : résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « vidéo seul ». Les triangles rouges	

<i>correspondent à la perception des stimuli, les ronds noirs indiquent les catégories d'attitude.</i>	<i>126</i>
Figure 29 : <i>résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « vidéo seul ».</i>	<i>127</i>
Figure 30 : <i>résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « audio-vidéo ». Les triangles rouges correspondent à la perception des stimuli, les ronds noirs indiquent position des catégories d'attitude.</i>	<i>129</i>
Figure 31 : <i>résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « audio-vidéo ».</i>	<i>130</i>
Figure 32 : <i>résultats de l'ACP sur les paramètres prosodiques des différentes attitudes du locuteur A. Ces deux premières dimensions de l'ACP expliquent 64% de la variance.</i>	<i>132</i>
Figure 33 : <i>résultats de la classification hiérarchique menée sur les données prosodiques des 12 attitudes du corpus audio-visuel.</i>	<i>133</i>
Figure 34 : <i>résultats de l'ACP menée sur les AU, montrant la dispersion des attitudes obtenue à partir des informations faciales.</i>	<i>134</i>
Figure 35 : <i>résultats de la classification hiérarchique des 12 attitudes à partir de leur description en AUs.</i>	<i>135</i>
Figure 36 : <i>taux de reconnaissance des attitudes pour les phrases de 2 et de 5 mores. Les attitudes sont présentées en abscisse, et le taux de reconnaissance au dernier gate en ordonnée, pour les 31 auditeurs japonais.</i>	<i>139</i>
Figure 37 : <i>évolution du taux de reconnaissance toutes attitudes confondues pour les phrases de 2 et 5 mores. Les abscisses indiquent le gate, l'ordonnée indique le taux de reconnaissance moyen.</i>	<i>140</i>
Figure 38 : <i>taux de reconnaissance par gate pour politesse-simple (PO), arrogance-impolitesse (AR), admiration (AD), pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.</i>	<i>141</i>
Figure 39 : <i>taux de reconnaissance par gate pour question-simple (QS) en haut à gauche, doute-incrédulité (DO) en haut à droite et irritation (IR), pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.</i>	<i>142</i>
Figure 40 : <i>taux de reconnaissance par gate pour déclaration (DC), autorité (AU) et exclamation de surprise (EX), pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.</i>	<i>143</i>
Figure 41 : <i>taux de reconnaissance par gate pour les expressions de politesse des phrases de 2 et 5 mores. Le graphe de gauche pour le stimulus de kyoshuku, celui de droite la sincérité-politesse) L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.</i>	<i>143</i>
Figure 42 : <i>taux de reconnaissance par gate pour les expressions de politesse regroupées ensemble, pour les phrases de 2 et 5 mores. Le graphe de gauche pour le stimulus de kyoshuku, celui de droite la sincérité-politesse) L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.</i>	<i>144</i>
Figure 43 : <i>croissance progressive de la reconnaissance globale pour toutes les attitudes confondues sur les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais.</i>	<i>146</i>

- Figure 44 :** croissance linéaire du taux d'identification pour quatre attitudes dans lesquelles l'effet du gate est identifié : doute-incrédulité [DO] (en haut à gauche), irritation [IR] (en haut à droite), politesse-simple [PO] (en bas à gauche) et arrogance-impolitesse [AR] (en bas à droite) pour les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais. 147
- Figure 45 :** Le mouvement du taux d'identification en chaque position du gate pour trois autres attitudes dans lesquelles l'effet du gate est identifié : question-simple [QS] (en haut à gauche), déclaration [DC] (en haut à droite) et exclamation de surprise [EX] (en bas) pour les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais. 148
- Figure 46 :** Trois attitudes pour lesquelles nous n'avons pas pu identifier l'effet du gate : admiration [AD] (en haut à gauche), évidence [EV] (en haut à droite) et autorité [AU] (en bas) pour les phrases de 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais. 150
- Figure 47 :** La reconnaissance pour les expressions culturelles de la politesse pour les phrases de 8 mores. Le graphe à gauche pour le stimulus de kyoshuku, et sincérité-politesse à droite. La ligne bleue indique la phrase de 8 mores sans accent lexical, et celle en rouge indique la phrase avec un accent sur la 1^{ère} more. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais. 151
- Figure 48 :** la reconnaissance globale de la politesse pour les phrases de 8 mores. Le graphe à gauche pour le stimulus de kyoshuku, et sincérité-politesse à droite. La ligne bleue indique la phrase de 8 mores sans accent lexical, et celle en rouge indique la phrase avec un accent sur la 1^{ère} more. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais. 151
- Figure 49 :** taux de reconnaissance des attitudes pour trois longueurs différentes (la phrase de 2, 5 et 8 mores respectivement). L'axe des abscisses indique les attitudes. L'axe des ordonnées indique le taux d'identification moyenne de 33 auditeurs japonais. 155
- Figure 50 :** croissance linéaire de la reconnaissance globale pour toutes les attitudes confondues sur les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais. 156
- Figure 51 :** croissance linéaire de la reconnaissance de cinq attitudes (admiration, arrogance-impolitesse, déclaration, doute-incrédulité et évidence) pour les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais. 157
- Figure 52 :** attitudes perçues dès le premier gate : autorité et exclamation de surprise, pour les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais. 158
- Figure 53 :** attitudes reconnues à un gate précis (le dernier, sauf l'irritation de 8 mores) : interrogation et irritation, phrases de 2, 5 et 8 mores. Les abscisses indiquent le gate et les ordonnées le taux de reconnaissance moyen des 33 auditeurs japonais. 158

- Figure 54** : attitudes de politesse pour les phrases de 2, 5 et 8 mores. En haut à gauche : kyoshuku ; en haut à droite : politesse-simple ; en bas : sincérité-politesse. Les abscisses indiquent le gate et les ordonnées le taux de reconnaissance moyen. 158
- Figure 55** : reconnaissance globale de la politesse pour les phrases de 2, 5 et 8 mores. Le graphe en haut à gauche pour le stimulus de kyoshuku, politesse-simple en haut à droite et la graphe pour le stimulus de sincérité-politesse est en bas). L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est un taux de reconnaissance moyenne de 33 auditeurs japonais. 159
- Figure 56** : distribution des points perceptifs et des points stimuli en deux dimensions calculées par l'analyse de correspondance. Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 162
- Figure 57** : Deux dimensions principales de l'ACP pour les paramètres prosodiques. Les étiquettes en rouge indiquent les paramètres pris en compte pour l'ACP. La distribution des attitudes qui sont abrégées et indiquée en noir. En ce qui concerne les vecteurs rouges, ils indiquent l'importance de paramètre. Paramètres pris en compte : la pente de F0 (F0slope), la pente en dernière voyelle de F0 (F0SlopelastVoy), l'écart type de F0 (F0Std), la moyenne de F0 (F0Mean), allongement de la dernière more (DURalgt), l'écart type de la durée (DURStd), la durée moyenne (DURMoy), la pente d'intensité (INTSlope), l'écart type de l'intensité (INTStd) et la moyenne d'intensité (INTMean). 164
- Figure 58** : Catégories des attitudes à partir de la similarité des valeurs de paramètres prosodiques selon la classification hiérarchique (AR). 165
- Figure 59** : deux dimensions principales de l'ACP pour les paramètres de qualité de voix. Les étiquettes rouges indiquent les paramètres de l'ACP. La distribution des attitudes qui sont abrégées et indiquée en noir. En ce qui concerne les vecteurs rouges, ils indiquent l'importance de paramètre. Breathly (= soufflée), Harsh (= rauque). 166
- Figure 60** : Catégories des attitudes à partir de la similarité des valeurs de paramètres de la qualité de voix selon la classification hiérarchique. 167
- Figure 61** : Catégories des attitudes à partir de la similarité des valeurs de paramètres prosodiques et de la qualité de voix selon la classification hiérarchique. 168
- Figure 62** : Contour de F0 et d'intensité pour question-simple japonaise sur la phrase de 2 mores présenté en haut à gauche, celle de 5 mores en haut à droite et 8 mores en bas. La ligne noire indique le contour de F0 présenté en demi ton, l'épaisseur de la ligne représente l'intensité et la taille des points rouges indique également la force d'intensité de l'attaque de chaque point perceptif (P-center). La ligne pointillée en bleue indique la frontière d'un segment. 169
- Figure 63** : Contour de F0 et d'intensité pour irritation japonaise sur la phrase de 2 mores présenté en haut à gauche, celle de 5 mores en haut à droite et 8 mores en bas. La ligne noire indique le contour de F0 présenté en demi-ton, l'épaisseur de la ligne représente l'intensité et la taille des points rouges indique également la force d'intensité de l'attaque de chaque point perceptif (P-center). La ligne pointillée en bleue indique la frontière d'un segment. 170
- Figure 64** : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs français et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO (doute-incrédulité), EV (évidence), EX (exclamation de

- surprise), IR (irritation), KYO (kyoshuku), PO (politesse-simple), QS (question-simple) et SIN (sincérité-politesse) 184
- Figure 65 :** Comportement perceptif de 15 auditeurs français pour les 12 attitudes japonaises présenté en deux dimensions (figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Signification des étiquettes : les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 186
- Figure 66 :** Graphe de Confusion pour 15 auditeurs français (niveau 0): pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO (doute-incrédulité), EV (évidence), EX (exclamation de surprise), IR (irritation), KYO (kyoshuku), PO (politesse-simple), QS (question-simple) et SIN (sincérité-politesse). 188
- Figure 67 :** Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs américains, français et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : AD (admiration), PO (politesse-simple), KYO (kyoshuku), SIN (sincérité-politesse), AR (arrogance-impolitesse), AU (autorité), IR (irritation), DO (doute-incrédulité), EX (exclamation de surprise), QS (question-simple), EV (évidence) et DC (déclaration)..... 191
- Figure 68 :** Comportement perceptif de 20 auditeurs américains pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4. Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles..... 193
- Figure 69 :** Graphe de Confusion pour 20 auditeurs américains (niveau 0) : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Abréviations : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO(doute-incrédulité), 194
- Figure 70 :** Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais et américains. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incrédulity), SU (surprise), SE (seduction) et PO (politeness). 203
- Figure 71 :** Comportement perceptif de 20 auditeurs américains pour les 11 attitudes anglaises (la figure du haut présente les dimensions 1 et 2 ; celle du bas les dimensions 3 et 4). Les étiquettes rouges correspondent aux distributions des stimuli, les étiquettes noires indiquent les distributions des catégories d'attitudes. Les cercles bleus indiquent la bonne perception des attitudes, les disques violets indiquent les confusions des attitudes qui sont à l'intérieur..... 204
- Figure 72 :** Graphe de Confusion pour 20 auditeurs américains : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de

- reconnaissance des attitudes en valeur relative. Signification des étiquettes: DC (déclaration), IN (interrogation), SU (surprise), DO (doubt-incredulity), EV (evidence), CO (command-authority), IR (irritation), PO (politeness), AL (scorn-aloofness), SE (seduction) et SI (sarcastic irony)..... 206
- Figure 73 :** Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais, américains et français. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (déclaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness)..... 208
- Figure 74 :** Comportement perceptif de 30 auditeurs français pour les 11 attitudes anglaises présenté en deux dimensions (la figure du haut présente les 1^{ère} et 2^{ème} dimensions. Celle du bas illustre les 3^{ème} et 4^{ème} dimensions. Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 210
- Figure 75 :** Graphe de Confusion pour 30 auditeurs français : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes: DC(déclaration), IN(interrogation), EV(evidence), IR(irritation), CO(command-authority), AL(scorn-aloofness), SI(sarcastic irony), DO(doubt-incredulity), SU(surprise), SE(seduction) et PO(politeness). 212
- Figure 76 :** Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais, américains, français et japonais. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (déclaration), IN (interrogation), EV (evidence), IR(irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness). 214
- Figure 77 :** Comportement perceptif de 34 auditeurs japonais pour les 11 attitudes anglaises (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les étiquettes rouges correspondent aux distributions des stimuli, les étiquettes noires indiquent les distributions des catégories d'attitudes. Les cercles bleus indiquent la bonne perception des attitudes, les disques violets indiquent les confusions d'attitudes qui sont à l'intérieur..... 216
- Figure 78 :** Graphe de Confusion pour 34 auditeurs japonais : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : DC(déclaration), IN(interrogation), EV(evidence), IR(irritation), CO(command-authority), AL(scorn-aloofness), SI(sarcastic irony), DO(doubt-incredulity), SU(surprise), SE(seduction) et PO(politeness). 218
- Figure 79 :** Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais et américains. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO .. 225
- Figure 80 :** Comportement perceptif de 21 auditeurs américains pour les 12 attitudes françaises présenté en deux dimensions (la figure du haut présente les 1^{ère} et 2^{ème} dimensions. Celle du bas les 3^{ème} et 4^{ème} dimensions. Les étiquettes en rouge

correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 227

Figure 81 : Graphe de Confusion pour 21 auditeurs américains : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Abréviations : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration). 229

Figure 82 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs français, américains et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration). 231

Figure 83 : Comportement perceptif de 38 auditeurs japonais pour les 12 attitudes françaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 233

Figure 84 : Graphe de Confusion pour 38 auditeurs japonais : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration). 235

Figure 85 : Taux d'identification de chaque attitude chez les 16 auditeurs français de niveau 1 vs. Français naïfs vs. natifs. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse), SIN (sincérité-politesse) et KYO (kyoshuku). 249

Figure 86 : Comportement perceptif de 16 auditeurs français du niveau 1 (débutant) pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les étiquettes avec le triangle rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles. 251

Figure 87 : Graphe de confusions chez les auditeurs français du niveau 1, le chiffre encadré indique le taux de confusion. L'autre chiffre en dessous de l'étiquette d'attitude représente le taux d'identification. Les attitudes qui sont dans les cercles verts sont les attitudes mal reconnues. Le chiffre qui se trouve à côté de la flèche indique le taux de confusion importante. 254

Figure 88 : Comparaison du taux d'identification des attitudes japonaises chez les Français du niveau 0 vs. Français du niveau 1 vs. Français du niveau 2 vs. natifs. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse), SIN (sincérité-politesse) et KYO (kyoshuku)..... 257

Figure 89 : Comportement perceptif de 16 auditeurs français du niveau 2 (intermédiaire) pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les 258

Figure 90 : Graphe de confusions chez les 16 apprenants français du niveau 2. Le chiffre encadré indique le taux de confusion en pourcentage. L'autre chiffre en dessous de l'étiquette d'attitude représente le taux d'identification. Les attitudes qui sont dans les cercles verts sont les attitudes mal reconnues. Le chiffre qui se trouve à côté de la flèche indique le taux de confusion importante..... 260

Liste des Tableaux

Tableau 1 : Corpus des attitudes françaises : quatre phrases de longueur variée.	49
Tableau 2 : Résultats du test de χ^2 sur les réponses moyennes par attitude (toutes longueurs confondues) par rapport au hasard.	51
Tableau 3 . Matrices de confusion des 30 auditeurs français. Les valeurs présentées en gras dans le sens diagonal sont identifiées au dessus de hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).	52
Tableau 4 : Corpus des attitudes anglaises, basé sur douze phrases de longueurs variées avec les différentes positions de l'accent lexical marquées par un astérisque.	60
Tableau 5 : Matrices de confusion des 10 auditeurs anglais. Les valeurs présentées en gras sont identifiées au dessus de hasard (9%). Signification des étiquettes : DC(declaration), IN(interrogation), SU(surprise), DO(doubt-incrédulity), EV(evidence), CO(command-authority), IR(irritation), PO(politeness), AL(scorn-aloofness), SE(seduction) et SI(sarcastic irony).	63
Tableau 6 : Corpus des attitudes japonaises : sept phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque.	92
Tableau 7 : Résultats du test de χ^2 sur les distributions moyennes des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.	95
Tableau 8 : Matrice de confusion des 15 auditeurs japonais. Signification des étiquettes : AD (admiration), PO (politesse-simple), KYO (kyoshuku), SIN (sincérité-politesse), AR (arrogance-impolitesse), AU (autorité), IR (irritation), DO (doute-incrédulité), EX (exclamation de surprise), QS (question-simple), EV (évidence) et DC (déclaration).	98
Tableau 9 : liste des indices prosodiques calculés à partir des mesures acoustiques de F0, durée moraïque et intensité et de qualité vocale, ainsi que les abréviations utilisées pour les désigner par la suite et le type de ces indices (c-à-d, s'il s'agit d'une mesure globale à la phrase ou d'une mesure permettant de juger de l'évolution d'un paramètre, de la forme d'un contour)	104
Tableau 10 : Proportion de la variance expliquée par chaque composante de l'ACP sur les indices prosodiques.	107
Tableau 11 : Proportion de la variance expliquée par chaque composante de l'ACP sur la qualité de voix.	109
Tableau 12 : résultats de l'ANOVA pour les sujets japonais, sur les réponses catégorielles. Les facteurs utilisés sont Grp pour les deux groupes d'ordre de passation des trois conditions ; Spk pour les deux locuteurs ; Mod pour les trois modalités. Les facteurs ayant un effet significatif sur les résultats (à $p < .01$) sont marqués d'une étoile.	119
Tableau 13 : phrases du test de dévoilement progressif.	138
Tableau 14 : phrases sélectionnées pour le test de dévoilement progressif.	145

- Tableau 15** : réponses des 10 auditeurs japonais à chaque gate, pour le stimulus d'admiration exprimé avec la phrase de 8 mores sans accent lexical. Les chiffres en gras italique indiquent le taux de reconnaissance, et les chiffres en gras désignent un degré de confusion important. 149
- Tableau 16** : Phrases sélectionnées pour le 3^{ème} test de dévoilement progressif. 154
- Tableau 17** : résultats des 33 auditeurs japonais pour chaque gate, pour la politesse exprimée avec la phrase de 8 mores. Les chiffres en gras italique indiquent le taux de reconnaissance, et les chiffres en gras désignent un degré de confusions important. 160
- Tableau 18** : L'importance des composants sur chaque dimension pour les résultats de l'ACP pour les paramètres acoustiques. Les deux premières dimensions expliquent 73% de la variance. 164
- Tableau 19** : L'importance des composants sur chaque dimension pour les résultats de l'ACP pour les paramètres de la qualité de voix. 166
- Tableau 20** : Corpus des attitudes japonaises : sept phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque. 180
- Tableau 21** : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 182
- Tableau 22** : Les matrices de confusion des 15 auditeurs français. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : AD (admiration), PO(politesse-simple), KYO (kyoshuku), SIN (sincérité-politesse), AR (arrogance-impolitesse), AU (autorité), IR (irritation), DO (doute-incrédulité), EX (exclamation de surprise), QS (question-simple), EV (évidence) et DC (déclaration). 187
- Tableau 23** : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 189
- Tableau 24** : Matrices de confusion des 20 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : AD(admiration), PO(politesse-simple), KYO (kyoshuku), SIN(sincérité-politesse), AR(arrogance-impolitesse), AU (autorité), IR (irritation), DO(doute-incrédulité), EX (exclamation de surprise), QS(question-simple), EV (évidence) et DC (déclaration). 194
- Tableau 25** : Corpus des attitudes anglaises : 12 phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque. 200
- Tableau 26** : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 202
- Tableau 27** : Matrices de confusion des 20 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC(declaration), IN(interrogation), SU(surprise), DO(doubt-incredulity), EV(evidence), CO(command-authority), IR(irritation), PO(politeness), AL(scorn-aloofness), SE(seduction) et SI(sarcastic irony). 206
- Tableau 28** : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 207
- Tableau 29** : Matrices de confusion des 30 auditeurs français. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-

authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness). 211

Tableau 30 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 213

Tableau 31 : Matrices de confusion pour 34 auditeurs japonais. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (déclaration), IN (interrogation), EV (évidence), IR (irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness). 218

Tableau 32 : Corpus des attitudes français : quatre phrases de longueur variée. 222

Tableau 33 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs confondues) par rapport au hasard. 224

Tableau 34 : Matrices de confusion pour 21 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration). 228

Tableau 35 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs confondues) par rapport au hasard. 230

Tableau 36 : Matrices de confusion des 38 auditeurs japonais. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration). 234

Tableau 37 : Les informations linguistiques des 32 sujets à partir de la 4^{ème} colonne : la moyenne de taux de réussite des deux tests « SPOT », la catégorie de niveau du japonais à laquelle ils appartiennent, leur âge, le nombre d'heures d'enseignement qu'ils ont suivi, le nombre d'heures de pratique orale, le nombre d'heure d'enseignement d'une autre langue étrangère, l'éventualité d'un séjour au Japon et leur possibilité de rencontre avec la langue japonaise dans leur vie quotidienne. 247

Tableau 38 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues). 247

Tableau 39 : Matrice de confusion des 16 auditeurs français, niveau 1. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse-simple), SIN (sincérité-politesse) et KYO (kyoshuku). 252

Tableau 40 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard. 255

Tableau 41 : Matrices de confusion des 16 français niveau 2. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse-simple), SIN (sincérité-politesse) et KYO (kyoshuku). 260

My major difficulty was the way I processed emotion through vocal variety. I am second-generation Italian-American and I am accustomed to STRESS-ing certain words to add COLOR and EMPH-asis. Unfortunately, this habit of mine unintentionally altered [Taiwanese] tones and consequently changed what I said.

[Ma difficulté majeure était la façon dont je traitais l'émotion à travers la variété de voix. Je suis la seconde génération Italiano-Américaine, et je suis habituée à accentuer certains mots afin d'ajouter la couleur et d'insister. Malheureusement, cette habitude de ma part a altéré, sans intention, les tons (taïwanais), et par conséquent cela a changé ce que je venais de dire.]

(Petrucelli, 2000: p.162, missionnaire américaine à Taïwan, cité par Pavlenko 2005, p.52. traduit par l'auteur.)

1. Avant Propos : l'Expressive Speech Project

1.1. Présentation générale

Le projet « *Expressive Speech Project* » (ESP) a permis la collaboration entre deux laboratoires de recherche : GIPSA-lab, département Parole et Cognition (ex-Institut de la Communication Parlée) et le laboratoire japonais ATR (Advanced Technology Research). Ce projet a abouti à être financé grâce à un contrat CREST établi en 2001 pour une période de 5 ans.

Ce projet a pour propos l'analyse et la simulation des paramètres acoustiques émotionnels et attitudinaux de la parole pour différentes langues maternelles / étrangères.

Il vise à montrer l'importance et le rôle des informations émotionnelles véhiculées dans la communication parlée puisque ces informations cohabitent de fait avec les informations linguistiques de la parole. En outre, le projet ESP permet de s'interroger sur la communication homme-machine par le biais des simulations d'expressions émotionnelles. Des voix de synthèse dotées d'émotions pourraient transformer les relations homme-machine encore peu naturelles et peu efficaces. Ainsi, des énoncés synthétisés sur la base d'expressions naturelles faciliteraient le décodage humain. On sait que tout énoncé naturel est inconsciemment accompagné de données émotionnelles : ne pas prendre en compte l'expression des émotions pourrait conduire à une désorientation de l'auditeur humain.

Le projet ESP porte sur l'analyse de la fonction expressive dans trois langues : le français, l'anglais et le japonais. Le japonais est principalement traité par l'équipe japonaise.

1.2. La place de nos travaux au sein du projet ESP

Les travaux de thèse que nous présentons ici s'inscrivent dans le cadre du projet ESP.

Selon le modèle développé par V. Aubergé au sein de GIPSA-lab, département Parole et Cognition, trois types de fonctions prosodiques distinctes sont postulées, selon la nature du contrôle exercé par le locuteur : l'expressivité émotionnelle (expression indirecte des émotions), les expressions attitudinales (expression directe des affects par

un contrôle volontaire et intentionnel du locuteur) et les fonctions linguistiques (fonction d'encodage des informations linguistiques liées au discours).

L'ensemble de ces fonctions prosodiques ont déjà été l'objet de travaux effectués au sein du GIPSA-lab. À la suite des études consacrées aux attitudes du français (Morlec, 1997 ; Morlec et al., 1997 ; Grépillat, 1996) et de l'anglais (Diaferia, 2002), ce sont les attitudes du japonais que nous traitons principalement. Le projet ESP vise à obtenir des résultats pouvant être utilisés dans l'enseignement systématique des langues étrangères et nourrir de nouvelles perspectives de recherche concernant la didactique de l'interculturel.

2.Introduction

2.1. Communiquer

Tentons de retourner concrètement à une description, au sens figuratif du terme, de ce qu'est la communication à travers deux exemples épiques : l'un emprunté au monde animal, l'autre au monde humain plus précisément, à la société adolescente japonaise.

Le premier cas concerne deux singes : le premier se fait chercher les puces par le second. Le temps passe lentement ; ils ne se regardent même pas mais le « skinship »¹ a valeur d'entraide sociale et de plaisir qu'aucun singe du groupe n'aurait envie de refuser. Nous choisissons d'utiliser le mot japonais « skinship » pour caractériser ce contact physique, terme que les Japonais ont emprunté à l'anglais pour désigner un comportement culturellement inhabituel au Japon mais qui entre doucement dans les mœurs en tant que mode de communication.

Second cas enfin : deux lycéennes rentrant chacune chez elles, venant à peine de se séparer après les cours, se parlent sur leur téléphone portable. Les sujets de conversation s'enchaînent et se ressemblent mais qu'importe. Leur parler d'une époque sans téléphone portable est peine perdue : c'est déjà le prolongement naturel de leur main...

L'objectif commun à ces deux types de communication physique et verbale est d'entretenir un contact au sein de la communauté et il est clair que ces échanges sont loin de répondre à un besoin de survie, mais visent plutôt à tisser des liens « tampons » qui relayent les échanges laconiquement informatifs.

Nous avons montré ici deux exemples de mode de communication et, dans le même ordre d'idée, mais à un niveau micro, la parole - mode de communication verbale - en plus de véhiculer des discours, est le reflet des sentiments, des émotions, des états mentaux et des intentions de chaque interactant.

Cette compétence communicative que nous avons acquise depuis l'enfance peut être à l'origine de problèmes de communication à l'étranger ou lorsqu'on communique avec des gens qui ne partagent ni la même langue ni la même culture car la façon (ou le mode) de communiquer peut être tout à fait différent de celle de l'interlocuteur.

¹ « Communication affective à travers un contact physique » (Japanese language dictionary, Daijirin, 2006, traduit par l'auteur).

C'est pourquoi, lorsqu'on apprend une langue étrangère, il est indispensable d'apprendre comment on exprime un sentiment, une intention par un comportement gestuel, facial et/ou verbal adapté.

2.2. L'enseignement de la communication verbale

Le mode verbal de communication en langue étrangère est enseigné sur la base de la description de leur fonctionnement linguistique. A l'oral, si la correction phonétique a donné lieu à nombre de méthodes, très peu sont proposées pour l'enseignement de la prosodie. On ne peut pas s'en étonner puisque le statut même de la prosodie est sujet de discussions encore abondantes en linguistique et phonétique, et que les connaissances sur son fonctionnement sont sans doute trop peu objectives pour alimenter le champ didactique de l'enseignement.

On constate qu'en français par exemple, les méthodes les plus abouties de l'enseignement de la prosodie proposé par G. Calbris et D. Callamand s'insèrent clairement dans une approche théorique très « gestaltiste » de la prosodie, ouverte par P. Delattre puis Y. Fónagy. Les fonctions attitudinales de la prosodie y sont mises en avant : le locuteur apporte à son discours oral des informations sur ses intentions pendant son interaction verbale.

L'hypothèse que nous proposons est que ces valeurs ne sont pas à proprement parler du domaine des émotions, mais sont propres à la langue elle-même. En conséquence les expressions de ces attitudes, voire les valeurs de ces attitudes, doivent être acquises par l'apprenant étranger pour éviter le danger d'un faux savoir implicite qui consisterait à projeter sur la langue cible les prosodies attitudinales de sa langue maternelle, avec tous les problèmes d'incompréhension, voire, de confusion que cela implique.

Notre objectif global dans cette thèse est d'envisager des caractéristiques prosodiques des attitudes prototypiques de plusieurs langues-cibles : le japonais, le français et l'anglais, et d'identifier de potentiels « décalages » de la perception des attitudes parmi ces trois groupes de sujets qui ne partagent pas ni la même langue ni la même culture.

Notre travail se base sur deux études préalables établies à Gipsa-lab, département Paroles et Cognition (ex-ICP), sur les attitudes de l'anglais (Diaferia, 2002) et du français (Grépillat, 1996 ; Aubergé et al., 1997). C'est dans ce cadre méthodologique que nous envisageons d'étudier les attitudes du japonais en les comparant avec celles de l'anglais et du français à travers des expérimentations perceptives.

Dans la première partie de la thèse, nous présenterons brièvement la phylogenèse de la terminologie « attitude » dans le domaine de la psychologie sociale en la comparant à celle d'« émotion », puis nous en ferons la description dans le cadre linguistique telles

qu'elle est exprimée par la prosodie. Par la suite, nous exposerons trois fonctions de la prosodie dans le système de la communication parlée, et surtout la conception de la fonction attitudinale de la prosodie sur laquelle nos recherches sont basées.

Dans la deuxième partie, nous présenterons les travaux effectués sur les attitudes du français (Grépillat, 1996 ; Aubergé et al., 1997) et sur celles de l'anglais (Diaferia, 2002) en complétant certaines analyses statistiques.

Dans la troisième partie, nous aborderons les attitudes japonaises : nous avons d'abord établi un repérage prosodique japonais à partir de deux types de littératures : des données provenant de domaines variés tels que la linguistique, la sociolinguistique et la phonétique et des données issues de la didactique du japonais langue étrangère. Après la sélection de 12 attitudes représentatives (y compris des attitudes spécifiques de la culture japonaise), puis leur validation auprès de natifs (de langue maternelle japonaise), nous avons orienté nos recherches en vue de comprendre le rôle et l'influence des deux modalités communicationnelles que sont l'audition et la vision, dans la perception des attitudes. Parallèlement, nous avons effectué un test de perception dans le paradigme du gating (ou de dévoilement progressif) aux contours prosodiques de phrase afin de tester l'« anticipation » des auditeurs et surtout d'identifier à quel moment (point d'unicité) ils sont capables de reconnaître une attitude. Les résultats obtenus sont comparés avec les valeurs acoustiques extraites du corpus validé.

Dans la quatrième partie de cette thèse, nous nous sommes concentrés sur une série de tests perceptifs des attitudes en français, en anglais et en japonais effectués auprès de publics non natifs de chacune de ces langues : des Français, des Américains et des Japonais. L'objectif de cette comparaison interculturelle des attitudes prosodiques était de mettre à jour les points communs et les décalages d'interprétations entre ces trois publics de langue et culture différentes.

Enfin, la dernière partie de cette thèse est consacrée à la description du comportement perceptif d'attitudes prosodiques dans un cadre didactique restreint. Il s'agissait d'examiner l'évolution du comportement perceptif des attitudes japonaises chez des sujets français étudiants en japonais langue étrangère ayant différents niveaux d'acquisition.

3. Analyse de l'état de l'art

3.1. Les émotions et les attitudes

Dans de nombreux ouvrages sur l'expression affective, les deux termes « émotion » et « attitude » sont confondus ou parfois utilisés comme synonymes. La cause de cette confusion doit être due aux définitions de ces deux termes qui varient selon les auteurs car ces deux notions sont, en fait, très complexes et difficiles à évaluer de manière objective.

Il nous semble utile de présenter brièvement les études effectuées sur l'émotion et l'attitude dans le champ de la psychologie afin de mieux les cerner.

Ensuite, nous allons discuter des étiquettes des attitudes.

3.1.1. Les émotions

Voyons d'abord quatre principaux courants théoriques dans la discipline de la psychologie : les perspectives évolutionnistes, physiologique, cognitive et la perspective du constructivisme social.

3.1.1.1. La perspective évolutionniste

Les premières études effectuées sur le statut des émotions remontent à 1872 lorsque Darwin publie « The Expression of Emotion in Man and Animals » qui pose les fondements de la théorie évolutionniste. Darwin considère l'émotion sur les plans biologique et adaptatif et spécialise ses recherches sur la fonction communicative des émotions et surtout sur les expressions faciales. Les recherches de Darwin établissent que les émotions sont à envisager comme des phénomènes évolués qui trouveraient racines dans les fonctions de survie. Finalement, l'aspect primitif des émotions prend tout son sens autant du point de vue phylogénétique (évolution de notre espèce) qu'ontogénétique (expérience personnelle).

A la suite de Darwin, d'autres théoriciens appelés les néo-darwiniens, comme Tomkins, Izard, Ekman ou encore Plutchik, ont complété la taxonomie des émotions et ont tenté d'en montrer l'aspect universel. Qu'il s'agisse « d'émotion de base » (Ekman, 1999), « d'émotions fondamentales » (Izard, 1971) ou encore « d'émotions primaires » (Plutchik, 1984), six émotions (étudiées chez la majorité des chercheurs mais de façons différentes) apparaissent incontournables au point d'être appelées « The Big Six » par Cornelius (2000) : la joie, la colère, la tristesse, la peur, le dégoût et la surprise.

3.1.1.2. L'approche physiologique des émotions

W. James tente d'associer les états émotionnels aux perceptions corporelles.

Selon W. James (1902), les émotions ne sont pas assimilées à une cause (le stimulus), mais elles constitueraient plutôt un effet (les changements physiologiques). Autrement dit, une fois que l'homme perçoit un stimulus venant de l'extérieur, c'est la perception des changements physiologiques périphériques qui constituerait l'émotion. Dans cette optique, l'émotion n'est pas déterminée par un centre spécifique du cerveau, mais par différents patrons physiologiques. Par conséquent, W. James ignore totalement les processus cognitifs et situationnels. Quant à C. Lange (1895), il considère, comme James, que l'émotion est une résultante des changements physiologiques, à la suite d'un stimulus externe. Cependant, contrairement à James, C. Lange admet l'existence d'un centre spécifique émotionnel. Malgré cette différence, ces deux auteurs insistent sur les manifestations physiologiques qui accompagnent systématiquement les émotions.

3.1.1.3. L'approche cognitive des émotions

En ce qui concerne les recherches faites sur les émotions, l'approche cognitive sous-tend que cognition et émotion sont inséparables. De fait, le déroulement et le déclenchement de la séquence émotionnelle constituent le point d'orgue de cette perspective. Arnold (1960) a établi un schéma récapitulant les liens entre le sujet et la situation qui l'environne. Le sujet est considéré selon son expérience, son état physiologique etc., tandis que la situation est traduite selon toutes les caractéristiques qui la composent. Ce schéma correspond aux processus cognitifs mis en place par l'individu pour appréhender les événements extérieurs (sont-ils bons ou mauvais ?)

En 1962, Schachter et Singes proposent leurs théories de l'évaluation. Selon eux, c'est l'action physiologique accompagnant une activité cognitive qui mène à un processus émotionnel. Ainsi, une activation physiologique pourrait être le moteur d'une émotion en considérant également la cognition du sujet. Arnold (1960) reprend cette notion d'évaluation dans un module comprenant divers mécanismes psychologiques qui transforment l'appréhension d'une situation ou d'un événement en un stimulus de chercheurs d'évaluation.

D'autres recherches de cette approche tentent de décrire les évaluations correspondantes aux émotions les plus exprimées.

3.1.1.4. La perspective du constructivisme social.

D'un point de vue social et culturel, les émotions n'ont été que peu étudiées.

Averill (1980) est l'un des théoriciens à penser les émotions dans le cadre du « constructivisme social », pour lequel les émotions s'intègrent dans le contexte social environnant.

Les constructions sociales permettent à l'individu de trouver une signification et une cohérence à son environnement. Dans cette optique, l'émotion est le résultat de normes sociales existantes dans un contexte social précis.

3.1.2. Les attitudes

Voyons la phylogénèse de la notion d'attitude au sein de la psychologie sociale, du modèle unidimensionnel au modèle multidimensionnel.

3.1.2.1. La perspective unidimensionnelle

Le concept d'attitude a été introduit au sein de la psychologie sociale à la fin du siècle dernier. Le terme « attitude » est alors utilisé pour pouvoir expliquer les réactions des individus face à une situation donnée. Les attitudes sont envisagées à partir de processus mentaux qui déterminent une réaction motrice (Thomas & Znaniecki, 1918). Selon cette théorie, les attitudes sont considérées comme des dispositions comportementales. Il s'agit, autrement-dit, d'une tendance générale vers ou contre certaines actions sociales. Dans cette perspective unidimensionnelle, l'attitude peut être mesurée selon des critères positif/négatif ou favorable/défavorable. De ce point de vue comportementaliste, les processus mentaux n'ont été évalués qu'à partir de l'acte effectué envers un objet ou une action, et ce qui n'est ni observable ni mesurable à travers une réaction (par exemple, l'attitude) n'appartient pas au domaine scientifique.

Ainsi, peu à peu, apparaît un « décalage » entre une attitude et la disposition à passer à l'action réelle qu'on mesure (voir par exemple les travaux conduits par LaPiere, 1934). Mais la relation entre l'attitude évaluée et l'action qui la suit n'a pas été clairement prouvée.

Dans cette situation, Allport (1935) a réclamé, par exemple, la nécessité d'une conception qui tiendrait compte de l'analyse et de l'évaluation qualitative des attitudes. Il définit le terme « attitude » comme suit :

An attitude is a mental and neural state of readiness which exerts a directing influence upon the individual's response to all objects and situations with which it is related²

(Allport 1935, p.810)

Dans cette perspective, la notion d'attitude est fondée par l'expérience d'un individu, et ce concept ne fait que guider un comportement envers un objet ou un événement, en étant tout de même indispensable afin d'expliquer un comportement. Selon cette définition, il conçoit l'attitude comme un état mental et nerveux d'une part et, d'autre part, comme une notion distincte de l'émotion ou de l'instinct avec qui elle est

² « Une attitude est un état mental et neuronal de disposition à passer à l'action, exerçant une influence directive sur les réponses de l'individu à tous les objets ou situations auxquels il est confronté » (traduit par l'auteur).

souvent confondue car l'attitude est un concept fondé par l'expérience, individuelle au sein de la société à laquelle l'individu appartient. Il reste cependant une part encore incomplète dans la notion d'état mental et nerveux. Ainsi, qu'en est-il du support physiologique et psychologique des attitudes ?

3.1.2.2. La théorie multidimensionnelle

Le concept d'attitude est principalement défini par rapport à sa représentation en mémoire ainsi qu'à son évaluation dans le temps. De fait, toutes les tentatives de définition du concept d'attitude du point de vue de ses propriétés intrinsèques ont été évacuées.

Il en résulte que les recherches actuelles traitent la notion d'attitude dans un cadre strictement opérationnel et formel (Stoezel, 1963) :

- une attitude caractérise un comportement dont l'orientation dépend d'une action liant un sujet à un objet. Une attitude est donc extrêmement versatile et ne peut être confondue avec un trait de caractère.
- l'orientation d'une attitude varie d'un extrême positif à un extrême négatif.
- une attitude est le résultat d'une acquisition au sein d'une société et dépend donc de l'environnement du sujet.

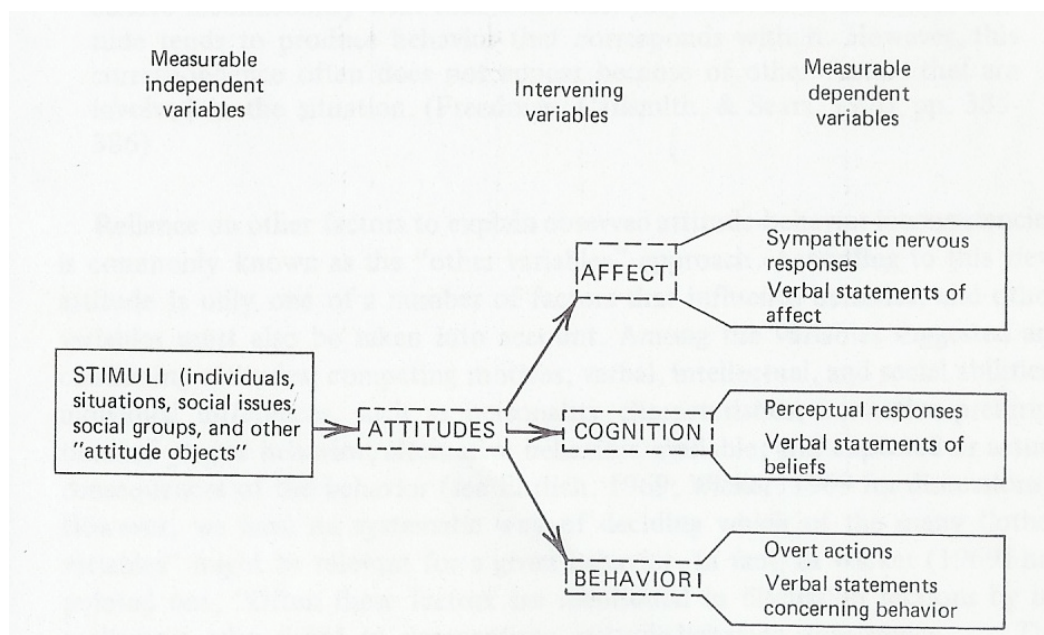


Figure 1 : Trois composants pour l'attitude selon Rosenberg & Hovland, *Attitude Organization and Change*. (Ajzen & Fishbein, 1980 p.19)

Rosenberg & Hovland (1960) ont adopté le modèle multidimensionnel afin d'évaluer l'attitude. Dans cette conception, l'attitude s'articule selon un processus complexe de trois paramètres interagissant entre eux : l'affect (sentiment), la cognition et la conation, et on utilise ces trois paramètres pour évaluer systématiquement les objets suivants. L'affect évalue un sentiment envers cet objet, la cognition, quant à elle,

est un composant qui évalue la croyance à propos de cet objet. Enfin, la conation est un composant qui nous sert à évaluer les diverses tendances d'action envers cet objet (Figure 1)

Afin d'évaluer la multidimensionalité de l'attitude selon ces trois composants, Osgood (1966) a développé un modèle nommé « semantic differential ». Il s'agit d'une évaluation systématique de l'attitude qui peut être observée selon les trois dimensions suivantes : évaluation (positive / négative), activation (forte / faible) et contrôle (volontaire / involontaire). Cette échelle d'évaluation de l'attitude est exploitée par Pakosz (1982) dans ses travaux traitant des caractéristiques de l'intonation pour des attitudes.

Chacun semble donc admettre l'acquisition des attitudes chez chaque sujet selon son expérience au sein de son environnement, ce qui les distingue des émotions ou de l'instinct. Elles se construisent au contact de situations externes ou de stimuli. Elles présupposent un tandem sujet-objet opérant selon trois axes : cognitif, affectif et comportemental. L'axe cognitif traite de l'idée que le sujet se fait de son environnement, comment il l'évalue. L'axe affectif procède des ajustements affectifs du sujet. Enfin, l'axe comportemental se rapporte à une réaction adoptée par le sujet, corrélative aux stimuli physiques et sociaux.

3.1.2.3. L'approche linguistique

Comme il a été vu précédemment, les émotions et les attitudes sont, selon les psychologues et les sociologues, toujours exprimées par les sujets au moyen de deux types de canaux comportementaux : physique (expressions du visage, gestualité) et langagier (langage et parole). De fait, les émotions et les attitudes font partie intégrante du langage.

On remarque alors que les recherches sur les attitudes et les émotions nécessitent d'explorer les rapports existants entre l'état des émotions et des attitudes et l'expression de celles-ci dans le langage. Il en résulte une grande variabilité : les états émotionnels et attitudeaux sont bien plus variés et complexes qu'on ne le postule souvent dans les modèles de base d'émotions fondamentales et discrètes (Scherer, 1996, 1998). En réponse à ce foisonnement, nombreux sont les spécialistes des attitudes et des émotions qui tentent de se rassembler au tour d'un « continuum attitude-émotion » dont l'intentionnalité et l'universalité seraient les critères majoritaires d'expression des attitudes et des émotions.

Ohala (1996), lui, tient compte de l'aspect ethnologique des attitudes et des émotions :

« Finally, at another level, one can recognize ways that humans convey their attitudes about the receiver, about the content or referent of their utterance, or about themselves »³

Cette définition démontre que l'expression attitudinale, réservée aux humains, a pour objet de signaler ce que l'émetteur pense d'un objet, de son interlocuteur, du contenu du message, ou de lui-même. Contrairement aux expressions qui portent des valeurs de survie (par exemple, montrer les dents pour la menace vs. non menace : sourire afin de montrer que l'émetteur n'est pas dangereux), les expressions attitudinales telles que l'ironie, le sarcasme, etc. ne semblent pas servir à la survie de l'individu. Aussi, Ohala suppose que ce niveau expressif fait l'objet d'un apprentissage et qu'il varie par conséquent d'une culture à l'autre, voire d'un individu à l'autre.

En outre, seulement ce troisième type d'expression émotionnelle aurait besoin d'une grande quantité de contexte pragmatique et linguistique de haut niveau pour être communiqué adéquatement.

3.1.3. Définition des attitudes et des émotions

A travers les diverses présentations des théories existantes sur les émotions et les attitudes, les caractéristiques de ces deux termes ont émergé :

Les émotions sont innées chez l'individu. Ce sont donc des caractéristiques universelles que possèdent les hommes pour la survie. Ce type d'affect émerge de manière involontaire.

A contrario, le terme « attitude » est utilisé essentiellement pour déterminer l'intention, et cette intention est en elle-même déterminée par la croyance, l'affect et le comportement. Ces trois facteurs qui encadrent l'intention d'un individu sont fortement influencés par (1) la société à laquelle appartient l'individu et par (2) la situation qui l'environne. Ce fait nous rappelle que les attitudes sont acquises, et non pas innées. C'est pourquoi les attitudes peuvent changer d'une culture à une autre. Les attitudes sont donc caractérisées par un contrôle volontaire, une intention encadrée par les aspects socioculturels.

3.1.4. Problème des étiquettes données aux attitudes

Une fois définie la différence entre émotions et attitudes, nous pouvons classer les étiquettes utilisées au sein de chacune de ces deux catégories.

Comme il existe une confusion entre les termes « émotion » et « attitude » parmi les chercheurs, les étiquettes utilisées en tant qu'attitude varient selon les auteurs. O'Conner (1973) parle du manque d'accord au niveau des catégorisations et des termes afin de traiter les attitudes. Il n'est pas rare de trouver des étiquettes telles que

³ « Enfin, à un autre niveau, on peut déterminer les moyens par lequel les humains transmettent leurs attitudes par rapport à l'interlocuteur, par rapport au contenu ou par rapport au référent de l'énoncé ou encore par rapport à eux-mêmes » (Ohala, 1996, p.1815. traduit par l'auteur).

l'angoisse, la tristesse ou la joie dans la catégorie « attitude » plutôt que dans la catégorie « émotions ».

Les définitions que nous avons vues au paragraphe précédent (*supra*) mènent à diviser les étiquettes en deux classes différentes.

Nous considérons que la joie, la tristesse, l'angoisse, la colère, le dégoût etc. sont des étiquettes réservées à la catégorie des émotions en accord avec la définition citée ci-dessus.

En revanche, l'irritation, l'ironie, la sincérité, le doute, la politesse etc. sont des étiquettes évaluées dans le cadre des attitudes. Les étiquettes des attitudes peuvent être nombreuses et parfois, une langue peut posséder des étiquettes attitudinales spécifiques (par exemple, la séduction pour l'anglais, une des formes de politesse en japonais etc.).

En ce qui concerne les deux étiquettes que sont la surprise et l'admiration, elles peuvent être classées dans les deux catégories car celles-ci peuvent être aussi bien volontaires qu'involontaires.

3.2. La prosodie

Historiquement, les phénomènes prosodiques ont été étudiés dans les domaines de la phonétique, de la phonologie et de la stylistique. Les linguistes du début du siècle les ont considérés comme des phénomènes « paralinguistique », « extralinguistique », ou de « style vocal ». Ces termes nous indiquent que la prosodie était perçue comme un fait marginal, sans importance pour les linguistes de l'époque. De nos jours, nous estimons que la prosodie joue un rôle essentiel dans la communication parlée, en tant qu'élément à part entière de la structure linguistique d'une part, et en tant qu'expression vocale des émotions et des attitudes d'autre part.

Nous présenterons d'abord sa définition, puis son rôle dans le système de communication et enfin une approche fonctionnelle des faits prosodiques (Aubergé, 2002).

3.2.1. Tentative de définition de la prosodie

Le terme « prosodie » est largement utilisé parmi les linguistes, les phonéticiens, les phonologues etc.. Cependant, sa définition est parfois changeante. L'origine de cette diversité est due à la difficulté de distinguer ce qui relève du niveau prosodique de ce qui émerge du niveau segmental ou phonématique.

Malgré ce désaccord sur la définition de la prosodie, les spécialistes utilisent trois paramètres physiques afin de donner les caractéristiques prosodiques d'une langue.

C'est ainsi qu'au niveau phonétique, la prosodie est mesurée par l'évolution temporelle de quatre paramètres acoustiques : **la fréquence fondamentale (F0), la**

durée, l'intensité et la qualité de voix. Ces quatre paramètres se définissent comme suit :

***La fréquence fondamentale** : fréquence de base qui donne la périodicité d'un son périodique complexe dont les harmoniques sont les multiples entiers. La fréquence fondamentale d'un son périodique complexe dont on connaît les harmoniques est égale au plus grand commun diviseur de ces harmoniques (PGCD). La fréquence est mesurée en Herz (Hz) ou nombre de périodes par seconde.*

(Rossi 1999, p.206)

***La durée** correspond à la mesure d'un intervalle de temps nécessaire pour émettre le signal de parole; elle concerne donc l'organisation temporelle du message et comprend le débit de parole (nombre de syllabes par seconde), le tempo (accélération ou ralentissement du débit à l'intérieur d'un groupe prosodique) et les pauses.*

(Lacheret-Dujour et al. 1999, p.12)

***L'intensité** : l'intensité mesure ce qu'on appelle le volume dans le langage courant ; elle est une mesure logarithmique de l'énergie du signal. L'intensité se mesure en décibels (dB). Le seuil différentiel d'intensité pour la parole est d'environ 3 dB : augmenter une voyelle de 3 dB équivaut à augmenter son amplitude de 40% et à doubler sa puissance.*

(Rossi 1999, p.206)

***La qualité de voix** : Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, [...] Both laryngeal and supralaryngeal features will be seen as contributing to voice quality.⁴*

(Laver 1980, p.1)

Parmi ces quatre paramètres, c'est la fréquence fondamentale qui est la plus étudiée car la F0 est la plus significative et la plus facile à percevoir. Cependant ces quatre paramètres prosodiques ne varient pas indépendamment au niveau phonétique. Au contraire, ils montrent une corrélation entre eux : ils sont dépendants les uns par rapport aux autres du fait des contraintes physiologiques.

Au niveau linguistique, le ton, le stress, le rythme et l'intonation naissent de la modulation des trois premiers paramètres physiques. Certains théoriciens incluent le ton et le stress dans la prosodie, alors que d'autres (Aubergé, 1997) considèrent qu'ils

⁴ La qualité de voix est conçue ici dans un sens large, comme la caractéristique auditive colorée par la voix d'un locuteur, [...] les traits laryngés et supralaryngés contribuent à la qualité de voix. (traduit par l'auteur).

relèvent de l'ordre lexical. Nous signalons ici que le présent travail de recherche a été effectué dans cette dernière perspective. En ce qui concerne la qualité de voix, elle est étroitement liée avec les affects sociaux (Campbell & Mokhtari, 2003 ; Brown & Levinson, 1986 ; Ito, 2001 ; Laver, 1980 ; Fónagy, 1991 etc.).

3.2.2. La Prosodie et les attitudes

Il faut remonter à la fin du XIXème siècle pour trouver les premières recherches effectuées sur les attitudes du français. Dans son étude sur les contours mélodiques du mot « oui », P. Passy fait état de 7 contours différents (Léon, 2000).

Voici quelques exemples de ces attitudes :

<i>Oui↓</i>	<i>[C'est mon avis].</i>
<i>Oui↓↓</i>	<i>[J'affirme cela].</i>
<i>Oui↑</i>	<i>[Est-ce vrai ?]</i>
<i>Oui↑↑ montée forte</i>	<i>[Pas possible !]</i>
<i>Oui↓↑</i>	<i>[C'est possible mais j'en doute].</i>
<i>Oui↑↓</i>	<i>[C'est bien clair].</i>
<i>Oui↑↓↓</i>	<i>[Sans doute, au premier abord ; mais...]</i>

(P. R. Léon 2000, p.134)

Il est à noter que cette description a été établie uniquement à partir du trait mélodique alors que les autres paramètres perceptifs tels que la longueur ou l'intensité n'étaient pas pris en compte. Bien que cette description des attitudes du français nous semble trop simpliste, cette approche nous paraît significative : elle signe les débuts des recherches sur les attitudes.

Comment les travaux sur la prosodie des attitudes ont-ils évolué par la suite ? L'approche didactique de D. Callamand et P. Delattre traite des attitudes comme moyen de communication. Dans cette perspective, leurs recherches sont centrées sur la stratégie pragmatique au cours de l'acte de parole. Cette conception de l'attitude est complétée par l'approche biologique de Fónagy (1991) ou neurophysiologique de Damasio (1994).

3.2.3. L'approche didactique

L'approche globale de la prosodie a été notamment exploitée par des didacticiens afin de pouvoir enseigner aux apprenants l'aspect pragmatique de la prosodie. Les contours globaux sont attribués et classés selon des fonctions linguistiques des énoncés. Nous présentons d'abord les travaux de D. Callamand, puis P. Delattre.

Callamand (1987) s'est intéressé aux aspects pragmatiques des attitudes du français dans un but didactique. Pour la description des attitudes du français, elle a choisi une approche empirique. C'est-à-dire qu'elle a interprété des attitudes prosodiques par impression auditive, puis leur a attribué des valeurs communicatives pour des

raisons pragmatique et didactique correspondant à l'enseignement de la langue française aux apprenants étrangers.

Après l'observation perceptive de la conversation ou du discours de différentes sources (par exemple, une interview à la radio ou à la télévision), D. Callamand a évalué les contours prosodiques globaux qui expriment chaque attitude dans un contexte donné. C'est ainsi qu'elle montre cinq attitudes qui permettent au locuteur d'informer autrui sur ses réactions personnelles et d'exprimer une position face à un message :

- *le [ton bas – ton haut – ton bas] caractérise l'ironie,*
- *le ton mélodique est indissociable de l'expression des sentiments,*
- *le ton suraigu traduit l'évidence ou l'indignation,*
- *le ton de justification,*
- *le [ton bas + chuchotement] dévoile une attitude méprisante.*

(Callamand 1987 p.6)

S'ensuivent des caractéristiques individuelles qui peuvent constituer des indices d'appartenance socioculturelle ou témoigner du désir de produire un effet :

- l'allongement de la syllabe finale et la montée de la hauteur d'où naît une impression de légèreté calculée pour la séduction,
- on peut ajouter l'allongement sur la dernière syllabe du groupe manifestant un ton affecté.

P. Delattre (1966) est un précurseur des travaux sur l'intonation du français. Sa théorie est basée sur une approche globale. Cette théorie a pour principe le classement des contours globaux selon des critères fonctionnels et perceptifs qu'il nomme « intonèmes ». Delattre (1966) a procédé à des tests perceptifs de discrimination auditive entre deux courbes et a identifié dix intonations de base, distinctives, à valeur linguistique qui servent à exprimer l'interrogation, l'exclamation etc.. Ces intonations sont représentées par des contours globaux dessinés à partir du décalage du niveau de hauteur du plus grave (1) au plus aigu (4). Chaque intonation est porteuse de sens sans se référer au contenu sémantique de la phrase. P. Delattre considère que pour l'identification de chaque contour global, il suffit de discerner les informations suivantes : la direction d'un contour (montant, descendant ou plat), le nombre et le type de niveaux traversés par les contours (par exemple, l'exclamation passe du niveau 4 au niveau 1) et la forme d'un contour (concave ou convexe). C'est ainsi que les dix intonations sont catégorisées en 7 intonèmes différents ; (1) continuation mineure, (2) continuation majeure, (3) question, (4) implication, (5) finalité, (6) interrogation, commandement et exclamation, (7) parenthèse et écho.

Selon les caractéristiques des traits distinctifs, ces intonèmes sont classés en trois types. (1)-(4) se trouvent dans la classe ascendante. (5) et (6) appartiennent à la classe descendante. En ce qui concerne l'interrogation, le commandement et l'exclamation, Delattre considère qu'ils ne sont pas distinctifs entre eux. Quant au (7), il est supposé

comme une classe en plateau. C'est ainsi que la parenthèse et l'écho sont en distribution complémentaire.

L'approche didactique concerne surtout les études pragmatiques visant des actes de parole. Les méthodes de description utilisées dépendent principalement de l'impression auditive de l'auteur (didacticien). Cette méthode intuitive n'a pas été validée perceptivement dans le cadre expérimental.

3.2.4. L'approche « biologique »

C'est ainsi que Y. Fónagy a effectué des tests perceptifs des énoncés avec des attitudes différentes afin de définir les prosodies attitudinales en rapport physiologique.

Y. Fónagy (1991) attribue un caractère biologique au phénomène de la fonction expressive. Les différents types de mouvements des organes phonatoires se greffent sans perturber ou décaler le sens du message verbal pour que ce message soit expressif. Y. Fónagy considère que la glotte est l'organe phonatoire indispensable pour transmettre non seulement un message verbal articulé, mais aussi les affects dans la voix en montrant les variations de tension subglottique. Par exemple, la contraction musculaire corporelle dans une situation tendue peut entraîner des effets sur la prosodie et amener à exprimer sa colère involontairement. En revanche, la décontraction musculaire dans une situation agréable, détendue génère une « tendresse » dans la voix. Ce fait implique que les organes phonatoires sont avant tout destinés à satisfaire des fonctions biologiques (respiration, alimentation etc.) avant de transmettre des messages linguistiques.

3.2.5. La prosodie dans le système de communication parlée

Certains modèles théoriques des affects considèrent que les expressions affectives peuvent être contrôlées dans les différents niveaux des processus cognitif (Scherer & Ellring, 2007), du contrôle involontaire au contrôle intentionné, délibérément des expressions contrôlées des attitudes du locuteur (Daneš, 1994 ; Ohala, 1996). Notre position (Aubergé, 2002) est de distinguer attitudes vs. émotions comme une différence en terme de nature du contrôle de la part du locuteur (volontaire vs. involontaire), ce qui n'est pas la valeur affective véhiculée par les expressions. Selon Damasio (1994), les émotions spontanées sont produites dans une « body loop », et peuvent être effectuées sans impliquer de changements physiologiques. Notre hypothèse est que cette compétence d'imitation des émotions peut être utilisée pour les émotions simulées et que des processus de nature différente existent pour des expressions authentiques vs. simulées. Dans cette perspective, une émotion produite à l'extérieur de la « *body loop* », que le locuteur soit sincère ou non, est un type de contrôle similaire aux affects attitudinaux produits spécifiquement par / pour la langue et la culture. C'est pourquoi tous les outils encodés socialement, tels que les expressions attitudinales, doivent être acquis pendant les phases développementales des enfants, et doivent être appris par les

apprenants des langues étrangères lorsque ces attitudes ne sont pas partagées par la langue maternelle (Daneš, 1994).

3.3. Un modèle de système de communication

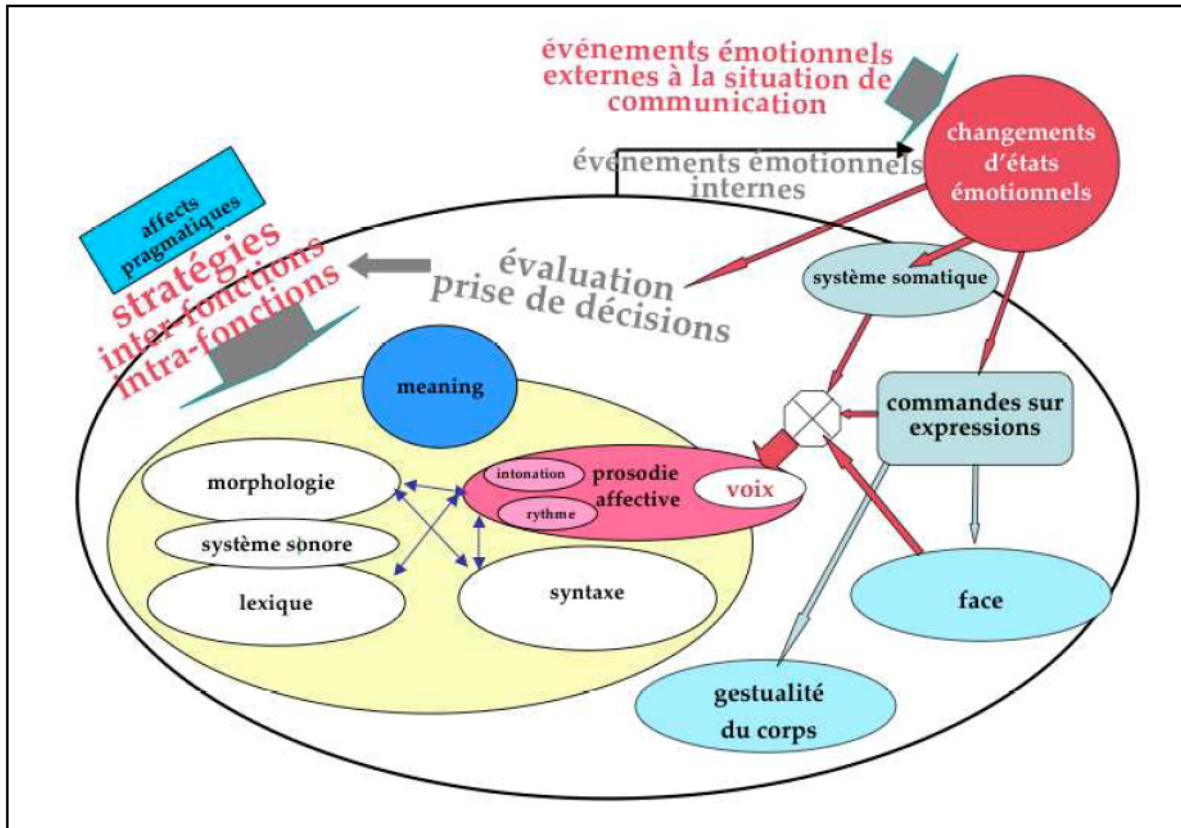


Figure 2 : Architecture cognitive du système de communication (Aubergé, 2004)

Selon le modèle sur lequel nos recherches sont basées (Aubergé, 2002), le système de la communication parlée utilise un ensemble de fonctions évaluées globalement pour le système (Figure 2). Ce système se compose de modules dont l'organisation interactive est basée sur la coopération entre les modules. Il s'agit, ici, de « rendez-vous » structurels (Aubergé, 1999).

Dans ce système de communication, les fonctions linguistique et émotionnelle de la prosodie émergent dans l'interaction parallèlement à des agents tels que la syntaxe, la morphologie, le lexique, la pragmatique, les expressions faciales...

3.3.1. Les trois niveaux prosodiques du système de communication parlée

Selon ce modèle, la prosodie réalise trois types de fonctions (figure 3).

3.3.1.1. Les fonctions linguistiques

Les fonctions linguistiques de la prosodie prennent en charge la segmentation et la hiérarchisation de l'énoncé. Ici, la prosodie joue un rôle d'indicateur des frontières syntagmatiques et paradigmatisques d'ordre syntaxique. La prosodie entretient également la distinction entre le thème et le rhème. Il s'agit donc d'une focalisation en tant que marque discursive.

La modélisation telle que la question totale, la question partielle, la déclaration ou encore l'exclamation, est aussi considérée comme une des fonctions linguistiques

Ces fonctions de la prosodie sont étroitement liées (et partagées) avec les autres agents des modules de la communication.

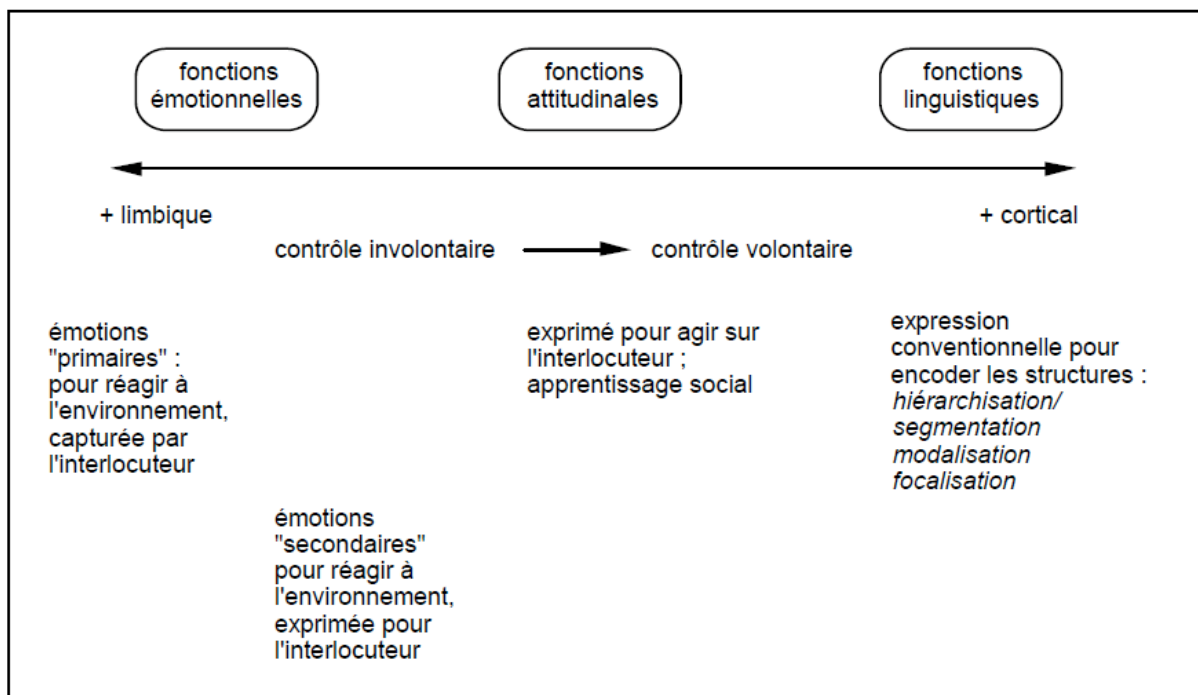


Figure 3 : Les trois fonctions de la prosodie (Aubergé, 2000)

3.3.1.2. Les fonctions émotionnelles

Les fonctions émotionnelles de la prosodie sont caractérisées par le contrôle direct et involontaire de la parole dans la sphère du locuteur. Les valeurs émotionnelles données par la fonction prosodique sont primordiales parmi tous les modules de la communication car la prosodie qui caractérise l'état d'esprit, l'émotion, se situe en dehors du cadre linguistique. Dans cette perspective, la prosodie joue un rôle fondamental accompagné des changements physiologiques et faciaux.

3.3.1.3. Les fonctions attitudinales

Les fonctions attitudinales de la prosodie prennent en charge des attitudes intentionnées de la part du locuteur dans une situation donnée. Ces fonctions sont

définies selon des critères issus expérimentalement du domaine de la didactique mais qui n'ont jamais été validées sur le plan perceptif ni même linguistique. Elles sont caractérisées par un contrôle direct et volontaire dans le domaine pragmatique et socioculturel (Callamand, 1987). Des énoncés véhiculés par les fonctions attitudinales de la prosodie peuvent être le fait d'états affectifs, cognitifs, ou d'une stratégie d'interaction par le biais de sa composante comportementale. Cette prosodie attitudinale est acquise assez tardivement dans l'interaction sociale (Clément, 1999). La prosodie n'est pas le seul agent permettant l'encodage de cette fonction (par exemple, la morphosyntaxe y participe également).

Dans cette perspective, les attitudes sont définies comme affects construits par et pour le langage, et ces faits de langage sont acquis dans l'interaction sociale. Puisque les attitudes sont acquises au sein de la société au cours de l'enfance (ou de l'adolescence), les attitudes d'une langue donnée reflètent les aspects culturels forts. La figure 4 illustre le paradigme des études sur les attitudes selon le modèle sur lequel nos travaux sont basés.

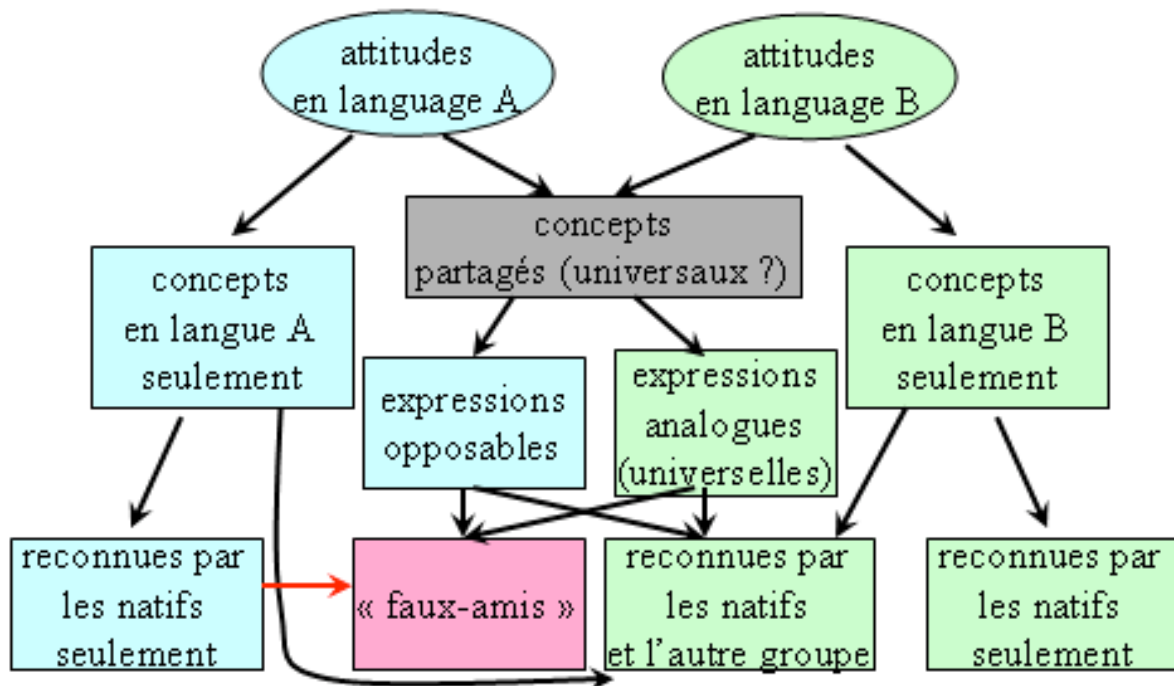


Figure 4 : Paradigme d'études sur les attitudes selon le modèle développé au Gipsa-lab.

Dans ce paradigme, certains concepts des attitudes de la langue A peuvent ne pas exister ou ne pas être conventionnalisés dans une autre langue⁵ sur le plan cognitif. Même si les langues / cultures partagent le même concept des attitudes telles que la surprise, le doute etc., ces attitudes peuvent être produites d'une façon tout à fait différente sur le plan expressif. Enfin, au niveau de la reconnaissance de ces attitudes, l'expression d'un

⁵ Par exemple, le concept de la politesse existe dans chaque culture, mais le concept de certains types de la politesse d'une langue / culture n'existe peut-être pas d'une manière conventionnelle dans une autre culture.

concept spécifiquement culturel ou d'un concept partagé peut être reconnue autant chez les natifs que les non natifs. Dans ce cas, cette expression peut être une attitude universelle.

A contrario, une expression d'un concept spécifiquement culturel peut être identifiée uniquement chez les natifs. De même, un concept partagé parmi plusieurs langues / cultures peut être identifié paradoxalement, uniquement chez les natifs malgré l'existence du même concept dans ces cultures concernées. Dans ce cas, cette expression renvoie en fait à un concept tout à fait différent chez les non natifs. Il s'agit, ici, de « faux-amis ». Ce terme désignant différentes formes (lexicale, prosodique, gestuelle) renvoie à un concept cognitif différent selon la langue et la culture. Par exemple le terme « déception » possède la même forme lexicale en français et en anglais, mais « *deception* » en anglais signifie « mensonge » en français. Un autre exemple est l'expression d'une des politesses japonaises réalisée avec la voix rauque mais perçue chez les Chinois comme de l'indifférence. Ce « décalage » de la perception est particulièrement appuyé dans la didactique des langues car ce décalage peut créer une perturbation considérable lorsqu'on communique en langue étrangère à cause de l'interférence de la langue maternelle.

Enfin le troisième cas que l'on peut imaginer est qu'une expression d'un concept spécifiquement culturel ou d'un concept partagé soit reconnue seulement par les natifs mais que cette expression n'interfère avec aucun concept dans l'autre culture. Dans ce cas, cette attitude est supposée comme inconnue dans cette langue et culture.

Les études interculturelles sont une façon idéale d'identifier des attitudes universelles (mais non innées comme les émotions), des « faux-amis » et des attitudes inconnues chez les non natifs (Kramer, 1964 ; Sherer et al., 2001 ; Maekawa & Erickson, 2001 ; Erickson, 2006 ; Albelin & Allwood, 2000, 2002 ; Nakamichi et al., 2003 ; Shigeno, 2001). Les aspects interculturels de la reconnaissance des attitudes sont particulièrement étudiés dans le chapitre 3 de cette thèse, mais avant cela, nous allons regarder dans le chapitre suivant les travaux sur les expressions des attitudes traitées initialement au sein de GIPSA-lab afin d'envisager les caractéristiques prosodiques des attitudes françaises (Morlec, 1997 ; Aubergé et al., 1999 ; Grépillat, 1996) et anglaises (Diaferia, 2002).

4. Les attitudes françaises et anglaises à l'I.C.P.

Les fonctions attitudinales de la prosodie française ont d'abord été étudiées par Grépillat (1996), Morlec (1997), Aubergé et al. (1997) et Morlec et al. (1999) qui ont effectué des tests de reconnaissance sur quatre attitudes et deux modalités (la déclaration, l'évidence, l'exclamation de surprise, le doute/incrédulité, ironie de soupçon et la question simple) en utilisant des stimuli naturels et synthétiques. Dans ce chapitre, nous allons d'abord nous intéresser rapidement aux travaux de T. Grépillat qui ont permis de valider les expressions attitudinales françaises précitées. Puis, nous nous pencherons plus attentivement sur la validation d'un nouveau corpus contenant ces six attitudes accompagnées de six autres. Ces 12 attitudes sont celles utilisées dans le chapitre 4 de cette thèse.

Nous rappellerons ensuite les travaux menés sur les attitudes anglaises par M.L. Diaferia (2002) qui ont permis de valider l'expression de 11 attitudes de l'anglais britannique. Nous compléterons ce travail par des analyses statistiques supplémentaires.

4.1. Études des attitudes prosodiques du Français

T. Grépillat (1996) a travaillé sur six attitudes du français. Elle a présenté, dans son mémoire de Mastère, deux études perceptives d'attitudes françaises. La première expérimentation a permis d'évaluer l'identification des six attitudes véhiculées par la prosodie. La seconde expérimentation consiste à dévoiler progressivement ces expressions attitudinales afin de confirmer la prédictibilité et la stabilité de l'identification précoce de la prosodie attitudinale.

Nous présentons ici sa première expérimentation perceptive qui est en rapport direct avec les tests perceptifs que nous présentons plus tard dans cette thèse.

Come nous l'avons déjà dit, T. Grépillat a travaillé sur quatre attitudes et deux modalités (c-à-d, ironie de soupçon, doute-incrédulité, évidence, exclamation de surprise, déclaration et question-simple), choisies sur la base de travaux en didactique du français langue étrangère (Callamand, 1973 ; Calbris & Porcher, 1989).

Ces six expressions ont été produites par un locuteur masculin de langue maternelle française, expert en prosodie. Il a enregistré 19 phrases différentes avec pour chacune les six expressions attitudinales, soit un total de 114 stimuli.

Ces 19 énoncés sont construits sur des contenus lexicaux relativement neutres tout en respectant une variabilité systématique de leur structure syntaxique.

Lors du test de perception, les stimuli sont présentés dans un ordre aléatoire différent pour chaque sujet. Chaque phrase est présentée deux fois aux sujets.

Ainsi, un nombre total de 228 stimuli (114 x 2 fois) ont été testés auprès de 12 auditeurs (étudiants entre 20 et 27 ans, tous de langue maternelle française). Afin de tester un éventuel effet d'apprentissage⁶, ces auditeurs sont répartis en deux groupes de 6 : le premier groupe effectue une phrase d'apprentissage tandis que le second commence le test directement.

Les résultats montrent que la plupart des attitudes ont été correctement identifiées (taux de bonnes réponses en dessus de 69%) sauf l'ironie de soupçon pour les sujets sans entraînement.

La « déclaration », l'« évidence » et la « question simple » ont été très bien reconnues, même sans apprentissage. Elles n'entretiennent aucune confusion avec les autres attitudes.

En revanche, le « doute-incrédulité », l'« exclamation de surprise » et l'« ironie de soupçon » ont été perçues différemment par les auditeurs entraînés et ceux non-entraînés. En effet, les auditeurs non-entraînés ont moins bien perçu ces trois attitudes (70% pour l'exclamation de surprise, 69% pour le doute-incrédulité et 51% pour l'ironie de soupçon). L'écart du taux d'identification entre les auditeurs entraînés et non-entraînés est d'environ 20%.

Parmi les trois attitudes les moins bien perçues, le « doute-incrédulité » et l'« ironie de soupçon » sont fréquemment confondues entre elles. L'« exclamation de surprise » a été confondue avec le « doute-incrédulité », mais ce phénomène n'a pas été identifié dans l'autre sens. C'est-à-dire que le « doute-incrédulité » n'a pas été confondu avec l'« exclamation de surprise ».

4.2. Validation des attitudes françaises

Ces travaux permettent donc de valider l'expression de six attitudes largement enseignées en didactique du français langue étrangère : *ironie de soupçon*, *doute-incrédulité*, *évidence*, *exclamation de surprise*, *déclaration* et *question-simple*. Il nous semble maintenant nécessaire d'élargir ce choix d'attitudes à un ensemble plus large, reflétant mieux les expressions utilisées par les français dans la vie courante. Cette nouvelle étape va permettre d'étudier la morphologie des nouvelles attitudes d'une part, et d'autre part, l'observation des phénomènes de confusion parmi ce large choix d'attitudes afin d'identifier les tendances de comportements similaires.

⁶ Que l'on pourrait aussi appeler l'effet d'enquête.

4.2.1. Objectif du test perceptif

A la suite des travaux sur les attitudes françaises développés au sein du GIPSA-lab, nous avons donc ajouté six nouvelles attitudes, soit un corpus de 12 attitudes françaises afin de les valider et d'identifier des catégories d'attitudes à travers l'observation de comportements perceptifs similaires.

4.2.2. Choix des attitudes du français

En plus des six attitudes françaises validées précédemment : ironie, doute-incrédulité, évidence, exclamation de surprise, déclaration et question-simple, nous avons ajouté les six attitudes suivantes : autorité, irritation, mépris, séduction et politesse à la liste des attitudes traitées dans cette expérimentation. Ces six étiquettes sont tirées du corpus de l'anglais britannique. En effet, il nous a semblé pertinent de les reprendre afin de faciliter et consolider nos travaux interculturels que nous verrons au chapitre quatre. De même, lors de la discussion avec la locutrice du corpus français, ses élicitations étaient satisfaisantes au niveau de l'expressivité en vue des natifs. Certaines de ces attitudes nous semblent spécifiques ou particulières à la culture occidentale, notamment celle de séduction de la part des hommes qui peut être difficile à éliciter en japonais. De plus, l'ironie sarcastique, la politesse ou le mépris peuvent aussi être les attitudes qui possèdent des aspects culturels forts.

4.2.3. Définition des attitudes

Voici les définitions des 12 attitudes françaises.

- Déclaration :

Le locuteur fait part d'une simple information, sans exprimer aucun point de vue.

- Question-simple :

Le locuteur demande une information, sans exprimer de point de vue, et sans attendre autre chose qu'une simple réponse.

- Exclamation de surprise :

Le locuteur exprime son étonnement tout en parlant.

- Evidance :

Le locuteur parle de quelque chose dont il est certain.

- Doute-incrédulité :

Le locuteur exprime son incertitude ou son manque de conviction concernant une information donnée par l'interlocuteur. Il exprime sa suspicion en reprenant dans son énoncé cette même information.

- Autorité :

Le locuteur veut influencer ou imposer son avis ou son désir à l'interlocuteur.

- Irritation :

Le locuteur est fortement mécontent par ce qui a été dit et proteste à travers cette contrariété.

- Ironie sarcastique :

Le locuteur remet en question (voire s'oppose) à l'affirmation de l'interlocuteur. Il veut montrer son désaccord et même la contrariété occasionnée par ce qu'il a entendu. Il réproouve des paroles ou une situation en utilisant une intonation contraire à ce qu'il pense et qui reviendrait à : «Oui, c'est exactement ça... »

- Mépris :

Le locuteur parle de façon arrogante et méprisante, montrant qu'il considère ce qui a été dit comme nul.

- Politesse :

Quelque soit le statut social du locuteur par rapport à l'interlocuteur, cette attitude est utilisée pour exprimer la politesse et la courtoisie.

- Séduction :

Le locuteur s'exprime de manière à attirer son interlocuteur (cas d'une relation de séduction)

- Admiration :

Le locuteur exprime son sentiment de plaisir, émerveillement et approbation.

4.2.4. Composition du Corpus

Le corpus est construit sur un ensemble de phrases isolées, comme c'est le cas pour d'autres travaux sur les attitudes françaises (Grépillat, 1996 ; Morlec, 1997), sur le hollandais (Van Heuven & al, 1997) ou sur le danois (Thorsen, 1980).

Toutes les phrases utilisées pour les tests de perceptions sont choisies parmi les phrases utilisées dans les travaux précédemment effectués (Grépillat 1996). Ces phrases sont soit un mot isolé, soit un Groupe Nominal (GN) suivi d'un Groupe Verbal (GV) réduit à la forme simple du verbe. Enfin la structure peut être composée d'un GN suivi d'un GV avec un Complément en fin de phrase. Voici les trois structures syntaxiques différentes pour les énoncés :

- *Structure A* : Mot isolé
- *Structure B* : N + V
- *Structure C* : Det + N + C

Quant à la longueur syllabique de l'énoncé, pour les phrases monosyllabiques, dissyllabiques, elle sera de cinq et de sept syllabes pour le test perceptif.

Tableau 1 : *Corpus des attitudes françaises : quatre phrases de longueur variée.*

Syllabes	Phrase	Structure
1	L'eau	mot isolé
2	L'eau bout	N + V
5	L'auto dérapait	N + V
7	Les gars coupaient des rondins	N + V + C

4.2.5. Enregistrement

Locuteur

Un chercheur féminin, expert de la prosodie française, a été choisi comme locutrice. Il s'agit d'une locutrice française native, chargée de recherche au CNRS.

Dispositif et procédure expérimentale

Le corpus a été enregistré sur une cassette DAT (échantillonné à 44,1 kHz, 16bits) dans la chambre sourde du GIPSA-lab, département parole et cognition. Les textes des phrases à enregistrer ainsi que l'attitude ont été envoyés de l'extérieur de la chambre sourde grâce à un fichier Power Point sur un écran faisant face à locutrice. Toutes les phrases ont ainsi été enregistrées, attitude par attitude, dans l'ordre suivant :

1. Déclaration
2. Question simple
3. Exclamation de surprise
4. Evidence
5. Doute-Incrédulité
6. Autorité
7. Irritation
8. Ironie sarcastique
9. Mépris
10. Politesse
11. Séduction
12. Admiration

Nous avons par la suite procédé au transfert des données audio grâce au logiciel de traitement du son « Cool Edit », puis découpé les énoncés produits sur le logiciel « Praat ver.4.1. ».

4.2.6. Validation perceptive

4.2.6.1. Méthodologie expérimentale

a- Stimuli

Les stimuli du test de perception sont donc constitués des quatre énoncés de longueurs différentes produits avec les 12 attitudes, soit un nombre total de 48 stimuli.

b- Matériel

L'interface du test perceptif a été réalisée grâce au logiciel « Revolution ». Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 12 attitudes possibles et les définitions de ces attitudes.

c- Procédure

Le test perceptif a été conduit sujet après sujet, individuellement. Lors du test, le sujet est assis en face d'un écran, et écoute les stimuli selon un ordre de présentation aléatoire différent pour chaque sujet. Pendant le test, le sujet doit juger quelle attitude il perçoit, à l'intérieur du choix fermé des 12 attitudes possibles. Chaque stimulus ne peut être écouté qu'une seule fois.

Lors du commencement du test perceptif, les sujets lisent la consigne affichée sur l'écran. Ensuite, des explications plus détaillées leur sont données à l'oral, ainsi que la définition des 12 attitudes et des explications supplémentaires pour s'assurer que chaque attitude soit bien comprise. Enfin, à la fin du test perceptif, nous avons interrogé le sujet à propos des attitudes afin de savoir quelles attitudes étaient difficiles à distinguer ou s'ils avaient perçu d'autres attitudes qui n'étaient pas proposées.

d- Sujets

Trente sujets (14 sujets masculins et 16 sujets féminins) ont participé à ce test perceptif. L'âge moyen des sujets est de 22,5 ans. Aucun sujet n'a fait état de troubles de l'audition.

Tous les sujets ayant participé à ce test sont de langue maternelle française, et habitent dans l'agglomération grenobloise. Certains sujets ont séjourné à l'étranger avant la passation du test perceptif, mais la période de ce séjour ne dépasse pas deux ans.

4.2.7. Résultats du test perceptif

4.2.7.1. Distribution des réponses par rapport au hasard

Nous allons commencer la description des résultats par l'analyse des distributions des réponses par rapport au hasard avant d'entreprendre de plus amples études statistiques. Pour comparer cette répartition par rapport à une répartition au hasard, le test du χ^2 est utilisé. Nous observons que la répartition diffère significativement du hasard avec une probabilité $p. < 0.01$ (Tableau 2).

Ce premier résultat montre qu'en formulant leur jugement, les auditeurs utilisent une stratégie cohérente : nous allons maintenant tenter de mieux comprendre ce qu'ils ont perçu. Or, comme chaque attitude reçoit un score d'identification supérieur au seuil du hasard ($110 / 12 < 8,4\%$), on peut en inférer en première approximation que chaque attitude a été en partie reconnue par les auditeurs français. Nous allons maintenant décrire plus précisément l'effet des différents facteurs contrôlés durant ce test de perception, sur les réponses des sujets.

Tableau 2 : Résultats du test de χ^2 sur les réponses moyennes par attitude (toutes longueurs confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	131.2	(ddl : 11, $p < 0.01$)
Autorité (AU)	521.8	(ddl : 11, $p < 0.01$)
Déclaration (DC)	522.2	(ddl : 11, $p < 0.01$)
Doute-incrédulité (DO)	294.4	(ddl : 11, $p < 0.01$)
Evidence (EV)	356	(ddl : 11, $p < 0.01$)
Exclamation de surprise (EX)	465.4	(ddl : 11, $p < 0.01$)
Irritation (IR)	246.6	(ddl : 11, $p < 0.01$)
Mépris (ME)	209.8	(ddl : 11, $p < 0.01$)
Politesse (PO)	335.4	(ddl : 11, $p < 0.01$)
Question-simple (QS)	289	(ddl : 11, $p < 0.01$)
Ironie-sarcastique (SA)	116.2	(ddl : 11, $p < 0.01$)
Séduction (SE)	301.8	(ddl : 11, $p < 0.01$)

4.2.7.2. Effet de la longueur des énoncés

Pour observer si la longueur de la phrase a pu avoir un effet sur la stratégie de réponses des sujets, nous avons effectué un test de χ^2 sur les résultats moyens obtenus par chaque longueur de phrases, et ceci entre les phrases de 1 et 2 syllabes, de 2 et 5 syllabes et de 5 et 7 syllabes, toutes attitudes confondues.

Le test de χ^2 sur les résultats moyens entre les phrases de 1 et 2 syllabes, toutes attitudes confondues, ne donne pas d'effet significatif (ddl = 11, $\chi^2 = 9.25$, $p > 0.05$). En revanche pour les phrases de 2 et 5 syllabes, toutes attitudes confondues, le test de χ^2 montre des distributions significativement différentes entre les deux longueurs de phrase (ddl = 11, $\chi^2 = 36.70$, $p < 0.05$). Enfin pour les phrases de 5 et de 7 syllabes, toutes attitudes confondues, le test χ^2 ne donne pas d'effet significatif (ddl = 11, $\chi^2 = 18.97$, $p > 0.05$).

Ce résultat montre un effet réduit de la longueur de l'énoncé sur la stratégie des auditeurs pour donner leur réponse : seules les phrases de 2 et 5 syllabes obtiennent des distributions de résultats significativement différentes.

4.2.7.3. Taux d'identification des attitudes

La figure 5 présente le pourcentage de reconnaissance de chacune des attitudes évaluées, et les résultats complets sont disponibles dans la matrice de confusion présentée au tableau 3. La ligne pointillée de la figure 5 indique le seuil du hasard. Cette figure indique que les auditeurs francophones ont reconnu toutes les attitudes au-dessus du seuil du hasard, même si certains scores restent particulièrement bas. Il ne faut cependant pas négliger la complexité de la tâche pour les auditeurs, et il sera surtout intéressant de détailler quelles ont été les confusions montrées par les auditeurs entre les attitudes, au delà du score brut de reconnaissance.

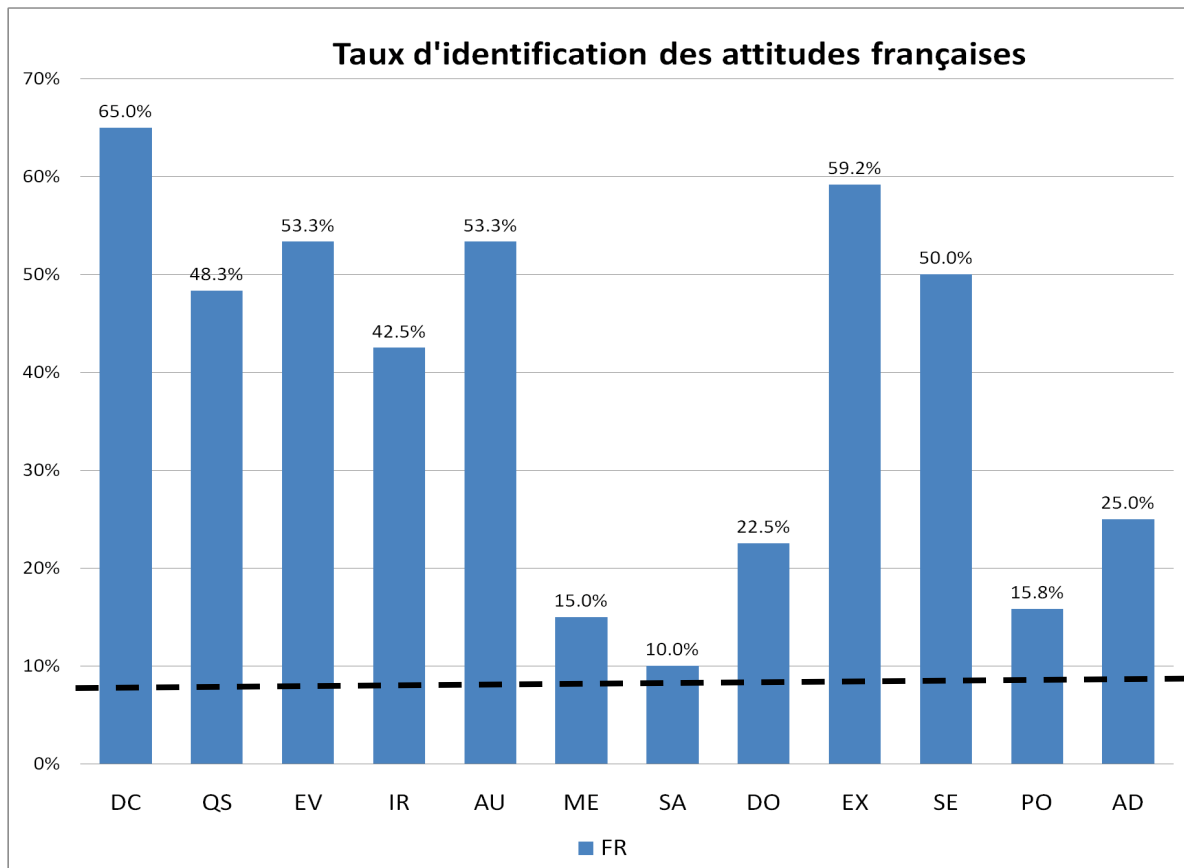


Figure 5 : Taux d'identification (pourcentage) de chaque attitude par 30 auditeurs français. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

Tableau 3. Matrices de confusion des 30 auditeurs français. Les valeurs présentées en gras dans le sens diagonal sont identifiées au dessus de hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

Attitudes perçues	Attitudes présentées											
	AD	AU	DC	DO	EV	EX	IR	ME	PO	QS	SA	SE
Admiration	25%	0%	0%	1%	1%	3%	0%	2%	5%	1%	1%	1%
Autorité	1%	53%	3%	0%	13%	1%	13%	3%	0%	2%	12%	0%
Déclaration	4%	2%	65%	3%	14%	0%	12%	41%	51%	18%	15%	8%
Doute-Incrédulité	10%	1%	3%	23%	3%	13%	0%	0%	3%	9%	3%	13%
Evidence	8%	2%	8%	0%	53%	0%	7%	16%	17%	4%	23%	5%
Surprise	26%	0%	0%	32%	0%	59%	0%	2%	2%	3%	2%	1%
Irritation	1%	40%	5%	0%	7%	0%	43%	4%	1%	5%	11%	1%
Mépris	0%	3%	11%	0%	4%	1%	19%	15%	3%	4%	23%	2%
Politesse	1%	0%	4%	1%	0%	0%	1%	8%	16%	0%	1%	11%
Question	14%	0%	0%	37%	3%	19%	1%	0%	0%	48%	0%	4%
Ironie-Sarcastique	3%	0%	1%	3%	2%	3%	6%	7%	2%	7%	10%	4%
Séduction	8%	0%	1%	1%	0%	1%	0%	4%	3%	0%	1%	50%

On peut remarquer sur cette figure que les expressions de *déclaration*, *question-simple*, *évidence*, *irritation*, *autorité*, *exclamation de surprise* et *séduction* reçoivent les meilleurs scores de reconnaissance. Par contraste, les taux d'identification obtenus par *mépris*, *ironie sarcastique*, *doute-incrédulité*, *politesse* et *admiration* restent faibles. *Ironie sarcastique* semble, selon cette figure, l'attitude la plus difficile à percevoir. Sur ce point, nous nous rappellerons que le pourcentage de reconnaissance d'*ironie sarcastique* était déjà le plus faible chez les sujets non entraînés dans les résultats obtenus par Grépillat (1996).

4.2.7.4. Note sur l'analyse statistique des données perceptives

Afin d'observer le comportement perceptif des sujets français, il semble particulièrement important d'analyser les confusions perceptives montrées par les auditeurs entre les attitudes. Pour cela, nous ne pouvons pas nous permettre de travailler uniquement sur le résultat de reconnaissance de chaque attitude (c-à-d, les résultats présentés par la figure 5), mais nous devons au contraire mener une analyse statistique sur les matrices de confusions obtenues en sortie du test de perception. Les données contenues dans une telle table sont de beaucoup plus riches, car elles permettent de mieux comprendre les erreurs de perception, et d'avoir une idée plus précise des proximités perceptives existant entre les différentes expressions attitudinales. Malheureusement, de telles données ne peuvent pas s'analyser à l'aide d'une analyse de variance (ANOVA). Ce problème est récurrent à l'ensemble de nos travaux de thèse, et il nous a fallu recourir à des méthodes d'analyses de données multidimensionnelles plus complexes, telles que l'Analyse en Composantes Principales, la classification hiérarchique ou l'Analyse des Correspondances, en fonction du type d'information que nous chercherons à obtenir des données, et bien entendu en fonction du type des données sur lesquelles nous aurons à travailler. On trouvera par exemple dans l'ouvrage de Benzecri (1973) une excellente description de ces deux dernières méthodes. L'analyse des données des tests de perception sera faite grâce à une Analyse des Correspondances et à une classification hiérarchique des différentes expressions attitudinales, car les matrices de confusion obtenues lors du test sont des tableaux de contingence où l'on compte le nombre de fois qu'une attitude est reconnue sous une certaine étiquette.

L'Analyse des Correspondances (AC) permet de calculer les distances perceptives existant entre les différentes attitudes présentées, et entre l'image que se font les auditeurs des 12 étiquettes attitudinales qui leur ont été présentées. Dans les graphiques présentant la distribution des résultats obtenus grâce à l'AC, nous aurons donc à chaque fois deux séries de points. Les uns, figurés par des triangles rouges, représentent les stimuli présentés aux auditeurs – et donc les réalisations des attitudes par chaque locuteur. Les seconds, figurés par des cercles noirs, représentent les étiquettes attitudinales proposés aux auditeurs. La position des triangles rouges par rapport aux cercles noirs permet donc de représenter graphiquement la distance perceptive existante entre la réalisation d'une attitude par un locuteur (triangles rouges) et la représentation

cognitive qu'ont de cette même attitude un ensemble de sujet à partir de son étiquette et de sa définition (cercle noirs). Plus ces deux points sont proches, plus la réalisation est conforme aux attentes perceptives des sujets. Cela permet également d'observer des catégories des comportements similaires par rapport aux attitudes perçues. Il est cependant nécessaire de noter que ces représentations graphiques sont réalisées selon des dimensions abstraites, calculées à partir des données des tests de perception, comme expliquant la plus grande part de la variance. Les positions des points sur un nombre réduit de ces dimensions montrent uniquement une projection de la distribution réelle des données. Lors de nos tests de perception, les données forment un espace à 12 dimensions (ou 11 pour les 11 attitudes de l'anglais), et nous présenterons généralement les résultats pour les quatre premières dimensions de l'analyse, à l'aide de deux graphiques en deux dimensions. Certaines proximités présentes sur les graphiques peuvent donc être tout à fait fortuites, et dues à la projection de certains points au même endroit sur une dimension, alors qu'une autre de ces dimensions permettrait de les différencier. C'est pour cela que l'analyse ne peut se passer des données réelles contenues dans les matrices de dispersion, et qui seules permettent de dire si une proximité sur un graphique correspond réellement à une proximité perceptive entre attitudes. La classification hiérarchique permet aussi de mieux répondre à cette question du regroupement des attitudes en « *clusters* ». La création d'arbres de classification hiérarchique permet donc d'obtenir une autre vision des données, complémentaire des distributions dans l'espace obtenues grâce à l'AC. Notons pour finir que l'analyse en Composantes Principales (ACP) permet le même type d'analyse de l'AC, mais doit être réalisée sur des mesures continues et non pas sur un tableau de contingence. Nous l'utiliserons donc pour traiter l'analyse des données prosodiques.

4.2.7.5. Analyse des résultats grâce à l'A.C.

Selon l'AC menée sur ce test de perception, les quatre premières dimensions expliquent 81% de la variance. La figure 6 représente les résultats de ce test respectivement projetés sur les deux premières dimensions, puis sur les dimensions 3 et 4 de l'AC. On peut y observer que sept attitudes (*déclaration*, *séduction*, *évidence*, *question-simple*, *exclamation de surprise*, *autorité*, et *irritation*) sont assez bien reconnues par les auditeurs francophones. En particulier, les expressions de *déclaration*, d'*évidence* et de *séduction* sont parfaitement reconnues, sans confusions particulières.

On observe par contre des rapprochements entre les attitudes d'*exclamation de surprise*, de *doute-incrédulité*, d'*admiration* et de *question-simple* d'une part et les expressions de l'*autorité* et de l'*irritation* d'autre part. Ces regroupements sont intéressants car ils témoignent d'une certaine proximité cognitive entre ces expressions, même si cela n'empêche pas les auditeurs de les différencier.

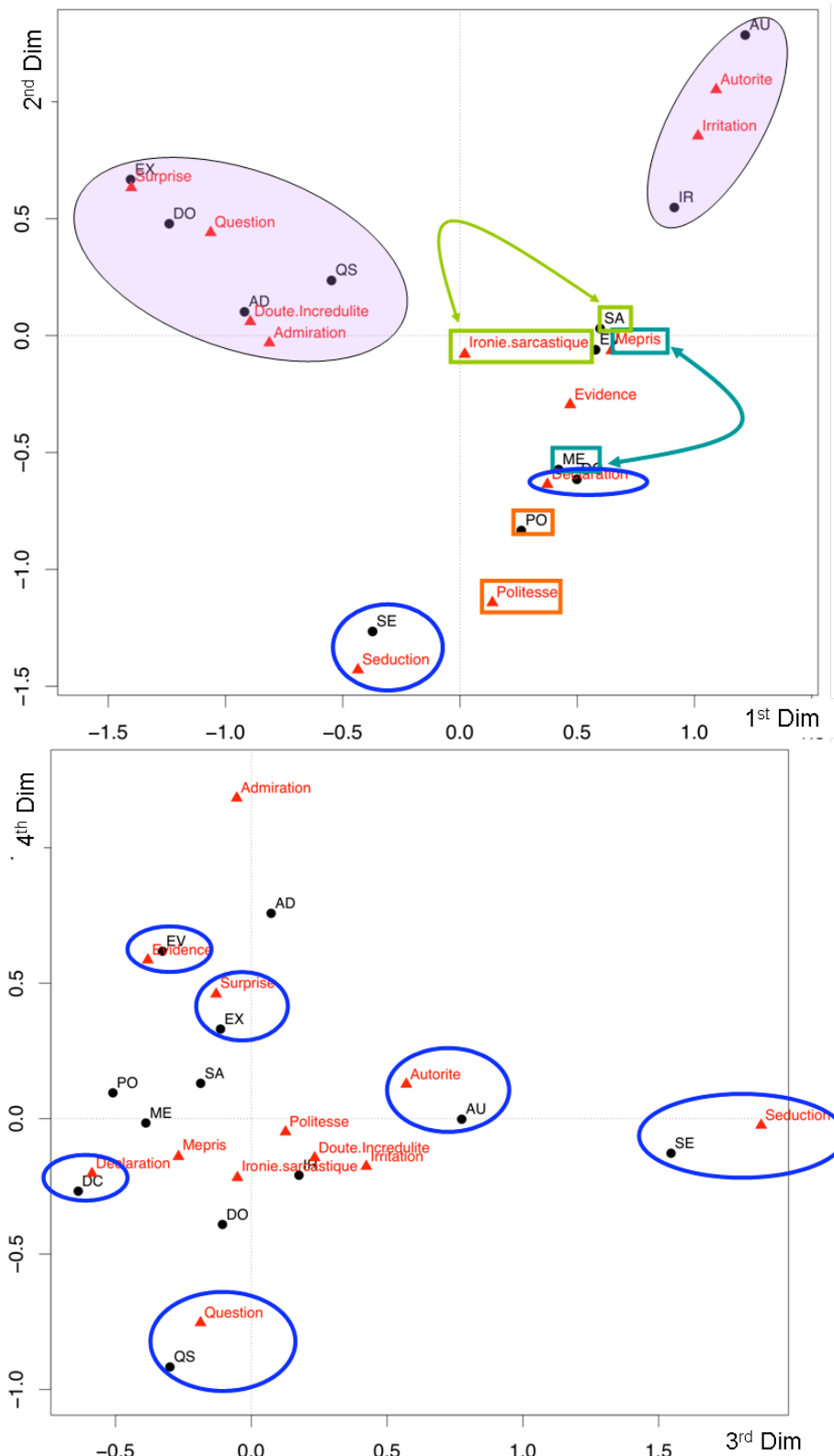


Figure 6 : distribution des résultats des 30 auditeurs français pour les 12 attitudes françaises obtenus par l'AC (la figure du haut présente les dimensions 1 et 2 et celle du bas les dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitudes. Les cercles bleus indiquent les attitudes les mieux reconnues, tandis que les disques violets soulignent les principales confusions entre attitudes. Les autres couleurs indiquent les principales confusions.

Les confusions les plus importantes concernent les attitudes de *mépris*, d'*ironie sarcastique* et de *politesse*. À la fois *mépris* et *politesse* sont principalement perçus comme une simple *déclaration*. Ce qui souligne la difficulté pour les auditeurs à donner un sens précis à ces expressions. Quant à *ironie sarcastique*, le point de stimulus étant localisé près de l'origine, il semblerait que les sujets perçoivent une expression, mais ont du mal à la qualifier. On notera cependant la relative proximité des expressions d'*ironie sarcastique*, de *mépris* et d'*évidence*, ce qui fait ressortir les similitudes qui existent entre ces notions et ces expressions.

4.2.7.6. Classification hiérarchique

L'analyse des résultats grâce à la classification hiérarchique (cf. figure 7) fait apparaître les regroupements suivants. Au plus haut niveau, on obtient trois grands groupes d'expressions attitudinales :

- *autorité, évidence, irritation, et ironie sarcastique*
- *déclaration, mépris et politesse*
- *question-simple, admiration, doute-incrédulité et exclamation de surprise*

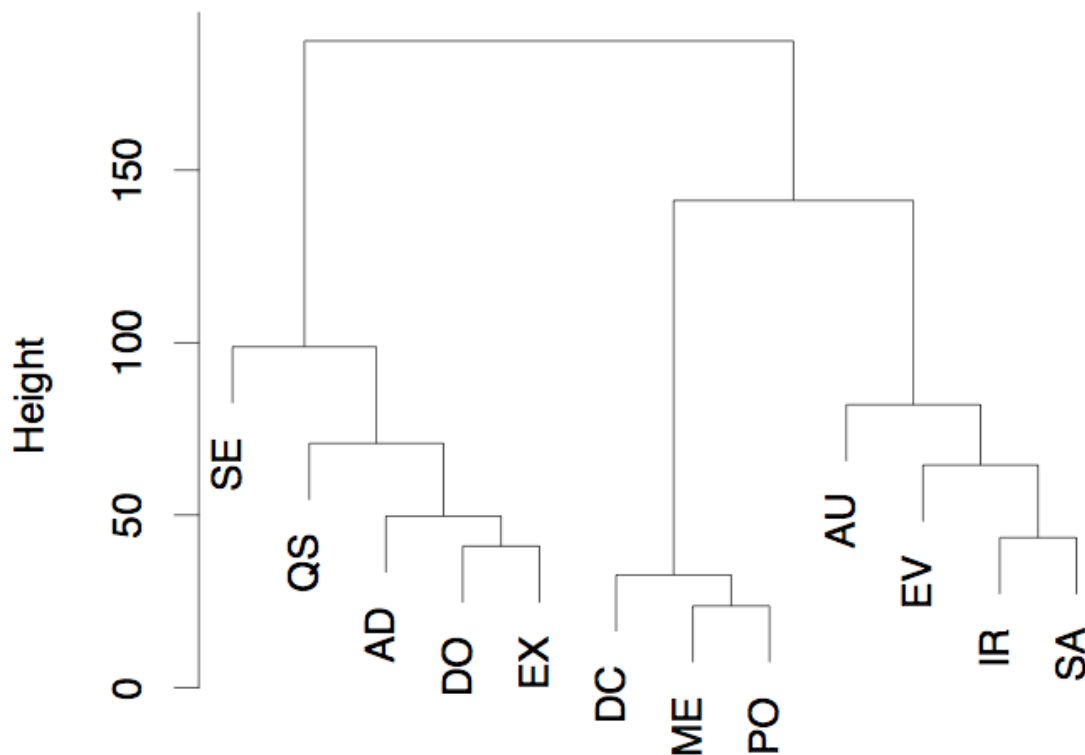


Figure 7 : dendrogramme résultant de la classification hiérarchique des expressions attitudinales du corpus de français, obtenue en utilisant la distance de Ward pour calculer les distances interclasses.

L'expression de *séduction* reste à part. Chacun de ces trois groupes montrent des expressions ayant des points communs : soit un degrés d'imposition de son opinion à son

interlocuteur fort (premier groupe) ou faible (troisième groupe) ; ou un taux élevé de confusions avec la *déclaration* dans le cas du second groupe.

Si l'on s'intéresse aux expressions montrant les plus faibles distances entre elles, on retrouve, mis à part le cas de *mépris* et de *politesse* déjà traité, (1) la perception d'*ironie-sarcastique* comme d'*évidence* ou d'*irritation* et (2) les confusions des expressions de *doute-incrédulité* et d'*exclamation de surprise* entre elles et avec l'expression de *question-simple* et la perception d'*admiration* comme d'*exclamation de surprise* ou de *question simple*.

4.2.8. Conclusion :

À la suite des travaux sur les attitudes françaises développés au sein du GIPSA-lab, il nous a semblé nécessaire d'augmenter le nombre d'attitudes du corpus afin de mieux refléter les différentes expressions que les francophones utilisent quotidiennement, attitudes qu'il nous a fallu valider perceptivement. L'observation des confusions au sein de ce plus large choix d'attitudes nous a permis d'identifier les expressions les plus proches et celles les mieux différenciées.

De cette étude, il est possible de conclure tout d'abord que la distribution moyenne des attitudes est différente d'une distribution au hasard, selon un test de khi deux. On peut par ailleurs noter un effet de la longueur de l'énoncé existant uniquement entre les phrases de 2 et de 5 syllabes. Finalement, le pourcentage de reconnaissance obtenu par chaque attitude étant supérieur au seuil du hasard, on peut supposer que les auditeurs ont perçu chacune des attitudes, même si certaines reçoivent des scores particulièrement bas. Les analyses multidimensionnelles permettent de mieux comprendre comment sept attitudes sont bien identifiées par les auditeurs ; comment certaines expressions se regroupent en ensembles d'expressions plus génériques et cohérents ; et comment certaines attitudes entretiennent d'importantes confusions avec d'autres – mais ces confusions restent logiques au vu des regroupements d'attitudes : il n'y a pas de contresens dans la reconnaissance des expressions proposées.

Le regroupement des expressions d'*irritation*, d'*ironie-sarcastique*, d'*évidence* et d'*autorité* montre que les auditeurs français perçoivent au travers de ces expressions l'imposition de son opinion personnelle. Les confusions observées à l'intérieur de ce groupe peuvent donc s'expliquer ainsi.

Ce corpus de 12 attitudes françaises est ainsi validé. L'évaluation des comportements perceptifs des auditeurs non natifs (Américains et Japonais) pour ces attitudes françaises sera présentée plus loin afin d'extraire les similitudes et les décalages d'interprétations entre ces trois publics de langue et culture différentes.

4.3. Validation des attitudes anglaises

4.3.1. Objectif du test perceptif

A la suite de la validation des attitudes françaises, nous allons suivre ici les travaux de M.L. Diaféria (2002) sur la validation des attitudes anglaises en ajoutant des analyses statistiques supplémentaires. L'objectif de ses travaux était d'analyser et d'évaluer la production et la perception d'un ensemble d'expressions attitudinales de l'anglais britannique.

4.3.2. Choix des attitudes de l'anglais

Un ensemble de 11 attitudes représentatives de l'anglais (voir Diaferia, 2002 pour les détails d'élaboration du corpus) a été sélectionné à partir de la littérature sur les attitudes anglaises (Crystal, 1997; Trask, 1996). Un corpus a été élaboré à l'aide d'un expert de langue maternelle anglaise, enseignant d'anglais langue étrangère et spécialiste de prosodie. Ce sera ce corpus qui va nous servir à tous les tests perceptifs menés sur les expressions attitudinales de l'anglais britannique dans la suite de ce travail.

Sur l'ensemble de ces onze expressions attitudinales (cf. infra), certaines sont spécifiques à la culture britannique, notamment l'expression de séduction masculine, que l'on pourrait difficilement transposer dans un contexte culturel japonais. De plus, l'ironie sarcastique, la politesse ou le mépris constituent aussi des attitudes fortement marquées culturellement.

4.3.3. Définition des attitudes

Les attitudes sélectionnées sont les suivantes (les noms sont donnés en anglais avec la traduction et leurs abréviations entre parenthèses)⁷ :

Doubt-incredulity (doute-incrédulité, DO) :

Le locuteur exprime son incertitude ou son manque de conviction concernant une information donnée par l'interlocuteur. Il exprime sa suspicion en reprenant dans son énoncé cette même information.

Evidence (évidence, EV) :

Le locuteur parle de quelque chose et dont il est certain.

Surprise (exclamation de surprise, SU) :

Le locuteur exprime son étonnement tout en parlant.

⁷ La définition originale, en anglais est présentée en annexe.

Command-authority (autorité, CO) :

Le locuteur veut influencer ou imposer son avis ou son désir à l'interlocuteur.

Irritation (irritation, IR) :

Le locuteur est fortement mécontent par ce qui a été dit et proteste à travers cette contrariété.

Sarcastic irony (ironie sarcastique, SI) :

Le locuteur remet en question (voire s'oppose) à l'affirmation de l'interlocuteur). Il veut montrer son désaccord et même la contrariété occasionnée par ce qu'il a entendu. Il réproouve des paroles ou une situation en utilisant une intonation contraire à ce qu'il pense et qui reviendrait à : «Oui, c'est exactement ça... »

Scorn-aloofness (mépris, AL) :

Le locuteur parle de façon arrogante et méprisante montrant qu'il considère ce qui a été dit comme nul.

Seduction (seduction, SE) :

Le locuteur s'exprime de manière à attirer son interlocuteur (cas d'une relation de séduction)

Politeness (politesse, PO) :

Quelque soit le statut social du locuteur par rapport à l'interlocuteur, cette attitude est utilisée pour exprimer la politesse et la courtoisie.

En plus de ces attitudes se trouvent deux modalités : l'interrogation et la déclaration. Elles peuvent être assimilées à des attitudes n'ayant pas de composante affective ou en ayant une totalement neutre. Elles servent également de référence.

Declaration (déclaration, DC) :

Le locuteur donne l'information simplement, sans exprimer aucun point de vue.

Interrogation (question-simple, IN) :

Le locuteur demande une information, sans exprimer de point de vue, et sans attendre autre chose qu'une réponse simple.

4.3.4. Composition du Corpus

Le corpus est construit sur un ensemble de phrases isolées, sur le même principe que pour les travaux sur les attitudes françaises (Grépillat, 1996 ; Morlec, 1997), ou sur le hollandais (Van Heuven & al, 1997) et sur le danois (Thorsen, 1980).

Toutes les phrases utilisées pour les tests de perceptions sont constituées d'un Groupe Nominal (GN) suivi d'un Group Verbal (GV) réduit à la forme simple du verbe. La structure syntaxique du GN (N, Det + N, Det + Adj + N) a été systématiquement

variée. Suite à ces opérations, quatre structures syntaxiques différentes ont été obtenues pour les énoncés :

- *Structure A* : Mot isolé, si GN ou GV = $\emptyset$
- *Structure B* : N + V
- *Structure C* : Det + N + V
- *Structure D* : Det + Adj + N + V

Tableau 4 : Corpus des attitudes anglaises, basé sur douze phrases de longueurs variées avec les différentes positions de l'accent lexical marquées par un astérisque.

Syllabes	Utterance	Structure
2	Ma [*] ry	mot isolé
2	Beware [*]	mot isolé
2	Men [*] work [*]	N + V
2	Man [*] counts [*]	N + V
5	Li [*] beralizes	mot isolé
5	Unpre [*] cedented	mot isolé
5	Interroga [*] tion	mot isolé
5	Man [*] commu [*] nicates	N + V
5	A man [*] ma [*] nages	Det + N + V
5	A new [*] man [*] ar [*] gued	Det + Adj + N
5	A new [*] hu [*] man counts [*]	Det + Adj + N
5	Tele [*] phony works [*]	N + V

De plus, l'existence d'un accent lexical en anglais peut perturber la prosodie. Ainsi le nombre et la position de l'accent lexical ont été systématiquement contrôlés et distribués. Finalement, la longueur syllabique de l'énoncé a été systématiquement variée dans le corpus complet. Pour le test de perception, un sous-ensemble de phrases de deux et de cinq syllabes a été sélectionné (cf. tableau 4).

4.3.5. Enregistrement

4.3.5.1. Locuteur

Le locuteur enregistré pour ce corpus est un homme de langue maternelle anglaise, enseignant d'anglais à l'université. Il est spécialiste de la prosodie et des phénomènes intonatifs. Ses travaux de recherche portent sur l'enseignement de la prononciation et de l'intonation de l'anglais.

4.3.5.2. Dispositif et procédure expérimentale

Le corpus a été enregistré sur une cassette DAT (numérisé à 44,1 kHz, 16 bits) dans la chambre sourde du GIPSA-lab, département parole et cognition. De l'extérieur de la chambre sourde, la phrase à produire ainsi que l'attitude étaient affichées sur un écran faisant face au locuteur. Les attitudes ont été réalisées une à une pour l'ensemble des phrases, selon l'ordre suivant :

Sarcastic irony
 Surprise Command / Authority
 Politeness
 Evidence
 Scorn / Aloofness
 Doubt / Incredulity
 Irritation
 Seduction
 Declaration
 Interrogation

Le corpus a été par la suite transféré sur un support informatique grâce au logiciel de traitement du son par « Cool Edit », puis les énoncés produits ont été découpés à l'aide du logiciel « Praat ver.4.1. ».

4.3.6. Validation perceptive

4.3.6.1. Méthodologie expérimentale

Nous rappelons que ce test perceptif est détaillé ici, pour des raisons de clarté de l'exposé, mais qu'il a été mené à bien par M.L. Diaferia (2002): nous renvoyons donc le lecteur à ce travail pour de plus amples détails. Nous avons par contre repris les analyses statistiques afin de présenter les résultats suivant la même méthodologie que pour les autres tests perceptifs.

1.1.1.1. Stimuli

Les douze phrases présentées dans la table 4, produites avec chacune des onze attitudes ont servi de stimuli, soit un nombre total de 132 stimuli qui ont été utilisés pour le test de perception.

4.3.6.2. Matériel

L'interface du test perceptif a été réalisée grâce au logiciel « SuperCard ». Cette interface présente aux sujets de manière permanente la consigne du test perceptif, le choix fermé des 11 attitudes possibles et les définitions des attitudes présentées (cf. figure 8).

4.3.6.3. Procédure

Le déroulement expérimental du test de perception est le suivant. Chaque sujet a passé le test de manière individuelle. En face de l'écran d'un ordinateur, chaque sujet, après lecture des consignes écrites et une explication orale supplémentaire fournie par l'expérimentatrice, doit écouter (à l'aide d'un casque) chaque stimulus et juger de l'attitude qu'il a perçue, parmi les onze attitudes possibles qui lui sont présentées. L'ordre de présentation des stimuli est aléatoire et change pour chaque sujet. Chaque stimulus ne peut être écouté qu'une seule fois. À la fin du test perceptif, les sujets

répondent à un bref interrogatoire à propos des attitudes afin de savoir lesquelles leur semblaient difficiles à distinguer ou s'ils avaient perçu d'autres expressions qui n'étaient pas proposées.

Hello!
 You are going to listen to a series of sentences pronounced with different types of attitudes on the part of the speaker (i.e. the speaker expresses an intention when saying the sentence), by using intonation. Do not pay attention to the *content* or *meaning* of the sentence itself: focalize your attention on the *intonation*. This time, please choose one of the 11 attitudes proposed, which you think best corresponds to the intention of the speaker.

Listen

Attitude :

☐ Sarcastic irony

☐ Surprise

☐ Command / authority

☐ Politeness

☐ Evidence

☐ Scorn / Aloofness

☐ Doubt / Incredulity

☐ Irritation

☐ Seduction

☐ Declaration

☐ Question

Next sound

Sarcastic irony :
 The speaker calls into question (and even contradicts) a statement his interlocutor has said. He wants to show he doesn't agree and even that he's annoyed with what he's hearing, he condemns what is said or what happened, by using an intonation contour which could mean : 'Yeah, sure! I believe you ... (meaning, I don't)'. (ex: 'Yeah, sure! he's a vegetarian...')

Surprise :
 The speaker expresses his surprise at the moment he speaks (ex: 'Oh! he's a vegetarian!')

Command / authority :
 The speaker expresses himself in order to influence other people or to impose his view or wish. He acts with authority and gives orders that must be obeyed. (ex: 'You must absolutely be a vegetarian!!!')

Politeness :
 The speaker says something in a polite way. (ex: 'I hope you don't mind: I am a vegetarian')

Evidence :
 The speaker says something he feels is self-evident. (ex: 'he's a vegetarian, for sure')

Scorn / aloofness :
 The speaker says something in an arrogant, aloof way, showing he gives no credit to what has been said. (ex: 'Poor stupid guy! he is a vegetarian. How pathetic!')

Doubt / Incredulity :
 The speaker expresses a feeling of being uncertain about something or of not believing something; he expresses a suspicion about what has previously been expressed (ex: 'I'm not real sure he's a vegetarian')

Irritation :
 The speaker expresses his irritation towards what has been said; he protests because of this irritation (ex: 'He 's a vegetarian, and it really irritates me!!')

Seduction :
 The speaker says something to seduce his interlocutor (seduction relationship) (ex: 'I am a vegetarian, and I know you can not resist it : be mine!')

Declaration : neutral declarative sentence (ex: 'he's a vegetarian')

Question : neutral interrogative sentence (ex: 'he's a vegetarian?')

Figure 8 : interface du test de perception en anglais pour les sujets britanniques (Diaferia, 2002).

4.3.6.4. Sujets

Dix auditeurs de langue maternelle anglaise britannique (avec un accent londonien) ont participé à la validation de ce corpus (6 sujets masculins et 4 sujets féminins). Leur âge moyen est de 31 ans. Aucun sujet n'a fait état de trouble auditif. La plupart des sujets habitaient en France depuis quelques années lors de la passation du test de perception.

4.3.7. Résultats du test perceptif

4.3.7.1. Distribution des réponses par rapport au hasard

De la même manière que pour l'analyse des résultats du test précédent, le test du chi-deux est utilisé pour vérifier si les distributions des résultats correspondent ou non à une distribution au hasard. La répartition observée (cf. les résultats dans la table 5) est bien significativement différente du hasard avec une probabilité $p. < 0.01$. Comme les sujets ne donne pas leur réponse au hasard, et que chacune des attitudes reçoit un score d'identification supérieur au seuil du hasard (qui est ici de 9%), on peut en inférer que

les sujets ont significativement reconnu chacune des expressions attitudinales. Nous verrons par la suite lesquelles sont les mieux reconnues, et lesquelles montrent le plus de confusions.

Tableau 5 : Matrices de confusion des 10 auditeurs anglais. Les valeurs présentées en gras sont identifiées au dessus de hasard (9%). Signification des étiquettes : DC(declaration), IN(interrogation), SU(surprise), DO(doubt-incredulity), EV(evidence), CO(command-authority), IR(irritation), PO(politeness), AL(scorn-aloofness), SE(seduction) et SI(sarcastic irony).

Attitudes perçues	Attitudes présentées										
	CO	DC	DO	EV	IN	IR	PO	SI	SC	SE	SU
Command Authority	53%	3%	0%	3%	0%	8%	2%	12%	1%	3%	2%
Declaration	15%	57%	0%	30%	3%	31%	16%	15%	7%	5%	0%
Doubt	0%	0%	43%	2%	20%	2%	16%	2%	1%	2%	27%
Evidence	12%	13%	2%	42%	1%	33%	9%	15%	7%	2%	5%
Interrogation	0%	0%	20%	1%	48%	0%	13%	1%	0%	1%	7%
Irritation	14%	5%	0%	7%	2%	17%	8%	19%	10%	1%	1%
Politeness	0%	16%	2%	13%	6%	2%	24%	3%	3%	3%	3%
Sarcastic Irony	2%	3%	12%	0%	4%	8%	7%	16%	25%	8%	5%
Scorn Aloofness	4%	1%	3%	1%	1%	2%	3%	15%	43%	3%	3%
Seduction	0%	3%	0%	0%	2%	0%	0%	3%	3%	73%	0%
Surprise	0%	0%	19%	2%	14%	0%	3%	0%	1%	0%	48%

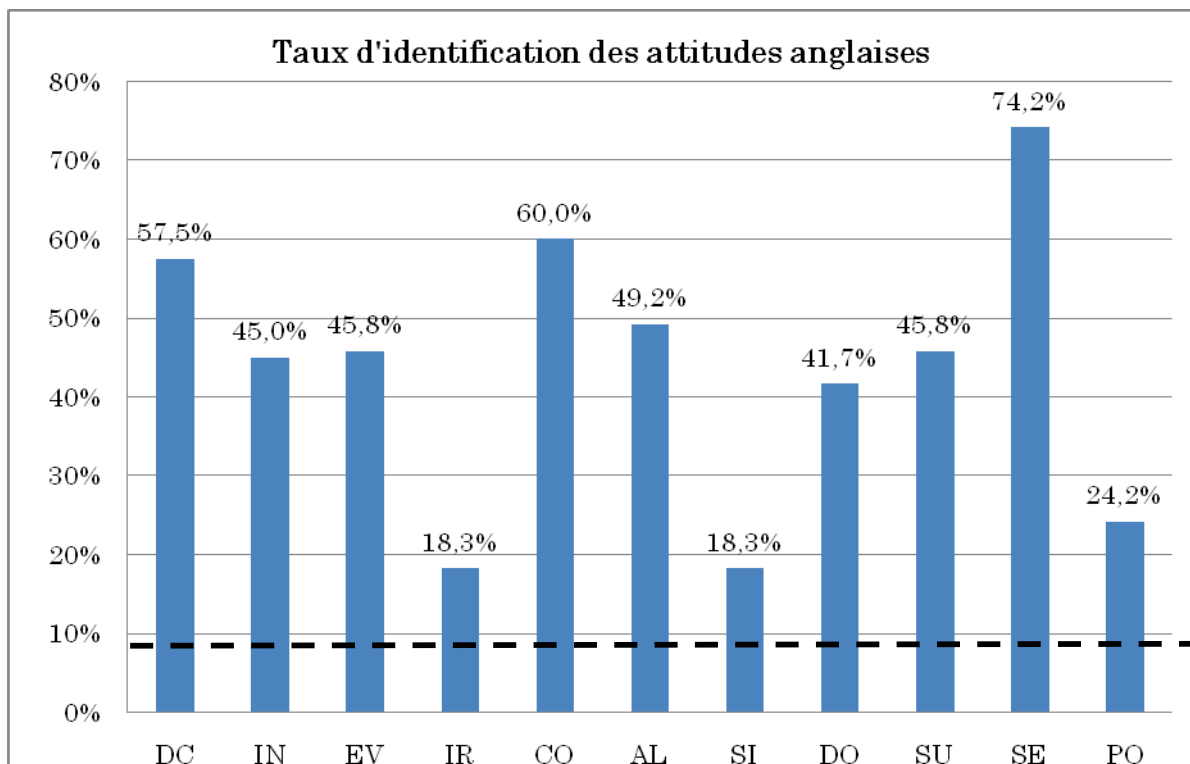


Figure 9 : pourcentage d'identification de chaque attitude par les 10 auditeurs anglais. La ligne pointillée indique le seuil du hasard (9%) (Diaferia, 2002)

4.3.7.2. Taux d'identification des attitudes

La figure 9 présente le pourcentage de reconnaissance de chacune des attitudes, d'après les travaux de M.L. Diaferia (2002). La ligne pointillée horizontale indique le seuil du hasard. Cette figure indique que les auditeurs anglophones britanniques ont reconnu chaque attitude au-dessus du seuil du hasard, ce qui permet de valider le choix et la réalisation des expressions attitudinales par ce locuteur, pour des auditeurs anglophones. Cependant, la large variabilité des scores de reconnaissance obtenus par les différentes attitudes nous pousse détailler cette analyse.

Certaines expressions obtiennent des pourcentages de reconnaissance élevés : *declaration*, *interrogation*, *evidence*, *command-authority*, *scorn-aloofness*, *doubt-incredulity*, *surprise* et *seduction*. L'attitude de *seduction* est particulièrement bien reconnue.

En revanche, *irritation*, *sarcastic irony* et *politeness* sont les attitudes les moins bien reconnues par les auditeurs anglais même si ces attitudes ont reçu un taux d'identification supérieur au seuil du hasard.

Diaferia (2002) donne comme explication possible à cette faible reconnaissance le fait que certaines attitudes telles qu'*irritation* ou *sarcastic irony* ont pu être produites par le locuteur avec une intensité moindre, du fait de l'effet de liste provoqué lors de l'enregistrement. Elle relève également l'influence du contenu sémantique des phrases utilisées. Certains sujets se sont dits gênés par le décalage sémantique entre le mot « *beware* » et certaines des attitudes exprimées.

Nous allons maintenant procéder à des analyses statistiques supplémentaires, selon la même procédure que celle utilisée pour les attitudes françaises (par analyse des correspondances et classification hiérarchique – cf. supra), et qui sera aussi appliquée dans toute la suite de notre travail.

4.3.7.3. Analyse des résultats grâce à l'A.C.

Selon l'AC menée sur les résultats, les quatre premières dimensions de l'analyse expliquent 85% de la variance. La figure 10 présente les distributions des expressions attitudinales et des catégories d'attitudes telles qu'elles sont projetées sur les première et deuxième dimensions (en haut) et les troisième et quatrième dimensions (en bas).

La plupart des expressions attitudinales sont proches de la catégorie correspondante, ce qui indique une bonne reconnaissance des attitudes par les locuteurs, sauf pour l'expression de *sarcastic irony* qui semble la moins bien reconnue. Cette attitude est majoritairement reconnue comme une expression de *politeness* (politesse) ou de *scorn-aloofness* (mépris).

Les plus importantes confusions inter-attitudes sont observées entre les expressions d'*interrogation*, de *surprise* et de *doubt-incredulity*, toutes ces expressions

exprimant une dimension de questionnement. On peut relever aussi que la *seduction* est effectivement l'attitude la mieux différenciée de toutes les autres expressions.

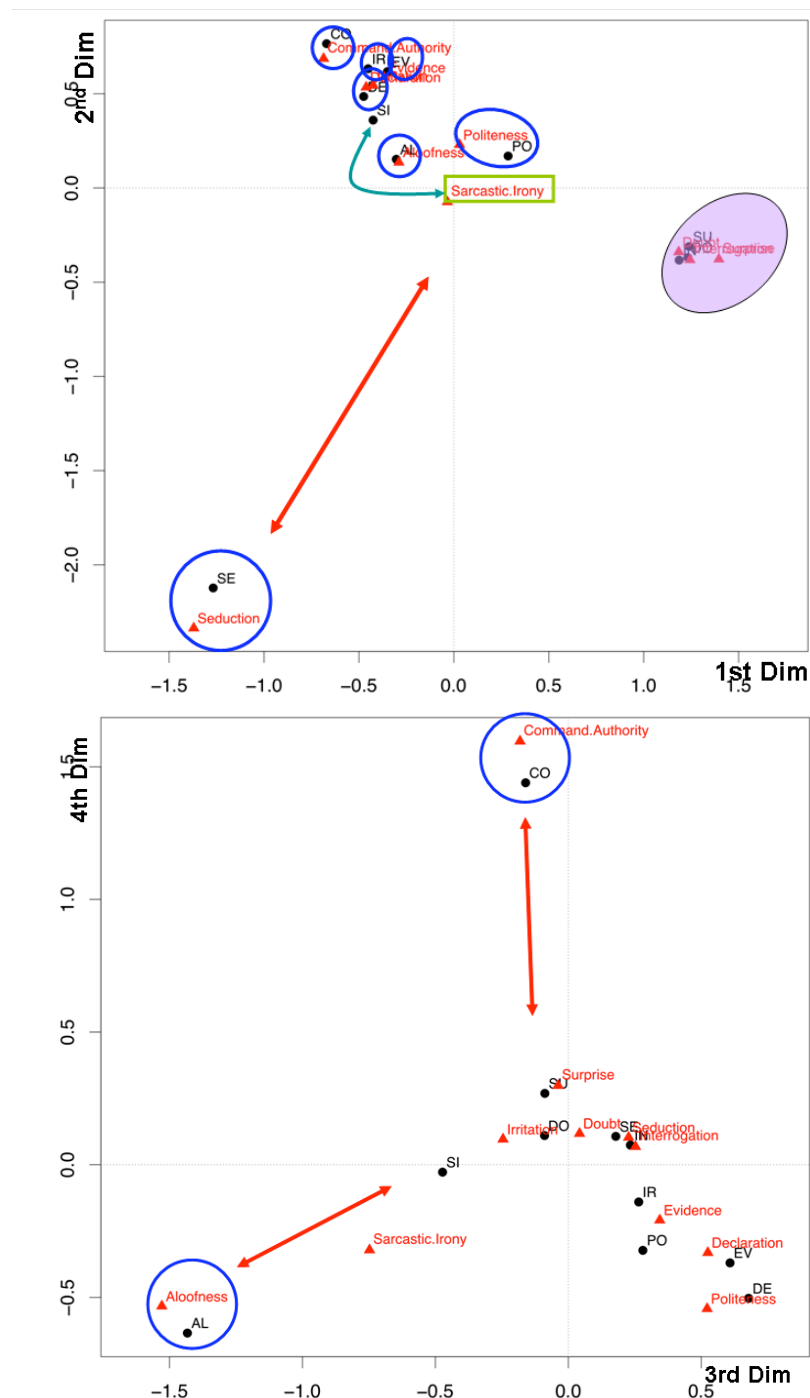


Figure 10 : distribution des résultats des 10 auditeurs anglais pour les 11 attitudes anglaises par l'AC (figure du haut : dimensions 1 et 2 ; figure du bas : dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitude. Les cercles bleus indiquent les attitudes les mieux reconnues ; les disques violets soulignent les principales confusions inter-attitudes. Les autres couleurs indiquent les principales confusions.

4.3.7.4. Classification hiérarchique

L'analyse des résultats (cf. figure 11) grâce à la procédure de classification hiérarchique permet de mettre en évidence les regroupements suivants. Au plus haut niveau, deux grands groupes d'attitudes sont distingués, en plus de l'attitude de *seduction*, reconnue sans confusions avec les autres expressions.

- *declaration, command, sarcastic irony, aloofness, evidence et irritation*
- *doubt-incredulity, surprise, interrogation et politeness*

On retrouve entre ces deux groupes une distinction, déjà relevée pour les expressions françaises, entre des expressions fondées sur le degrés d'imposition de son opinion par le locuteur : ce degrés est élevé dans le premier groupe, tandis qu'il est bas dans le second groupe.

Dans le premier groupe, les expressions de *sarcastic irony* et de *scorn-aloofness* montrent beaucoup de confusions entre elles et avec l'expression de *command-authority*, qui est quant à elle bien reconnue ; et l'expression d'*irritation* est confondue avec la *declaration* et l'*evidence*, elle même confondue avec la *declaration*.

Pour le second groupe, les expressions de *doubt-incredulity*, de *surprise* et d'*interrogation* entretiennent de nombreuses confusions entre elles, tandis que la *politeness* est confondue, mais à un degrés un peu moindre avec ces expressions.

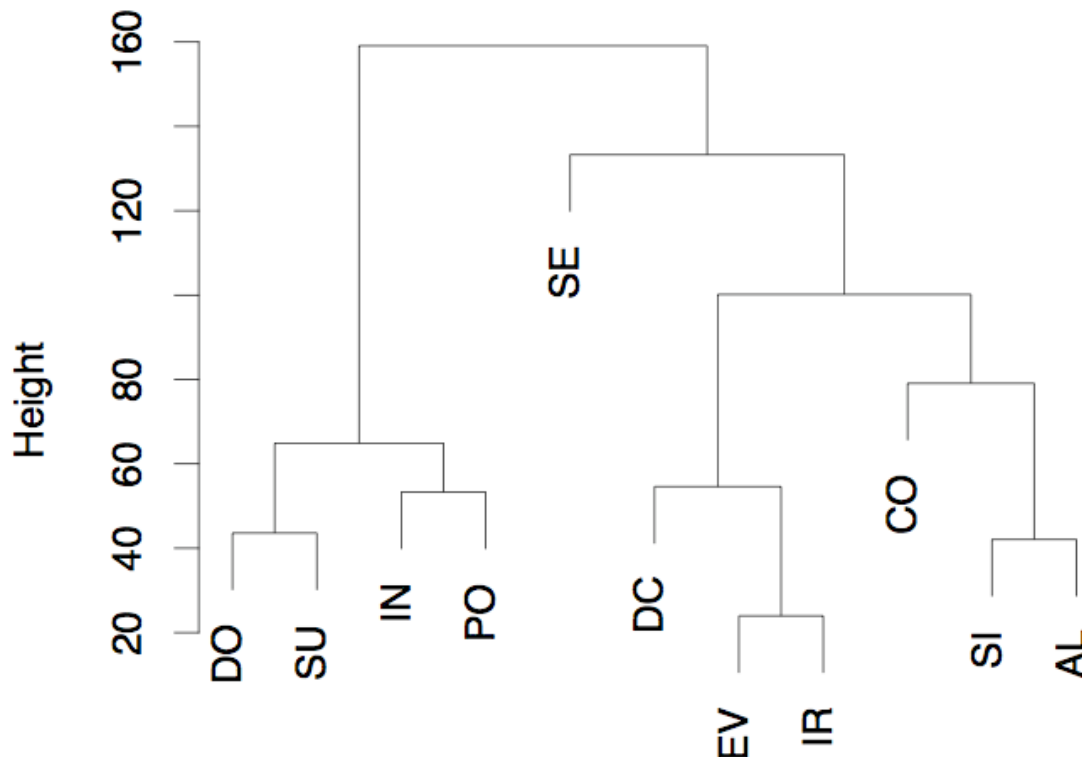


Figure 11 : dendrogramme résultant de la classification hiérarchique des expressions attitudinales du corpus anglais, obtenue en utilisant la distance de Ward pour calculer les distances interclasses.

4.3.8. Conclusion

Nous avons récapitulé les travaux de Diaferia (2002) en ajoutant les analyses supplémentaires afin d'évaluer des comportements perceptifs des auditeurs anglais pour 11 attitudes anglaises britanniques.

En évaluant ce corpus des 11 attitudes anglaises britanniques, Diaferia relève les points suivants. Tout d'abord, la distribution des réponses par les sujets est différente d'une distribution au hasard, selon un test khi-deux. Ensuite, aucun effet de la longueur de l'énoncé n'est enregistré sur ces stimuli.

Le taux de reconnaissance de chaque attitude indique une bonne compréhension des attitudes par les sujets, ce qui implique une validation des expressions choisies et de leur réalisation par le locuteur. Par ailleurs, les confusions réalisées entre les attitudes sont cohérentes vis-à-vis de la sémantique de chaque ensemble d'expressions : on a vu que le critère du degré d'imposition permettait de séparer les expressions en deux groupes, si l'on excepte la *seduction*, qui ne montre pas de confusion particulière.

5. Étude des attitudes japonaises

Après avoir validé et étudié les attitudes françaises et anglaises, nous proposons au sein de ce chapitre d'aborder les attitudes japonaises. Pour commencer, nous en ferons le repérage à partir de deux types de littératures : des données provenant de domaines variés tels que la linguistique, la sociolinguistique et la phonétique et des données provenant de la didactique du japonais langue étrangère. Après la sélection de 12 attitudes représentatives (y compris des attitudes spécifiques à la culture japonaise), puis leur validation auprès de natifs (Japonais), nous orienterons nos recherches en vue de comprendre le rôle et l'influence des deux modalités communicationnelles que sont l'audition et la vision, dans la perception des attitudes. Parallèlement, nous avons effectué un test de perception suivant le paradigme du gating, ou du dévoilement progressif, au niveau des contours prosodiques de phrase afin de tester l'« anticipation » des auditeurs et afin surtout d'identifier à quel moment (point d'unicité) ils sont capables de reconnaître une attitude. Les résultats obtenus sont comparés avec les valeurs acoustiques extraites du corpus validé.

5.1. Introduction

Les fonctions prosodiques du japonais n'ont été étudiées que tardivement. C'est à partir des années 1980 que de nombreux chercheurs (Sugito, 1997 ; Kori, 1997 ; Fujisaki, 1989 etc.) commencent à s'y intéresser. La plupart des travaux effectués sur la prosodie concernent surtout la fonction linguistique (par exemple, segmentation, hiérarchisation, focalisation etc.). En ce qui concerne les études sur la fonction attitudinale de la prosodie, ce champ est longtemps resté plus ou moins abandonné (Maekawa, 2004). Cette fonction prosodique est souvent traitée comme une sous-catégorie de la modalité ou dans le cadre socioculturel. Pour le travail présent, nous proposons préalablement de faire la description de la structure phonétique du japonais (c'est à dire présenter la notion de *more* et le système accentuel du japonais). Nous compléterons cette description par la présentation de travaux effectués sur certaines expressions attitudinales du japonais traitées initialement dans la méthode d'enseignement du Japonais Langue Etrangère par Mizutani & Mizutani (1979).

5.2. Description de la structure phonologique du japonais.

5.2.1. Présentation générale

La langue japonaise est la langue officielle de environ 120 millions d'habitants dans l'archipel japonais. Cette langue japonaise est, en effet, constituée de divers dialectes. Lorsqu'on se focalise sur le dialecte de Tokyo, cette langue est souvent décrite comme une langue tonale (Abe, 1998) qui est caractérisée par l'intonation qui porte le rôle distinctif au niveau morphologique (un mot portant la même structure syllabique, mais prononcé différemment renvoie un sens différent). Cependant, le trait distinctif du ton n'est pas réalisé à l'intérieur d'une syllabe comme la langue chinoise.

5.2.2. La structure phonologique du dialecte de Tokyo

5.2.2.1. La notion de « more »

La « more » est l'unité phonologique minimale du dialecte de Tokyo. La structure d'une more est [(C)V]. La more possède les caractéristiques suivantes, qui le distinguent de la syllabe : une consonne nasale en position de coda dans la structure syllabique forme une more indépendante. Par exemple, le mot « San » [monsieur] comprend une syllabe mais deux mores. De plus, la dernière moitié d'une voyelle longue ou d'une diphtongue, de même que la première consonne dans une structure CCV de deux mêmes consonnes sont aussi perçues comme une more supplémentaire. Dans cette perspective, l'unité moraïque est vue comme unité minimale pour créer la rythmicité du dialecte de Tokyo (Kubozono 1997).

5.2.2.2. Le système accentuel de Tokyo

Sur le plan du système (ou de la structure) accentuel, on peut les regrouper selon cinq types différents : « accent de Tokyo », « accent variant de Tokyo », « accent du Kansai », « accent variant du Kansai » et « sans accent lexical ».

L'accent de Tokyo, le plus répandu, ne désigne pas les formes accentuelles utilisées uniquement chez les habitants de la ville de Tokyo. Il s'agit ici du système accentuel disséminé dans la province de « Chugoku » (Ouest du Japon), de « Chubu » (centre) et la moitié de l'Ouest du « Kanto » (région de Tokyo) pour des raisons historiques.

La structure accentuelle du dialecte de Tokyo est dite « l'accent libre » (Shibata 1961). Autrement-dit, l'accent peut se situer sur n'importe quelle more du mot.

Cet accent lexical est réalisé par la disposition relative de deux tons (haut et bas) perceptibles phonologiquement pour chaque mot lexical (Kindaichi 1956). Par exemple, le mot « hashi » ne correspond pas à la même entrée lexicale selon que « ha » est réalisé

avec un ton haut et « shi » avec un ton bas (dans ce cas il signifie « une baguette ») ou que « ha » est réalisé avec un ton bas et « shi » avec un ton haut (« un pont »).

Par ailleurs, selon Hattori (1955), pour la description phonologique de l'accent de Tokyo, il suffit d'identifier le « nucleus accentuel », porteur du ton « haut », qui entraîne l'abaissement du ton suivant. Ce « nucleus accentuel » est donc la dernière more avec un ton haut du mot.

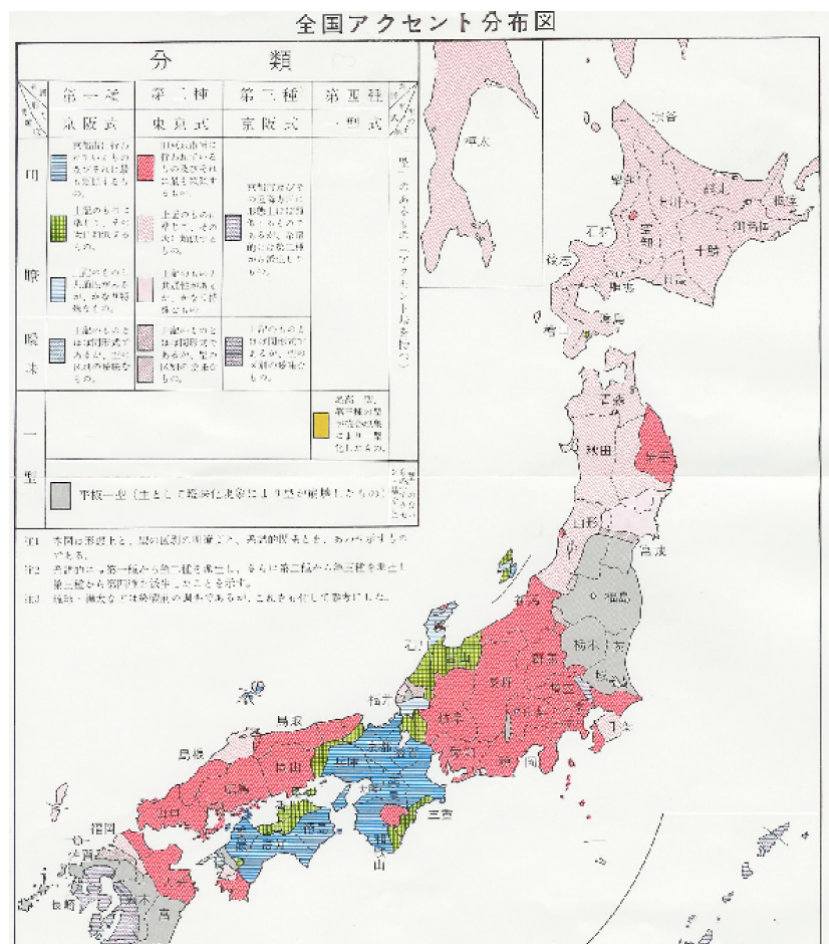


Figure 12 : La distribution géographique des 4 types principaux de formes accentuelles (Hirata 1960). La couleur qui encadre la région signifie : rouge ; la forme accentuelle de Tokyo, bleue ; la forme accentuelle de Kansai, vert ; forme accentuelle variante de Kansai et gris ; l'absence totale d'accent lexical.

L'accent de Tokyo se caractérise par les règles suivantes ;

L'accent est une combinaison de tons : haut et bas.

Les tons haut et bas ne peuvent être intervertis à l'intérieur d'une seule more.

Lorsque le ton haut n'est pas aligné sur la première more du mot, cette première more porte toujours le ton bas, toujours suivi du ton haut.

Un mot donné, ne peut porter qu'un seul accent. En d'autre termes, le passage du ton haut au ton bas ne peut avoir lieu qu'une fois dans un mot.

Une fois que le ton haut fait place au ton bas, ce ton bas ne remontera jamais dans l'intérieur du mot.

Toutes les mores spécifiques⁸, ainsi que les mores dans lesquelles la voyelle est dévoisée ne peuvent pas être porteuses de l'accent.

Un mot peut recevoir un nombre $N+1$ (où N est le nombre de mores dans le mot) de formes accentuelles distinctes. Par exemple, un mot constitué de deux mores peut avoir trois types de formes accentuelles différentes (1^{ère} more accentuée, 2^{ème} more accentuée, pas d'accent).

Parmi ces règles, la règle 3 ne nous semble pas être une caractéristique de l'accent lexical. En effet, lorsqu'on étudie un syntagme ou une phrase, l'occurrence de la montée du ton bas sur la première more au ton haut sur la deuxième more n'apparaît pas systématiquement. Cette montée du ton apparaît plutôt en début de syntagme afin de marquer une frontière syntagmatique (Uwano 1989)⁹. Ce phénomène caractérise donc plutôt un fait prosodique que lexical.

Au niveau phonétique, Fujisaki & Sudo (1971) ont travaillé sur la réalisation de l'accent japonais produite par six locuteurs japonais homme et femme (Figure 13). La réalisation phonétique de l'accent japonais est similaire pour tous les locuteurs étudiés : le contour de F_0 monte du début jusqu'à la more accentuée où la valeur de F_0 est maximale, puis le contour baisse progressivement jusqu'à la fin de l'énoncé. Il est à noter que le pic de F_0 n'est pas toujours situé à la position finale de la more accentuée, même si la transition du ton haut vers le ton bas est perçue à cette position.

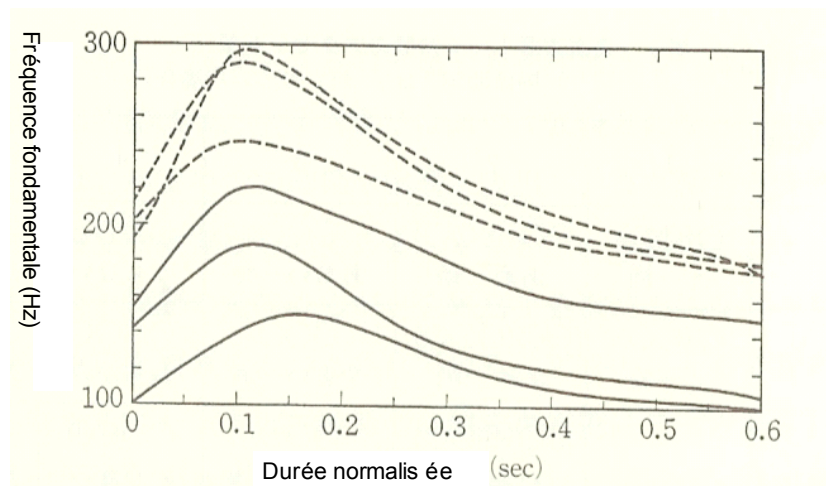


Figure 13 : Réalisation phonétique de l'accent japonais avec la phrase /amearare/ par six locuteurs japonais. L'accent lexical est localisé en 1^{ère} more. Le contour de F_0 en ligne

⁸ Une consonne nasale en position de coda, la dernière moitié d'une voyelle longue et d'une diphtongue et la double consonne sont considérés comme des mores spécifiques.

⁹ Ex. « kono akai sakana » [ce poisson rouge] se compose de trois mots lexicaux, cependant la montée du ton apparaît seulement en début de phrase (entre ko et no) sauf dans le cas où « akai » ou « sakana » sont focalisés.

pointillée indique la voix féminine, le contour en noir signifie la voix masculine. (Fujisaki & Sudo, 1971)

En prenant compte de ces aspects lexicaux et linguistiques, voyons maintenant les travaux effectués sur les expressions attitudinales du japonais.

5.3. Les expressions attitudinales du japonais

5.3.1. Six « attitudes » dans l’enseignement du japonais langue étrangère.

Initialement la description impressionniste des attitudes japonaises fut traitée dans la didactique du Japonais Langue étrangère bien avant que les phonéticiens et les linguistes s’y intéressent comme objet de recherche car les enseignants du japonais savaient que l’enseignement du lexique (incluant l’enseignement de l’accent lexical), et de la syntaxe n’était pas suffisant pour pouvoir communiquer dans le bain linguistique. L’importance a été donnée à l’enseignement des attitudes dans le cadre d’actes de parole pour des raisons pragmatiques, c’est-à-dire, pour que les apprenants puissent parler en langue cible avec l’expressivité appropriée.

Dans leur manuel d’enseignement du japonais, Mizutani & Mizutani (1979) présentent six « attitudes »¹⁰. Il s’agit d’« admiration », « déception », « suspicion », « indifférence », « neutre » (ou déclaration), « focalisé ». Bien qu’il n’y ait ni description phonétique de chacune des attitudes ni définition claire pour ces six expressions, cette méthode qui exige des apprenants de pouvoir les discriminer à partir d’une perception audio était très innovante dans le monde de la didactique de l’époque.

5.3.2. Caractéristiques phonétiques des six attitudes

Ces six attitudes¹¹ ont été reprises comme véritables objets d’analyse afin d’être mesurées phonétiquement dans les travaux de Maekawa (1998, 2004).

Pour ces travaux, Maekawa a demandé à trois enseignants natifs (deux hommes et une femme) de langue japonaise parlant le dialecte de Tokyo d’éliciter trois phrases n’ayant aucune connotation particulière selon les six « attitudes » précitées, en chambre sourde. Les locuteurs ont répété 10 fois chaque énoncé selon un ordre aléatoire.

¹⁰ Dans le manuel, les auteurs nomment les phrases portant des valeurs communicatives comme des phrases « émotives ».

¹¹ Il emploie le terme « information paralinguistique » emprunté à Fujisaki (2004), défini ainsi: “paralinguistic information is not inferable from the written counterpart and is deliberately added by the speaker. Second, the strength of a given paralinguistic information type can vary continuously within one and the same category” (Maekawa 1998 p.635). Cette définition ressemble à celle du terme “attitude”, nous employons donc « attitude » à la place de « information paralinguistique » afin de faciliter la lecture.

Un test perceptif a ensuite été effectué auprès de 17 japonais avec tous les énoncés enregistrés. Les auditeurs avaient pour tâche d'identifier quelle était l'attitude exprimée pour chaque énoncé, parmi les six attitudes possibles.

Selon les résultats obtenus, toutes les « attitudes » sauf « focalisé » ont obtenu un taux d'identification de plus de 80% en moyenne. Parallèlement, l'analyse acoustique a montré que la durée d'« admiration », « déception » et « suspicion » est plus longue que « indifférence », « neutre » et « focalisé ». On observe aussi une grande variation de la durée entre la première et la dernière more dans tous les énoncés et chez tous les locuteurs.

En ce qui concerne le mouvement de F0, K. Maekawa a identifié une large amplitude de F0 pour « admiration », « suspicion » et « focalisé », mais celle de « déception » reste plate. Quant à l'attaque de F0, « admiration » et « suspicion » montrent une attaque très basse tandis l'énoncé « neutre » a une montée de F0 nettement plus tardive.

De même, l'apparition temporelle du pic de F0 (c-à-d, F0 max.) est significativement en retard pour « admiration », « déception » et « suspicion » par rapport à « neutre » ou « focalisé » ($p < .001$). L'attitude « d'admiration » est surtout caractérisée par une longue durée de la dernière more.

Enfin K. Maekawa compare le timbre de la dernière voyelle de chacune des attitudes : « admiration » et « déception » montrent des valeurs de F1 et F2 plus basses comparées à « suspicion » et « indifférence ». K. Maekawa identifie notamment une différence systématique du deuxième formant pour toutes voyelles et toutes consonnes entre « suspicion » et « admiration ». La valeur du deuxième formant est significativement plus élevée pour « suspicion » plutôt que pour « admiration ». C'est-à-dire que la position de la langue est localisée à l'avant du conduit vocal pour la « suspicion », et plutôt en arrière pour « admiration ». Cette différence de gestualité articulatoire est confirmée par le système EMA (Maekawa & Kagomiya, 2000). On observe également la saturation du signal (changement de la qualité de voix) pour la première more de « suspicion », occasionnellement pour celui d'« admiration » et « déception ».

A la suite de K. Maekawa, Kori (1997) s'intéresse à la relation entre le contour de f0 en fin de phrase et les attitudes exprimées. Il décrit ainsi une typologie fonctionnelle des attitudes japonaises par l'indice du contour final de f0.

5.3.3. L'approche fonctionnelle à partir du contour final

S. Kori est l'un des premiers à avoir décrit les formes et les fonctions de l'intonation japonaise. Dans son ouvrage « Japanese intonation – forms and fonctions » (1997), sont présentées trois fonctions différentes de la prosodie : la fonction linguistique, la fonction émotive et la fonction socioculturelle.

Selon lui, la « modalité »¹² ou attitude est considérée comme une des fonctions linguistiques de l'intonation. Pour la description des différents types de « modalités », Kori se focalise sur le mouvement du contour prosodique en fin de phrase. En japonais, l'attitude prosodique du locuteur est souvent exprimée par des particules finales (ka / ne / yo / yone / wa) qui, sont des marqueurs d'informations individuelles telles que le sexe du locuteur ou/et diverses « modalités ». Cependant, Kori note bien que les attitudes exprimées uniquement par la prosodie sans influence des particules finales existent en japonais. Il considère que ces attitudes sont facilement identifiables perceptivement par les caractéristiques prosodiques de fin de phrase. C'est ainsi qu'il présente 5 types de traits intonatifs différents en fin de phrase : contour montant, contour montant emphatique, contour descendant, contour montant-descendant et contour plat.

- Le contour « montant » est utilisé principalement pour la **question simple**. Lorsque ce contour est utilisé pour l'**exclamation de surprise**, la montée de F0 est plus large et la durée de la dernière syllabe est plus longue.
- Le contour « montant emphatique » en fin de phrase est généralement utilisé pour l'**insistance** ou l'**affirmation** d'un droit.
- Le contour « descendant » est utilisé principalement pour l'étonnement face à des informations inattendues. La **déception** ou l'« **amaé** »¹³ peuvent être exprimés également par ce contour.
- Le contour « montant-descendant » en fin de phrase sous-entend une incitation.
- Le contour plat ne porte pas d'attitude particulière.

En ce qui concerne l'aspect socioculturel de la prosodie, Kori considère que la prosodie joue un rôle important dans la politesse. Sur ce point, il montre un exemple. Il s'agit d'expressions conventionnelles (par exemple, les salutations, les remerciements, les excuses etc.) liées directement à un contour de F0 prototypique. Ce contour est ainsi figé. Toutefois, ces mouvements de F0 conventionnels peuvent changer d'une région à une autre, selon la culture locale. C'est pourquoi, si un sujet venant de Tokyo emprunte un contour global de F0 sur la phrase conventionnelle dans la région d'Osaka, sa parole risque d'être perçue comme manquant de sincérité ou comme impolie.

En outre, Kori (1989) explique brièvement la contribution de différentes qualités de voix, en particulier la voix pressée. Selon Kori, la voix pressée peut être associée aux émotions (ou attitudes) d'**exclamation**, **surprise**, **dégoût**, **irritation** et **colère** » (Kori 1989 p.331 traduit par moi-même).

Nous avons suivi jusqu'ici la description des attitudes selon les trois fonctions prosodiques proposées par S. Kori. Il a envisagé que le mouvement de F0 en fin d'énoncé ainsi que la qualité de voix jouent un rôle important. Cependant cette description est

¹² Kori (1997) définit la modalité comme « l'attitude mentale intentionnée du locuteur. Dans cette perspective, la fonction attitudinale est intégrée dans les deux fonctions (linguistique et socioculturelle).

¹³ Il faut noter que l'« amae » est une attitude particulière, utilisée pleinement chez les Japonais. T. Doi, psychologue, définit dans son ouvrage intitulé de « Le jeu d'indulgence », que cette attitude confère un sentiment de sécurité identique à celui qu'éprouve l'enfant dans le giron de sa mère (Doi, 1976).

limitée. Par conséquent, il reste à savoir s'il existe des effets prosodiques à l'intérieur de l'énoncé pour déterminer chaque attitude.

5.3.4. Le mouvement du contour intonatif de l'interrogation japonaise

D'un point de vue perceptif, la montée du contour intonatif en fin de phrase prend en charge la fonction de « question simple¹⁴ » en japonais. Étonnamment, la description du contour global de F0 pour plusieurs types d'interrogations, plus objective et mesurable, n'a attiré l'attention des chercheurs que récemment (Ayusawa 1992). En fait, les linguistes se sont focalisés sur le contour intonatif montant en fin de phrase lié directement à la fonction linguistique de l'interrogation. Par conséquent, ils ont plus ou moins négligé les variations du contour global de F0. A ce propos, examinons plus en détails les travaux de deux phonéticiens japonais, T. Ayusawa (1992, 2001) et M. Sugito (1990) qui ont envisagé la description des phrases interrogatives d'un point de vue phonétique / phonologique.

5.3.5. Le rôle de la fréquence fondamentale pour 4 types d'interrogations

T. Ayusawa (1992) a interrogé 15 Japonais (six hommes et neuf femmes) parlant le dialecte de Tokyo et qui ne sont pas des acteurs. L'expérimentation a été effectuée dans une chambre sourde. Les sujets avaient pour tâche d'oraliser un dialogue dans lequel quatre types de phrases interrogatives (question simple, question « Tag », question à choix multiples et question avec un mot interrogatif) étaient présentées. L'objectif de cette recherche était de déterminer les caractéristiques acoustiques des quatre types de phrases interrogatives.

Les résultats montrent qu'en fin de phrase, la courbe et le point de départ de la montée de F0 varient selon les sujets. Cependant, aucune différence significative entre les quatre types de phrases interrogatives n'a été identifiée en fin de phrase. De plus, le pattern du mouvement de F0 en fin de phrase n'est pas corrélé avec la longueur de l'ensemble de la phrase.

En ce qui concerne la durée de la dernière more, elle peut être presque deux fois plus importante dans certains cas. En fait, lorsque la dernière more de la phrase interrogative porte la chute de F0 due à l'accent ainsi que la montée de F0 due à l'interrogation, la durée de cette more passe presque du simple au double. La chute de F0 a été identifiée dans la première moitié de cette more allongée, alors que l'autre moitié présente une montée de F0.

Quant à la phrase interrogative du type « Tag », l'intérieur de la phrase est marqué par une F0 plus élevée. C'est-à-dire que, lorsqu'une more au milieu du mot porte l'accent, on observe une F0 élevée sur cette more accentuée, mais cela n'a pas d'influence

¹⁴ Ce terme est utilisé comme équivalent de « Yes-No question » en anglais.

significative sur la réalisation de la montée de F0 en fin de phrase. Cette amplitude de la F0 pour la question « Tag » est mentionnée également dans les travaux de Sugito (1990). Par ailleurs, ce phénomène est réservé à la question « Tag » et aucune valeur significative du mouvement de F0 à l'intérieur de la phrase dans les trois autres types de phrases interrogatives n'a été identifiée.

5.3.6. Etude comparative de la phrase interrogative (anglais vs. japonais)

Sugito (1990) a travaillé sur le contour interrogatif de F0 en deux langues cibles (l'anglais américain et le japonais). L'objectif de cette recherche était d'étudier les caractéristiques acoustiques d'une phrase véhiculée par des modalités différentes qui peuvent être réalisées différemment en japonais et en anglais américain. Pour cette expérimentation, une Japonaise parlant le dialecte de Tokyo et deux Américaines (Tennessee, Nebraska) qui parlent couramment en japonais ont été choisies comme sujets. Ces trois sujets ne sont pas des acteurs. L'expérimentation a été effectuée par la lecture oralisée d'une phrase sémantiquement neutre « *Mie wa Nara wo aruku* » [Mie se promène à Nara] pour le sujet japonais et les sujets américains, et la phrase « *Are you going to America* » pour les sujets américains seuls. Les sujets japonais et américains ont oralisé ces énoncés avec trois modes différents (affirmatif, question « Tag » et question simple) dans une chambre sourde.

Le résultat de cette expérimentation montre que les sujets japonais produisent une plus grande amplitude de F0 dans la question « Tag » par rapport à la question simple, aussi bien à l'intérieur de la phrase qu'en fin de phrase. Cela est significatif de l'absence de « catathesis¹⁵ » dans la phrase. Quant aux sujets américains, elles montrent, en anglais, une grande amplitude de F0 dans la question « Tag » par rapport à la question simple. Le contour de question simple en japonais produite par les sujets américains est relativement plat du début jusqu'à la moitié de la phrase par rapport à celui des natifs qui montrent plus de relief dû aux accents. Une fois que F0 baisse brutalement, il monte en fin de phrase.

Par la suite, Sugito a modulé les valeurs de F0 à la fin des phrases en japonais produite par les deux Américaines. Pour la manipulation, elle a réduit la valeur de F0 jusqu'à la moitié de l'original sans moduler la durée et l'intensité. Ces phrases synthétisées ont été présentées à des auditeurs japonais natifs, et ces auditeurs les ont perçues comme des questions simples, naturelles malgré le contour plat du début de phrase. Cela implique que l'intonation en fin de phrase joue un rôle important pour la modalité interrogative.

¹⁵ C'est un processus phonologique qui réduit les valeurs de tons après chaque accent. Autrement-dit, cette règle implique la compression de l'écart de F0 sur chaque accent. (J. B. Pierrehumbert & M. E. Beckman 1988)

5.3.7. Discussion

Les linguistes japonais ont voulu, depuis longtemps, lier l'interrogation uniquement au phénomène de montée de F0 en fin de phrase. Ayusawa (1992) confirme que, quel que soit les différents types de l'interrogation, le contour de F0 monte systématiquement en fin de phrase. Cependant, Ayusawa (1992) ainsi que Sugito (1990) ont découvert une absence de « catathesis » (c-à-d, l'absence de la compression de la valeur du ton après chaque accent) pour la question du type « Tag ». Ayusawa a également envisagé que d'autres fonctions phonologiques telles que la focalisation ou les accentuations peuvent influencer le contour global de F0 et la durée de la dernière more dans la phrase interrogative. Sugito (1990), quant à elle, confirme également que le contour montant de F0 en fin de phrase est l'indicateur déterminant de l'interrogation japonaise. Pour justifier cette importance du contour de F0 en fin de phrase, elle a utilisé une phrase en japonais produite par deux Américaines. Elle a modulé uniquement la fin de ce contour de F0 pour que le contour final de F0 ressemble à celui du japonais. Après l'expérimentation perceptive, on a découvert que même si le contour de F0 tout au long de la phrase ne semble pas être celui du japonais natif, les auditeurs japonais considèrent que cette phrase leur semble naturelle.

5.3.8. La politesse japonaise

5.3.8.1. Présentation générale

Parmi les travaux effectués sur les expressions attitudinales, nous portons une attention particulière aux recherches qui traitent du lien entre la politesse et la prosodie en japonais. En effet la politesse joue un rôle important dans la société japonaise. Outre le contenu sémantique qui détermine le degré de politesse, la prosodie attitudinale sous-tend aussi le degré de politesse selon l'âge, le statut social, le sexe et la situation contextuelle du locuteur lui-même ainsi que ceux de l'interlocuteur.

5.3.8.2. « Code de fréquence » (J. J. Ohala)

J. J. Ohala (1984, 1996) s'intéresse aux similarités interculturelles d'utilisation de F0 dans la perspective éthologique.

Une des hypothèses suggérées par J. J. Ohala (1984) est qu'une moyenne de F0 élevée est liée avec la subordination sociale, la déférence et l'absence de menace alors qu'une F0 plus basse correspond, elle, à de la menace, de l'agression, de la dominance etc... Ohala considère que ce « code des fréquences » est inné¹⁶ et suppose alors que cet usage est partagé par toutes les langues du monde.

Sans aborder l'aspect universel de ce « code de fréquence », si on suppose, au moins, qu'il est applicable à la langue japonaise, les japonais venant de toutes les régions

¹⁶ Ohala définit l'innéisme comme "having a genetic predisposition which, however, may require extensive postnatal stimulation for its full development and implementation" (Ohala 1984 p.10).

et de n'importe quel âge devraient associer l'attitude polie (c-à-d, la subordination sociale, la déférence et l'absence de menace) à une F0 élevée. En revanche, une attitude plutôt impolie (c-à-d, la menace, l'agression, la dominance) sera exprimée par une F0 basse.

5.3.8.3. Les attitudes de courtoisie et l'impolitesse/l'arrogance

Il s'agit maintenant de savoir si le « code des fréquences » est applicable en japonais. Reportons-nous aux travaux de deux phonéticiens, T. Kawano (1993) et M. Ko (1993) qui ont traité les attitudes « Teinei » (polie) et « Zondai » (impolie ou arrogant).

Kawano (1993) s'est concentré sur les caractéristiques acoustiques (F0 et durée) de la phrase « polie ». Il a enregistré quinze sujets japonais de différentes régions (cinq femmes de Tokyo, Osaka et Nagoya : un public tout-venant c'est-à-dire que ce ne sont pas des acteurs) en chambre sourde. Trois phrases différentes « doomo » (excusez-moi), « doozo » (je vous en prie) et « dooro » (la route) sont préalablement choisies. On a demandé aux sujets d'oraliser ces phrases avec trois degrés de politesse différents tels que poli (teinei), ordinaire et impoli (zonzai).

L'analyse PARCOR¹⁷ a montré que le degré de politesse (poli, ordinaire et impoli) présente une corrélation avec la fréquence fondamentale maximale et avec la durée de la phrase, ceci pour les trois dialectes (excepté pour les mots « doozo » et « dooro » du dialecte d'Osaka). Ceci implique que lorsque une phrase relativement neutre au niveau sémantique est prononcée poliment, la F0 est plus élevée et la durée plus importante par rapport à la politesse ordinaire. D'un autre point de vue, l'attitude « impolie » est réalisée par une F0 plus basse et une durée plus courte par rapport à la politesse ordinaire.

L'expérimentation perceptive a porté sur la valeur moyenne de la F0 de chaque phrase de degré de politesse « ordinaire » dans chaque dialecte. Au total 25 stimuli ont été synthétisés à partir de ces phrases « ordinaires ». La durée a été modifiée à cinq niveaux espacés d'un intervalle de 20 ms (+ 20 ms, + 40 ms, - 20 ms et - 40 ms) sur la première et la deuxième voyelle. Simultanément, la F0 a été modifiée à cinq niveaux espacés d'un intervalle de 15Hz (+ 15 Hz, + 30 Hz, - 15 Hz et - 30 Hz) tout au long de la phrase.

C'est avec ces 75 stimuli (C = 25 stimuli×3 phrases) que Kawano a procédé à l'expérimentation perceptive auprès de 17 étudiants résidant à Tokyo pour les stimuli du dialecte de Tokyo, 21 étudiants résidant à Kyoto pour ceux du dialecte d'Osaka et 20 étudiants qui résident à Nagoya pour ceux du dialecte de Nagoya. Lors de l'expérimentation, les auditeurs ont noté le degré de politesse exprimé par chaque stimulus de -2 à +2.

Le résultat confirme les effets acoustiques identifiés dans l'analyse précédente pour les trois régions différentes. En effet une durée plus longue et une F0 plus élevée donnent aux auditeurs une impression « plus polie ». Entre ces deux paramètres, la durée de l'énoncé influence davantage le choix de la politesse. Parallèlement, une autre

¹⁷ L'analyse PARCOR (Partial Auto Correlation) est un algorithme qui mesure la durée, F0 max., F0 min. et F0 range.

expérimentation perceptive avec les stimuli du dialecte de Tokyo a été effectuée auprès des seize sujets étudiants de la région d'Osaka. Le résultat montre que les auditeurs du dialecte d'Osaka donnent une valeur de politesse nettement plus faible que les auditeurs du dialecte de Tokyo pour les stimuli du dialecte de Tokyo à l'exception des stimuli avec « dooro » (la route). Seuls ces stimuli avec la phrase « dooro » avaient un effet similaire pour les deux groupes d'auditeurs de différentes provenances.

Cependant, les caractéristiques envisagées dans les travaux de Kawano concernent uniquement des corpus de femmes. C'est pourquoi M. Ko (1993) a tâché d'envisager les caractéristiques acoustiques (F0 et durée) de la phrase polie parmi les deux sexes. Une conversation quasi spontanée de six étudiants japonais (soit 3×2 étudiants en binôme) a été enregistrée. Soit trois femmes et trois hommes à qui on a imposé deux situations : formelle/informelle¹⁸. Parmi ces données, six phrases utilisées par tous les sujets pour les deux situations différentes ont été sélectionnées. Ensuite, à partir des 80 stimuli, une expérimentation perceptive a été effectuée auprès de 314 Japonais de la région de Tokyo (177 hommes et 137 femmes, 19,7 ans en moyenne). Lors de l'expérimentation, les sujets ont choisi un degré de politesse parmi quatre choix proposés (1.impoli, 2.légèrement poli, 3.poli, 4.très poli) pour chaque stimulus entendu.

L'analyse des résultats montre que les énoncés qui portent les caractéristiques suivantes ont été jugés comme phrase « polie » auprès des 314 sujets ; (1) moyenne de F0 plus élevée, (2) durée plus longue, (3) pause plus longue et (4) contour intonatif en fin de phrase qui respecte le contour conventionnel lié avec l'acte de parole (par exemple l'énoncé de remerciement, salutation etc...). Ces caractéristiques de politesse ont été identifiées non seulement dans les signaux des femmes, mais aussi dans ceux des hommes.

Ces résultats confirment l'hypothèse du « code des fréquences » d'Ohala : les femmes et les hommes utilisent une F0 plus élevée pour marquer la déférence. Cette F0 plus élevée est corrélée avec l'augmentation de la durée de la phrase.

En outre, Ko montre que le contenu sémantique du type « expression conventionnelle » telle que les salutations est directement lié à un contour de F0 prototypique. Ce contour est figé : utiliser un autre type de contour global de F0 sur cette phrase est immédiatement perçu comme « Zonzai » (impoli).

5.3.8.4. Les travaux des sociolinguistes pour le rôle de F0

A contrario, d'autres études tendent à prouver qu'il existe une différence de stratégie entre les hommes et les femmes japonais lorsque ceux-ci veulent exprimer un degré élevé de politesse. Ces travaux ont été réalisés par des sociolinguistes : L. Loveday ainsi que R. Van Bezooijen et Y. Ohara qui s'intéressent surtout à la valeur sociosémiotique de F0.

¹⁸ Cependant, il n'a y a pas de précision s'ils ont parlé entre hommes, entre femmes ou en binômes mixtes.

L. Loveday (1981) a analysé la parole produite par les deux sexes, et affirme que les hommes et les femmes utilisent une stratégie différente pour exprimer la politesse en japonais.

Dans le cadre de recherches sur la contribution de la hauteur de F0 pour la politesse dans deux langues (l'anglais et le japonais) il a mis au point un dialogue en japonais et un dialogue en anglais pouvant être pratiqués dans des situations relativement formelles : des conversations entre deux personnes qui ne se connaissent pas bien, en l'absence de hiérarchie sociale définie. Cinq locuteurs japonais (trois hommes et deux femmes) pour deux dialogues en anglais/japonais, et cinq locuteurs anglais (trois hommes et deux femmes) pour le dialogue en anglais seulement ont été enregistrés. Les phrases examinées « aa, kon'nichiwa » (oh bonjour), « arigato gozaimas » (merci) et « sayonara » (au revoir) ont été extraites.

L'analyse acoustique montre que les femmes japonaises, les anglaises et les hommes anglais adoptent une F0 élevée pour exprimer la politesse. Notamment, les femmes japonaises adoptent une F0 extrêmement élevée (fréquence minimale de 280 Hz et maximale de 377 Hz en moyenne) par rapport aux femmes anglaises (fréquence minimale de 163 Hz et maximale de 250 Hz en moyenne). Cela signifie que contrairement aux femmes anglaises, dans une situation polie, les femmes japonaises adoptent un niveau de F0 extrêmement plus élevé pour exprimer leur politesse. Quant aux hommes japonais, ils adoptent un niveau de F0 nettement plus bas (écart entre 80 - 120 Hz), et ils utilisent probablement une stratégie différente afin d'exprimer la politesse. Quant aux sujets anglais, le niveau de F0 que les femmes anglaises adoptent pour exprimer la politesse n'est pas tellement différent de celui des hommes anglais. Il en résulte qu'une F0 plus élevée, utilisée dans une situation polie est le moyen d'exprimer adéquatement la politesse pour les Anglais des deux sexes.

Mais Loveday confirme seulement l'application de la déférence telle que définie par le « code des fréquences » pour les femmes japonaises. Y. Ohara (1992) a également observé des phrases en japonais et en anglais produites par six hommes et six femmes japonaises pour des résultats similaires à ceux de Loveday.

Y. Ohara (1992) s'intéresse aussi au rôle des différents niveaux de F0 utilisés par les hommes et les femmes japonais. Elle considère que la hauteur de F0 est étroitement liée à la culture à laquelle les sujets appartiennent ainsi qu'au sexe des sujets. Afin de déterminer ces aspects socioculturels de l'utilisation de la F0, elle a enregistré la lecture oralisée de dix phrases en japonais et de dix phrases en anglais américain qui ont été produites par douze Japonais (six hommes / six femmes) qui étudient à Hawaï.

Les résultats montrent que les femmes ont utilisé une F0 plus élevée pour les phrases en japonais par rapport à celles en anglais. Cet écart variait entre 16 et 49 Hz.

Cette tendance n'a pas été confirmée chez les hommes japonais. En fait, parmi les six hommes, seuls trois ont utilisé une F0 plus élevée. Les autres ont plutôt montré une F0 plus basse pour les phrases en japonais.

Ohara conclut ainsi que “l’idéal de la féminité dans la société japonaise est ainsi d’une importance cruciale pour une locutrice au point que le pattern féminin du comportement devient une partie de sa personnalité” (Ohara 1992, p.474 traduit l’auteur).

Bien que les travaux de ces deux sociolinguistes aient pu mettre en évidence l’utilisation d’une F0 élevée pour exprimer la déférence chez les femmes japonaises, en accord avec l’hypothèse du « code des fréquences », une telle tendance n’a pas pu être observée de façon aussi claire chez les hommes japonais. Ces auteurs concluent donc que cet usage d’une F0 élevée n’est ni universel ni inné, et que le niveau de F0 est nettement distinct entre les deux sexes car bien défini par la société à laquelle nous appartenons.

Bien que la pratique du niveau de F0 est distinct entre les deux sexes dans la société japonaise, R. Van Bezooijen (1995) montre que les hommes et les femmes partagent, d’un point de vue perceptif, la valeur du « code des fréquences ».

R. Van Bezooijen (1995) a essayé de cerner les aspects socioculturels de l’usage de différents degrés de la F0 chez des Japonaises et des Néerlandaises.

Dans cette étude, l’auteur a tout d’abord tâché de valider le « Code de fréquence » dans les deux cultures. En d’autres termes, son premier objectif était de retrouver le lien entre une fréquence F0 élevée et les connotations telles que petitesse, faiblesse, dépendance, modestie etc.. Ensuite, elle prédisait qu’en ce qui concerne l’idéal masculin (puissant) et féminin (faible), il y aurait une forte différence entre les deux cultures. C’est ainsi qu’il précise que les femmes japonaises préfèrent utiliser une F0 élevée et les femmes néerlandaises, quant à elles, plutôt une F0 moyenne ou basse.

Huit femmes néerlandaises parlant le néerlandais standard et ayant 33 ans en moyenne et huit femmes japonaises parlant le dialecte de Tokyo et ayant 29 ans en moyenne ont été sélectionnées.

L’enregistrement a été effectué à partir de la lecture oralisée d’un texte sémantiquement neutre (la description d’une maison) qui dure treize secondes. La F0 moyenne des signaux produits par les Néerlandaises étaient 180Hz (147Hz – 209Hz) et celle des Japonaises était de 185Hz (164 Hz – 209Hz). En plus de ces données originales, deux versions (F0 plus élevée et F0 plus basse) ont été créées.

La moyenne de F0 des trois versions chez les Néerlandaises était de 150 Hz pour la F0 la plus basse, 180Hz pour l’original, 212Hz pour la F0 la plus élevée. La F0 moyenne des trois versions chez les Japonaises était de 155 Hz pour la F0 la plus basse, 185Hz pour l’original, 218Hz pour la F0 la plus élevée.

C’est ainsi que 48 stimuli (8 locutrices x 2 cultures x 3 niveaux de F0) ont été évalués auprès de 30 étudiants néerlandais (15 hommes et 15 femmes) et de 30 étudiants japonais (15 hommes et 15 femmes). Tout d’abord, les auditeurs devaient juger chaque stimulus selon cinq critères, sur une échelle allant de 1 à 7. Les critères utilisés étaient les suivants : la taille de la locutrice (petit – grand), la puissance (fort – faible), la

dépendance (indépendant – dépendant), la modestie (modeste – arrogant) et l'attirance (attirant – inintéressant). Ensuite, les auditeurs devaient choisir un des 2 extrêmes définissant le mieux le stimulus entendu pour chacun des critères mentionnés précédemment, excepté l'attirance. Les quatre qualificatifs obtenus devaient former le type idéal masculin puis le type idéal féminin.

Cette expérimentation valide l'hypothèse du « code des fréquences ». Ainsi une F0 basse est associée aux impressions suivantes : « grand », « fort », « indépendant » et « arrogant ». En revanche, une F0 élevée est associée aux impressions suivantes : « petit », « faible », « dépendant » et « modeste ». Cette valeur est confirmée parmi les deux sexes des deux cultures différentes sans aucune valeur contrastive significative.

En ce qui concerne l'étude des idéaux-types féminin et masculin pour les deux cultures, le résultat montre qu'un large écart est identifié dans les trois critères (fort – faible, indépendant – dépendant, arrogant – modeste) parmi les sujets japonais selon qu'ils sont hommes ou femmes, ce qui n'est pas le cas pour les sujets néerlandais. La comparaison interculturelle des idéaux-types pour un sexe donné ne fait pas apparaître de différence significative entre le type idéal des japonaises et celui des néerlandaises. En revanche le type idéal des hommes japonais est fortement différent de celui des néerlandais. Van Bezooijen considère que c'est la société japonaise qui insiste, de nos jours, sur la masculinité chez les hommes au lieu d'insister sur la féminité chez les femmes.

En fait, ce travail de R. Van Bezooijen montre que le « code des fréquences » est réalisé parmi les deux sexes dans la société japonaise. Les sociolinguistes tels que L. Loveday, Y. Ohara et Van Bezooijen ont décodé ce « code des fréquences », et ils ont donné les étiquettes suivantes. Une moyenne de F0 plus élevée est liée avec « déférence » et « féminité » (Loveday et Ohara), « petit », « dépendant », « faible » et « modeste » (Van Bezooijen). En revanche, une moyenne de F0 plus basse est associée aux aspects « grand », « indépendant », « fort » et « arrogant » (Van Bezooijen).

Cependant, considérant ce « code des fréquences » d'un point de vue plus pragmatique, les sociolinguistes se sont rendu compte de l'exigence de la société dans la détermination de son usage, et par conséquent, les femmes ont tendance à utiliser naturellement une moyenne de F0 plus élevée dans la vie quotidienne. Or les hommes ont plus de facilité à utiliser une F0 basse.

Discussion

Les phonéticiens, M. Ko et T. Kawano, ont envisagé les caractéristiques acoustiques de la phrase polie en japonais. Ils ont conclu qu'une moyenne de F0 plus élevée associée à une durée plus longue est l'indicateur déterminant de la perception de la politesse en japonais.

Du point de vue des sociolinguistes, les recherches se complètent. R. Van Bezooijen (1995) s'est interrogée sur la reconnaissance du « code de fréquence » tandis

que Y. Ohara (1992, 2001) et L. Loveday (1981) ont fait des recherches sur l'utilisation du « code de fréquence ».

Bien que la reconnaissance du « code de fréquence » soit partagée parmi les deux sexes, seules les femmes japonaises empruntent le niveau d'une F0 élevée pour exprimer l'attitude polie ou l'image de la féminité (Loveday 1981, Ohara 1992, 2001). Ce résultat implique que les hommes et les femmes utilisent une stratégie différente pour la politesse en japonais. C'est-à-dire que les femmes ont tendance à exprimer la politesse avec une F0 plus élevée. En revanche, les hommes ont tendance à éviter une stratégie de F0 élevée pour exprimer la politesse.

5.3.8.5. La stratégie de politesse chez les hommes japonais

Si la tendance de l'utilisation de la moyenne de F0 plus élevée associée à la politesse chez les hommes japonais n'est pas confirmée, y a-t-il d'autres stratégies pour l'exprimer ?

Pour répondre à cette question, voyons les travaux d'Ofuka et al. (2000) qui se sont focalisé sur les caractéristiques acoustiques de la prosodie pour la politesse à partir de voix masculines.

E. Ofuka et al. (2000) ont tâché d'envisager les caractéristiques acoustiques de la politesse en japonais à partir de données de voix masculines. Dans cette optique, six hommes japonais de différentes régions dans lesquelles le système accentuel est équivalent à celui du dialecte de Tokyo, et n'ayant pas eu d'entraînement particulier ont été enregistrés en chambre sourde. Les auteurs ont procédé à l'expérimentation par jeu de rôle selon les deux scénarii suivants :

- un jeune inspecteur de la douane adresse la parole à un gentleman (poli) ou à un jeune étudiant (amical) dans une ambiance tendue.
- un jeune employé du service public adresse la parole à un gentleman ou à un jeune étudiant au téléphone.

Après avoir expliqué ces consignes, les sujets qui se connaissaient bien travaillent en binôme, l'un prend le rôle du locuteur et l'autre prend le rôle de l'interlocuteur. Après un moment de conversation spontanée sur le thème du jeu de rôle, on donne aux sujets un dialogue écrit à oraliser. Dans ce dialogue apparaissent deux phrases cibles : 1. « Nimotsuwa koredake deska ? » (Est-ce tout comme bagages ?) et 2. « Moshimoshi, Akagi-san no otaku deska ? » (Allô, je suis bien chez M. Akagi ?)

L'analyse acoustique de ces données n'a pas permis d'identifier une tendance dans la relation entre la moyenne de F0 et le degré de politesse. Un seul des six sujets a adopté une moyenne de F0 plus élevée pour la phrase polie dans les deux situations. Ce résultat est en accord avec les résultats de Loveday (1981) et Ohara (1992). Cependant, ils ont identifié une différence significative entre le degré de politesse et le débit de la phrase. Les phrases polies de tous les sujets étaient réalisées avec un débit plus lent.

Pour les aspects temporels de la politesse, ce résultat est en accord avec les résultats de Kawano (1993) et Ko (1993).

Ensuite, une expérimentation a été menée afin d'identifier le rôle de la F0 en fin de phrase et celui du débit de parole pour la phrase polie.

La réalisation des deux niveaux de politesse (polie et ordinaire) sur la phrase « Nimotsuwa koredake deska ? » produite par deux sujets venant de Kyushu et de Yokohama ont été utilisées comme sources. En ce qui concerne la production de la phrase « polie » du premier locuteur, la F0 initiale de la dernière voyelle est de 110 Hz, puis elle diminue à la vitesse de 54 demi-tons par seconde sur 120 ms. Pour le deuxième locuteur, la F0 initiale de la dernière voyelle de la phrase « polie » est également de 110 Hz, puis elle augmente à la vitesse de 28 demi-tons par seconde sur 70 ms. Quant à la phrase amicale, la F0 initiale de la dernière voyelle est de 117 Hz, puis elle diminue à la vitesse de 9 demi-tons par seconde sur environ 70% de la durée totale de la voyelle (240 ms), avant d'augmenter de 56 demi-tons par seconde sur le reste. Pour le deuxième locuteur, la F0 initiale de la dernière voyelle de la phrase « amicale » est de 101 Hz, puis elle augmente de 20 demi-tons par seconde sur 130 ms. Cette dernière more /ka/ de la phrase est resynthétisée en manipulant la durée (longue : 240 ms pour le premier locuteur et 130 ms pour le deuxième / courte : 120 ms pour le premier locuteur et 70 ms pour le deuxième) et la F0 (montante : + 25 demi ton par seconde du début jusqu'à la fin / descendante : - 25 demi ton par seconde jusqu'à la fin) sur la voyelle finale [a] à partir des deux phrases "polie / amicale" des deux locuteurs.

Au début de l'expérimentation perceptive, les deux phrases, "polie et amicale", produites par les deux locuteurs sont présentées à vingt Japonais (douze hommes et huit femmes de 20 à 36 ans) pour qu'ils puissent en évaluer le degré de politesse. Ensuite, les sujets devaient écouter les 164 stimuli et évaluer le degré de politesse.

Dans la plupart des cas, un stimulus ayant une durée plus courte et un contour montant de F0 en fin de phrase est perçu comme plus poli pour les deux types « poli » et « amicale ». Les auteurs concluent que la prosodie (surtout la durée) de la voyelle finale de l'énoncé a une forte influence sur l'identification de la politesse. Autrement dit, si on allonge la dernière voyelle et qu'on réalise un contour descendant dans la phrase prononcée poliment (c'est-à-dire avec un débit plus lent), cela risque d'être perçu comme « impoli ». Par ailleurs, dans certains cas, la durée et le mouvement de la F0 sur les dernières 100 ms pouvaient changer le degré de politesse attribué à l'ensemble de l'énoncé.

Après ces deux expérimentations, Ofuka et al. (2000) ont conclu que le mouvement de F0 sur la partie finale (dernier 100 ms) de l'énoncé ainsi que le tempo de l'énoncé, étaient les ultimes déterminants dans les deux scénarii. Ces deux facteurs sont ainsi des indicateurs importants pour différencier la phrase polie et la phrase ordinaire parmi tous les sujets.

Discussion :

Ofuka et al. (2000) se sont concentrés sur les stratégies utilisées par les hommes japonais pour exprimer la politesse japonaise.

De même que Loveday (1981) et Ohara (1992), Ofuka et al. (2000) n'ont pas identifié de lien entre la moyenne de F0 et le degré de la politesse chez les hommes japonais. Par contre, ils ont confirmé une relation significative entre le degré de la politesse et le débit de la phrase : la phrase polie est prononcée plus lentement que la phrase « ordinaire » et elle est donc plus longue. Ce résultat renforce ceux qui ont été mis en évidence par Kawano (1993) et Ko (1993) sur l'aspect temporel de la phrase.

Ofuka et al. (2000) ont également étudié l'effet de la durée et l'effet de la F0 en fin de phrase. A ce propos, ils affirment que la prosodie (surtout la durée) de la voyelle finale de l'énoncé influence fortement l'identification de la politesse. C'est ainsi que l'allongement de la dernière voyelle dans un énoncé dont le contour de F0 est descendant est perçu comme poli.

5.3.8.6. Stratégie de politesse des hommes japonais en parole spontanée

Les caractéristiques prosodiques des expressions de la politesse que nous avons vues jusqu'ici ont été étudiées en laboratoire uniquement. Les corpus plus ou moins manipulés (un enregistrement à partir de jeux de rôle, des phrases synthétisées, de la lecture oralisée etc.) posent le problème du « naturel » de la parole (Campbell 2000). C'est pourquoi M. Ito a procédé à l'analyse de phrases extraites d'une conversation spontanée.

Ito (2002) a étudié les caractéristiques prosodiques des expressions de la politesse chez les hommes japonais. Quatre sujets masculins parlant le dialecte de Tokyo ont été préalablement sélectionnés. Ces quatre sujets travaillent dans une même équipe de recherche. Ils sont intimes mais respectent la hiérarchie sociale entre eux. On leur a donné une tâche à résoudre en binôme. D'abord, un sujet supérieur dans la hiérarchie prend le rôle de navigateur tandis que l'autre sujet, inférieur dans la hiérarchie, prend le rôle de celui qui suit les instructions de navigation du supérieur avant d'inverser les rôles. La conversation durant la réalisation de cette tâche a été enregistrée dans une chambre sourde. Ito a extrait des phrases cibles de deux sujets.

Afin d'évaluer le corpus, elle a mesuré le niveau de F0 moyen, le débit de parole ainsi que le débit articulatoire (mores/sec) des phrases cibles. Les résultats montrent qu'un sujet a eu recours à des stratégies paralinguistiques utilisant le niveau moyen de F0 et le débit de parole pour marquer la politesse. Ce sujet a marqué sa politesse envers le supérieur en élevant la F0 et en articulant plus lentement. En revanche, on n'a pas pu identifier d'effet paralinguistique dans les énoncés de l'autre sujet. Cela signifie que la moyenne de F0 et le débit de parole n'ont pas été identifiés comme les indicateurs déterminants de la politesse dans la parole spontanée.

5.3.8.7. La stratégie de l'aspiration ?

M. Ito (2002) n'ayant pas établi le rôle des caractéristiques prosodiques mentionnées ci-dessus (la moyenne de la F0 et le débit de parole) pour exprimer les degrés de politesse, elle a orienté ses recherches du côté de la qualité de voix.

Au niveau perceptif, Laver (1980) note sept différentes qualités de voix : voix modale, voix soufflée, voix pressée, voix craquée, voix rauque, voix nasale et voix tendue. Ce changement de qualité de voix est réalisé principalement par un contrôle des muscles laryngés et supralaryngés.

En fait, des recherches précédentes montrent qu'il existe des langues présentant une fonction contrastive au niveau lexical¹⁹ ou une fonction démarcative en changeant la qualité de voix.

En ce qui concerne la langue japonaise, elle n'a pas ces deux caractéristiques, mais ses caractéristiques de qualité de voix peuvent affecter les attitudes du locuteur afin d'exprimer ses intentions (Campbell & Mokhtari 2003). Quant à la contribution de la qualité de voix à l'expression de la politesse, peu de travaux ont été réalisés²⁰.

Ito, donc, dans le but de déterminer la contribution de la qualité de voix à l'expression de la politesse japonaise, a sélectionné sept sujets masculins parlant le dialecte de Tokyo. Ces sept sujets se connaissaient tous, et ils avaient tous plus ou moins la même formation ainsi que la même expérience professionnelle. Ils étaient tous membres du même groupe de recherches. Les corpus étaient enregistrés dans une chambre sourde selon la consigne de « map task » dans laquelle les sujets supérieurs dans la hiérarchie prennent le rôle de navigateur et donnent des instructions.

Ito a extrait 16 phrases incluant le mot cible /hidari/, mélangées avec 16 phrases de contrôle, soit un total de 32 phrases produites par cinq locuteurs. Ces phrases ont été utilisées comme stimuli pour le test perceptif dans lequel le degré de politesse est évalué par 22 sujets (15 sujets sans expérience professionnelle, et 7 sujets avec). Parallèlement, elle a extrait la voyelle /a/ à partir de chaque phrase cible, puis elle a mesuré la bande passante du premier formant (BW1) et la pente spectrale (H1-A3) calculée comme la différence entre l'amplitude du premier harmonique (H1) et celle du troisième formant (A3). Indépendamment, elle a filtré les voyelles autour du troisième formant en utilisant un filtre passe-haut de 400 Hz. Ensuite, elle a demandé à deux auditeurs d'évaluer le taux de bruit sur une échelle de 0 (pas de bruit) à 10. Il y a peu d'évidence de la périodicité.

Les résultats du test perceptif et de l'analyse acoustique montrent que le degré de politesse évalué par les auditeurs et le niveau du BW1 sont fortement corrélés pour quatre locuteurs sur cinq. C'est-à-dire que l'augmentation de la valeur de BW1 entraîne

¹⁹ Par exemple, un des dialectes du Mazatec qui s'appelle Jalapa de Diaz présente une paire minimale entre la voix modale, la voix craquée et la voix soufflée (Ladefoged & Maddieson, 1996).

²⁰ Voir par exemple Campbell & Mokhtari (2003) pour le degré du « soin » avec la voix soufflée.

un degré de politesse perçue plus élevé. Cette variation de BW1 peut être un indicateur de la voix soufflée, de la voix chuchotée et de la nasalité, mais ce résultat seul ne nous permet pas d'identifier quelle qualité de voix est corrélée avec le degré de politesse. C'est pourquoi Ito a analysé la corrélation entre le taux de bruit perçu et les valeurs de pente spectrale (H1-A3). Les résultats montrent une corrélation entre le taux de bruit et les valeurs de pente spectrale pour les productions des cinq locuteurs. Le taux de bruit et H1-A3 ont un effet significatif sur les valeurs de BW1 pour un seul des deux locuteurs. L'autre locuteur adopte plutôt la voix tendue et modale pour exprimer un degré de politesse plus élevé. Elle conclut que ces deux qualités de voix (la voix soufflée et la voix tendue) peuvent être utilisées afin d'exprimer un degré élevé de politesse chez les hommes japonais.

5.3.8.8. Stratégie de « Kyoshuku » et « sincérité/sérieux »

Parallèlement à M. Ito, T. Sadanobu, linguiste spécialisé en sémantique, a envisagé les fonctions linguistiques de la voix pressée chez les Japonais des deux sexes.

Dans les travaux de Sadanobu (2003, 2004a), treize heures de conversation spontanée de onze japonais natifs (quatre hommes et sept femmes de dix ans à soixante-dix ans) de différentes régions ont été enregistrées, locuteur après locuteur dans une salle calme. Ce corpus a été analysé par dix Japonais (y compris l'auteur) selon la méthode d'introspection de leur propre parole.

Le résultat montre que la voix pressée est pratiquée fréquemment dans la conversation spontanée des quatre sujets suivants ; une femme de 70 ans, un homme de 60 ans, une femme de 40 ans, et un homme de 30 ans. Cependant, trois sujets (un homme de 40 ans, une femme de 30 ans, une femme de 20 ans) n'ont pas utilisé la voix pressée lors de l'enregistrement. Puisque ces trois sujets étaient issus de régions différentes de celle des autres, habitants la région du Kansai, on peut peut-être conclure que dans ces régions on n'utilise pas la voix pressée pour exprimer certaines attitudes. Toutefois, ils étaient tous capables d'identifier la présence de la voix pressée.

Ensuite, une expérimentation perceptive a été conduite à partir des treize stimuli composés des phrases dans lesquelles la voix pressée est utilisée. Parallèlement, Sadanobu a enregistré les mêmes phrases avec la voix modale, soit un total de 26 stimuli (13 phrases avec la voix pressée et 13 phrases avec la voix modale). Ces stimuli ont été présentés aux dix sujets précédents, tout d'abord en voix modale puis en voix pressé. Il était ensuite demandé aux sujets de dire quelle était l'intention du locuteur, parmi cinq choix proposés.

Le résultat montre que les attitudes exprimées par la voix pressée étaient parfaitement identifiées par les sujets des régions autres que le Kansai. Seules deux phrases produites par deux sujets ont été jugées comme exprimant des attitudes différentes, mais les 128 autres stimuli présentaient une correspondance parfaite. Sadanobu a donc conclu que les sujets, quelles que soient leurs régions d'origine, sont

sensibles au sens apporté par la voix pressée dans les phrases, même si eux-mêmes ne l'utilisent pas.

Sadanobu (2004a) a observé les données recueillies dans Sadanobu (2003). Après l'analyse linguistique de ces données, il a tenté de montrer une des caractéristiques de la voix pressée. Cette caractéristique a un rapport avec la stratégie de politesse japonaise, qu'il nomme l'expressivité de « *kyoshuku* ».

Cette attitude qui est un mélange de souffrance honteuse et d'embarras est une stratégie utilisée pleinement par les Japonais (les hommes et les femmes) de statut inférieur (par exemple, employé vs. client, employé vs. son supérieur) afin d'atténuer l'aspect arrogant et impoli dans la situation de demander une faveur ou de refuser l'opinion de l'interlocuteur socialement supérieur. Cette attitude est exprimée par la voix pressée et un visage exprimant la difficulté.

Par ailleurs, la voix pressée est utilisée dans un autre contexte de « *kyoshuku* ». Il s'agit du contexte d'admiration, par exemple devant un interlocuteur très proche au niveau relationnel et sachant bien jouer au tennis. Cette attitude « d'admiration » est exprimée par la phrase « *umainaa* » (vraiment bien joué !) prononcée avec la voix pressée. Rejoignant l'attitude de « *kyoshuku* », cette attitude est utilisée par le locuteur qui veut exprimer sa profonde sincérité. Dans un autre contexte, par exemple un locuteur racontant une anecdote personnelle et souhaitant profondément que ses interlocuteurs soient convaincus de ses dires, utilisera aussi cette voix pressée. Il est à noter que ces trois attitudes sont interconnectées les unes aux autres dans la langue japonaise du fait de la polysémie lexicale des items qui portent ces attitudes (par exemple le terme « *unaru* » porte deux sens : l'admiration et la souffrance).

Enfin, la voix pressée s'utilise aussi pour l'attitude « sincérité/sérieux » qui fait émerger des effets variables (admiration, politesse de *kyoshuku* ou sincérité/sérieux) dans l'interaction entre les contenus linguistiques et les contextes sociaux. C'est-à-dire qu'en utilisant la voix pressée, le locuteur montre qu'il est si sincère que sa parole ne peut être mise en doute.

5.3.8.9. Les particules honorifiques vs. la prosodie

Le japonais du dialecte de Tokyo possède quelques particules finales telles que /ne/, /yo/, /yone/ qui marquent la modalité d'un ensemble de phrase. La politesse est exprimée soit par des particules (par exemple, « o », ou « go » dans « o-namae » (votre nom), « go-ryooshin », (vos parents)), soit par la prosodie. En revanche, il n'existe pas de particule finale qui marque le degré de politesse. K. Maekawa (1999) s'est intéressé au dialecte de Kumamoto qui est un des rares dialectes n'ayant pas d'accent lexical et possédant trois particules finales de politesse.

Les travaux de K. Maekawa (1999) ont pour but d'étudier la contribution du contenu sémantique et les facteurs prosodiques dans la perception de la politesse japonaise dans le dialecte de Kumamoto.

Pour cette expérimentation, un japonais d'une cinquantaine d'année s'exprimant en dialecte de Kumamoto et, n'étant pas acteur, a été enregistré dans une chambre sourde. Il est à noter qu'en l'absence d'accent lexical dans le dialecte de Kumamoto, le pic de F0 peut être aligné sur n'importe quelle syllabe à l'intérieur de la phrase. Le recueil de données a été effectué grâce à la lecture oralisée d'une phrase interrogative (WH-question) « dore ni noboru » (Où voulez-vous monter ?) avec trois particules qui marquent le degré de la politesse (/na/ est la plus polie, suivie de /to/, /ya/ est la moins polie).

Sont ainsi enregistrées différentes compositions de dix huit stimuli : Les trois particules ont été combinées avec six contours de F0, pour un total de 18 stimuli. Ces stimuli ont été présentés en paires selon chacun des 2 ordres possibles et les sujets ont choisi dans chaque paire le stimulus qui leur semblait être le plus poli. 306 paires au total ont ainsi été testées auprès de 99 Japonais du dialecte de Kumamoto, se répartissant en trois groupes (20 adultes salariés, 29 étudiants et 50 collégiens).

Concernant l'usage des trois particules, la particule /na/ est peu utilisée chez les jeunes sujets qui l'ont considéré lors de l'entretien comme démodée. L'expérimentation perceptive a permis d'identifier une forte influence de l'intonation et des particules dans la détermination de la politesse pour les trois groupes d'âge examinés. En effet, les contours intonatifs qui montent en fin de phrase sont perçus comme plus polis que les contours descendants. Quant aux particules, l'ordre de la politesse est /to/ (la plus polie) > /na/ > /ya/ (la moins polie) par l'influence de l'intonation. Rappelons que le degré de la politesse du terme lexical était dans l'ordre de /na/ (la plus polie) > /to/ > /ya/ (la moins polie). On remarque donc que ces deux fonctions (prosodique et lexicale) ne montrent pas le même degré de politesse en réalité. L'utilisation des particules comme marqueur de la politesse est plus marquée dans le groupe des adultes que dans les deux autres groupes. En revanche, les jeunes sujets se servent surtout du contour intonatif afin d'exprimer les degrés de politesse.

Enfin, Maekawa conclut que le mécanisme de perception de la politesse intonative existe malgré l'existence des particules honorifiques, et que notamment les jeunes habitants de Kumamoto ont tendance à percevoir leurs degrés de politesses en montant la F0 en fin de phrase.

5.4. Validation des attitudes japonaises

À partir des enseignements tirés de la littérature et des travaux antérieurs réalisés sur les expressions attitudinales dans d'autres langues, nous avons sélectionné un certain nombre d'attitude pour le japonais, que nous allons maintenant décrire et évaluer. Le but de ce travail est d'obtenir un ensemble d'attitudes représentatives de la

langue japonaise, puis de valider ces expressions stéréotypiques pour des auditeurs japonais.

5.4.1. Choix des attitudes du japonais

Nous avons retenu un ensemble de 12 attitudes représentatives du japonais selon la littérature de ce domaine détaillée ci-dessus (Kori, 1989 ; Ofuka et al., 2000 ; Sadanobu 2003, 2004a ; Maekawa, 1998, 2004 ; Ito, 2003 etc.), mais surtout sur la base de méthodes d'enseignement du japonais (Mizutani & Ayusawa, 1991 ; Mizutani&Mizutani, 1979). Il s'agit des attitudes suivantes : *doute-incrédulité*, *évidence*, *exclamation de surprise*, *autorité*, *irritation*, *arrogance-impolitesse*, *sincérité-politesse*, *admiration*, *kyoshuku*, *politesse-simple*, *déclaration* et *question-simple*. Certaines de ces attitudes sont spécifiques ou particulières à la culture japonaise, notamment celles qui sont liées à la stratégie de politesse japonaise : *politesse-simple*, *sincérité-politesse* et *kyoshuku* vs. *arrogance-impolitesse*. L'attitude de *sincérité-politesse* apparaît lorsqu'un locuteur considéré comme inférieur communique avec un interlocuteur considéré comme supérieur dans la société japonaise. Le locuteur inférieur exprime que son intention est sérieuse et sincère à travers cette attitude prosodique. L'attitude de *kyoshuku* (il n'existe pas d'entrée lexicale pour traduire ce terme en français ou en anglais) est une attitude typique de la culture japonaise : même si des situations d'interactions sociales semblables se produisent dans toutes les cultures, la langue japonaise a choisi de coder cette situation particulière en tant qu'« attitudinème ». Elle apparaît quand le locuteur est dans une situation où son statut social est inférieur à celui de son interlocuteur, et quand il a de plus un avis contraire ou qu'il désire lui demander un service. Elle est décrite par T. Sadanobu comme "*a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker's consciousness of the fact that his/her utterance of request imposes a burden to the hearer*" (Sadanobu, 2004a p.34).

5.4.2. Définition des attitudes

Voici leurs définitions en français :

疑念 (doute - incrédulité) :

Le locuteur exprime son désaccord total ou partiel en reprenant dans son énoncé une affirmation que vient de donner l'interlocuteur.

自明 (évidence) :

Le locuteur donne une information qui lui semble vraie, naturelle, et dont il est très sûr.

驚き (exclamation de surprise) :

L'expression en réaction à quelque chose de soudain et d'inattendu.

高压的 (autorité) :

Le locuteur veut imposer son autorité à son interlocuteur.

苛立ち (irritation) :

Lorsqu'un événement contrarie fortement le locuteur, au point qu'il éprouve le besoin de l'exprimer.

感心 (admiration) :

L'expression utilisée pour manifester que l'on a beaucoup apprécié une chose vue, entendue ou encore goûtée.

ぞんざい (arrogance-impolitesse) :

L'expression impolie d'un sentiment de supériorité vis-à-vis de son interlocuteur.

誠意 (sincérité-politesse)

Cette attitude japonaise est en relation avec le statut social. Le locuteur est socialement inférieur à son interlocuteur et exprime par cette attitude que son intention est sérieuse et sincère.

恐縮 (kyoshuku) :

Cette attitude est typique de la culture japonaise. Elle apparaît quand le locuteur est dans une situation où son statut social est inférieur à celui de son interlocuteur, mais dans laquelle il a cependant un avis contraire à son interlocuteur, ou encore il désire lui demander un service. Dans ce contexte, le locuteur doit montrer sa souffrance honteuse et son embarras afin d'atténuer l'aspect arrogant et impoli qu'engendre sa désapprobation ou sa demande.

丁寧 (politesse-simple) :

Quelque soit le statut social du locuteur par rapport à son interlocuteur, cette attitude est utilisée pour exprimer la politesse et la courtoisie.

Parmi ces attitudes, se trouvent deux attitudes qui ne montrent aucune intention du locuteur : *déclaration* et *question-simple*.

平叙 (déclaration) :

Le locuteur donne l'information simplement, sans exprimer aucun point de vue.

疑問 (question-simple) :

Le locuteur demande une information, sans exprimer de point de vue, et sans attendre autre chose qu'une réponse simple.

Il est à noter que ces définitions sont présentées en langue maternelle (c-à-d, en japonais) lors du test perceptif mené sur les sujets japonais (voir en annexe les définitions en japonais).

5.4.3. Composition du Corpus

Le corpus est construit sur un ensemble de phrases isolées car la fonction attitudinale sera réalisée au niveau phrastique. On visera plus tard l'apprentissage des contours prototypiques porteurs de l'attitude par les apprenants du japonais.

Pour ce test, nous avons choisi sept phrases sans connotation particulière, quelle que soit la région du Japon. Lors de la sélection des phrases, nous avons évité les trois points suivants : d'abord, nous avons exclu toute phrase portant un contenu sémantique significatif « attitudinalement » afin de ne pas interférer dans la production et la perception des attitudes véhiculées. Nous avons également évité les phrases conventionnelles (par exemple, la salutation, le remerciement, l'excuse etc...) car contour prototypique et expression conventionnelle sont indissociables²¹ ; l'expression des attitudes en aurait été limitée. Notons de plus que les effets socioculturels liés à l'origine régionale des sujets japonais sont inévitables. Enfin, nous n'avons pas choisi d'énoncés constitués d'un agent (GN) suivi d'un prédicat (GV) (par exemple, « M. Tanaka mange. »). Contrairement à certaines langues européennes, une telle phrase en japonais ne nous semble pas naturelle : lorsque le contexte est explicite, le sujet de l'énoncé est souvent omis en japonais. C'est pourquoi nous avons choisi plutôt un lieu (GN) suivi d'un prédicat (GV) (par exemple, « Narae ikimas » ((...) vais / vas / va / allons / allez / vont à Nara). Quant à la longueur moraïque de l'énoncé, la longueur varie d'une à huit mores²².

Tableau 6 : *Corpus des attitudes japonaises : sept phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque.*

nb more	Phrase	Traduction
1	Me	L'oeil.
2	Na*ra	Nara.
5	Na*rade neru	Il dort à Nara.
8	Na*goyade nomimas [u]	Il boit à Nagoya.
8	Nara*shide nomimas [u]	Il boit dans la ville de Nara.
8	Matsuri*de nomimas [u]	Il boit à la soirée.
8	Naniwade nomimas [u]	Il boit à Naniwa.

Concernant l'accent lexical du japonais, sa présence peut perturber les caractéristiques prosodiques sur le reste de l'énoncé. De ce fait, il est préférable de prendre compte de l'influence de l'accent lexical à l'aide du dictionnaire de l'accent lexical japonais (Hirata, 1960). C'est ainsi que, en ce qui concerne des énoncés composés de 8 mores, un accent lexical est positionné sur la 1^{ère}, 2^{ème} ou 3^{ème} more, ou encore la phrase est construite sans accent. Afin d'exprimer certaines attitudes comme le doute ou la surprise, la voyelle [u] est insérée en fin de phrase, et dans ce cas un accent lexical est réalisé sur la septième more.

²¹ Voir par exemple les résultats de Ko (1997) et Kori (1997) pour le japonais.

²² Pour l'explication de la notion de more, voir *supra*.

5.4.4. Enregistrement

5.4.4.1. Locuteur

Les attitudes produites par un acteur risqueraient d'être en décalage avec la réalité (Campbell, 2000), c'est pourquoi un enseignant universitaire (un homme, originaire de la région de Yamaguchi, parlant avec l'accent de Tokyo) professeur de japonais langue étrangère, de langue maternelle japonaise, a été choisi comme locuteur de ce corpus. En effet, les apprenants d'une langue étrangère doivent acquérir les attitudes prosodiques de la langue et de la culture cible, et les enseignants de langue ont l'expérience de produire différents types d'attitudes pour des raisons didactiques et pragmatiques.

5.4.4.2. Dispositif et procédure expérimentale

L'enregistrement du corpus a été effectué dans la chambre sourde du GIPSA-lab. De l'extérieur de la chambre, les phrases à produire ainsi que l'attitude étaient envoyés sur un écran faisant face au locuteur. Les sept énoncés à oraliser étaient affichés dans un ordre aléatoire pour chaque attitude. Afin de contrôler au mieux le paramètre d'intensité un micro-casque a été utilisé pour éviter les variations de distance entre le micro et la bouche du locuteur. Ces données ont été enregistrées sur DAT (échantillonnées à 44,1 kHz, 16 bits). Parallèlement, la vibration des cordes vocales a été enregistrée à l'aide d'un électroglottographe (EGG) afin de faciliter de futures analyses de la qualité de voix, qui contribue largement à l'identification des attitudes japonaises d'après la littérature (voir Ito, 2002 ; Sadanobu 2004a).

L'ensemble du corpus a été enregistré 2 fois afin pallier d'éventuels problèmes durant l'enregistrement. Le corpus final est donc composé de 168 énoncés au total (2 fois x 7 énoncés x 12 attitudes) soit :

- 14 énoncés (7x2) * doute-incrédulité
- 14 énoncés (7x2) * évidence
- 14 énoncés (7x2) * exclamation de surprise
- 14 énoncés (7x2) * autorité
- 14 énoncés (7x2) * irritation
- 14 énoncés (7x2) * admiration
- 14 énoncés (7x2) * arrogance-impolitesse
- 14 énoncés (7x2) * sincérité-politesse
- 14 énoncés (7x2) * kyoshuku
- 14 énoncés (7x2) * politesse-simple
- 14 énoncés (7x2) * déclaration

14 énoncés (7x2) * question-simple

Nous avons par la suite procédé au transfert des données audio grâce au logiciel de traitement du son « Cool Edit », puis découpé et segmenté à la main les 168 énoncés produits sur le logiciel Praat 4.1²³, au niveau phonémique.

5.4.5. Validation perceptive

5.4.5.1. Problématique

Un test de perception a été effectué auprès de sujets japonais natifs parlant avec l'accent de Tokyo. L'objectif premier de ce test perceptif est de valider les 12 attitudes auprès des sujets japonais.

5.4.5.2. Méthodologie expérimentale

a- Stimuli

Parmi les 168 énoncés du corpus, la meilleure réalisation parmi les deux disponibles a été sélectionnée à l'aide d'une écoute informelle.

Le corpus est donc composé de sept énoncés différents produits avec les 12 attitudes, soit un nombre total de 84 stimuli.

b- Matériel

Le test perceptif a été effectué sur une interface développée à l'aide du logiciel « Revolution ». Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 12 attitudes possibles et les définitions des attitudes présentées. Cette interface a été créée en langue maternelle des sujets²⁴ (c-à-d, en japonais).

c- Procédure

La procédure expérimentale est la même que celle décrite lors des expériences de validation des corpus français et anglais (cf. supra). Chaque sujet passe le test sur un ordinateur dans une pièce calme. Après avoir lu les consignes et obtenu des informations supplémentaires à l'oral sur les attitudes et le déroulement du test, le sujet doit juger de l'attitude qu'il perçoit pour un stimulus présenté à l'aide d'un casque audio. Tous les stimuli sont présentés dans un ordre aléatoire, différent pour chaque sujet, et ne peuvent être écouté qu'une seule fois.

À la fin du test perceptif, nous avons interrogé le sujet sur leur perception des attitudes afin de savoir lesquelles étaient difficiles à distinguer ou s'il avait perçu d'autres attitudes qui n'étaient pas proposées.

²³ « Praat » est un logiciel gratuit qui permet d'analyser et synthétiser la parole sur un ordinateur. Ce logiciel est développé par Paul Boersma et David Weenink qui travaillent au sein du Département de Phonétiques à l'Université d'Amsterdam.

²⁴ L'interface utilisée est présentée en annexe.

5.4.5.3. Sujets

Quinze auditeurs japonais (3 sujets masculins et 12 sujets féminins) ont passé le test. L'âge moyen de ces 15 sujets est de 29,5 années. Aucun sujet n'a fait état de trouble auditif.

Les sujets sont tous de langue maternelle japonaise et parlent avec l'accent de Tokyo. Ils habitent tous dans la région grenobloise depuis moins d'un an. Deux sujets masculins ont habité 1 à 2 années aux États-Unis avant d'arriver en France et un sujet féminin y a habité 9 mois à l'âge de 20 ans, mais la plupart des sujets n'avaient pas séjourné à l'étranger avant leur arrivée en France. La plupart sont étudiants dans un centre d'apprentissage de la langue française.

5.4.6. Analyse des résultats

Les résultats obtenus lors de ce test de perception sont analysés selon la même procédure statistique que celle mise en place pour l'analyse des précédents tests (cf. supra) ; analyse des répartitions des réponses à l'aide du test du khi deux ; analyse des correspondances puis analyse de classification hiérarchique.

5.4.6.1. Distribution des réponses par rapport au hasard

La comparaison des réponses obtenues par chaque attitude par rapport à une distribution au hasard par le test du khi-deux donne les résultats résumés par le tableau 7. Toutes les distributions de réponse obtenues par chaque attitude sont significativement différentes du hasard – ce qui implique une stratégie cohérente de la part des sujets.

Tableau 7 : Résultats du test de χ^2 sur les distributions moyennes des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2
Admiration (AD)	111,3 (ddl : 11, p.>.001)
Arrogance-impolitesse (AR)	579,0 (ddl : 11, p.>.001)
Autorité (AU)	293,1 (ddl : 11, p.>.001)
Déclaration (DC)	478,0 (ddl : 11, p.>.001)
Doute-incrédulité (DO)	353,0 (ddl : 11, p.>.001)
Évidence (EV)	249,4 (ddl : 11, p.>.001)
Exclamation de surprise (EX)	392,9 (ddl : 11, p.>.001)
Irritation (IR)	838,2 (ddl : 11, p.>.001)
Kyoshuku (KYO)	127,3 (ddl : 11, p.>.001)
Politesse-simple (PO)	453,7 (ddl : 11, p.>.001)
Question simple (QS)	657,9 (ddl : 11, p.>.001)
Sincérité-politesse (SIN)	176,5 (ddl : 11, p.>.001)

Ce premier résultat, allié au fait que chaque attitude reçoit un score d'identification supérieur au niveau du hasard (8,3%), tend à montrer les capacités des sujets à bien reconnaître les expressions qui leur sont proposées.

Nous allons maintenant nous pencher sur l'analyse des facteurs ayant une action potentielle sur les réponses des sujets : la longueur de l'énoncé et la position de l'accent lexical.

5.4.6.2. Effet de la longueur des énoncés

Afin d'observer l'effet de la longueur pour le choix des attitudes, nous avons effectué un test de χ^2 sur les distributions moyennes des résultats, par longueur, toutes attitudes confondues. Le test donne un résultat significativement différent du hasard (ddl=33, $\chi^2=157.5$ p. < 0.05).

On peut donc constater que la longueur de l'énoncé influence le choix de l'attitude. Afin de savoir entre quelles longueurs on peut observer des différences, nous avons effectué un test de χ^2 sur les résultats moyens entre les phrases de 1 et 2 mores, de 2 et 5 mores et de 5 et 8 mores, toutes attitudes confondues. Du fait du petit nombre de réponses reçues par l'attitude d'admiration, les conditions d'application du χ^2 ne sont pas remplies. Nous avons donc mené le test du χ^2 en enlevant les réponses de l'attitude d'admiration.

Les résultats de cette analyse ne montrent aucun effet significatif entre les énoncés de 1 et 2 mores (ddl = 11, $\chi^2=10.0$, p. > 0.05). Par contre, entre les phrases de 2 et 5 mores et celles de 5 et 8 mores, un effet significatif de la longueur est observé (*2 et 5 mores* : ddl = 11, $\chi^2=21.3$, p. < 0.05 ; *5 et 8 mores sauf admiration* : ddl = 10, $\chi^2=34.5$, p. < 0.05). Ces résultats montrent que la longueur de l'énoncé peut influencer la distribution des réponses des sujets, même si toutes les différences de longueur ne reçoivent pas des distributions significativement différentes. Il sera donc intéressant de regarder les résultats du test de dévoilement progressif (cf. infra) pour tenter de mieux comprendre pourquoi de telles différences peuvent survenir.

5.4.6.3. Effet de l'accent lexical

L'influence de la position de l'accent lexical sur la perception des attitudes choisies a été évaluée à l'aide d'un test de χ^2 sur les distributions des résultats moyens toutes attitudes confondues, pour chaque position d'accent lexical, pour les phrases de 8 mores (celles qui présentent plusieurs positions de l'accent lexical). Ce test a été mené sur toutes les attitudes sauf « admiration » (car le nombre de réponses pour cette attitude est trop petit par rapport aux critères d'application du khi-deux) : aucun effet significatif (ddl = 10, $\chi^2=16.5$, p. > 0.05). On ne peut donc pas conclure à un effet significatif de la position de l'accent lexical dans le décodage des expressions attitudinales – ce qui revient à dire que les sujets semblent à même de séparer les informations relatives à la place de l'accent de celles qui concernent l'expressivité.

5.4.6.4. Taux d'identification des attitudes

La figure 14 montre le taux de reconnaissance des attitudes. La ligne pointillée indique le seuil du hasard. Cette figure permet de constater que les auditeurs japonais ont reconnu chacune des attitudes, ce qui permet de valider les expressions enregistrées dans ce corpus.

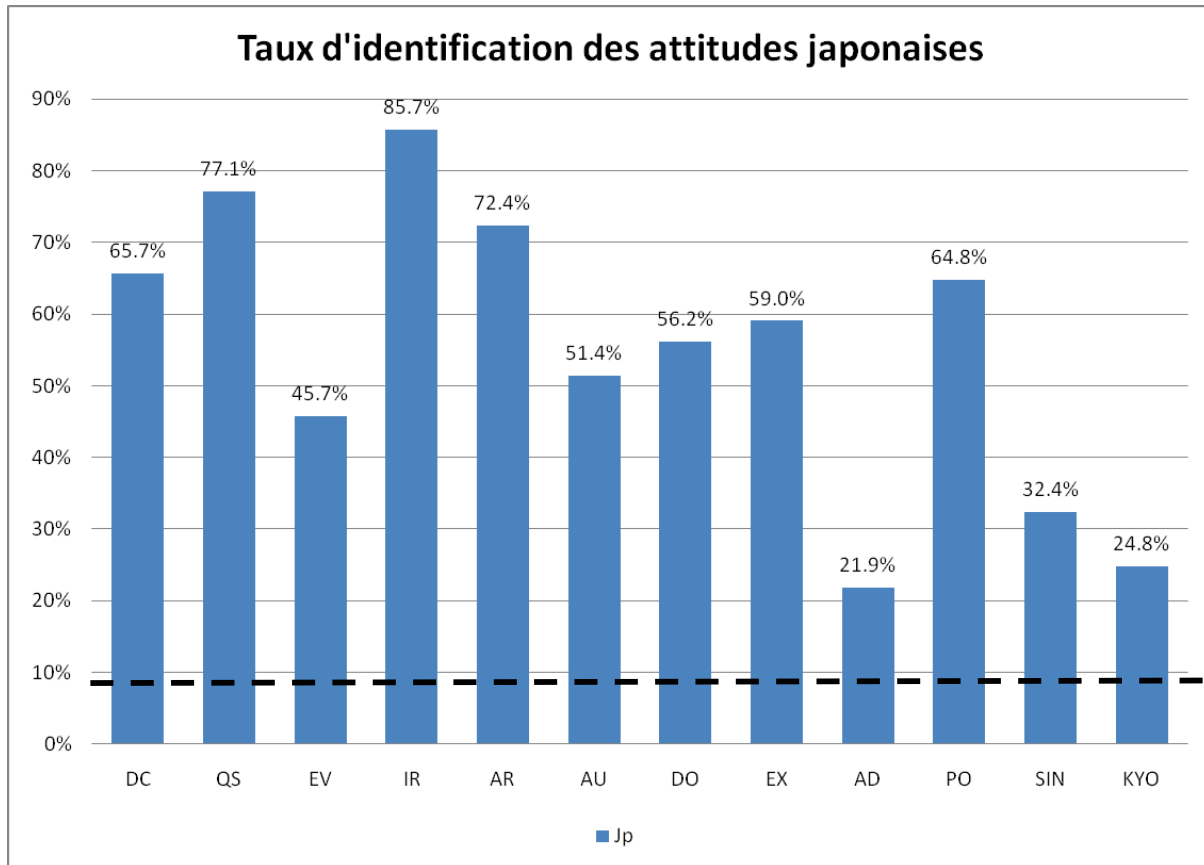


Figure 14 : pourcentage de reconnaissance de chaque attitude par les 15 auditeurs japonais. La ligne pointillée indique le seuil du hasard (8,4%)

Il existe cependant d'importantes disparités entre les taux de reconnaissance de ces attitudes. Ainsi l'attitude d'*irritation* est celle qui est le mieux identifiée. La plupart des autres attitudes sont elles-aussi largement reconnues : Il s'agit des expressions de *déclaration*, *question-simple*, *évidence*, *arrogance-impolitesse*, *autorité*, *doute-incrédulité*, *exclamation de surprise* et *politesse-simple*.

En revanche, les deux attitudes culturelles d'expression de la politesse (*kyoshuku* et *sincérité-politesse*) et l'attitude d'*admiration* reçoivent des scores largement plus bas que ceux des neuf autres attitudes.

Plusieurs explications sont envisageables pour les scores de reconnaissance les plus faibles. Les faibles taux de reconnaissance peuvent être expliqués par un effet des stratégies individuelles au niveau pragmatique. En effet, il est possible que les différents locuteurs d'une même langue choisissent des stratégies différentes pour exprimer des

attitudes comme le *kyoshuku* ou l'*admiration* (on pourra obtenir des éléments de réponse à cette suggestion au travers de l'analyse des résultats de l'expérience sur l'expression audio-visuelle des attitudes, qui est basée sur deux locuteurs – cf. infra). Lorsqu'un écart important peut être observé entre la stratégie prosodique véhiculée par le corpus et leur propre stratégie, les auditeurs peuvent avoir du mal à reconnaître ces attitudes.

Ces stratégies individuelles sont dissociables de l'effet dialectal lié à chaque auditeur. Sur ce point, Sadanobu (2004a) montre que certains locuteurs ayant participé à son expérimentation n'utilisaient pas la même stratégie que les autres afin d'exprimer la même attitude. Ce phénomène n'est pas étonnant car l'attitude est une stratégie acquise au sein de la société à laquelle on appartient.

Une seconde raison possible de cette variation des scores de reconnaissance est liée à un possible effet du contenu sémantique des phrases. Malgré les phrases que nous avons sélectionnées et qui ont été construites pour être le plus neutre possible sémantiquement, certains auditeurs ont ressenti un « décalage » entre le contenu sémantique (neutre) de l'énoncé et les attitudes exprimées. Autrement dit, les auditeurs n'arrivaient pas à conceptualiser la situation dans laquelle un énoncé associé à une attitude. Cette difficulté peut provenir justement du caractère neutre des phrases du corpus. On verra (cf. infra) qu'un changement des phrases utilisées pour l'enregistrement peut induire une modification des performances de reconnaissance.

Tableau 8 : Matrice de confusion des 15 auditeurs japonais. Signification des étiquettes : AD (*admiration*), PO (*politesse-simple*), KYO (*kyoshuku*), SIN (*sincérité-politesse*), AR (*arrogance-impolitesse*), AU (*autorité*), IR (*irritation*), DO (*doute-incrédulité*), EX (*exclamation de surprise*), QS (*question-simple*), EV (*évidence*) et DC (*déclaration*).

Attitudes perçues	Attitudes présentées											
	AD	AR	AU	DC	DO	EV	SU	IR	KYO	PO	QS	SIN
Admiration	22%	0%	0%	0%	2%	0%	1%	0%	2%	5%	0%	4%
Arrogance	2%	72%	6%	10%	6%	21%	0%	3%	2%	0%	1%	0%
Autorité	0%	10%	51%	3%	1%	10%	1%	11%	15%	0%	2%	1%
Déclaration	5%	6%	10%	66%	0%	12%	0%	0%	3%	9%	2%	10%
Doute-Incrédulité	0%	0%	0%	0%	56%	0%	14%	0%	0%	1%	4%	0%
Evidence	9%	6%	17%	8%	0%	46%	6%	0%	10%	2%	1%	4%
Surprise	10%	0%	0%	0%	14%	5%	59%	0%	1%	2%	1%	2%
Irritation	1%	6%	5%	1%	13%	3%	5%	86%	12%	0%	1%	1%
Kyoshuku	11%	0%	0%	0%	0%	0%	0%	0%	25%	10%	1%	28%
Politesse	27%	1%	5%	11%	0%	1%	0%	0%	3%	65%	9%	17%
Question	0%	0%	0%	0%	8%	3%	14%	0%	0%	1%	77%	2%
Sincérité-Politesse	14%	0%	6%	1%	0%	0%	0%	0%	27%	7%	2%	32%

Nous tenterons d'apporter des éléments de réponses à ces deux hypothèses par la suite, en réalisant de nouveaux tests de perception sur la base de données nouvelles qui

prennent en compte ces difficultés et permettent donc de les tester – ce qu’il n’est pas possible de faire maintenant.

Nous allons maintenant approfondir l’analyse des résultats en étudiant les confusions perçues entre les attitudes (cf. tableau 8), ce qui nous donnera des pistes de compréhension des difficultés de perception des attitudes les moins bien reconnues.

5.4.6.5. Analyse des résultats grâce à l’analyse des correspondances

Selon l’analyse des correspondances (AC), les deux tiers de la variance sont expliqués par les quatre premières dimensions (voir en annexe les résultats détaillés de l’AC).

La figure 15 est une représentation graphique due à l’AC. La partie haute de la figure représente la projection des points sur les deux premières dimensions de l’analyse. La partie basse représente la projection de ces points sur les dimensions 3 et 4. Les attitudes les mieux identifiées sont entourées par des cercles bleus. Il s’agit des sept attitudes suivantes : *arrogance-impolitesse*, *déclaration*, *doute-incrédulité*, *politesse-simple*, *exclamation de surprise*, *irritation* et *question-simple* qui n’ont pas eu de confusion particulière.

Les attitudes pour lesquelles des confusions se font jour sont désignées sur la figure par des disques violets. Il s’agit des cinq expressions suivantes : *autorité*, *évidence*, *kyoshuku*, *sincérité-politesse* et *admiration*. Nous allons maintenant détailler comment et avec quelle logique se sont produites ces confusions.

L’*autorité* a été confondue avec l’*évidence*, qui a elle-même été confondue avec l’*arrogance-impolitesse*. À la suite du test perceptif, deux sujets féminins nous ont indiqué avoir hésité entre les attitudes d’*évidence* et d’*arrogance-impolitesse* car l’expression d’*évidence* montre que le locuteur est sûr de lui-même et cette expression de certitude peut parfois être perçue comme irrespectueuse vis-à-vis de l’interlocuteur. En effet, dans la société japonaise, même si l’on est sûr de soi, il faut éviter d’afficher franchement cette certitude car cela risque d’être interprété comme une sorte d’imposition. Par conséquent, cette confusion de l’*évidence* avec l’*arrogance-impolitesse* est logique, de même que la confusion de l’*autorité* avec l’*évidence*, qui sont deux expressions d’imposition (on énonce avec autorité ce dont on est sûr).

Deux attitudes spécifiques de la culture japonaise, les expressions de *politesse sincérité-politesse* et *kyoshuku*, ont été largement confondues entre elles par les auditeurs. Bien que le contexte d’utilisation de ces deux attitudes soit différent, la *sincérité-politesse* et le *kyoshuku* ressortissent essentiellement de la même catégorie d’attitudes (Sadanobu, 2003, 2004a). Elles servent en effet essentiellement à montrer son humilité face à une personne supérieure dans la hiérarchie sociale. Il est à noter que la *sincérité-politesse* est également confondue avec la *politesse-simple*, tandis que

kyoshuku ne montre pas cette confusion, ce qui indique tout de même un traitement différencié de ces deux types d'expressions par les auditeurs.

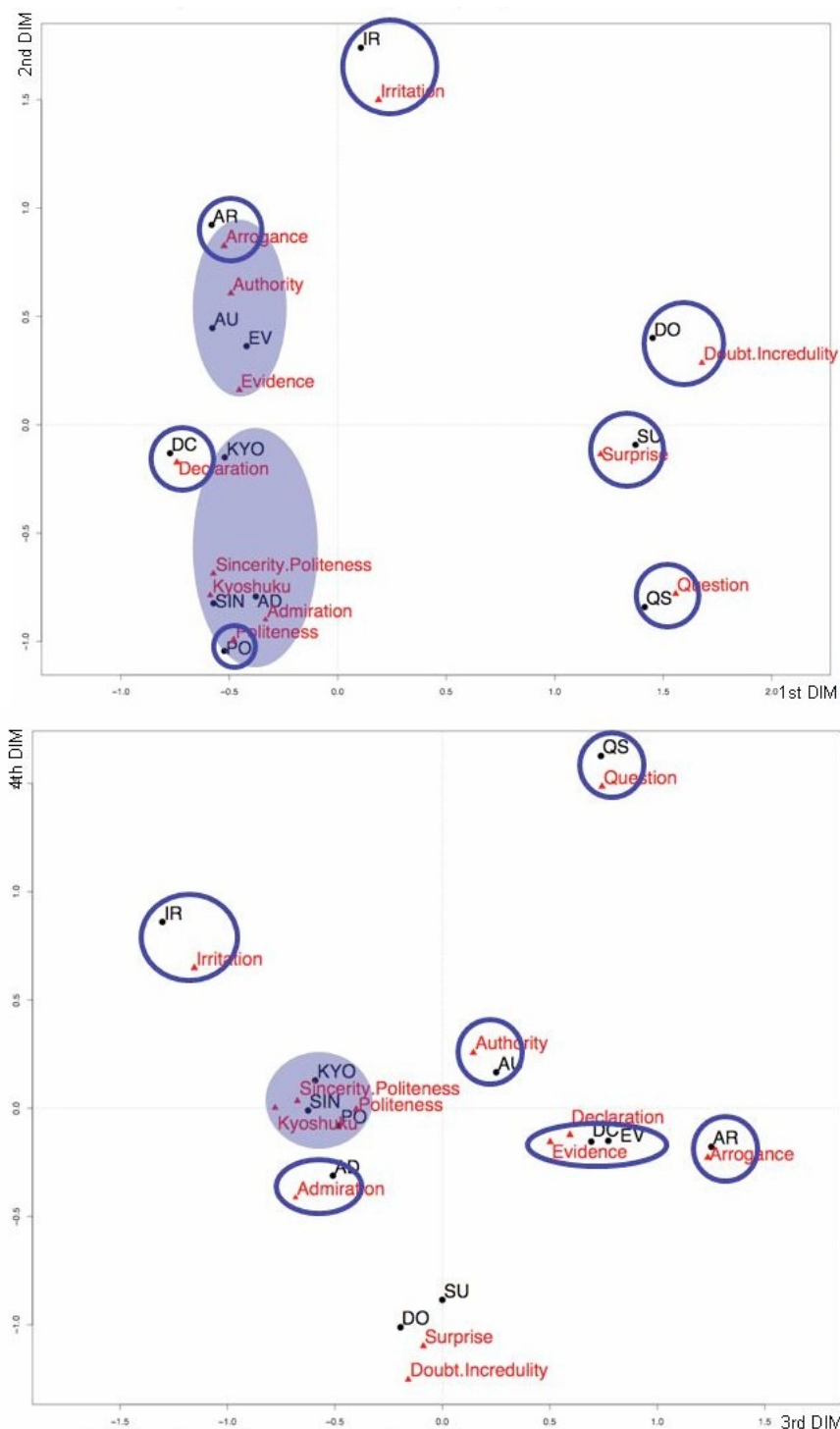


Figure 15 : distribution des résultats de 15 auditeurs japonais pour les 12 attitudes japonaises obtenus par l'AC (la figure du haut présente les dimensions 1 et 2, et celle du bas les dimensions 3 et 4). Les triangles rouges correspondent aux stimuli, les ronds noirs correspondent aux catégories d'attitudes. Les cercles bleus indiquent les attitudes les mieux reconnues, tandis que les disques violets soulignent les principales confusions entre attitudes.

L'attitude d'*admiration* est assez largement confondue avec la *politesse-simple*. L'expression de l'admiration et les expressions de politesse sont interconnectées les unes aux autres dans la société japonaise, ce que montre la polysémie lexicale des items « *Sonkee* » (admiration / politesse-simple) et « *Keifuku* » (admiration / politesse-simple).

Au-delà des regroupements qui viennent d'être analysés, l'AC oppose les expressions de *question-simple*, d'*exclamation de surprise* et de *doute-incrédulité* à l'ensemble des autres attitudes (cf. la première dimension de l'AC). Ces trois attitudes, qui sont identifiées sans confusion importante entre elles, partagent pourtant des caractéristiques similaires qui les rapprochent entre elles et les opposent aux autres. Toutes trois ont une intonation montante en fin de phrase. Elles véhiculent aussi un sens général similaire, qui renvoie à l'interlocuteur une interrogation qui se décline (1) en question par rapport à la sémantique de la phrase (dans le cas de la *question-simple*), ou en réaction à ce que vient de dire l'interlocuteur : (2) réaction de surprise ou (3) réaction de doute.

La deuxième dimension de l'AC oppose les attitudes exprimant l'« attention » du locuteur (*sincérité-politesse*, *politesse-simple*, *admiration* et l'expression de *kyoshuku*), aux expressions « d'imposition » de soi (*autorité*, *arrogance-impolitesse*, *irritation*). Il est important de noter que le point correspondant à la notion de *kyoshuku* (c-à-d, le vecteur de réponses *kyoshuku* donnée par les sujet en réponse à chaque type d'attitude) est situé entre les expressions « d'attention » et celles « d'imposition », indiquant un nombre comparable de réponses *kyoshuku* pour les deux types d'expressions. Cela souligne la complexité de cette attitude, et renvoie à la proximité acoustique existant entre *kyoshuku* et *irritation* du point de vue de la qualité vocale des deux types d'expressions (cf. infra). Cette opposition entre « attention » et « imposition » renvoie à la notion de « *care* » (c-à-d, d'attention) proposée par Campbell (2005).

5.4.6.6. Classification hiérarchique

L'analyse des résultats grâce à la classification hiérarchique (cf. figure 16) fait apparaître un certain nombre de regroupements, qui correspondent aux analyses déjà proposés (cf. supra). Au plus haut niveau, trois principaux groupes d'expressions attitudinales existent :

- *évidence*, *autorité*, *déclaration*, *arrogance-impolitesse*
- *kyoshuku*, *sincérité-politesse*, *admiration*, *politesse*
- *exclamation de surprise*, *doute-incrédulité*, *question-simple*

L'expression d'*irritation*, très bien reconnue, reste à part. Chacun de ces trois groupes correspond aux oppositions décrites grâce à l'analyse des correspondances, et donc à trois grandes catégories expressives : les expressions « d'imposition » de soi ; les expressions « d'attention » ; et les expressions « interrogatives » au sens large.

Les proximités les plus importantes entre deux attitudes que cette analyse permet de relever concernent les attitudes de *kyoshuku* et de *sincérité-politesse*, qui

renvoient toute les deux à une politesse marquée. Même si ces deux expressions correspondent à des contextes bien particuliers, il semblerait qu'à partir d'une phrase ou justement le locuteur n'a aucun moyen de retrouver le contexte d'énonciation, l'expressivité seule puisse difficilement suffire à réaliser une nette distinction entre les deux, même si les caractéristiques acoustiques sont clairement différentes (surtout en ce qui concerne la qualité de voix – cf. infra). On verra aussi que des indices dans d'autres modalités (e.g. expressions faciales) peuvent venir en aide aux auditeurs pour distinguer ces deux types d'expressions (cf. infra).

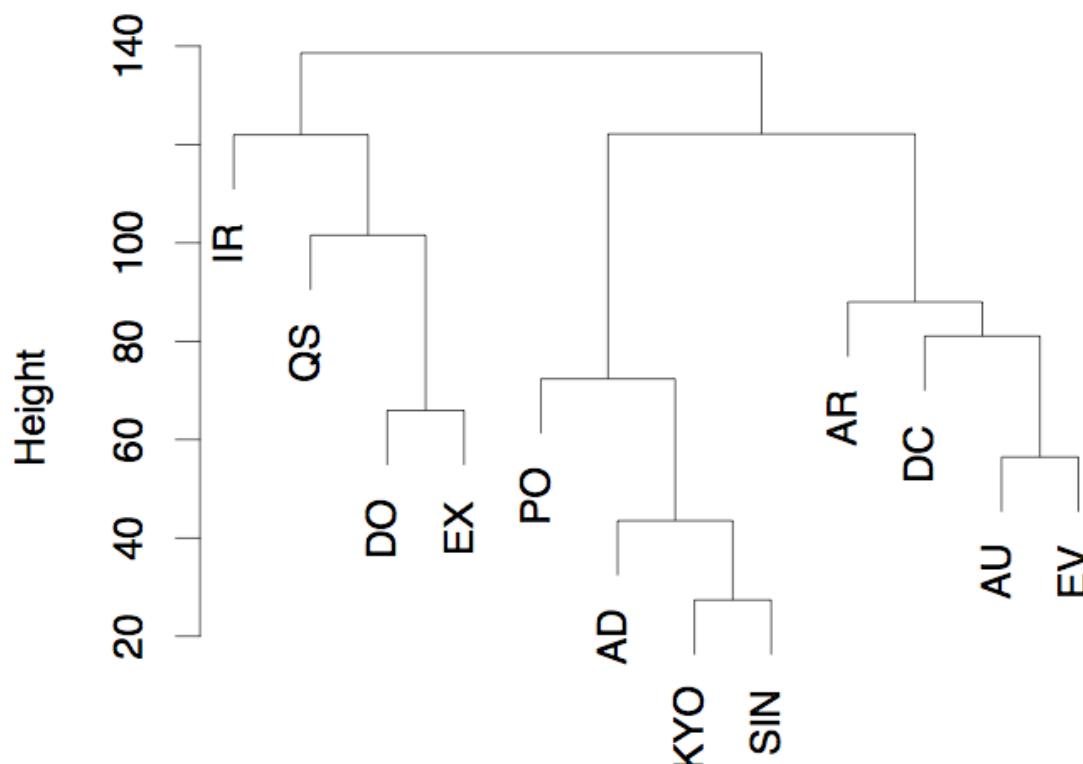


Figure 16 : résultats de l'analyse de classification hiérarchique, obtenue grâce à la distance de Ward sur les réponses moyennes de chaque attitude.

5.4.7. Analyse acoustiques

5.4.7.1. Paramètre prosodiques et analyse

Dans ce paragraphe, nous allons analyser le lien entre les caractéristiques acoustiques et les catégories perceptives que nous avons identifiées dans les paragraphes précédents.

Pour cela, nous avons extrait les valeurs acoustiques telle que la durée, l'intensité et la fréquence fondamentale de chaque more de chaque phrase du corpus d'une part, et d'autre part, une mesure d'indices perceptifs de la qualité de voix. À savoir une mesure de la présence des catégories de qualité de voix données par Laver (1980) : la voix modale, soufflée, rauque, tendue, nasale et dévoisée. Ces mesures ont été réalisées en notant

pour chaque voyelle de chaque phrase et pour chaque attitude, la présence (valeur 1) ou l'absence (valeur 0) de chacune de ces qualités de voix. Cette description perceptive de la qualité de voix des stimuli vient pallier la grande complexité d'une analyse acoustique précise et automatique de ces qualités de voix. Différents corrélats acoustiques sont proposés dans la littérature, comme la différence entre les deux premières harmoniques, ou le « spectral tilt » (pour une étude détaillée, voir d'Alessandro, 2006), mais la liaison entre les descriptions de qualité de voix faites par Laver (1980) et ces paramètres acoustiques est encore à l'heure actuelle un sujet de recherche à part entière (cf. Garnier et al., 2007). Par ailleurs, les procédures de traitement du signal permettant d'extraire ces paramètres sont complexes et peu (voir pas) automatisées. Ces corrélats acoustiques de la qualité de voix ne constituent donc pas à l'heure actuelle des paramètres ni faciles à extraire de manière fiable, ni exempts de variations importantes dues au contexte phonémique par exemple (voir par exemple, les travaux de Rossato et al., 2004, sur la mesure du paramètre de NAQ).

Nous avons donc choisi de baser cette analyse de la qualité de voix sur une mesure perceptive des qualités de voix proposées par Laver (1980), plutôt qu'utiliser des procédures d'analyse acoustique complexes et encore discutées par les spécialistes de ce domaine. Notons qu'il est, bien entendu, extrêmement intéressant (et cela est d'ailleurs prévu pour la suite de nos travaux) d'extraire des corpus de cette thèse un ensemble de paramètres acoustiques de la qualité de voix, afin de mieux comprendre leur pertinence. Cela est d'autant plus vrai que nous avons réalisé tous ces corpus en enregistrant les signaux d'électroglotogrammes pour ce but précis. Les contraintes temporelles liées à ce travail de thèse nous empêchent de le faire dans le présent document.

5.4.7.2. Extraction des paramètres prosodiques

Les paramètres acoustiques de la prosodie sont extraits automatiquement par les signaux étiquetés manuellement, avec les scripts Matlab en utilisant l'algorithme d'extraction de la fréquence fondamentale (f_0) « yin » (Cheveigné & Kawahara, 2002). Trois paramètres acoustiques sont extraits : la F_0 (en demi-tons avec comme valeur de référence 1 Hz), la durée moraïque (en secondes) et l'intensité (en dB). Pour le paramètre de F_0 , trois valeurs sont extraites pour chaque voyelle en début, milieu et fin de la voyelle – plus précisément à 10, 50 et 90% de la longueur de la voyelle, afin d'éviter les variations de F_0 dues aux phénomènes microprosodiques.

Pour chaque phrase un ensemble d'indices est calculé sur la base de ces trois paramètres acoustiques. Une partie de ces indices sont inspirés par les travaux de Scherer et de son équipe (Banse & Scherer, 1996 ; Scherer & Ellring, 2007) : il s'agit de la valeur moyenne et de l'écart type des valeurs de F_0 , de la durée moraïque et de l'intensité. Ces indices donnent un aperçu général des valeurs des paramètres acoustiques pour chaque phrase et chaque attitude. En plus, des indices portant sur la forme du contour des paramètres tout au long de la phrase ont été ajoutés. Il s'agit de :

- **la pente de F0**, calculée comme la différence entre la valeur moyenne de la dernière voyelle voisée de la phrase (si elle existe) et la valeur moyenne de la première voyelle de la phrase ;
- **la pente de F0 de la dernière voyelle voisée**, calculée comme la différence entre la dernière et la première valeur de F0 de la dernière voyelle voisée de la phrase (si elle existe) ;
- **l'allongement de la dernière more**, calculée comme la différence de durée entre la dernière et l'avant-dernière more ;
- **la pente d'intensité**, calculée comme la différence entre la valeur d'intensité de la dernière voyelle moins la valeur de la première voyelle.

Ces dix paramètres prosodiques (résumés dans la table 9) sont extraits pour chaque expression attitudinale et pour chaque phrase.

Tableau 9 : liste des indices prosodiques calculés à partir des mesures acoustiques de F0, durée moraïque et intensité et de qualité vocale, ainsi que les abréviations utilisées pour les désigner par la suite et le type de ces indices (c-à-d, s'il s'agit d'une mesure globale à la phrase ou d'une mesure permettant de juger de l'évolution d'un paramètre, de la forme d'un contour)

Paramètre	Abréviation	Mesure acoustique	Type
pente de F0	F0slope	F0	forme
pente de F0 de la dernière voyelle	F0slope_lastVoy	F0	forme
moyenne de F0	F0mean	F0	global
écart-type de F0	F0std	F0	global
allongement de la dernière more	AlgtFinal	durée moraïque	forme
moyenne de durée	DURmean	durée moraïque	global
écart-type de durée	DURstd	durée moraïque	global
pente d'intensité	INTslope	intensité	forme
moyenne d'intensité	INTmean	intensité	global
écart-type d'intensité	INTstd	intensité	global
modal	modal	qualité de voix	global
soufflée	breathy	qualité de voix	global
rauque	harsh	qualité de voix	global
tendu	tendu	qualité de voix	global
nasalisé	nasal	qualité de voix	global
dévoisé	dévoisement	qualité de voix	global

Les mesures de la qualité de voix se sont faite selon la procédure suivante. Dans un premier temps, nous avons fait écouter nos stimuli à un spécialiste travaillant sur l'analyse de la voix parlée et chantée, et s'intéressant particulièrement aux problématiques de qualité de voix et de modélisation de la source vocale. Celui-ci a donc noté pour chaque type d'expressions attitudeinales les types de qualité de voix qu'il percevait. À partir de ce premier travail, nous avons réécouté l'ensemble de nos corpus.

Pour chaque more de chaque phrase, réalisée dans chaque attitude, nous avons noté par 0 (absence) ou 1 (présence) de chacune des qualités de voix suivantes : modale, soufflée, rauque, tendue, nasale, dévoisée. Par la suite, pour chaque phrase, un score moyen est attribué, d'après la moyenne des scores obtenus pour chaque more de la phrase. On obtient ainsi une matrice donnant pour chaque phrase et pour chaque attitude une mesure de la qualité de voix.

L'ensemble de ces valeurs est utilisé en entrée des analyses en composantes principales (ACP) et la classification hiérarchique qui sont présentées ci-après.

5.4.7.3. Analyse en composantes principales & classification hiérarchique

En utilisant les indices prosodiques décrits dans le paragraphe précédent, nous avons mené des analyses statistiques pour tenter de mesurer l'apport de ces différents indices à la perception des différentes attitudes. Ces analyses sont basées d'une part sur une Analyse en Composantes Principales (ACP) qui permet de visualiser l'importance relative des différents indices acoustiques pour la réalisation de chaque attitude. D'autre part, l'analyse est faite grâce à une classification hiérarchique des attitudes d'après leurs caractéristiques acoustiques afin de mesurer les distances acoustiques existantes entre ces différentes expressions et de pouvoir vérifier si ces distances recoupent les distances perceptives.

Ces analyses sont faites sur les paramètres décrits ci-dessus. Or la nature des paramètres que nous avons à notre disposition diffère : les indices prosodiques tirés de la mesure acoustique de la F_0 , de la durée et de l'intensité sont des paramètres continus, tandis que l'évaluation de la qualité de voix est basée sur un code binaire. Afin de rendre ces deux types de mesure comparables, nous nous sommes basés sur la méthode utilisée par Scherer & Ellring (2007) dans un cas similaire, pour créer deux nouvelles variables (binaires) à partir de chaque indice prosodique, sur la base du z-score : pour chaque indice prosodique (par exemple F_0slope), une variable est créée pour les valeurs faibles de cet indice (par exemple $F_0slopePT$), et une autre variable pour les valeurs élevées (par exemple $F_0slopeGD$). Ainsi si le z-score de l'indice est inférieur à -0.5, la variable de notre exemple $F_0slopePT$ recevra la valeur 1, et elle recevra 0 si le z-score est supérieur à -0.5. De même, $F_0slopeGD$ recevra la valeur 1 si le z-score dépasse 0.5, et recevra 0 sinon. Dans le cas d'un z-score compris entre -0.5 et 0.5, les deux variables seront à 0.

Du fait de cette divergence dans la nature des indices mesurés, qu'ils soient prosodiques ou de qualité de voix, trois analyses différentes seront présentées. La première présentera l'apport des indices prosodiques (continus) seuls, la seconde présentera l'apport des seuls paramètres de qualité de voix, tandis que la troisième prendra en compte l'ensemble de ces deux types de paramètres, selon le recodage présenté ci-dessus. Cela est fait pour ne pas perdre l'information plus fine apportée par l'analyse acoustique des paramètres prosodiques.

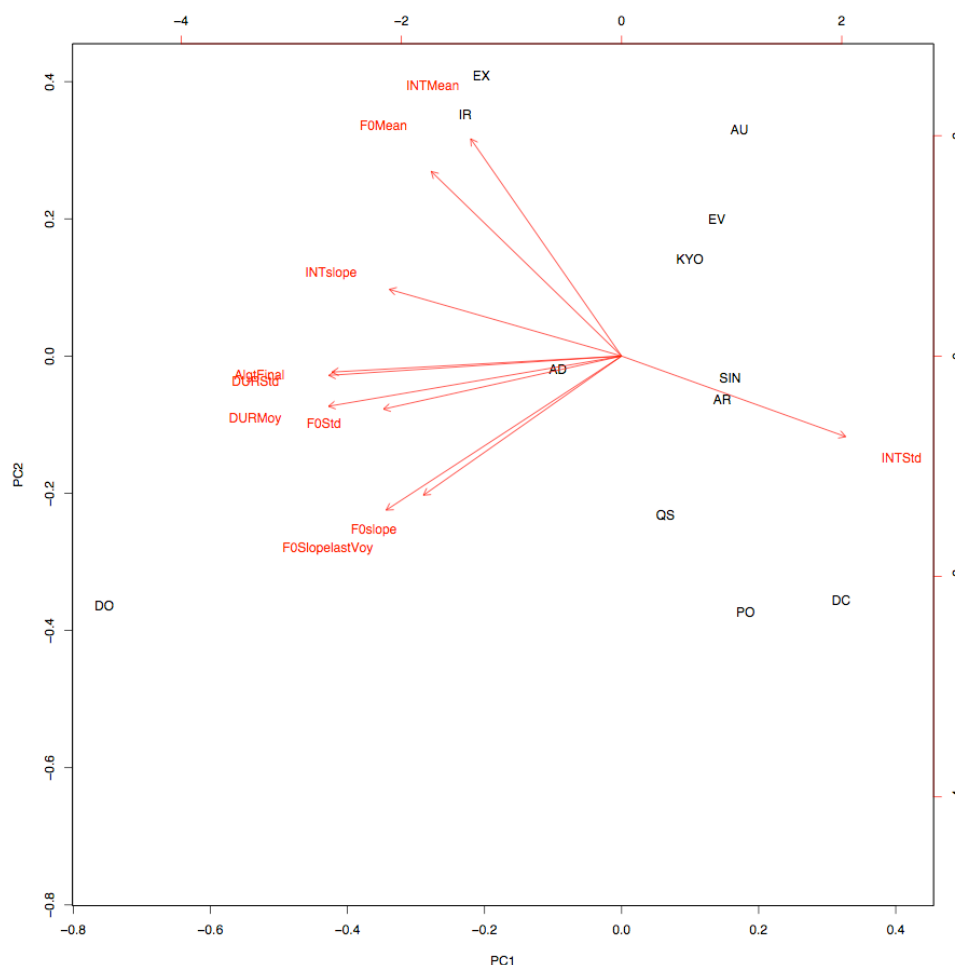


Figure 17 : Deux dimensions principales de l'ACP pour les indices prosodiques. Les étiquettes rouges représentent les projections des vecteurs d'indices sur les deux premières composantes de l'ACP. Les attitudes (abréviations noires) sont projetées sur ces deux principales composantes, qui représentent 76% de la variance.

5.4.7.4. Résultats de l'ACP sur les paramètres prosodiques

L'ACP basée sur les indices prosodiques continus montre les caractéristiques distinctives suivantes.

Un niveau élevé de la F0 moyenne et de l'intensité moyenne caractérisent l'*exclamation de surprise* et l'*irritation* ; tandis que les valeurs les plus basses sont obtenues la *politesse simple* et la *déclaration*.

L'écart type d'intensité, opposé à la pente d'intensité, permet de distinguer la *politesse simple* et la *déclaration* d'une part (qui montrent plus de variabilité de l'intensité) des expressions de *doute-incrédulité*, d'*irritation* et d'*exclamation de surprise* d'autre part (qui montrent la pente d'intensité la plus grande).

La pente de F0 (sur l'ensemble de l'énoncé comme sur la dernière syllabe) oppose très clairement l'expression de *doute-incrédulité* (qui montre une montée finale très

importante) aux expressions d'*autorité* et d'*évidence* (ayant une F0 clairement descendante)

Toutes les mesures d'allongement de la durée (durée moyenne et écart-type ou allongement final), de même que l'écart-type de la F0, opposent le *doute-incrédulité* (et à un niveau beaucoup plus faible avec l'*admiration*, l'*irritation* et l'*exclamation de surprise* et la *question*) aux expressions de *déclaration*, d'*autorité*, d'*évidence*, d'*arrogance-impolitesse*, de *kyoshuku*, de *sincérité-sérieux* et de *politesse*.

Tableau 10 : Proportion de la variance expliquée par chaque composante de l'ACP sur les indices prosodiques.

Importance of components:

Dimension	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2,48	1,23	1,07	0,69	0,61	0,44	0,39	0,08	0,07	0,03
Proportion of Variance	61%	15%	12%	5%	4%	2%	1%	0%	0%	0%
Cumulative Proportion	61%	76%	88%	93%	96%	98%	100%	100%	100%	100%

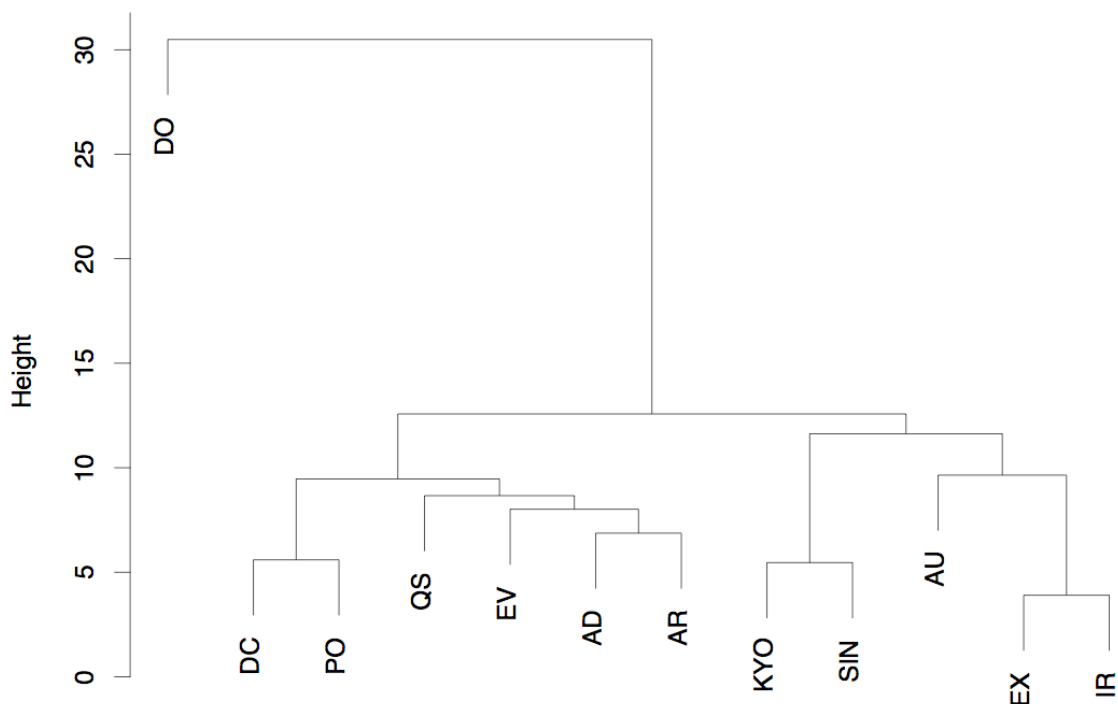


Figure 18 : classification hiérarchique des attitudes à partir de leurs indices prosodiques.

5.4.7.5. Classification hiérarchique des paramètres prosodiques

La classification hiérarchique des 12 attitudes d'après les valeurs indices prosodiques permet de faire les regroupements et les distinctions suivantes (cf. figure 18). La constatation la plus évidente et la différence majeure qui existent entre l'expression de *doute-incrédulité* et toutes les autres attitudes, ce qui confirme les résultats de l'ACP.

Les attitudes qui entretiennent le plus de confusions entre elles sont les expressions de *déclaration* et de *politesse simple* ; celles de *kyoshuku* et de *sincérité-politesse* ; et celles d'*exclamation de surprise* et d'*irritation*.

L'expression d'*autorité* est assez bien distinguée des autres, tandis que les expressions d'*arrogance-impolitesse*, d'*admiration*, d'*évidence* et de *question* montrent des oppositions assez claires, même si elles restent loin des valeurs extrêmes des indices prosodiques de l'attitude de *doute-incrédulité*.

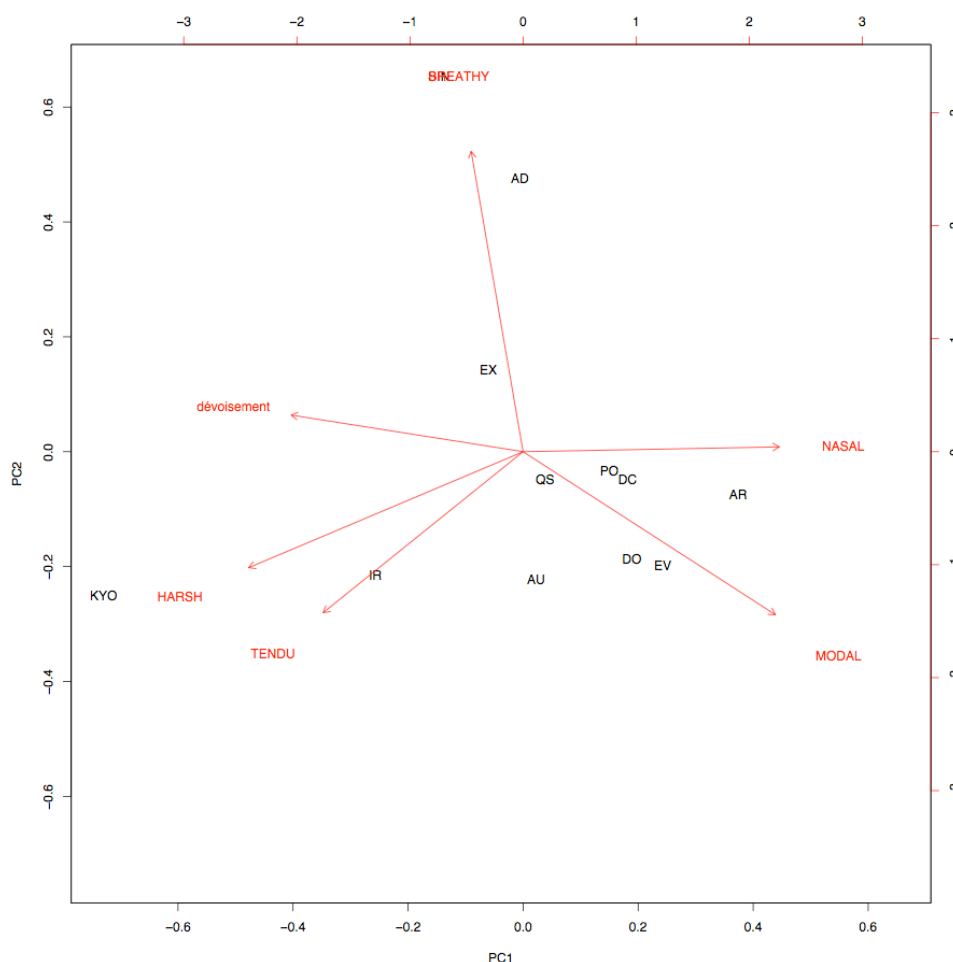


Figure 19 : représentation des deux premières dimensions de l'ACP pour la qualité de voix. Les étiquettes rouges indiquent la projection des paramètres de qualité de voix ; tandis que les attitudes (en noir) sont projetées sur ces deux mêmes dimensions qui expliquent presque 80% de la variance.

Tableau 11 : Proportion de la variance expliquée par chaque composante de l'ACP sur la qualité de voix.

Importance of components:						
Dimension	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1,751	1,268	0,893	0,62	0,3337	0,18953
Proportion of Variance	51%	27%	13%	6%	2%	1%
Cumulative Proportion	51%	78%	91%	98%	99%	100%

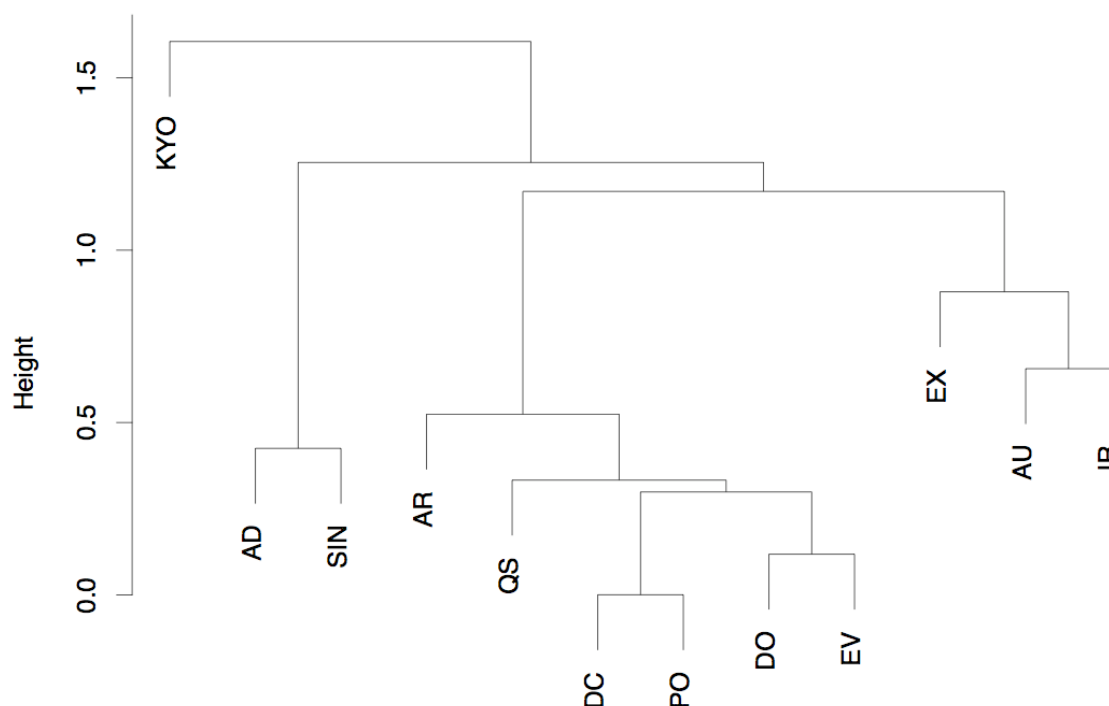


Figure 20 : catégorisation des attitudes à partir de la similarité des valeurs de qualité de voix, obtenue par une classification hiérarchique.

5.4.7.6. Analyse de la qualité de voix

Les six différentes qualités de voix mesurée au travers de nos corpus (modale, soufflée, rauque, tendue, nasale et dévoisement) permettent de différencier les attitudes de la manière qui suit.

Les résultats de l'ACP (cf. figure 19) montrent que la voix soufflée est principalement utilisée pour exprimer la *sincérité-politesse* et l'*admiration*. Elle apparaît aussi, mais très rarement pour l'*exclamation de surprise*. L'attitude de *kyoshuku* est caractérisée par une voix rauque et tendue, ainsi que l'*irritation* à un degré légèrement plus faible ; mais le *kyoshuku* est aussi fortement dévoisé, alors que l'*irritation* l'est beaucoup moins. L'*autorité* est marquée par une voix uniquement tendue. Quant à l'*arrogance-impolitesse*, elle est plus nasalisée que les autres attitudes. Les autres attitudes ne sont pas marquées par une qualité de voix particulière – si ce n'est qu'elles sont généralement perçues comme plus ou moins modales.

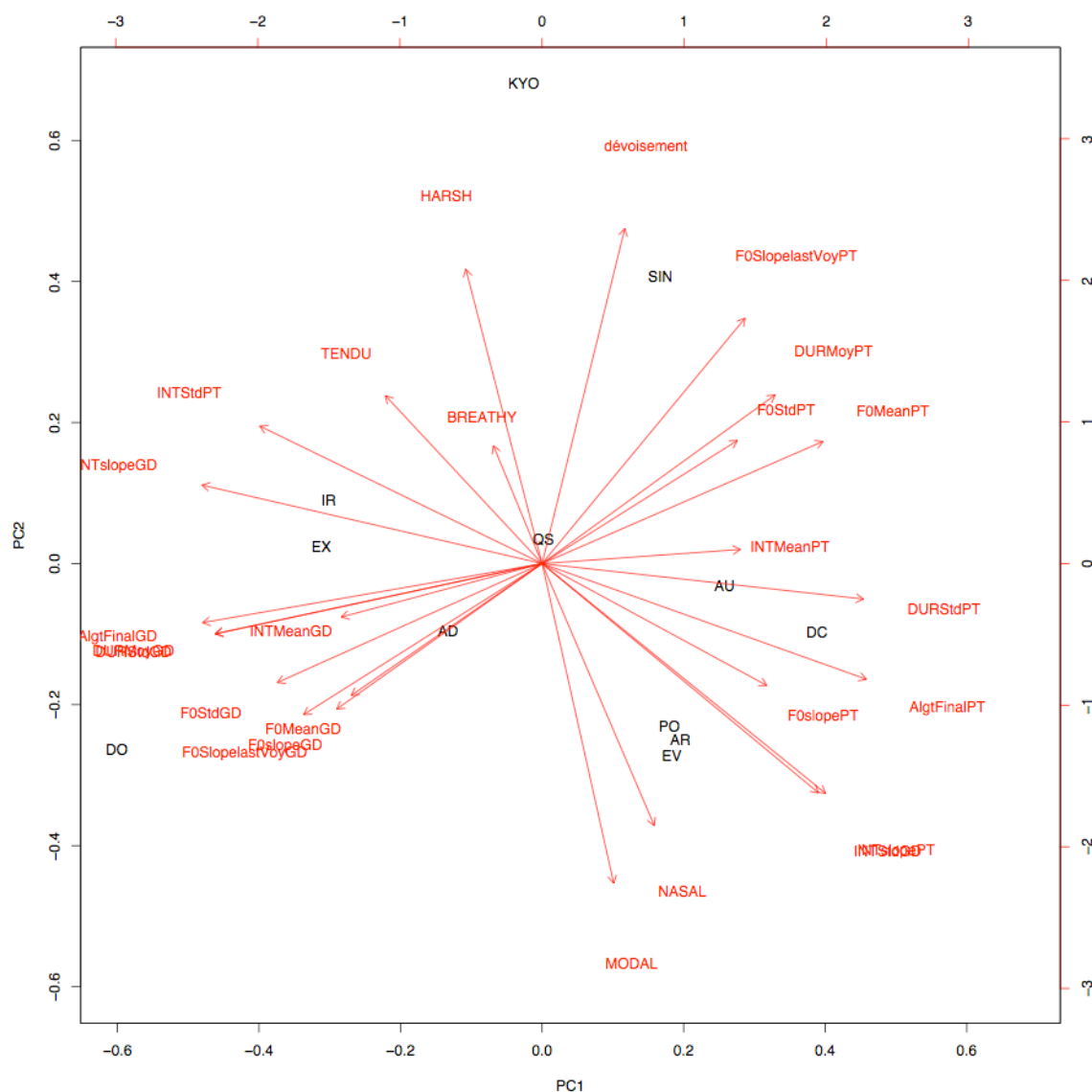


Figure 21 : représentation des deux premières dimensions de l'ACP pour les indices de prosodie et de qualité de voix. Les étiquettes rouges indiquent la projection des paramètres acoustiques ; tandis que les attitudes (en noir) sont projetées sur ces deux mêmes dimensions qui n'expliquent que 56% de la variance.

5.4.7.7. Classification hiérarchique de la qualité de voix

À partir de cette distribution des paramètres de qualité de voix pour chaque attitude, une classification hiérarchique (cf. figure 20) permet de distinguer avant tout l'expression de *kyoshuku*. Les deux expressions : *sincérité-politesse* et d'*admiration*, sont elles aussi très bien différenciées des autres, mais très proches l'une de l'autre. Les expressions d'*exclamation de surprise*, d'*autorité*, d'*irritation* et d'*arrogance-impolitesse* utilisent elles aussi des jeux de paramètres qui leur sont proches, et permettent de les séparer des autres. On retrouve finalement l'ensemble des attitudes modales (*question*, *déclaration*, *politesse-simple*, *doute-incrédulité* et *évidence*) mélangées ensembles.

5.4.7.8. Analyse prosodie et qualité de voix

Nous avons finalement regroupé les indices prosodiques avec la mesure de la qualité de voix, selon le principe expliqué ci-dessus, afin de mieux décrire ces expressions attitudinales. L'ACP (cf. figure 21) menée sur ces paramètres montre effectivement une meilleure distribution des 12 attitudes dans l'espace des deux premières composantes. Il serait cependant hasardeux de risquer ici une analyse de l'apport relatif de chaque paramètre à chaque attitude, étant donné la complexité de cet espace multiparamétrique : les deux premières composantes expliquent seulement 56% de la variance, et il faut tenir compte des 4 premières pour en expliquer 78%.

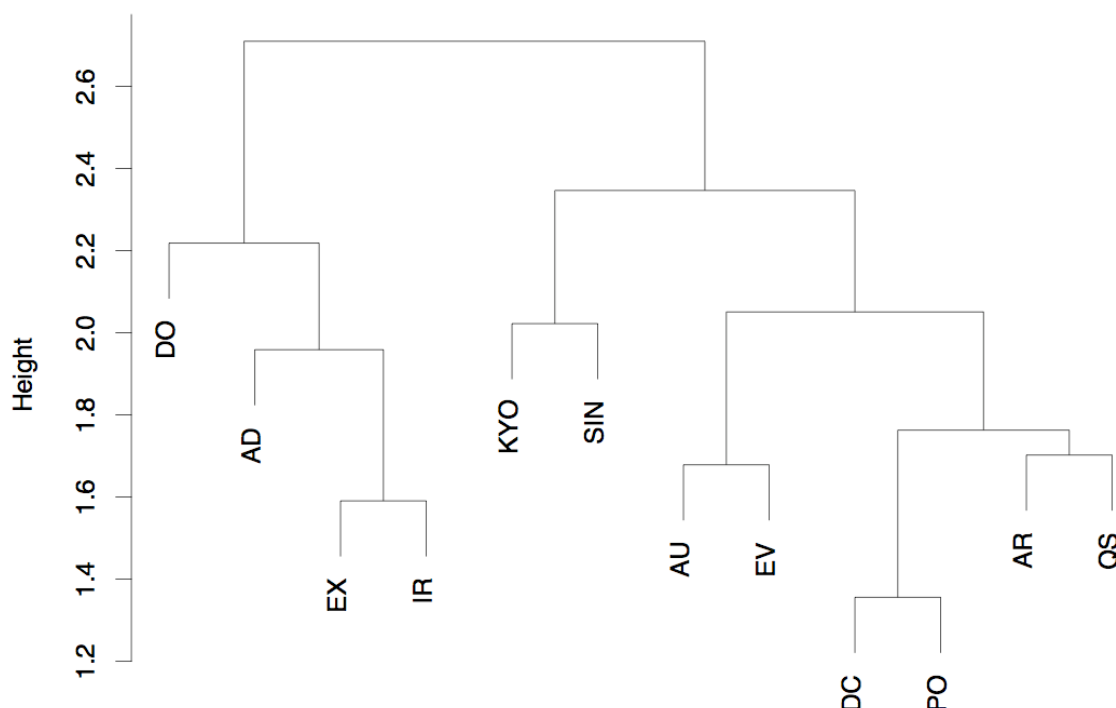


Figure 22 : catégorisation des attitudes à partir de la similarité des paramètres prosodiques et de qualité de voix par une classification hiérarchique.

Cependant l'utilisation conjointe de l'ensemble de ces paramètres permet une meilleure classification des expressions attitudinales (cf. figure 21 & 22). En effet, seules les attitudes de *déclaration* et de *politesse-simple* sont peu différenciées par les paramètres extraits du corpus. Toutes les autres attitudes sont assez bien différenciées les unes des autres.

Les mieux reconnues sont cependant les expressions de *doute-incrédulité*, celles de *kyoshuku* et de *sincérité-politesse*²⁵ et *l'admiration*. Les autres, même si elles sont assez bien distinguées, montrent des proximités qui sont dues à l'utilisation d'un (ou

²⁵ Ces deux attitudes, de manière intéressante, partagent des paramètres prosodiques semblables, mais sont bien différenciées par la qualité de voix.

plusieurs) indice prosodique ou de qualité de voix commun (comme des valeurs de F0 moyen et d'intensité moyenne similaires pour l'*autorité* et l'*évidence* et aussi pour l'*exclamation de surprise* et l'*irritation*).

Cette analyse permet donc d'extraire un ensemble pertinent d'indices prosodiques et de qualité de voix, qui permettent de qualifier les douze expressions attitudinales.

5.4.8. Conclusion

À la suite des recherches développées sur les attitudes prosodiques (Aubergé et al., 1997 ; Morlec, 1997 pour les attitudes françaises, Diaferia, 2002 pour l'anglais britannique), nous avons intégré les attitudes japonaises dans ce cadre de recherche. Si l'on compare aux résultats obtenus pour les attitudes occidentales²⁶ testées, les Japonais, qui proviennent d'une culture distincte, maîtrisent des attitudes qui ne sont pas conventionalisées dans la culture occidentale.

Après avoir effectué un large travail bibliographique sur les attitudes japonaises, et notamment sur les travaux extrêmement intéressants qui traitent de l'expression de la politesse en japonais (Sadanobu 2004a, Ofuka, 2002 ; Ito, 2003 etc.), nous avons sélectionné 12 attitudes représentatives de la langue japonaise, puis nous avons validé ces expressions auprès de quinze auditeurs japonais parlant le dialecte de Tokyo.

Une série de tests de khi-deux ont permis de relever les points suivants. D'abord, les distributions des réponses données pour chaque attitude par les sujets est significativement différente du hasard. Un second point important est que la position de l'accent lexical du japonais, qui est réalisé par une importante montée de fréquence fondamentale, n'a pas d'influence significative sur le comportement perceptif des auditeurs japonais. En revanche, ce test statistique montre un effet significatif de la longueur des énoncés. Certaines phrases reçoivent des scores significativement différents : c'est le cas entre les phrases de 2 et de 5 mores, ou de 5 et de 8 mores. Notons que cet effet « de longueur » pourrait aussi être dû au contenu lexical des phrases, même si celui-ci a été sélectionné pour être le plus neutre possible (cf. discussion supra sur l'influence possible de l'absence de contexte affectif sur les réponses des sujets)

Dans un premier temps, les analyses perceptives menées sur des auditeurs japonais valident globalement la qualité du corpus. Le taux d'identification de 12 attitudes étant systématiquement supérieur à celui du hasard.

L'analyse des distributions et des regroupements entre attitudes à l'aide des deux méthodes statistiques proposées (analyse des correspondances et classification hiérarchique) a permis de mettre en évidence les expressions les mieux reconnues et celles qui montrent le plus de confusions. Les sept attitudes suivantes sont bien reconnues par les sujets, sans confusions importantes avec d'autres expressions : *arrogance-impolitesse*, *déclaration*, *doute-incrédulité*, *politesse-simple*, *exclamation de*

²⁶ Les attitudes occidentales signifient, dans cette thèse, les attitudes françaises et anglaises.

surprise, *irritation* et *question-simple*. Par contre, les cinq attitudes suivantes ont été largement confondues avec d'autres attitudes : *autorité*, *évidence*, *kyoshuku*, *sincérité-politesse* et *admiration*.

Si nous prêtons une attention plus particulière aux expressions de politesse, le *kyoshuku* et la *sincérité-politesse* ont été largement confondus l'un avec l'autre par les sujets japonais. Pourtant ces deux attitudes sont réalisées, dans notre corpus, à l'aide de deux qualités de voix différentes (une voix rauque pour le *kyoshuku* et la voix soufflée pour la *sincérité-politesse*), cependant elles sont pragmatiquement reliées par la même notion d'humilité que leur utilisation en face d'un interlocuteur supérieur socialement implique. C'est cette similitude d'utilisation pragmatique qui a sans doute primé pour que les locuteurs regroupent ces deux catégories. L'attitude de *sincérité-politesse* est confondue avec l'expression de *kyoshuku*, mais aussi avec la *politesse-simple*. C'est-à-dire qu'elle montre des confusions avec l'ensemble des expressions sociales de la politesse japonaise. Ces trois expressions de politesse, plus l'expression d'admiration sont regroupées par la procédure de classification hiérarchique, montrant qu'elles entretiennent pour les auditeurs des proximités cognitives de haut niveau, au-delà de leur taux d'identification propre.

Les trois attitudes de *question-simple*, *d'exclamation de surprise* et de *doute-incrédulité*, chacune bien identifiée, sont cependant regroupées par les analyses, et clairement séparées de l'ensemble des autres attitudes. Les sujets japonais opposent donc clairement ces trois attitudes, ce qui permet de souligner qu'elles ont en commun une intention similaire de la part du locuteur qui « interroge » son interlocuteur. Cette similitude d'intention est soutenue par la présence pour ces trois attitudes d'une intonation montante en fin de phrase.

De manière similaire, les attitudes d'*autorité*, d'*arrogance-impolitesse* et d'*irritation*, bien reconnues individuellement, sont elles aussi regroupées dans une catégorie plus générale unie par la notion d'imposition de son avis à l'interlocuteur.

Parallèlement, l'analyse objective basée sur les indices prosodiques et de qualité de voix de l'ensemble de ces attitudes nous donne les informations suivantes.

- En premier lieu, l'*exclamation de surprise* et l'*irritation* sont caractérisées par le niveau élevé de l'intensité moyenne et de F0 moyenne, et par une voix tendue.
- L'expression de *doute-incrédulité* est caractérisée par les indices suivants : une importante pente montante de F0 sur l'ensemble de l'énoncé et sur la dernière more, de même qu'un écart type de F0 élevé, un allongement important de la dernière more et un grand écart type de la durée moraïque ainsi qu'une durée moyenne plus longue.
- Les attitudes de *kyoshuku* et de *sincérité-politesse* sont réalisées avec des caractéristiques prosodiques très proches, tandis que leur qualités de voix sont nettement distinctes : *kyoshuku* est caractérisé par une voix rauque et tendue, tandis que *sincérité-politesse* utilise une voix soufflée.

- L'*admiration* est elle aussi exprimée grâce à une voix soufflée.
- Quant à l'*arrogance-impolitesse*, elle montre d'importantes variations d'intensité, et utilise une qualité de voix nasalisée.
- Enfin, *déclaration* et *politesse-simple* sont réalisées avec beaucoup de variabilité de l'intensité et avec une voix modale.

5.5. Reconnaissance Audio-visuelle des attitudes

5.5.1. Introduction

Dans cette partie de la thèse, nous présentons les travaux effectués sur l'expression et la perception multimodale de ces 12 attitudes japonaises.

Les recherches sur les expressions multimodales des affects (en sens large) forment encore un domaine assez récent pour lequel les travaux de Scherer (par exemple Scherer, 2003 ; Scherer & Ellring, 2007) sont emblématiques. Dans cette littérature, seules quelques études se penchent spécifiquement sur la question de l'expression multimodale des affects sociaux, traités indépendamment des émotions (Granström & House, 2005). Pourtant la nature multimodale des informations affectives est reconnue, de même que la bimodalité (audio et visuel) de la prosodie (Swerts & Kramer, 2005). Dans cette perspective, Shigeno (1998) a effectué un test de perception des expressions japonaises dans deux modalités (audio seul, visuel seul et audio-visuel) avec des sujets natifs. Ses résultats montrent que les informations visuelles influencent davantage la perception des expressions affectives plutôt que les informations audio seules.

C'est pourquoi après la validation du corpus des attitudes japonaises, il nous semble important d'orienter nos recherches vers l'expression multimodales des attitudes japonaises, en vue de comprendre le rôle et l'influence de chacune de ces deux modalités que sont l'audition et la vision, pour la perception des attitudes japonaises, de même que les effets de synergie éventuelle que l'on pourrait observer entre les deux. Nous nous plaçons dans la continuité des travaux exposés jusqu'à maintenant, et donc nous conserverons les mêmes méthodologies expérimentales et continuerons à travailler sur le même ensemble de douze expressions attitudinales japonaises (cf. supra), mais enregistrées dans un nouveau corpus. Par ailleurs, comme nous l'avons signalé durant l'analyse des résultats du test de validation perceptive du corpus japonais (cf. supra), la stratégie individuelle du locuteur peut avoir un effet important dans la production et la reconnaissance des affects sociaux. Afin d'avoir une première évaluation de l'importance de cette stratégie, nous avons enregistré deux locuteurs différents dans ce nouveau corpus. Si ce nombre de deux locuteur reste, bien entendu, réduit pour analyser la variabilité possible des expressions attitudinales japonaises, il permet tout de même d'avoir une première évaluation des possibles divergences en choisissant deux locuteurs ayant des expériences très différentes dans ce domaine.

5.5.2. Enregistrement du corpus audio-visuel

Deux locuteurs masculins (A et B), de langue maternelle japonaise ont enregistré ce corpus. Locuteur A²⁷ est enseignant de langue japonaise étrangère, entraîné à

²⁷ C'est le même locuteur qui a produit le corpus japonais déjà validé (cf. supra).

l'expression des attitudes japonaise pour le public de sa classe (cf. supra). Le locuteur B est un homme de trente deux ans, parlant avec l'accent du Kansai, qui n'a aucune expérience particulière de l'enseignement de sa langue. Ces deux locuteurs ont produit les sept phrases utilisées pour le corpus précédent avec même douze attitudes dans une chambre sourde du LIMSI (à Orsay, France).

Un total de trois sessions ont été enregistrées pour chaque locuteur, avec des pauses entre chaque prise, durant lesquelles les locuteurs avaient le loisir de vérifier leur performance afin de corriger d'éventuel défauts expressifs. Durant l'enregistrement, les locuteurs répondent à une question de l'autre locuteur, selon un ensemble de petits scénarios mis au point pour fournir un contexte plausible à la production de chacune des attitudes enregistrées (cf. annexe). Cette méthodologie d'enregistrement a été préférée à la méthodologie précédemment utilisée, basée sur une lecture, et possible grâce à l'expérience du locuteur du premier corpus. Comme dans ce cas, l'un des deux locuteurs n'a pas d'expérience particulière d'enregistrement de corpus, il nous a semblé préférable de le mettre dans une situation où il pourrait plus simplement répondre à une sollicitation d'un interlocuteur, et en espérant une meilleure naturalité que ce que nous aurions pu obtenir à l'aide d'une simple tâche de lecture avec l'attitude à produire en consigne.

Pour l'enregistrement du corpus, les locuteurs sont debout devant la caméra, avec un micro omnidirectionnel AKG C414B placé à quarante centimètre de distance de leur bouche. Le microphone est connecté à un dispositif audio « USBPre » qui numérise le son (à 44,1kHz, 16b bits) et le transmet à un ordinateur situé à l'extérieur de la chambre sourde, sur lequel sont stockés les enregistrements sonores. La gestuelle faciale des locuteurs (la face et le haut du buste sont cadré par la caméra) est enregistrée grâce à la caméra (mini DV digital Canon XM1 3CCD). Un claquement des mains après la production de chaque énoncé permet de donner un indice de synchronisation fiable entre la caméra et le signal sonore. Ceci nous a permis, grâce à un post-traitement des données enregistrées, de remplacer le son de basse qualité enregistré par la caméra par le son de haute qualité enregistré par le microphone

Toutes les séquences vidéo sont numérisées dans une série de petites séquences vidéo, à l'aide du codec « cinepack » avec une résolution vidéo de 784 x 576 pixels.

5.5.3. Perception des attitudes audio-visuelles

Afin de comprendre le rôle et l'importance relative de chaque modalité dans la perception et la production des attitudes, un test de perception a été effectué auprès de sujets japonais natifs parlant avec l'accent de Tokyo. L'objectif du test perceptif étant de contrôler l'apport des deux modalités communicatives (c-à-d, audio et visuelle) en mesurant le degré d'identification de chacune des attitudes, soit sur des informations incomplètes (audio seule ou visuelle seule), puis de comparer les résultats de chaque modalité, soit sur les deux modalités ensemble afin de mesurer les effets de synergie.

5.5.3.1. Méthodologie expérimentale

a- Modalité

Afin d'évaluer l'influence de chacune des deux modalités et des deux ensemble, nous avons effectué ce test perceptif selon trois conditions expérimentales différentes, chacune présentant une modalité particulière : audio seule, visuelle seule et audio-visuelle. Dans la condition « audio seul », les auditeurs doivent juger de l'attitude exprimée à l'aide des seules informations véhiculées par la voix ; dans la condition « visuel seul », les sujets jugent de l'attitude exprimée grâce aux seules informations véhiculées par la gestualité faciale et du haut du buste ; enfin dans la condition « audio-visuel », les auditeurs forment leur jugement sur la base de ces deux informations combinées. Ces trois conditions expérimentales sont passées par les sujets dans les deux ordres suivants : un premier groupe de sujet passe en premier la condition « audio seul », puis la condition « visuel seul », pour finir avec la condition « audio-visuel » ; tandis qu'un second groupe de sujets passe en premier la condition « visuel seul », puis « audio seul », pour finir par « audio visuel ». Ces deux ordres de passation des trois conditions expérimentales servent à contrebalancer un éventuel effet de l'ordre de présentation des modalités.

b- Choix des stimuli

Parmi les trois séquences d'enregistrement du corpus audio-visuel, nous avons sélectionné la performance que nous avons jugée globalement la plus expressive pour chaque locuteur. Ensuite, une seule phrase a été retenue pour chaque locuteur, réalisée avec les douze attitudes : la phrase de 8 mores avec l'accent lexical réalisé sur la première more (*Nagoya de nomimasu* – il boit à Nagoya), car cette longueur de phrase a obtenu de bons résultats durant le précédent test, pour l'ensemble des attitudes. Nous avons choisi de n'utiliser qu'une seule phrase pour ce test perceptif car nous cherchons ici à tester uniquement l'effet de la modalité (et dans une certaine mesure l'effet du locuteur), sur des expressions attitudinales et des phrases qui ont déjà été validées précédemment. Ainsi, une phrase de 8 mores produite selon 12 attitudes et présentée sous trois modalités, pour deux locuteurs, donne déjà 72 stimuli, et il était par conséquent difficile de tester d'autres hypothèses sur cette même expérience de perception sans augmenter la charge cognitive des sujets.

c- Interface du test perceptif

Le test perceptif a été effectué sur une interface développée à l'aide du logiciel « Revolution », pour Windows XP. Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 12 attitudes possibles et les définitions de ces attitudes. De plus, avant la passation du test, nous avons distribué un document papier décrivant plus en détail chaque attitude, en ajoutant des exemples afin de faciliter l'évocation du contexte dans lequel ces attitudes peuvent être exprimées.

d- Procédure expérimentale

Le test perceptif a été passé par deux groupes de sujets, chaque groupe passant les trois conditions expérimentales dans un ordre précis (cf. supra). Pour chacun de ces groupes, les sujets sont assis devant l'écran d'un ordinateur et doivent juger, pour chaque stimulus qui leur est présenté, l'attitude qu'ils pensent avoir perçue, parmi les douze attitudes possibles. Pour cela, ils doivent sélectionner l'un des douze curseurs à leur disposition (celui qui correspond à l'attitude choisie) et le placer sur la position qui correspond à l'intensité perçue de l'attitude, sur une échelle libre allant de « difficilement perceptible » à « très marqué ». Cette intensité est encodée sur une échelle de 1 à 100, en donnant la valeur 0 aux onze attitudes non sélectionnées.

À l'intérieur de chaque condition expérimentale, l'ordre de présentation des stimuli est aléatoire et différent pour chaque sujet. Le sujet ne perçoit qu'une seule fois chaque stimulus pour chaque modalité, et doit répondre, sans limitation de son temps de réflexion, avant de passer au stimulus suivant. Au début du test perceptif, les sujets doivent lire la consigne ainsi que les définitions des attitudes. Ensuite, la tâche leur est expliquée plus en détail à l'oral. À la fin du test perceptif, nous avons interrogé le sujet à propos des attitudes afin de savoir lesquelles étaient difficiles à distinguer ou s'ils avaient perçu d'autres attitudes qui n'étaient pas proposées.

e- Sujets

Un total de 46 sujets, tous étudiant dans des universités de la région de Tokyo et parlant le dialecte de Tokyo, ont été divisés en deux groupes différents pour cette expérimentation. Les participants du premier groupe (28 sujets, dont 4 hommes et 24 femmes, d'âge moyen de 18,8 ans) ont passé les trois conditions dans cet ordre : « audio seul », puis « visuel seul », et enfin « audio-visuel ». Le second groupe (18 sujets, dont 7 hommes et 11 femmes, d'âge moyen de 18,6 ans) a passé les trois conditions expérimentales dans cet ordre : « vidéo seul », puis « audio seul », et enfin « audio-visuel ». Aucun participant n'a signalé de trouble de l'audition ni de la vision.

5.5.3.2. Résultats du test perceptif

Ainsi que nous l'avons déjà dit, les sujets donnent leur réponses en donnant deux informations différentes : un simple choix catégoriel sur l'attitude perçue ou l'intensité perçue pour l'expression de cette attitude. Ainsi, l'analyse statistique des résultats de ce test de perception peut se faire sur la base des mesures suivantes.

- Le taux de reconnaissance de chaque attitude, exprimé soit comme le pourcentage de bonne réponse au choix catégoriel, soit comme l'intensité moyenne obtenue par ces bonnes réponses.
- Des matrices de confusions, exprimées soit comme un tableau de contingence des choix catégoriels, soit comme la proportion des scores d'intensité obtenue par chacune des catégories de réponse possibles, par rapport à la somme des valeurs d'intensité donnée pour un stimuli.

Tableau 12 : résultats de l'ANOVA pour les sujets japonais, sur les réponses catégorielles. Les facteurs utilisés sont Grp pour les deux groupes d'ordre de passation des trois conditions ; Spk pour les deux locuteurs ; Mod pour les trois modalités. Les facteurs ayant un effet significatif sur les résultats (à $p < .01$) sont marqués d'une étoile.

Facteur	ddl	F	p	
Grp	1	0.08	0.779	
Spk	1	163.6	0.000	*
Grp:Spk	1	0.02	0.899	
Mod	2	44.8	0.000	*
Grp:Mod	2	1.2	0.299	
Att	11	23.5	0.000	*
Grp:Att	11	1.5	0.134	
Spk:Mod	2	6.8	0.000	*
Grp:Spk:Mod	2	0.15	0.856	
Spk:Att	11	11.0	0.000	*
Grp:Spk:Att	11	1.4	0.188	
Mod:Att	22	3.7	0.000	*
Grp:Mod:Att	22	1.2	0.227	
Spk:Mod:Att	22	7.3	0.000	*
Grp:Spk:Mod:Att	22	0.95	0.522	

Seuls les résultats exprimés à partir des réponses catégorielles seront détaillés dans la suite de ce travail, car aucune différence n'a été observée entre les deux types de mesure.

L'analyse des résultats de ce test est faite dans un premier temps grâce à une analyse de la variance (ANOVA), dont les résultats sont détaillés dans la table 12, qui va nous permettre de déterminer l'influence des différents facteurs du test perceptif sur les réponses des sujets. En particulier, nous regarderons comment l'ordre de passage des trois conditions expérimentales influence les réponses. Ensuite, nous analyserons l'effet des locuteurs, et enfin celui des modalités. Dans un second temps, et de manière similaire aux analyses réalisées précédemment, la dispersion des résultats sera analysée grâce aux méthodes statistiques d'analyses multidimensionnelles déjà présentées.

a- Effet de l'ordre de présentation des modalités

Les résultats de l'ANOVA montrent que les deux groupes de sujets, passant les trois conditions expérimentales dans deux ordres différents, obtiennent des résultats comparables. L'effet de l'ordre de présentation des modalités n'est pas significatif. La figure 23 montre le taux de reconnaissance obtenu par chacune des douze attitudes pour les deux groupes de sujets ; scores de reconnaissance similaires pour chaque attitude.

a- Effet du locuteur

Le second facteur pris en compte dans l'ANOVA est le locuteur ayant produit chacune des attitudes. Les résultats montrent l'influence significative de la performance des locuteurs sur le taux de reconnaissance des douze attitudes. La figure 24 présente le

taux de reconnaissance des douze attitudes pour chacun des deux locuteurs. On y observe clairement que les auditeurs ont globalement mieux identifié la performance du locuteur A que celle du locuteur B, pour une majorité d'attitudes, et en particulier pour celles d'*autorité* et d'*exclamation de surprise*.

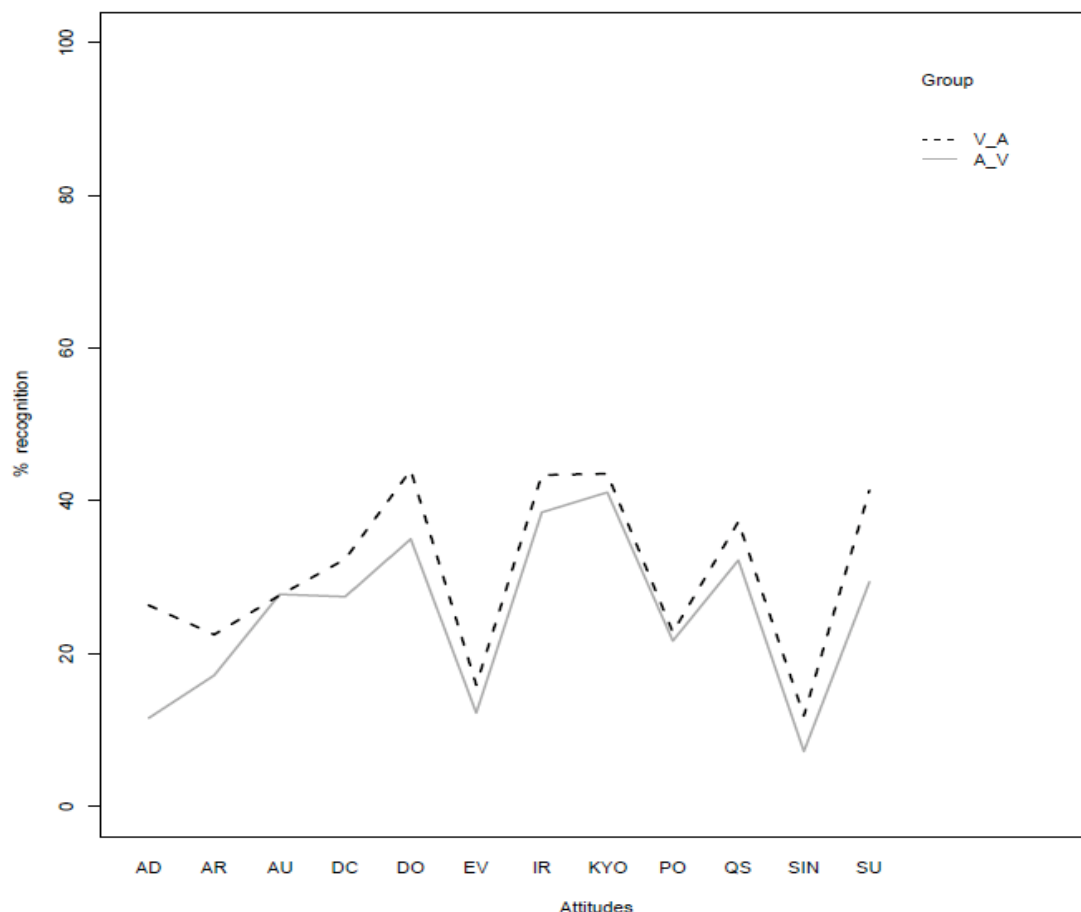


Figure 23 : taux de reconnaissance des 12 attitudes selon les groupes d'ordre de présentation des modalités. Les attitudes sont présentées sur l'axe des abscisses, et les taux d'identification en ordonnée.

b- Effet de la modalité

Le facteur modalité (c-à-d, celle des trois modalités avec laquelle le sujet perçoit les attitudes, parmi « audio seul », « vidéo seul » et « audio-vidéo ») montre une influence significative sur les choix catégoriels moyens des sujets. La figure 25 illustre ce taux de reconnaissance moyen obtenu par chacune des attitudes dans les trois modalités différentes. Selon cette figure, il est clair que les résultats obtenus en condition audio-visuelle montrent les meilleurs taux d'identification pour toutes les attitudes sauf l'*arrogance-impolitesse*. Le taux de reconnaissance de l'*arrogance-impolitesse*, du *doute-incrédulité* et de la *sincérité-politesse* en condition visuelle seule est presque identique à celui de la condition audio-visuelle, ce qui indique une prédominance de l'information visuelle par rapport aux informations audio pour ces expressions.

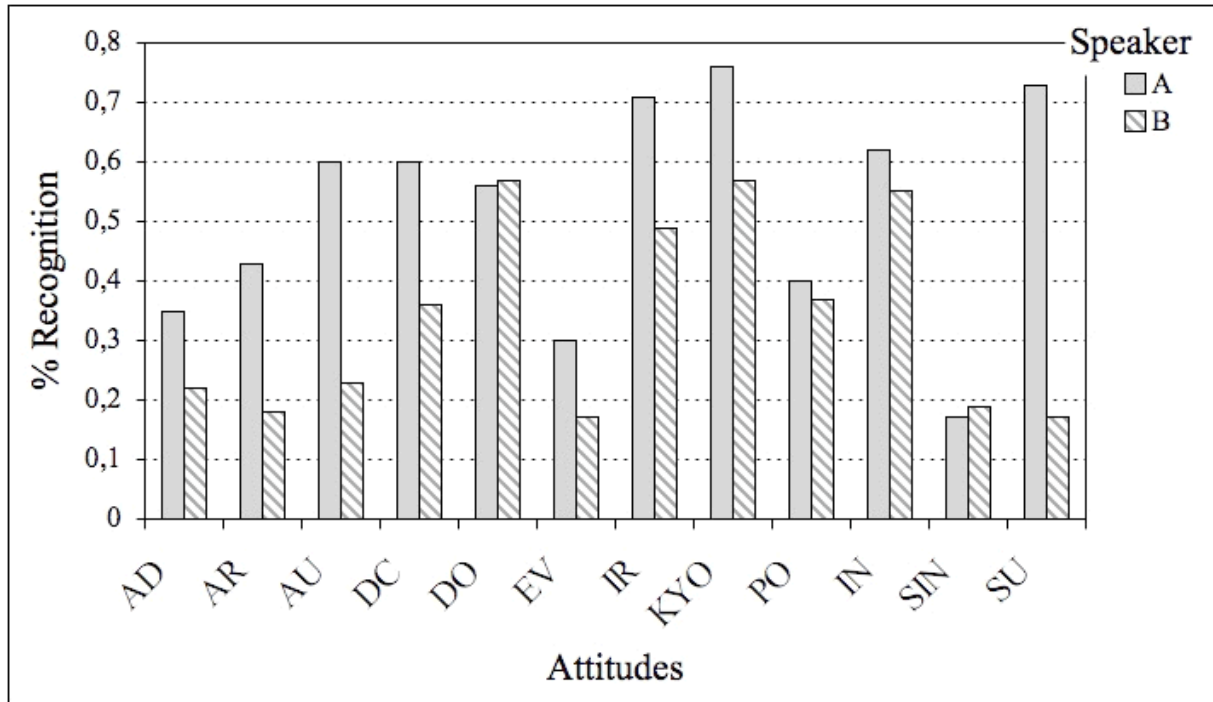


Figure 24 : effet de la performance des locuteurs sur la perception des attitudes. Les attitudes sont présentées en abscisse et le taux de reconnaissance moyen en ordonnée.

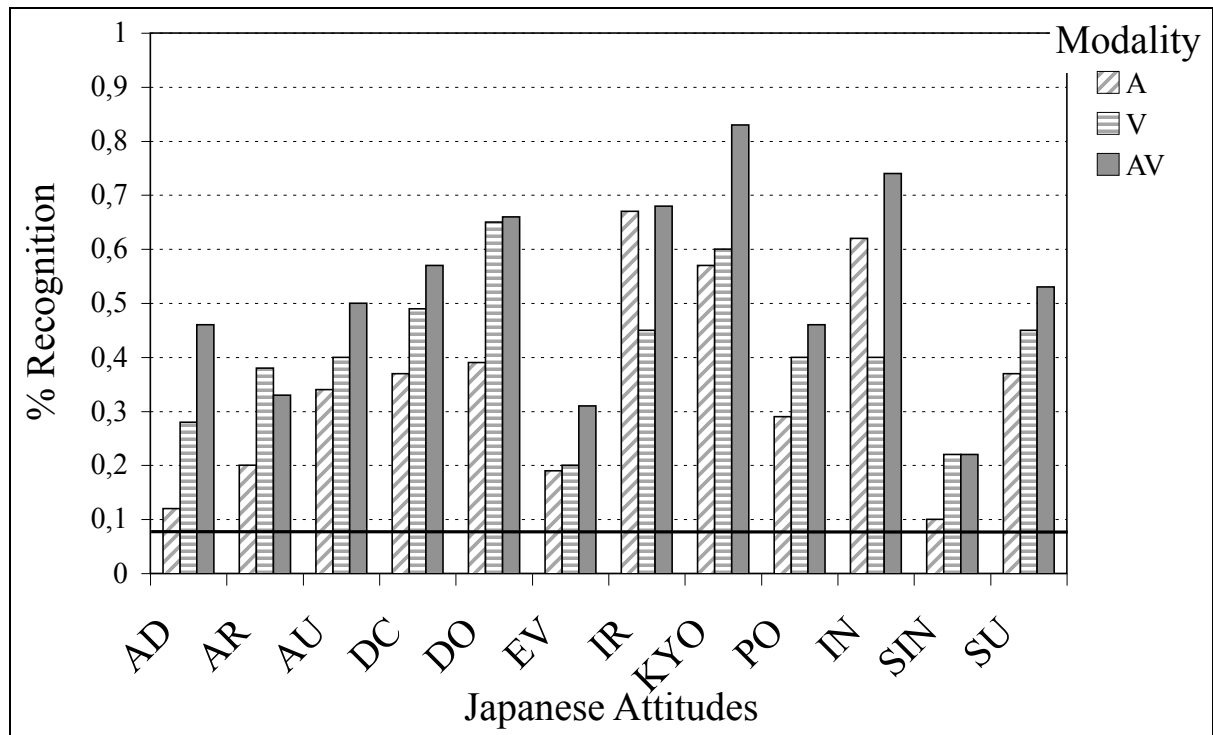


Figure 25 : effet du locuteur et de la modalité sur les taux d'identification des attitudes. Les attitudes sont présentées en abscisse et les scores de reconnaissance en ordonnée.

Inversement, l'attitude d'*irritation* est mieux reconnue grâce aux informations audio qu'avec les informations visuelles. Pour la plupart des attitudes : *admiration*, *autorité*, *déclaration*, *évidence*, *kyoshuku*, *politesse-simple*, *question-simple* et

exclamation de surprise, l'information audio et visuelle coopèrent, se complètent, comme l'indique le taux d'identification audio-visuel, meilleur que ceux de chacune des deux modalités.

c- Analyses des confusions

L'analyse de confusions et des regroupements d'attitudes est réalisée pour les 46 sujets japonais, selon chaque modalité et selon chaque locuteur par l'analyse de correspondance et de l'analyse factorielle (les résultats des deux analyses seront présentés conjointement, dans un souci de concision). Les résultats des deux groupes de sujets sont mélangés, étant donné qu'aucun effet significatif de l'ordre de présentation des modalités n'a été mis en évidence.

Modalité « audio seul »

La figure 26 représente la dispersion des résultats des 46 auditeurs pour les 12 attitudes dans la condition « audio seul » obtenus grâce à l'analyse des correspondances. La figure 27 présente les résultats de la classification hiérarchique sur les mêmes données. Les résultats de chaque analyse sont présentés séparément pour chaque locuteur. L'analyse de la dispersion et des principaux regroupements d'attitudes pour la modalité « audio seul » permet d'arriver aux conclusions suivantes.

Pour le locuteur A, les résultats sont les suivants. Chaque attitude est bien reconnue individuellement (figure 26), sauf les deux expressions d'*admiration* et de *sincérité-politesse*. Les douze attitudes sont regroupées, au plus haut niveau de différenciation, en trois catégories (figure 27). La première regroupe *sincérité-politesse*, *kyoshuku*, *politesse-simple* et *déclaration* – l'ensemble des expressions de politesse plus la déclaration. La seconde contient les expressions de *question-simple*, de *doute-incrédulité* et d'*exclamation de surprise*, c'est-à-dire les expressions d'interrogation. La troisième catégorie est composée de l'*irritation*, de l'*autorité*, de l'*évidence*, de l'*arrogance-impolitesse* et de l'*admiration*. Toutes ces attitudes sauf l'*admiration* (qui est mal reconnue), expriment l'imposition de l'opinion du locuteur.

Pour le locuteur B, les résultats sont les suivants. L'AC (figure 26) montre que les auditeurs reconnaissent bien toutes les attitudes sauf l'*autorité* et l'*arrogance-impolitesse*. La classification hiérarchique permet d'extraire au plus haut niveau trois catégories perceptives pour les 12 attitudes (figure 27), différentes des catégories obtenues par le locuteur A. La première catégorie regroupe *sincérité-politesse*, *politesse-simple*, *déclaration*, *autorité* et *arrogance-impolitesse*. Les expressions d'*autorité* et d'*arrogance-impolitesse* y sont regroupées avec celles de politesse moins le *kyoshuku*. La seconde catégorie contient la *question-simple*, le *doute-incrédulité* et l'*exclamation de surprise*, toutes les attitudes qui expriment de l'interrogation de la part du locuteur. La troisième catégorie est composée de l'*irritation*, de l'*évidence*, du *kyoshuku* et de l'*admiration*, et mélange donc aussi les expressions d'imposition et de politesse.

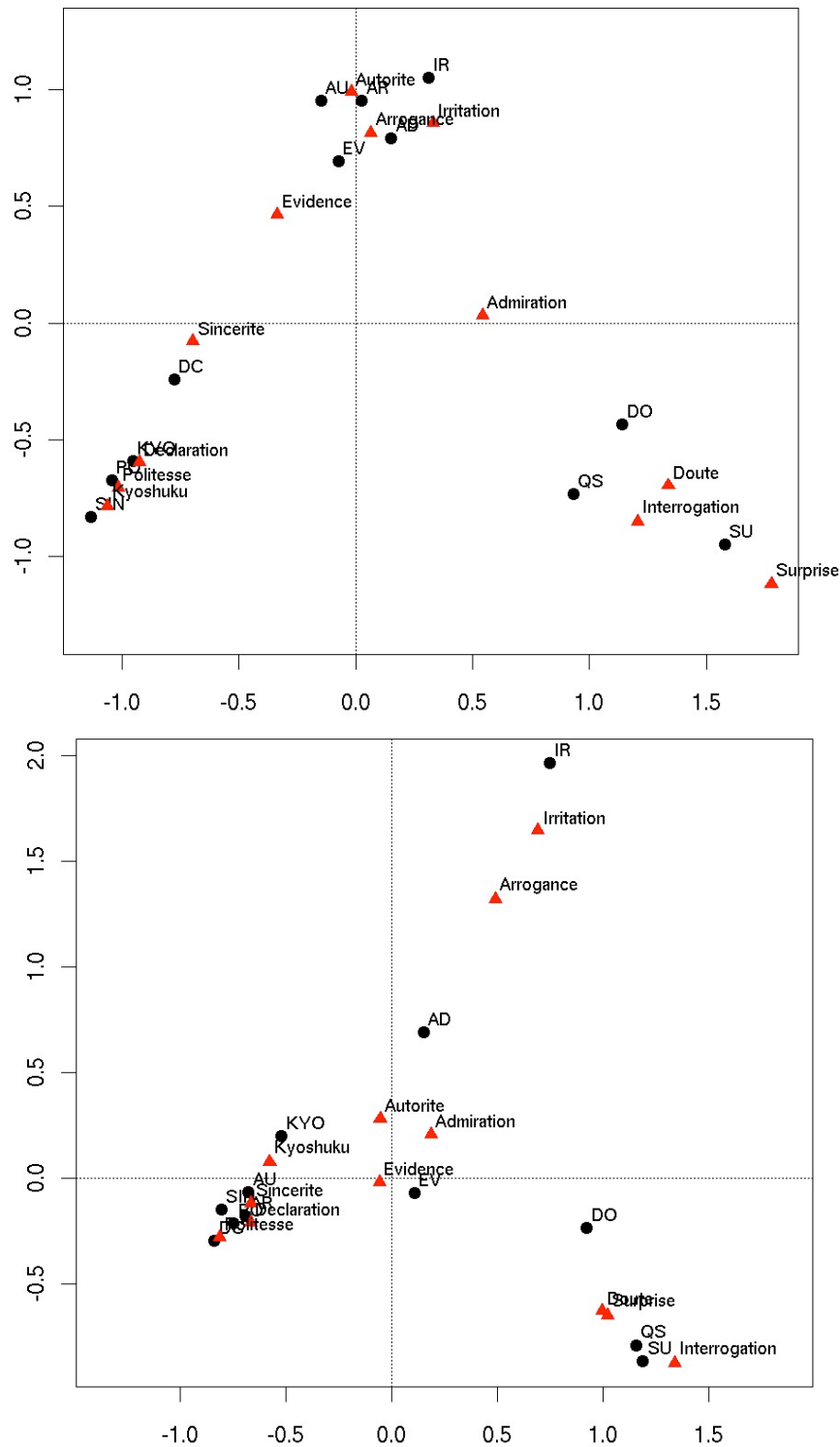


Figure 26 : résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « audio seul ». Les triangles rouges correspondent à la perception des stimuli, les ronds noirs indiquent les catégories d'attitude.

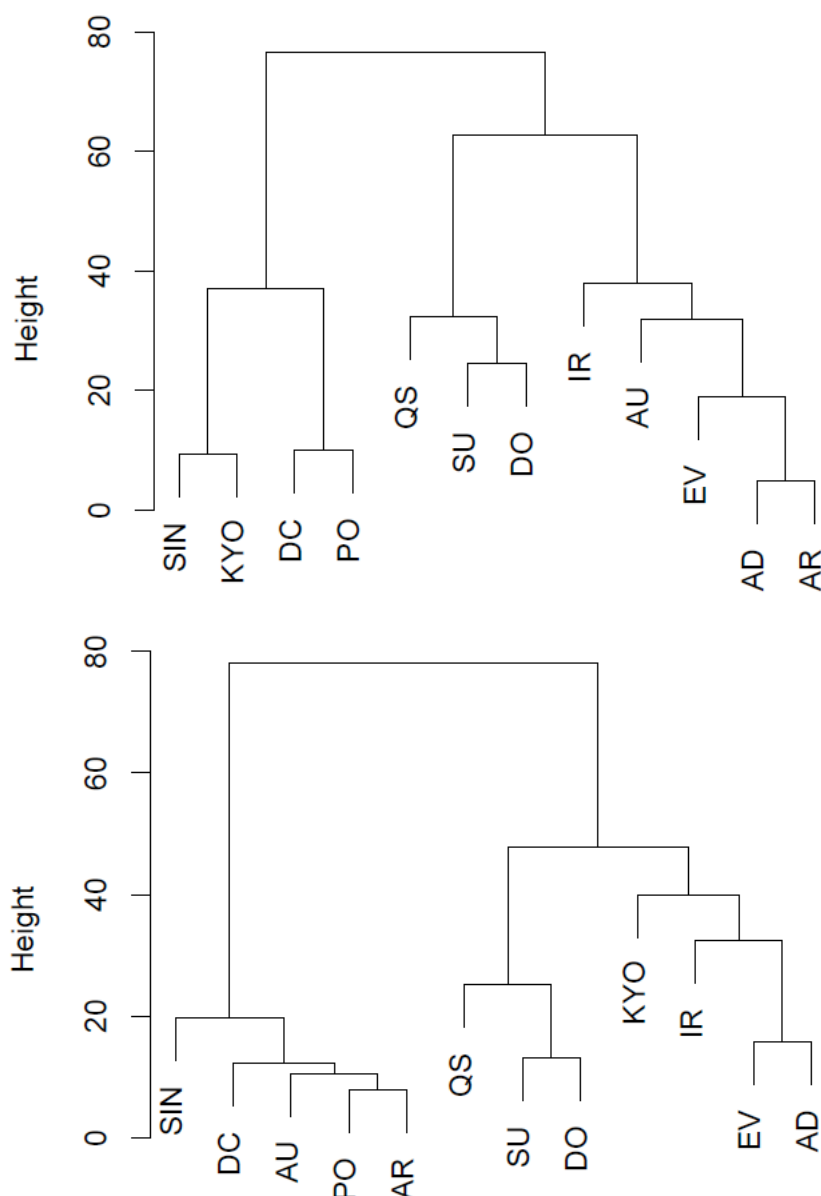


Figure 27 : résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « audio seul ».

Pour le locuteur B, les résultats semblent donc en nettement moins cohérents que ceux obtenus par le locuteur A : deux catégories regroupent des expressions qui sont opposées d'un point de vue pragmatique – et l'on ne peut mettre en cause la performance des auditeurs, qui reconnaissent bien les expressions réalisées par le locuteur A. Comme, de plus, la performance globale du locuteur B est significativement moins bonne que celle du locuteur A (cf. résultats de l'ANVOA, supra) et que le locuteur B n'avait aucune expérience particulière dans l'enregistrement de corpus affectifs, il semblerait que nous devions remettre en cause directement la qualité de ses performances. Vérifions cela pour les autres modalités.

Modalité « vidéo seul »

La figure 28 représente la dispersion des résultats des 46 auditeurs pour les 12 attitudes dans la condition « vidéo seul » obtenus grâce à l'analyse des correspondances. La figure 29 présente les résultats de la classification hiérarchique sur les mêmes données. Les résultats de chaque analyse sont présentés séparément pour chaque locuteur. L'analyse des dispersions et des principaux regroupements d'attitudes pour la modalité « vidéo seul » permet d'arriver aux conclusions suivantes.

Pour le locuteur A, les résultats sont les suivants. Chaque attitude est bien reconnue individuellement (figure 28), même si deux expressions (*surprise* et *kyoshuku*) particulièrement bien reconnues, expliquent la plus grande part de la variance pour les deux premières dimensions, ce qui resserre les points des autres expressions. Les douze attitudes sont regroupées, au plus haut niveau de différenciation, en trois catégories (figure 29). Ces trois catégories sont les mêmes que celles identifiées pour la modalité audio : expressions de politesse, de questionnement et d'imposition. La principale différence réside dans le fait que les deux expressions reconnues sans aucune confusion en modalité visuelle (*surprise* et *kyoshuku*) ne font partie d'aucune catégorie.

Pour le locuteur B, les résultats sont les suivants. La figure 28 montre que les expressions d'*évidence*, d'*exclamation de surprise*, d'*autorité* et de *déclaration* sont assez mal identifiées avec les seules informations visuelles, les autres restants correctement identifiées. Les douze attitudes sont regroupées, au plus haut niveau de différenciation, en trois catégories (figure 29). La première regroupe les expressions de *politesse-simple*, de *sincérité-politesse*, de *déclaration* et d'*admiration*. La seconde contient les expressions de *kyoshuku*, d'*évidence*, d'*interrogation*, et d'*exclamation de surprise*, et regroupe donc l'ensemble des expressions qui montrent le plus de confusions. La troisième catégorie est composée de l'*irritation*, de l'*autorité* et de l'*arrogance-impolitesse*. L'expression de *doute-incrédulité* ne montre pas de confusion particulière. Ces résultats pour le locuteur B confirment ses performances assez basses sur la modalité vidéo.

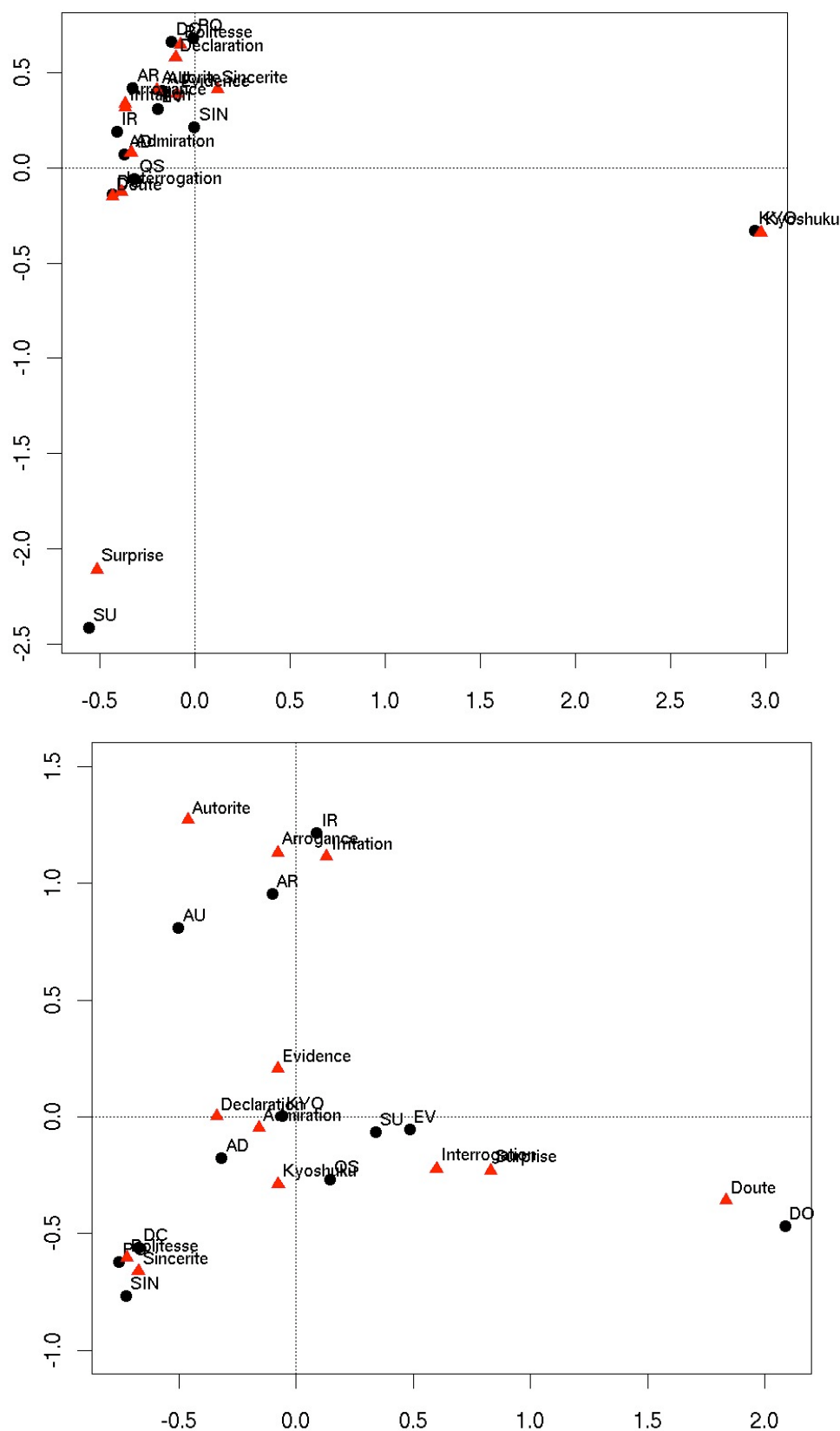


Figure 28 : résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « vidéo seul ». Les triangles rouges correspondent à la perception des stimuli, les ronds noirs indiquent les catégories d'attitude.

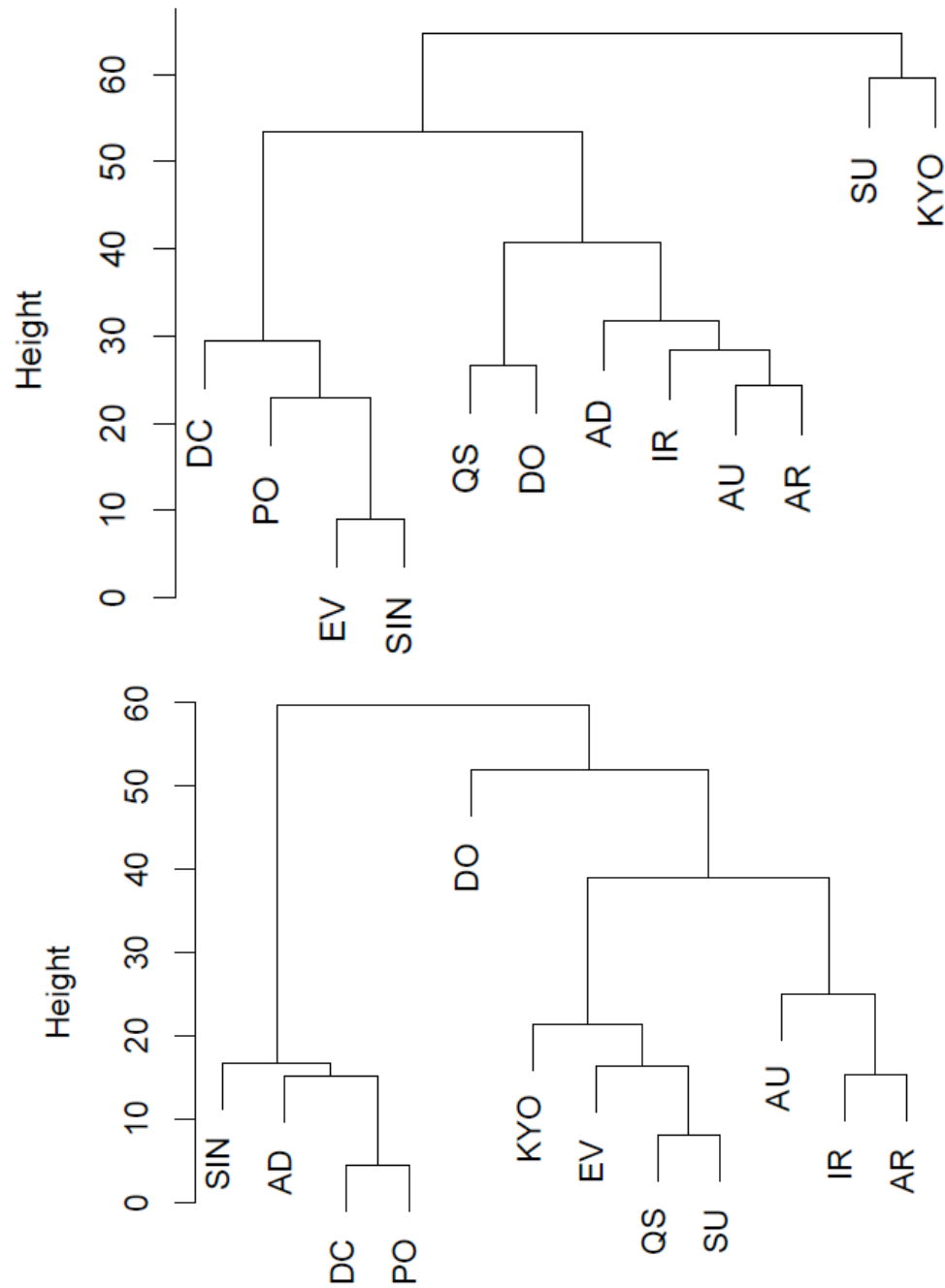


Figure 29 : résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « vidéo seul ».

Modalité « audio-visuel »

La figure 30 représente la dispersion des résultats des 46 auditeurs pour les 12 attitudes dans la condition « audio-vidéo » obtenus grâce à l'analyse des correspondances. La figure 31 présente les résultats de la classification hiérarchique sur les mêmes données. Les résultats de chaque analyse sont présentés séparément pour chaque locuteur. L'analyse des dispersions et des principaux regroupements d'attitudes pour la modalité « vidéo seul » permet d'arriver aux conclusions suivantes.

Pour le locuteur A, les résultats sont les suivants. Chaque attitude, sans exception, est bien reconnue individuellement en modalité audio-visuelle (figure 30). Les douze attitudes sont regroupées, au plus haut niveau de différenciation, en trois catégories (figure 31), qui sont les mêmes que celles obtenus dans la condition audio seule : une catégorie de d'expressions de politesse, une catégorie d'expressions de questionnement et une catégorie d'expression d'imposition, dans laquelle vient se rajouter l'expression d'admiration qui, même si elle est bien reconnue comme telle, montre toujours des confusions avec ces expressions.

En revanche pour le locuteur B, les résultats diffèrent de ceux observés pour les deux autres conditions. Tout d'abord, chaque attitude, sauf l'*arrogance-impolitesse*, est bien reconnue individuellement (figure 30). Les douze attitudes sont regroupées (figure 31), au plus haut niveau de différenciation de la même manière que pour les expressions en modalité audio seule, sauf pour les expressions de *kyoshuku* et d'*évidence* qui forment une quatrième catégorie, à part. La synergie des deux modalités permet donc aux sujets de mieux reconnaître les expressions du locuteur B, mais ne leur permet pas de réaliser des regroupements pragmatique auxquels on pourrait s'attendre : d'importantes confusions demeurent.

Par ailleurs, dans cette condition expérimentale qui mêle les deux modalités, on a pu observer une tendance à reproduire les résultats obtenus dans la condition « audio seul ». Cette possible prédominance perceptive de l'audio sur le visuel pour la perception de ce type d'attitude est intéressante, mais serait à analyser plus avant. Elle est tout de même cohérente avec les résultats observés sur les attitudes françaises audio-visuelles (Rilliard et al., 2008).

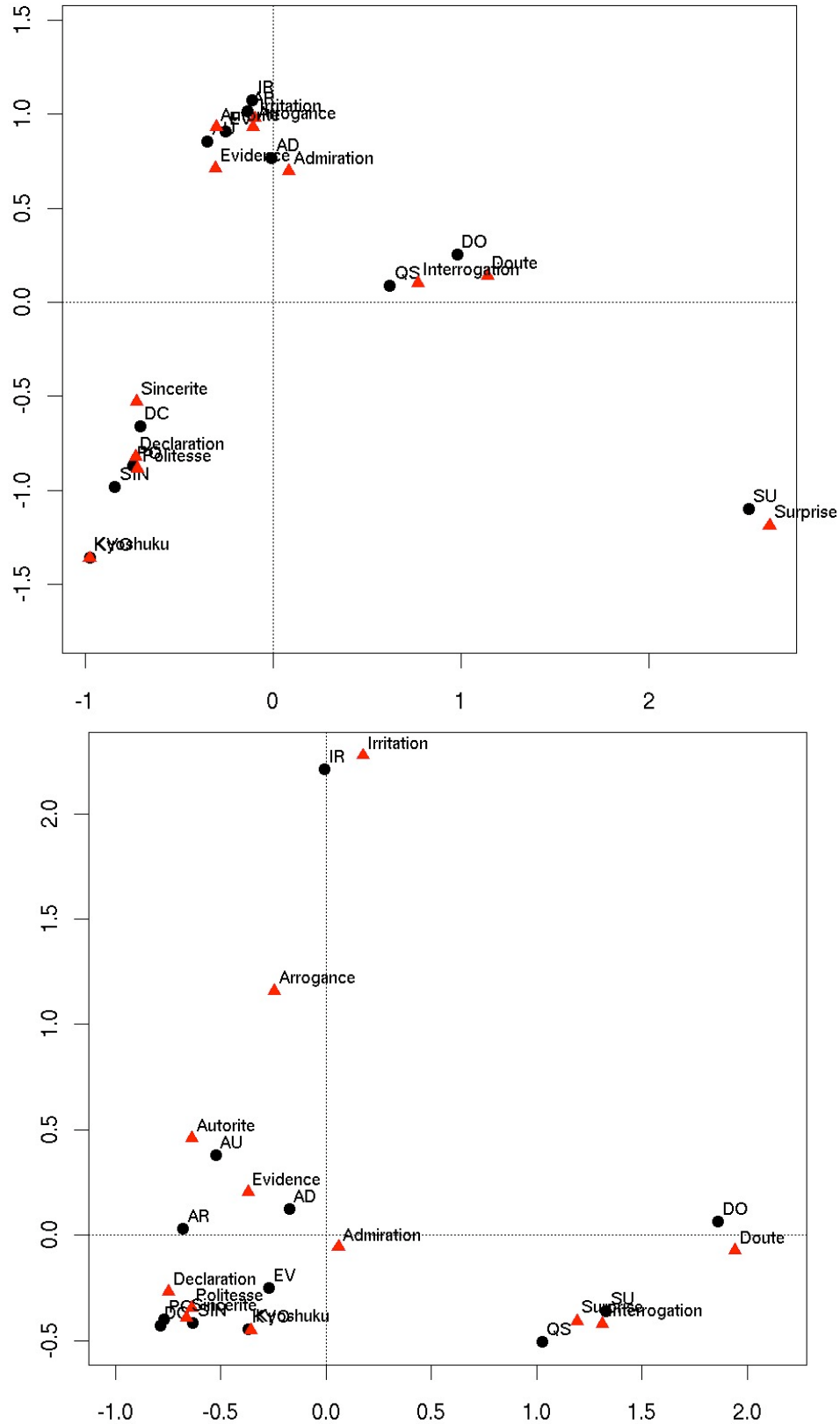


Figure 30 : résultats de l'AC pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) à la condition « audio-vidéo ». Les triangles rouges correspondent à la perception des stimuli, les ronds noirs indiquent position des catégories d'attitude.

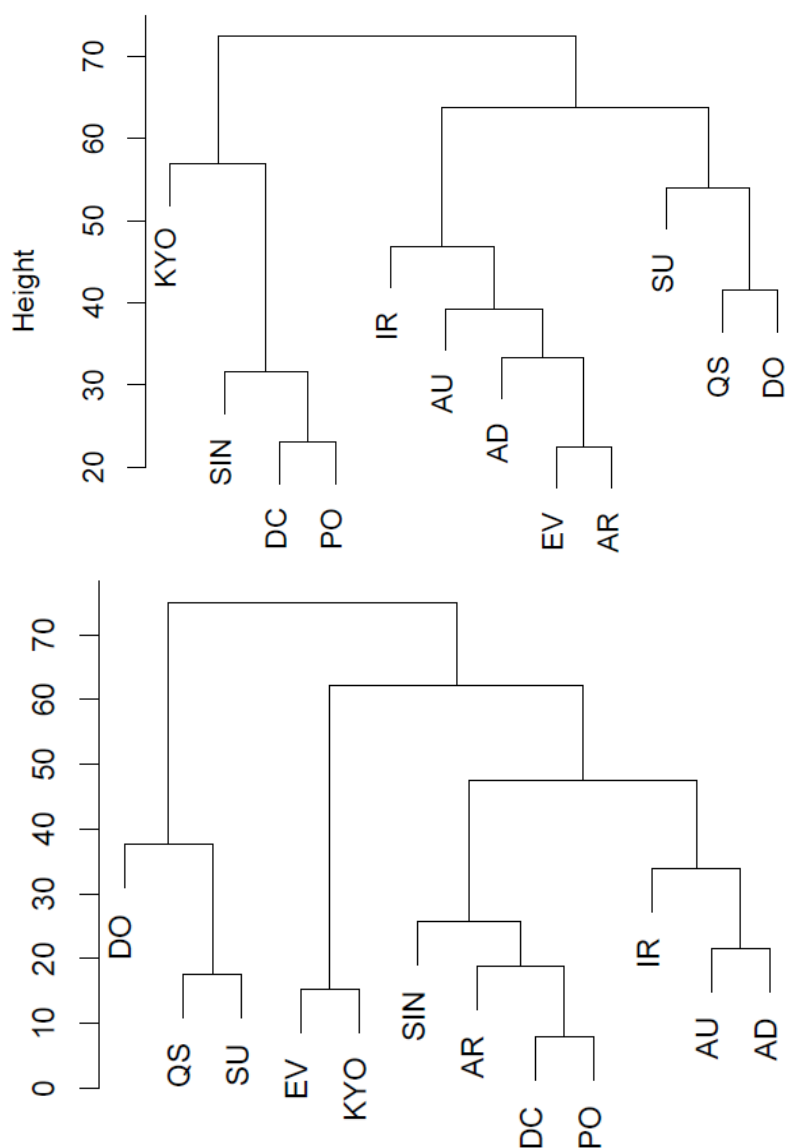


Figure 31 : résultats de la classification hiérarchique pour les attitudes japonaises produites par le locuteur A (en haut) et par le locuteur B (en bas) dans la condition « audio-vidéo ».

5.5.3.3. Paramètres objectifs – acoustiques et visuels

Nous allons maintenant analyser les réalisations des expressions attitudinales multimodales de ce corpus – à la fois les caractéristiques acoustiques de la prosodie et les gestes faciaux. Par la suite, nous tenterons de faire le lien entre ces caractéristiques objectives et les catégories perceptives qui viennent d’être identifiées.

Pour cela, seules les attitudes produites par le locuteur A seront analysées, car nous avons vu que les expressions produites par le locuteur B sont souvent mal reconnues par les auditeurs. Il est donc de peu d’intérêt d’analyser des paramètres qui ne sont pas pertinents perceptivement.

a- Paramètres objectifs

Extraction des paramètres acoustiques

Les paramètres acoustiques de la prosodie sont extraits automatiquement à partir des signaux étiquetés manuellement, à l'aide de scripts Matlab utilisant l'algorithme « yin » pour l'extraction de la fréquence fondamentale (de Cheveigné & Kawahara, 2002). Trois paramètres acoustiques sont extraits : la fréquence fondamentale (F0 – en demi-ton par rapport à 1Hz), la durée moraïque (en secondes) et l'intensité (en dB). Pour les deux paramètres de F0 et d'intensité, trois valeurs pour chaque voyelle sont calculées.

Pour chaque paramètre, nous avons calculé le maximum, le minimum, la moyenne et le registre (différence entre la valeur maximum et la valeur minimum), et également la pente de la courbe (c-à-d, l'écart entre la valeur moyenne de la première voyelle et la valeur moyenne de la dernière voyelle). Ces cinq valeurs pour chacun des trois paramètres acoustiques sont extraites pour chaque expression attitudinale, et elles sont utilisées en entrée d'une analyse en composantes principales (ACP).

Extraction des « Action Units »

Les paramètres de l'expressivité faciale ont été extraits en terme « d'Action Units » (Ekman et al., 2002), dans le cadre d'un travail collaboratif sur l'expression audio-visuelle des attitudes française et japonaise (cf. Rilliard et al., en révision). Un chercheur spécialisé dans les expressions multimodales s'est basé sur le système d'Ekman (« Facial Action Coding System (FACS) »²⁸) pour analyser les vidéos. Pour chaque séquence vidéo ont été annotés quelles « Action Units » (AU) apparaissaient en haut du visage et en bas du visage, en plus des positions de la tête.

b- Analyse des paramètres audio

En utilisant les paramètres objectifs de la prosodie décrits dans le paragraphe précédent, une ACP a été menée afin d'identifier l'importance des différents composants pour véhiculer les différentes attitudes.

Cette analyse (cf. figure 32) permet d'associer les caractéristiques suivantes aux différentes attitudes :

L'attitude d'*admiration* est principalement liée aux paramètres de durée : c'est cette attitude qui montre les plus longues mores, la plus importante différence de durée et aussi l'allongement le plus important tout au long de la phrase (paramètre de pente). Cette attitude montre aussi le plus important registre de F0. Les attitudes de *doute-incrédulité* et d'*irritation* montrent les mêmes caractéristiques, mais dans une moindre mesure. Ce constat diffère de celui fait à partir du corpus précédent (cf. supra), dans lequel c'est l'attitude de *doute-incrédulité* qui montrait (et de loin) les plus grandes variations de durée. L'expression de la *surprise* est caractérisée dans ce corpus par une

²⁸ Cependant, ce codeur n'est pas certifié en tant que codeur officiel de FACS.

F0 et une intensité élevée (F0 moyenne élevée, ainsi qu'une importante valeur de F0 maximum), ainsi que par une intensité moyenne élevée. L'*autorité* est principalement liée à une intensité élevée, mais avec une F0 basse et assez constante, ainsi qu'un débit élevé. Les expressions de *sincérité-politesse*, de *kyoshuku*, d'*interrogation* et d'*évidence* partagent un débit qui s'accélère légèrement, et une intensité et une F0 moyennes. La *déclaration*, l'*arrogance-impolitesse* et la *politesse* quant à elles ont une pente de F0 et d'intensité importante.

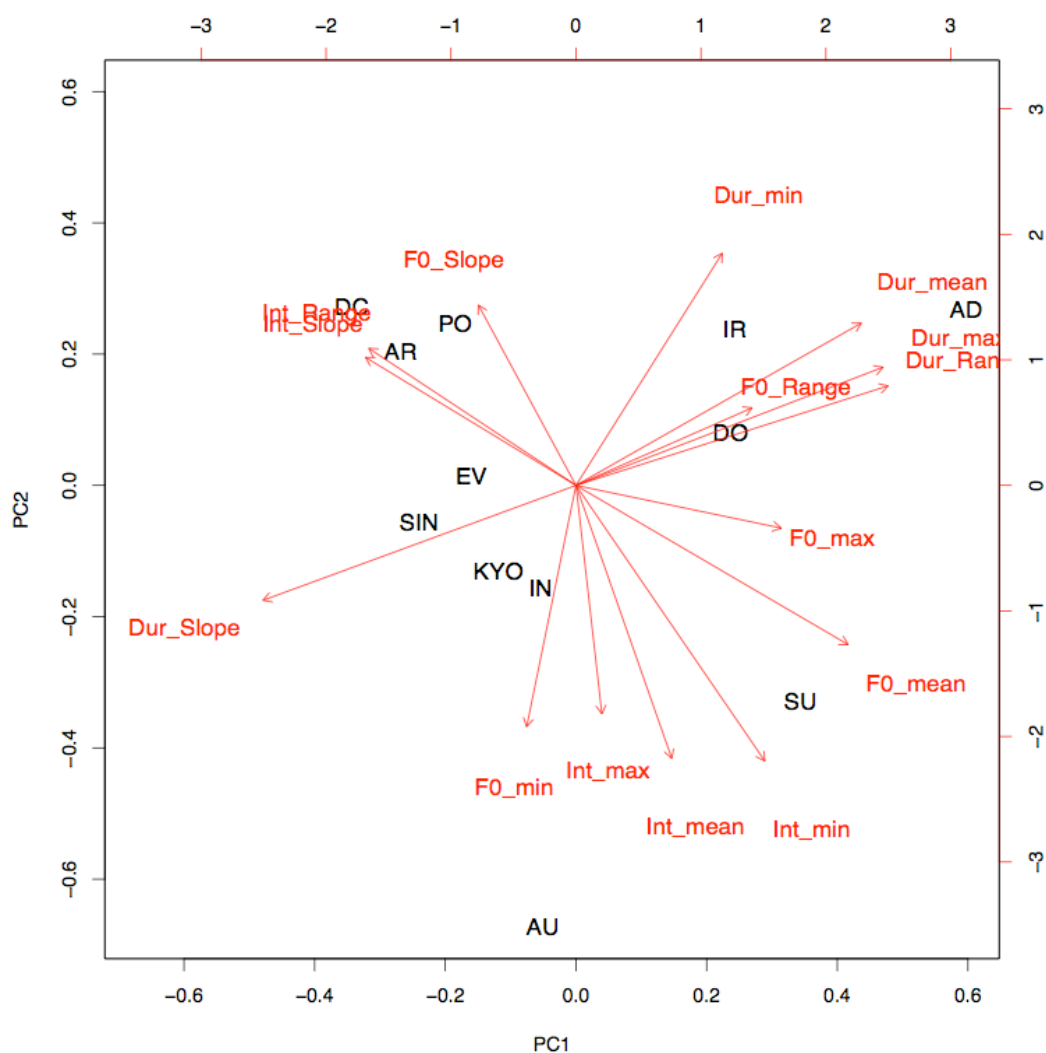


Figure 32 : résultats de l'ACP sur les paramètres prosodiques des différentes attitudes du locuteur A. Ces deux premières dimensions de l'ACP expliquent 64% de la variance.

Certains de ces paramètres s'expliquent par les variations de réalisation phonémique des différentes attitudes. Par exemple, une pente de durée positive correspond systématiquement à des énoncés pour lesquels la dernière more est composée d'un simple [s], sans y ajouter la voyelle [u] : *arrogance-impolitesse*, *autorité*, *déclaration*, *sincérité-politesse* et *kyoshuku*.

Les résultats de la classification hiérarchique (cf. figure 33) montrent la capacité des paramètres prosodiques calculés à séparer efficacement les expressions de *doute*-

incrédulité, de *surprise*, d'*autorité* et d'*évidence*. Dans une moindre mesure, l'*arrogance-impolitesse* et le *kyoshuku* sont aussi des expressions portant des caractéristiques prosodiques propres. Par contre, les expressions de *sincérité-politesse*, *déclaration* et *politesse* (notons la présence de deux expressions de politesse) d'une part, et celles d'*interrogation*, d'*admiration* et d'*irritation* d'autre part, sont moins efficacement différenciées sur la base de ces critères.

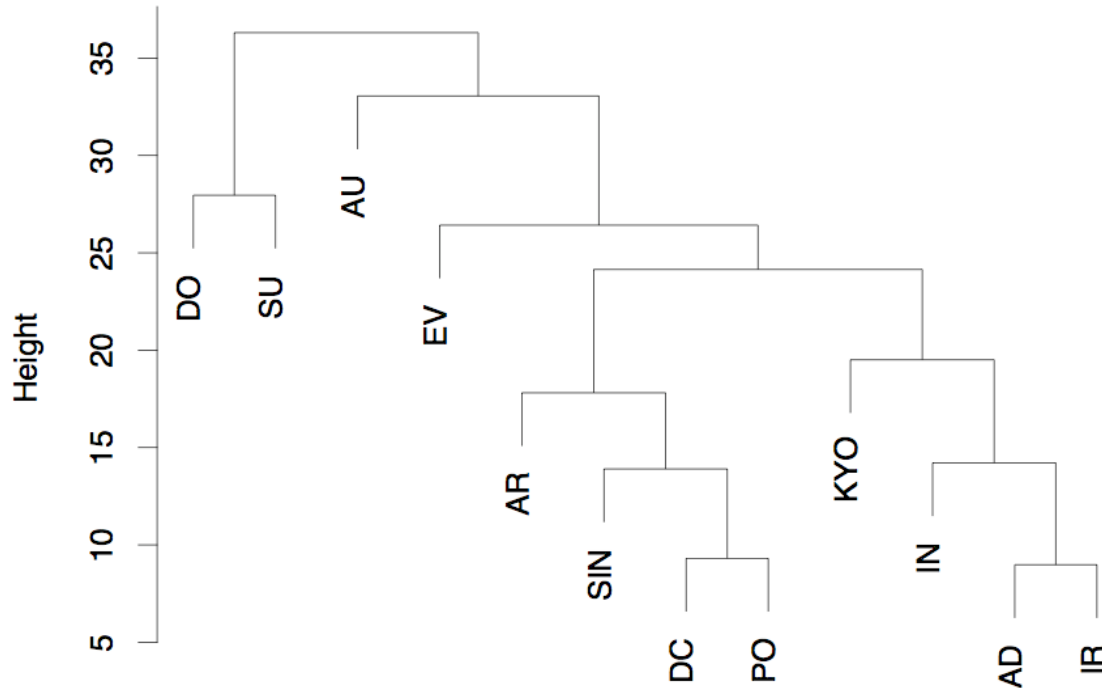


Figure 33 : résultats de la classification hiérarchique menée sur les données prosodiques des 12 attitudes du corpus audio-visuel.

Notons tout de suite que nous n'avons pas eu le temps de réaliser l'analyse de la qualité de voix de ces stimuli – ce qui nuit indubitablement à l'efficacité de la discrimination des 12 attitudes sur la base de critères audio seuls. On l'a vu (supra), la qualité de voix permet d'améliorer nettement les performances des seuls paramètres prosodiques pour différencier certaines attitudes. Par exemple, l'*irritation* est réalisée avec une qualité de voix qui la différencie de l'*admiration*.

c- Analyse des paramètres vidéo

L'analyse des paramètres expressifs vidéo, faite à l'aide de la description du corpus en termes d'AUs, permet d'obtenir les résultats suivants. Pour l'ACP (cf. figure 34) :

- Les expressions d'*irritation* et de *kyoshuku* sont largement séparées des autres par des AUs qui leur sont propres.
- Les expressions de *surprise*, de *doute-incrédulité* et d'*interrogation* (et l'*évidence* dans une moindre mesure) sont caractérisées par un ensemble d'AUs proches, et en ont plusieurs en commun.

- Les expressions d'*admiration*, de *politesse*, de *sincérité-politesse* et d'*arrogance-impolitesse* sont elles aussi très proches.
- Finalement, les expressions d'*autorité* et de *déclaration* ne sont pratiquement pas marquées par l'analyse des expressions faciales qui a été réalisée.

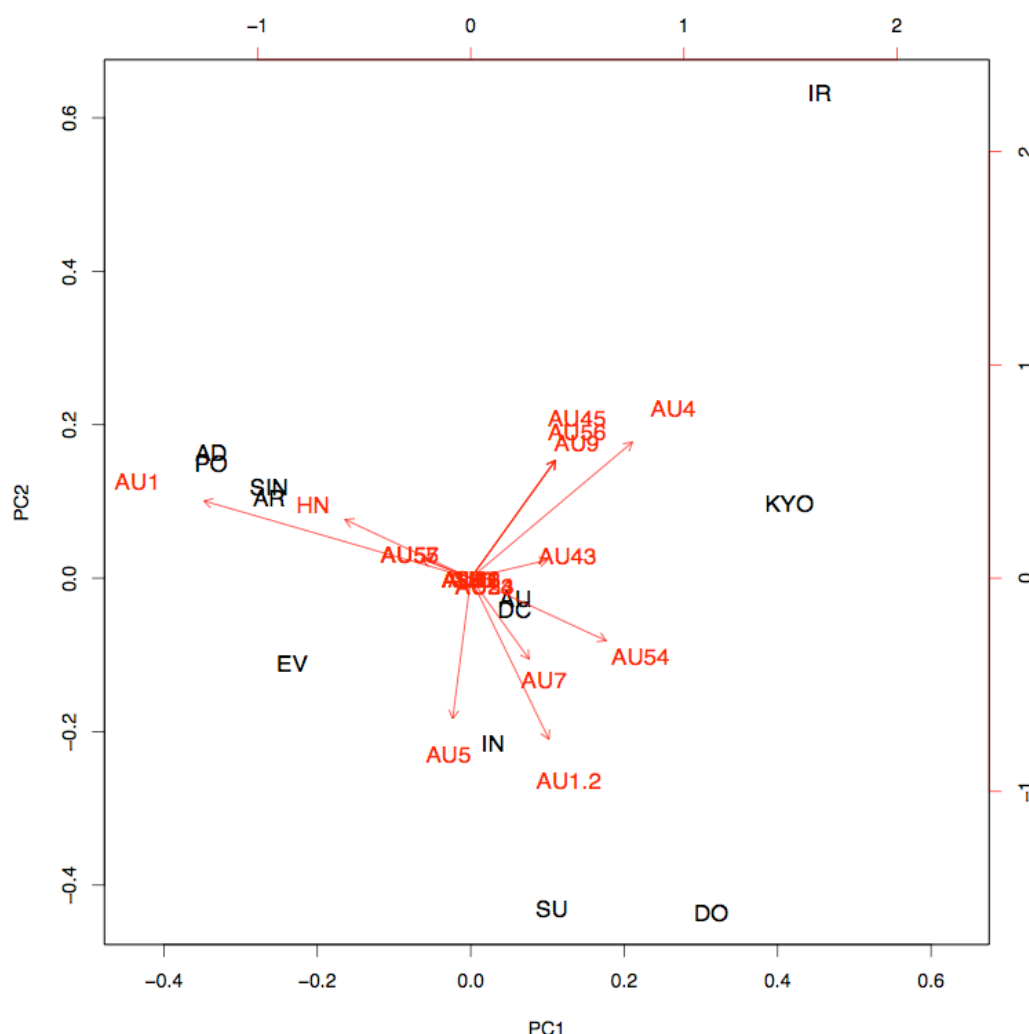


Figure 34 : résultats de l'ACP menée sur les AU, montrant la dispersion des attitudes obtenue à partir des informations faciales.

La classification hiérarchique de ces expressions faciales permet de montrer que les expressions d'*irritation*, de *kyoshuku* et de *doute-incrédulité* sont chacune parfaitement bien caractérisées. Les expressions de *sincérité-sérieux* et d'*arrogance-impolitesse* et celles d'*admiration* et de *politesse* ont exactement les mêmes caractéristiques, et ces deux groupes d'attitudes sont en outre assez proches (froncement des sourcils). Finalement, les expressions de *surprise*, d'*évidence*, et d'*interrogation* partagent l'AU5 tout en conservant certains traits caractéristiques.

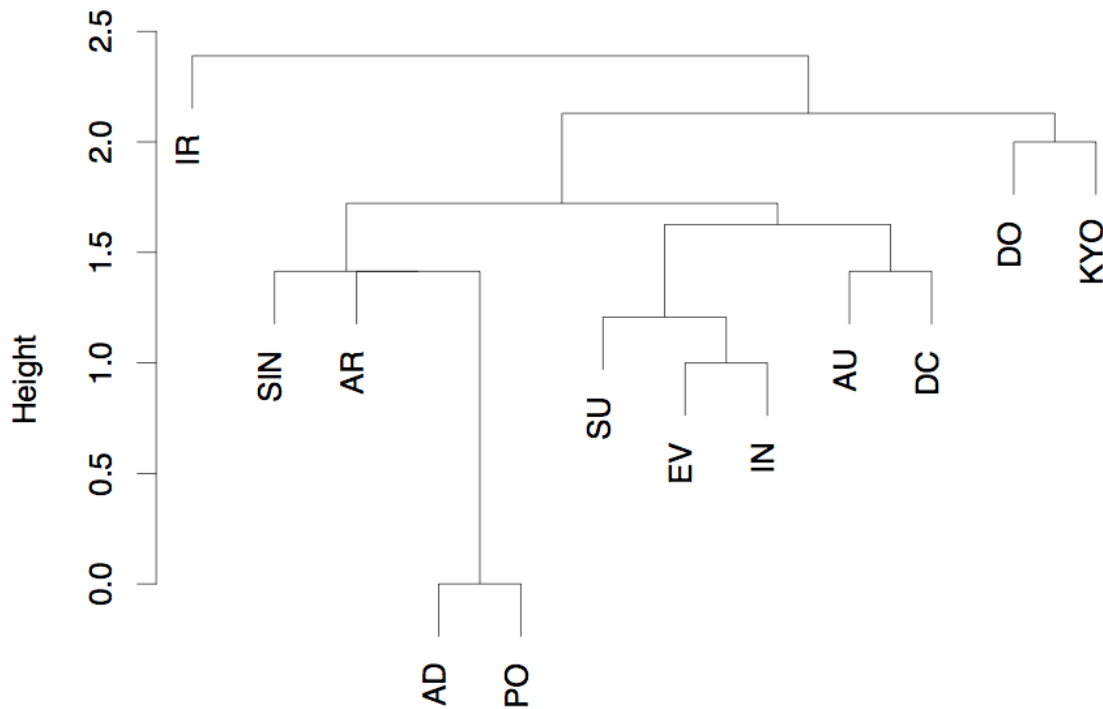


Figure 35 : résultats de la classification hiérarchique des 12 attitudes à partir de leur description en AUs.

d- Analyse et comparaison avec les résultats subjectifs

La liaison entre les évaluations objective et subjective de ce corpus audio-visuel n'est pas évidente. L'étude similaire menée sur six attitudes audio-visuelles du français (Rilliard et al., en révision) donne une liaison entre ces deux niveaux beaucoup plus claire. Dans ce travail, la F0, la durée et les AUs permettent de distinguer clairement ironie de soupçon des autres attitudes (*déclaration*, *question-simple*, *doute-incrédulité*, *exclamation de surprise* et *évidence*). Ils identifient également l'opposition entre les expressions déclaratives et interrogatives à l'aide de ces paramètres.

L'analyse de nos données japonaises donne quelques indices concernant les caractéristiques principales de la prosodie attitudinale, mais la description des paramètres prosodiques nous semblent insuffisante afin de cerner la complexité de toutes les expressions. Deux mesures devront être analysées : une mesure de la qualité de voix qui apporte des caractéristiques importantes pour la perception de certaines attitudes japonaises telles que *kyoshuku* (cf. supra pour plus de détails) ; et une description de l'intonation en terme de contours, indépendamment des valeurs extraites sur les énoncés (cf. supra les paramètres prosodiques utilisés). On notera que le résultat sensiblement meilleur obtenu sur le français avec le même ensemble de paramètres peut s'expliquer par (1) le plus petit nombre d'attitudes et (2) par l'usage beaucoup moins important de la qualité de voix pour les expressions attitudinales du français.

Nous avons toutefois identifié un ensemble d'AU caractéristique des expressions attitudinales du japonais. Curieusement, la *déclaration* n'utilise pas d'AU spécifique, mais elle est reconnue en modalité visuelle seule, ce qui pourrait être interprété comme

la reconnaissance par les sujets d'une expression sans expressivité particulière : l'expression « neutre » que l'on retrouve dans tant d'étude sur les affects. De plus, les attitudes « interrogatives » sont produites avec une montée de F0 à la fin de l'énoncé, et elles utilisent l'AU1+2 (c-à-d, une activation musculaire dans la portion intérieure et extérieure des sourcils).

5.5.3.4. Conclusion

L'objectif de cette partie de thèse est d'envisager comment les auditeurs japonais reconnaissent les expressions attitudinales dans les modalités audio et visuelle, d'une part ; et d'autre part de réaliser une première mesure de l'éventuelle redondance (ou de la complémentarité) de ces deux modalités pour chaque attitude représentée. Les résultats montrent l'importance de la performance du locuteur, de la modalité et une large variabilité due aux attitudes exprimées.

Presque toutes les attitudes sont reconnues au-dessus du seuil de hasard dans chacune des modalités de présentation, même si globalement, la bimodalité donne le meilleur taux d'identification pour toutes les attitudes, ce qui montre la bonne synergie existant entre ces deux modalités. Cependant, pour certaines attitudes, une seule modalité (c-à-d, soit l'audio seule soit le visuel seul) apporte déjà la même quantité d'information qu'en bimodal : l'autre modalité est alors redondante.

Généralement les auditeurs regroupent un ensemble de 12 attitudes en trois catégories perceptives (pour le locuteur A). Le premier groupe est alors composé des expressions de politesse, le second des attitudes « interrogatives », et le troisième des expressions d'imposition de l'opinion du locuteur. Les attitudes de *kyoshuku* et *exclamation de surprise* sont particulièrement bien reconnues grâce aux informations visuelles pour le locuteur A.

Les attitudes produites par le locuteur B sont nettement moins bien reconnues et montrent beaucoup de confusions, qui peuvent avoir plusieurs facteurs : l'inexpérience du locuteur B, les différences de stratégies expressives (les attitudes de B sont plus subtiles et moins intenses). Tout cela montre aussi l'importance du contexte discursif dans la reconnaissance des affects sociaux. La différence d'intensité expressive entre A et B reflète bien cela.

L'analyse objective du locuteur A montre que les paramètres observés semblent insuffisants pour expliquer les performances des auditeurs. Cela peut avoir plusieurs origines. D'abord les paramètres eux-mêmes, qui sont cependant assez proches de ceux utilisés dans l'analyse du premier corpus de japonais et donnaient de meilleurs résultats. Mais surtout le petit nombre de phrases impliquées dans ce test de perception – ce qui donne des mesures sans aucun doute moins fiables que celles obtenues à partir de l'analyse d'un corpus complet. Par ailleurs, l'analyse de la qualité de voix nous paraît nécessaire.

5.6. Test de gating sur les attitudes japonaises

5.6.1. Introduction

Nous venons de montrer dans les sections précédentes que les analyses statistiques de la perception des attitudes japonaises permettent de mieux comprendre le comportement perceptif des locuteurs japonais pour les 12 attitudes. L'analyse acoustique nous a également permis d'extraire des caractéristiques prosodiques pour chaque attitude du locuteur. L'enjeu essentiel de cette partie de thèse est d'étudier comment l'information linguistique est répartie sur ces contours stéréotypiques de phrase, en ayant conscience que certaines parties de ces contours peuvent porter plus d'information que d'autres.

Les auditeurs utilisent les informations passées et présentes, non seulement pour analyser la phrase jusqu'au point atteint par le locuteur, mais également pour prédire les informations à venir. Afin de tester cette « anticipation » des auditeurs et surtout afin d'identifier à quel moment (point d'unicité) ils seront capables de reconnaître l'attitude appropriée du locuteur, nous avons appliqué le paradigme du gating (c-à-d, du dévoilement progressif) aux contours prosodiques de phrase (cf. Grosjean, 1985).

Le paradigme de gating consiste à dévoiler progressivement aux auditeurs un stimulus de parole depuis son commencement jusqu'à sa présentation complète. Ce stimulus peut être constitué d'une more (dans le cas du japonais), d'un mot, d'un syntagme ou même d'une phrase complète suivant le type de test que l'on veut réaliser. Le découpage du stimulus peut être temporel (Grosjean 1985, Cotton & Grosjean 1984, Dupoux 1989) en présentant une portion du signal avec un intervalle de temps régulier, ou syllabique en découpant le signal à la fin de chaque more, syllabe voir par mot (Thorsen 1980, Grépillat 1996, Morlec et al. 1999). Quel que soit la façon de découper le stimulus, cette portion de signal sera nommée gate.

Aubergé et al. (1997), Grépillat (1996), Morlec et al. (1999) ont effectué un test de perception en utilisant ce paradigme de gating avec les six attitudes françaises validées dans les travaux déjà présentés (cf. supra). Pour ce test perceptif, ils ont sélectionné trois phrases de deux syllabes et six phrases de cinq syllabes prononcées avec six attitudes pour chaque phrase. Ensuite, ces phrases ont été découpées à la fin de chaque syllabe : jusqu'à la dernière syllabe d'une phrase. Le reste de la phrase était masqué par un bruit blanc d'une intensité similaire à celle de la phrase. Ensuite, l'ensemble des stimuli ont été présentés au sujet, par nombre de gate croissant (d'abord tous les stimuli d'un gate, puis ceux de deux gate...). Douze auditeurs ont cherché à identifier l'attitude véhiculée par ces stimuli. L'analyse montre que globalement les auditeurs ont mieux reconnu les attitudes sur les phrases de cinq syllabes plutôt que sur celles de deux syllabes. Les résultats montrent une identification progressive (une croissance continue du taux de

reconnaissance) pour cinq attitudes (*doute-incrédulité*, *évidence*, *exclamation de surprise*, *question-simple* et *ironie de soupçon*). Cependant, les auteurs observent une baisse du score (25-47%) au quatrième gate pour le *doute-incrédulité* sur la phrase de cinq syllabes, puis au cinquième gate, le taux d'identification monte à 80%. En ce qui concerne la *déclaration*, cette attitude est bien reconnue dès le premier gate, et ce taux d'identification demeure similaire jusqu'au dernier gate.

Nous allons maintenant présenter une série de tests perceptifs pour les attitudes japonaises basée sur ce paradigme du gating. Plusieurs tests ont été nécessaires, car la complexité de la tâche nous a poussé à modifier légèrement les stimuli ou le paradigme expérimental afin d'être à même de répondre aux questions soulevées par les différentes expérimentations.

5.6.2. Premier test de gating

5.6.2.1. Corpus

Nous avons réalisé ce test sur la base du premier corpus de douze expressions attitudinales japonaises présenté dans cette thèse (cf. *supra*). Dans cette expérimentation, nous avons sélectionné deux phrases extraites de ce corpus, respectivement de deux et de cinq mores (cf. tableau 13), qui ont été validées (*supra*). Ces phrases portent un accent lexical sur la première more (cf. Hirata 1960).

Ces deux phrases, produites avec les 12 attitudes du japonais, ont été découpées en deux ou cinq gates, more après more. La partie masquée de la phrase a été remplacée par un bruit blanc, d'une longueur calculée afin que tous les stimuli aient la même durée. Cela a été fait afin que les sujets ne puissent pas inférer la durée restante de la phrase à partir de la durée du stimulus. À la suite de cela, nous avons donc obtenu un ensemble de 84 stimuli (7 gates – 2 pour la phrase de 2 mores et 5 pour celle de 5 mores – x 12 attitudes) qui ont été utilisés pour le test perceptif.

Tableau 13 : phrases du test de dévoilement progressif.

nb de more	Phrase	Traduction
2	Na*ra	Nara
5	Na*ra de neru	Il dort à Nara

5.6.2.2. Protocole expérimental

Les stimuli sont présentés aux auditeurs par groupes croissant de gate (d'abord les stimuli de un gate, puis ceux de deux gates...), jusqu'à leur présenter la phrase en entier. Le test perceptif se déroule en face d'un écran, les sujets portant des casques HDA200 Sennheiser dans une salle calme. Le test est effectué grâce à une interface développée à l'aide du logiciel « Runtime revolution », avec un ordinateur PC Windows XP professionnel. L'interface du test présente les 12 attitudes avec leurs définitions. Avant le début du test, nous avons distribué une feuille aux sujets sur laquelle se trouve la définition de chaque attitude, additionnée d'exemples de contextes possibles, dans le

dessein de lever de possibles ambiguïtés de compréhension de ces 12 attitudes. Aucun sujet n'a rapporté avoir eu des difficultés pour comprendre les concepts sous-jacents aux attitudes.

Tous les sujets ayant participé à cette expérimentation ont écouté chaque stimulus une fois seulement. Pour chaque stimulus, ils doivent juger de l'attitude perçue parmi les 12 possibles. L'ordre de présentation des stimuli était aléatoire et différent pour chaque sujet.

Le test de perception est effectué auprès de 31 Japonais natifs (19 femmes et 12 hommes) qui habitent dans la région centrale du Japon (Chubu) depuis toujours. L'âge moyen des sujets est de 20,6 ans. Aucun sujet n'habitait à l'étranger avant la passation de ce test perceptif. Aucun auditeur n'a signalé de trouble de l'audition.

5.6.2.3. Résultats

a- Analyse de la variance

Une ANOVA à deux facteurs a été menée sur les résultats du test de perception : le facteur gate, qui rend compte de l'influence du nombre de gate sur la reconnaissance des attitudes, et le facteur « attitude », qui rend compte de l'influence de l'expression sur les performances des sujets. Les résultats montrent une influence significative du gate sur les réponses, pour les deux longueurs de phrases ($F = 44.4$ p. < .01 pour la phrase de 2 mores et $F = 13.311$ p. < .01 pour la phrase de 5 mores). Un effet significatif de l'attitude est également observé pour les deux phrases ($F = 11.125$ p. < .01 pour la phrase de 2 mores et $F = 18.621$ p. < .01 pour la phrase de 5 mores).

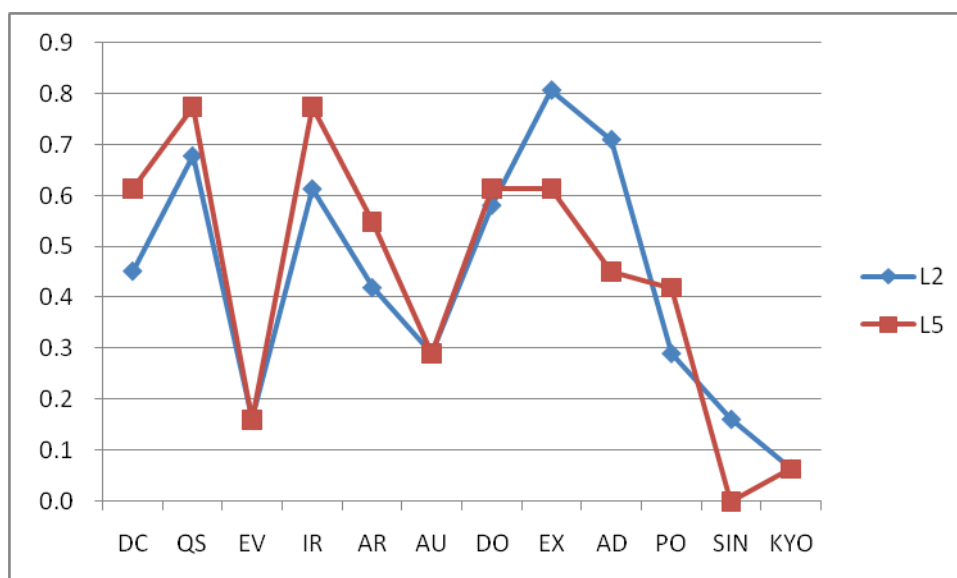


Figure 36 : taux de reconnaissance des attitudes pour les phrases de 2 et de 5 mores. Les attitudes sont présentées en abscisse, et le taux de reconnaissance au dernier gate en ordonnée, pour les 31 auditeurs japonais.

b- Taux de reconnaissance des 12 attitudes

La figure 36 présente le taux de reconnaissance de chacune des 12 attitudes pour les phrases complètes de 2 et de 5 mores. Les auditeurs montrent un comportement presque similaire pour toutes les attitudes entre deux phrases de la longueur différente sauf pour l'attitude d'*exclamation de surprise*, *admiration* et de *sincérité-politesse*. Parmi les attitudes testées, évidence, autorité, sincérité-politesse et kyoshuku sont généralement mal perçus.

c- Évolution du taux de reconnaissance en dévoilement progressif

La figure 37 illustre clairement la croissance progressive du taux de reconnaissance avec l'évolution du nombre de gates, toutes attitudes confondues. Ce résultat suggère que la reconnaissance des attitudes se fait progressivement tout au long de l'énoncé.

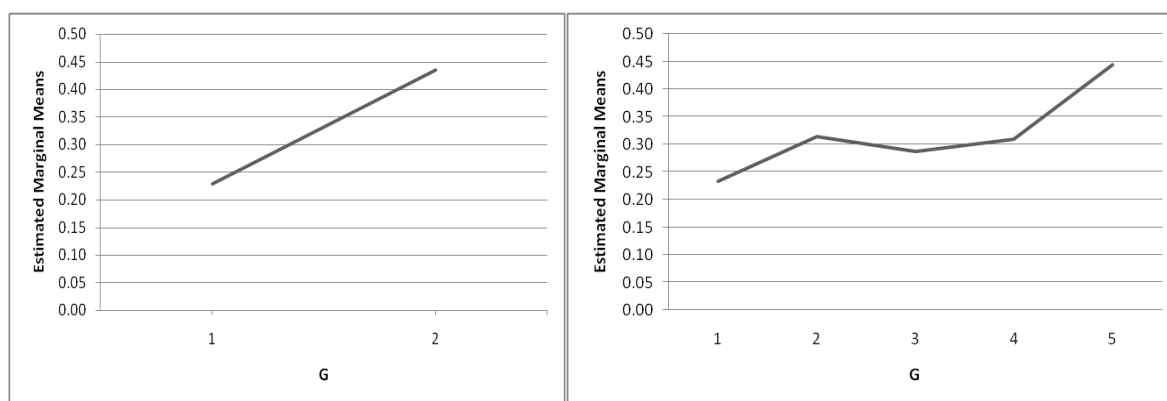


Figure 37 : évolution du taux de reconnaissance toutes attitudes confondues pour les phrases de 2 et 5 mores. Les abscisses indiquent le gate, l'ordonnée indique le taux de reconnaissance moyen.

Pourtant, cette augmentation du taux d'identification avec l'évolution du gate n'est pas vérifiée sur toutes les attitudes. Selon un test post-hoc (« *simple main effect analysis* »), l'effet du gate pour la phrase de deux mores est présent pour les expressions de *question-simple* ($F = 42.654$, $p < .00$), *doute-incrédulité* ($F = 21.762$, $p < .00$), *irritation* ($F = 13.928$, $p < .00$), *admiration* ($F = 11.703$, $p < .00$) et *arrogance-impolitesse* ($F = 11.703$, $p < .00$). Nous avons également effectué le même test statistique pour la phrase de 5 mores, et nous avons identifié un effet du gate pour les attitudes de *question-simple* ($F = 23.383$, $p < .00$), *doute-incrédulité* ($F = 9.788$, $p < .00$), *irritation* ($F = 10.347$, $p < .00$), *politesse-simple* ($F = 5.318$, $p < .00$) et *arrogance-impolitesse* ($F = 6.891$, $p < .00$). Ces résultats montrent que le gate influe sur un certain nombre d'attitudes (presque toujours les mêmes), sauf pour l'*admiration* et la *politesse-simple*.

d- Effet du gate par attitude

La figure 38 représente l'évolution progressive de la reconnaissance pour 3 attitudes pour lesquelles a été montré un effet du gate. Il s'agit des attitudes de : *politesse-simple*, *admiration* et *arrogance-impolitesse*. Pour ces expressions, plus une portion importante de la phrase est présentée, et mieux les auditeurs reconnaissent ces deux attitudes.

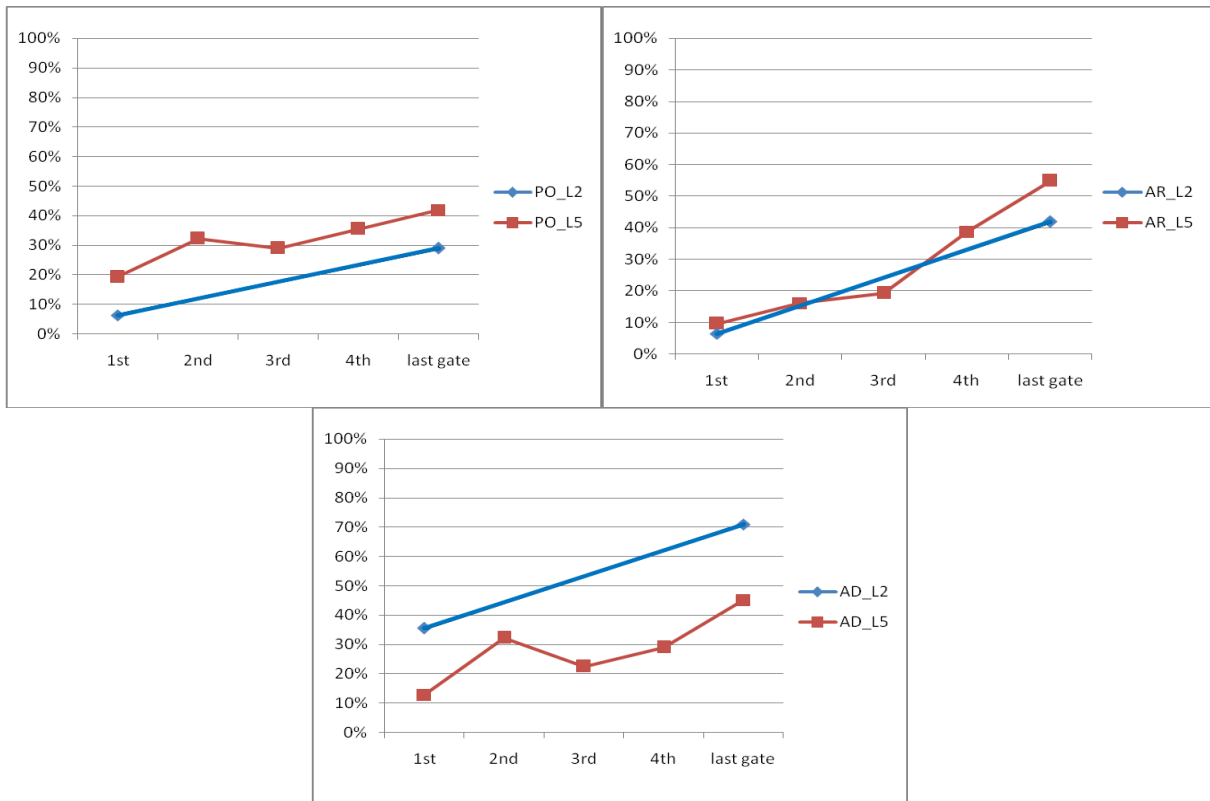


Figure 38 : taux de reconnaissance par gate pour *politesse-simple* (PO), *arrogance-impolitesse* (AR), *admiration* (AD), pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.

En ce qui concerne *question-simple*, *doute-incrédulité* et *irritation* (cf. figure 39), le taux d'identification pour ces trois attitudes monte brutalement au dernier gate. Cela implique que l'information prosodique portée par la dernière more joue un rôle décisif pour l'identification de ces trois attitudes.

En revanche, le taux d'identification des trois attitudes d'*autorité*, de *déclaration* et d'*exclamation de surprise*, et d'environ 40 à 60 % dès le premier gate, contrairement aux autres attitudes (*supra*). Ce résultat indique que les auditeurs reconnaissent ces deux attitudes très tôt (dès le début) de l'énoncé.

En ce qui concerne les deux expressions culturelles de la politesse (*kyoshuku* et *sincérité-politesse*), le taux de reconnaissance reste faible jusqu'à la fin de phrase (cf. figure 41). Pourtant cette figure ne nous permet pas de savoir si les auditeurs ont pu identifier globalement la notion de politesse car ces graphes ne prennent pas en compte

la confusion des attitudes perçues. C'est pourquoi nous avons analysé la perception globale de la notion de politesse²⁹ pour ces deux attitudes de politesse. La figure 42 montre la perception globale de la notion de politesse pour les stimuli de *kyoshuku* et *sincérité-politesse*. Selon ces graphes, le taux de la reconnaissance de la notion de politesse pour l'expression de *kyoshuku* reste similaire aux résultats de l'identification de cette attitude mentionnés dans le paragraphe précédent. Cette attitude est principalement confondue avec l'*irritation* ou l'*autorité*. En revanche, *sincérité-politesse* a bien été perçue comme une expression de politesse dès le premier gate avec le taux de 40 %, puis ce taux d'identification reste plus ou moins stable jusqu'à la fin de phrase. Ce résultat nous indique que même si les auditeurs reconnaissent mal *sincérité-politesse*, cette expression est perçue comme une expression de politesse dès le début de l'énoncé.

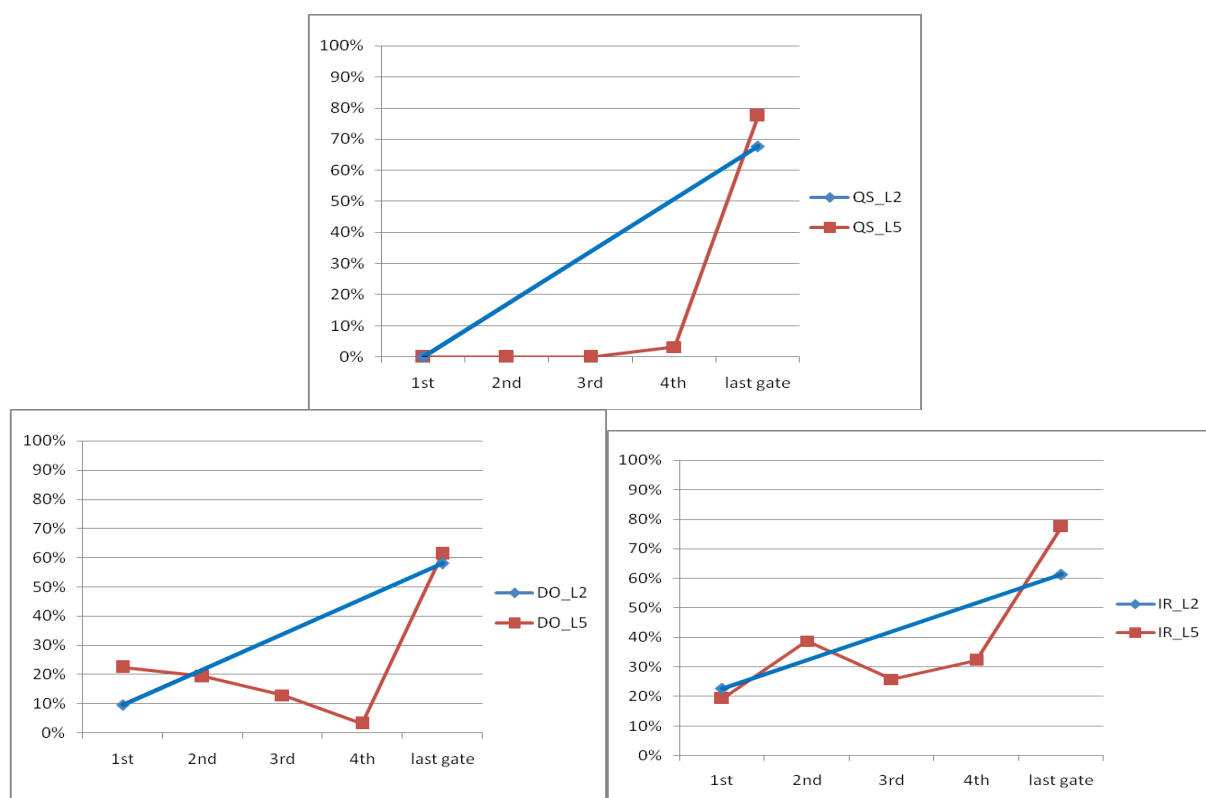


Figure 39 : taux de reconnaissance par gate pour question-simple (QS) en haut à gauche, doute-incrédulité (DO) en haut à droite et irritation (IR), pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.

²⁹ La réponse des sujets est pour les stimuli de *kyoshuku* et *sincérité-politesse* correspondent à politesse-simple, *kyoshuku* ou *sincérité-politesse*.

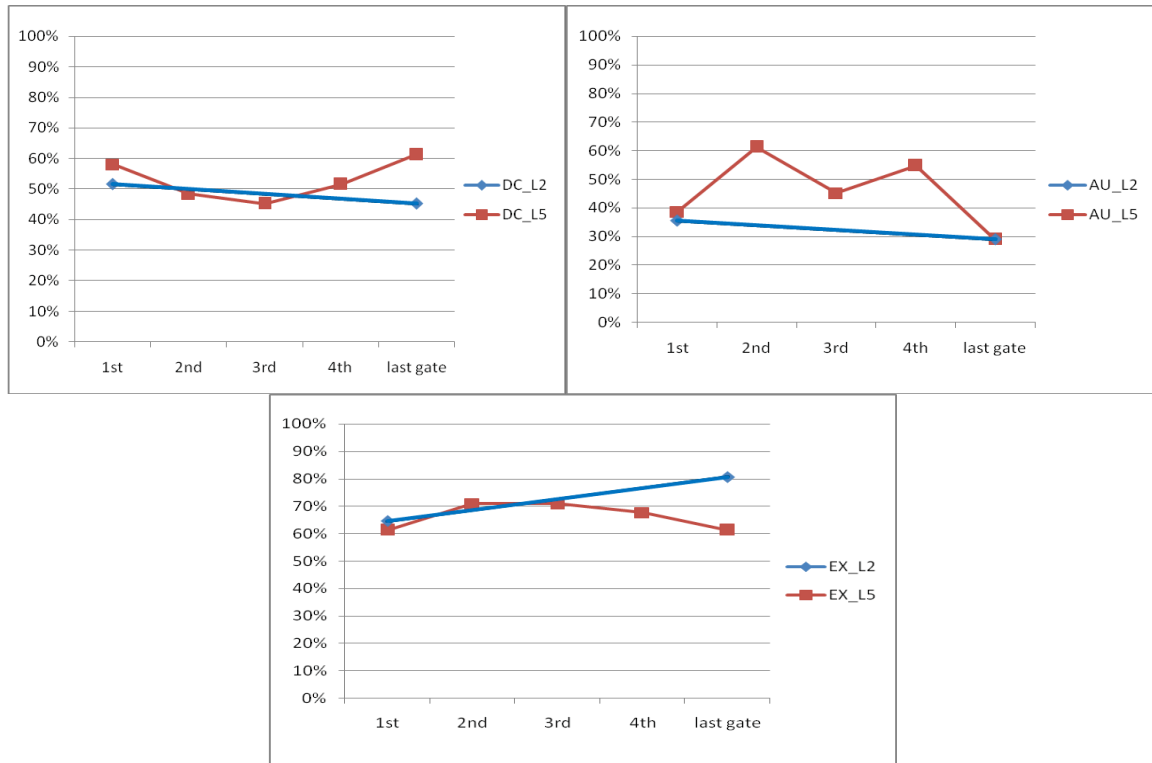


Figure 40 : taux de reconnaissance par gate pour déclaration (DC), autorité (AU) et exclamation de surprise (EX, pour les phrases de 2 et 5 mores. L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.

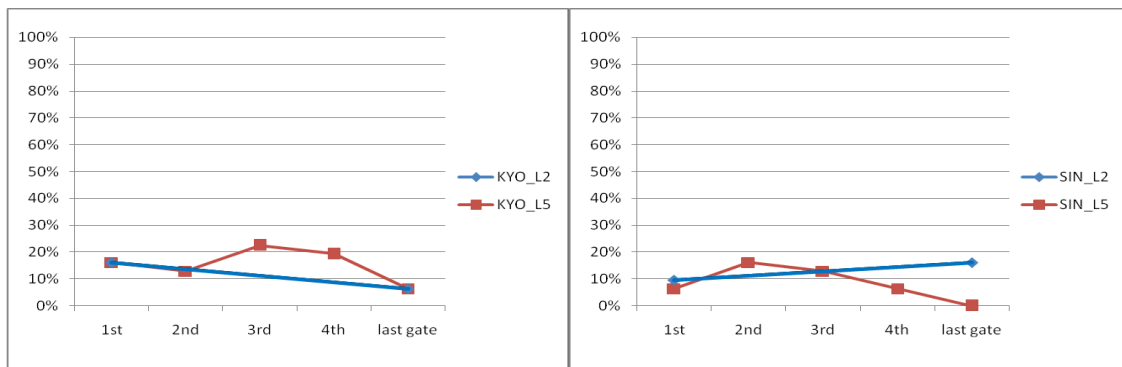


Figure 41 : taux de reconnaissance par gate pour les expressions de politesse des phrases de 2 et 5 mores. Le graphe de gauche pour le stimulus de kyoshuku, celui de droite la sincérité-politesse) L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.

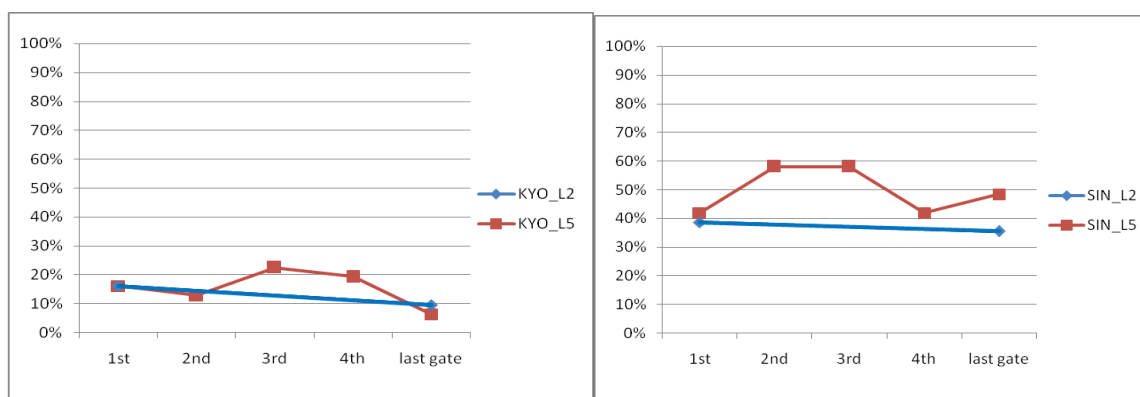


Figure 42 : taux de reconnaissance par gate pour les expressions de politesse regroupées ensemble, pour les phrases de 2 et 5 mores. Le graphe de gauche pour le stimulus de *kyoshuku*, celui de droite la *sincérité-politesse*) L'abscisse présente le gate, l'ordonnée le taux de reconnaissance.

5.6.2.4. Conclusion

L'objectif de ce test perceptif préliminaire était de savoir si les auditeurs japonais sont capables d'identifier l'attitude exprimée avant la fin de la phrase.

L'ANOVA sur les résultats montre qu'une influence significative du gate a été identifiée pour les deux phrases de la longueur différente ; un effet significatif du choix des attitudes est également observé pour les deux phrases. Pourtant, le même test statistique indique que l'effet du gate est présent uniquement pour *question-simple*, *doute-incrédulité*, *irritation* et *arrogance-impolitesse* pour deux phrases de longueurs différentes, et sur *admiration* pour la phrase de 2 mores ou *politesse-simple* pour celle de 5 mores.

Parmi les attitudes montrant un effet du gate, *question-simple*, *doute-incrédulité* et *irritation* sont reconnues seulement au dernier gate. Cela implique que l'information prosodique est portée par cette more. En ce qui concerne *politesse-simple*, *admiration* et *arrogance-impolitesse*, ces expressions montrent une croissance progressive du taux de reconnaissance avec l'évolution du gate.

En revanche, trois attitudes n'ont pas d'effet du gate : *autorité*, *déclaration* et *exclamation de surprise*, qui sont reconnues immédiatement dès le premier gate.

Les deux expressions culturelles de la politesse (*kyoshuku* et *sincérité-politesse*) pour lesquelles il n'y a pas d'effet du gate, montrent un taux de reconnaissance faible du début à la fin de la phrase. En effet, l'attitude de *kyoshuku* est fortement confondue avec *irritation* ou *autorité* du fait d'un biais lexical des phrases testées. Par contre, l'attitude de *sincérité-politesse* est perçue immédiatement comme une expression de politesse dès le début de l'énoncé.

5.6.3. Deuxième test de gating

Lors du premier test de gating avec les phrases de 2 et de 5 mores, nous avons observé la confusion importante entre l'attitude de *kyoshuku* et celles d'*irritation* ou d'*autorité*. Nous constatons que cela est dû à l'influence du contenu sémantique qui a gêné les auditeurs pour percevoir une telle attitude dans ce contexte. C'est pourquoi nous présentons maintenant ce second test, réalisé avec une nouvelle phrase dont le sens est plus facilement adaptable au contexte du *kyoshuku*.

5.6.3.1. Corpus

Afin de pouvoir comparer les résultats obtenus avec ceux du test précédent (phrases portant un accent lexical sur la première more), nous avons utilisé la phrase de 8 mores accentuée sur la première more (cf. tableau 14), validée *supra*. Nous avons également retenu la phrase de 8 mores sans accent lexical afin de tester une éventuelle influence de l'accent lexical (qui n'a pas été observée pour l'instant).

Tableau 14 : phrases sélectionnées pour le test de dévoilement progressif.

nb de more	Phrase	Traduction
8	Na* goya de nomimasu	Il boit à Naogya
8	Naniwa de nomimasu	Il boit à Naniwa

Les stimuli sont préparés selon la même procédure que pour l'expérience précédente : chaque phrase est découpée en gate, more par more, et la partie finale de l'énoncé est masquée par un bruit blanc. On obtient ainsi un ensemble de 192 stimuli (8 gates x 2 phrases x 12 attitudes x 1 locuteur) utilisés pour ce second test perceptif.

5.6.3.2. Protocole expérimental

Les stimuli sont présentés aux auditeurs par nombre croissant de gate, jusqu'à la phrase entière, et sont mélangés en ordre aléatoire à l'intérieur de chaque groupe de gate. Les tests perceptifs se déroulent devant un écran, les sujets portant des casques HDA200 Sennheiser dans une salle calme. Tout le test est effectué sur une interface développée grâce au logiciel « Runtime revolution » avec un ordinateur PC Windows XP professionnel. La procédure du test perceptif est exactement la même que celle du test précédent (*supra*).

Le test de perception est effectué auprès de 10 Japonais natifs (6 femmes et 4 hommes) qui habitent dans la région centrale du Japon (Chubu) depuis toujours. L'âge moyen des sujets est de 25,3 ans. Aucun sujet n'habitait à l'étranger avant la passation de ce test perceptif. Aucun auditeur n'a signalé de trouble de l'audition.

5.6.3.3. Résultats

a- L'analyse d'ANOVA (3 facteurs : gate x attitudes x accent lexical)

Tout d'abord, l'effet du gate sur les résultats de réponses est évalué par l'ANOVA. Les résultats montrent que l'influence de l'accent lexical sur les réponses n'est pas observée ($F = 0.031$, $p. > .01$), mais nous avons identifié un effet significatif du choix des attitudes ($F=13.237$, $p.<.01$). Un effet significatif du gate sur les réponses des auditeurs est également observé ($F = 11.965$, $p. < .01$). L'interaction entre deux facteurs (facteur gate et facteur d'accent lexical) ne donne pas d'effet significatif ($F = 1.118$, $p. > .01$), mais celle entre le facteur des attitudes et le facteur d'accent lexical donne un effet significatif ($F = 5.461$, $p. < .01$).

b- L'évolution du taux de reconnaissance en dévoilement progressif

La figure 43 représente l'augmentation du taux d'identification des attitudes avec l'évolution de la position du gate. Ce graphique illustre la croissance linéaire du taux global de reconnaissance pour toutes les attitudes confondues avec l'évolution de la position du gate. Ce résultat nous confirme l'effet du gate pour l'identification des attitudes, autrement dit, plus les auditeurs écoutent une portion importante de l'énoncé, mieux ils reconnaissent l'attitude exprimée.

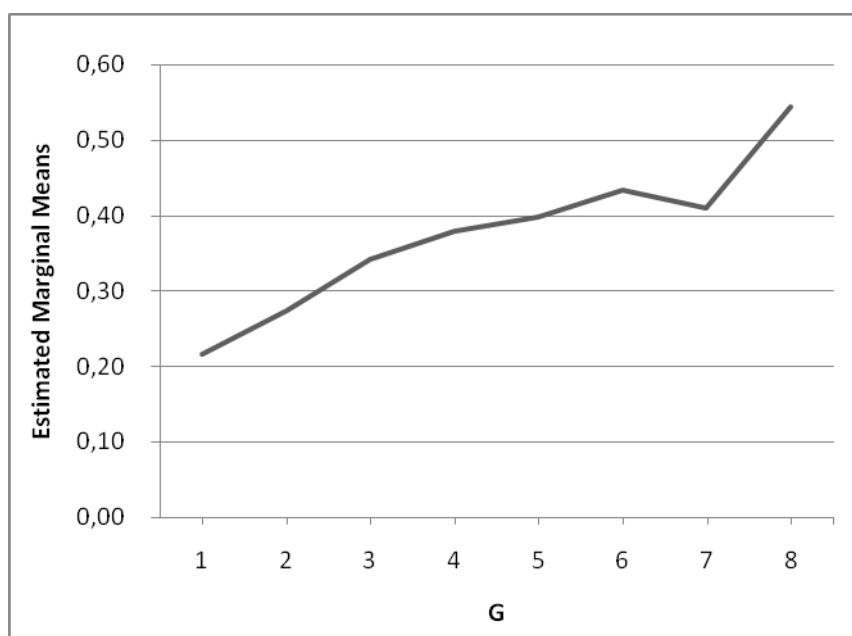


Figure 43 : croissance progressive de la reconnaissance globale pour toutes les attitudes confondues sur les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais.

Pourtant, cette augmentation du taux d'identification avec l'évolution du gate n'est pas justifiée pour toutes les attitudes. Nous avons alors effectué un test post-hoc

(*simple main effect analysis*) en mettant le facteur gate sur chacune des attitudes. Les résultats montrent que l'effet du gate est observé précisément pour sept attitudes suivantes : *déclaration* ($F = 3.026$, $p. < .01$), *question-simple* ($F = 8.166$, $p. < .01$), *exclamation de surprise* ($F = 2.878$, $p. < .01$), *doute-incrédulité* ($F = 6.392$, $p. < .01$), *irritation* ($F = 11.598$, $p. < .01$), *politesse-simple* ($F=5.074$, $p.<.01$) et *arrogance-impolitesse* ($F = 7.183$, $p. < .01$).

c- L'effet du gate sur la reconnaissance de chacune des attitudes

La figure 44 représente le taux de la reconnaissance de chaque position du gate pour quatre attitudes dans lesquelles nous avons identifiées l'effet du gate. Selon ces graphes, l'attitude du *doute-incrédulité*, *irritation*, *politesse-simple* et *arrogance-impolitesse* montrent la croissance linéaire du taux d'identification.

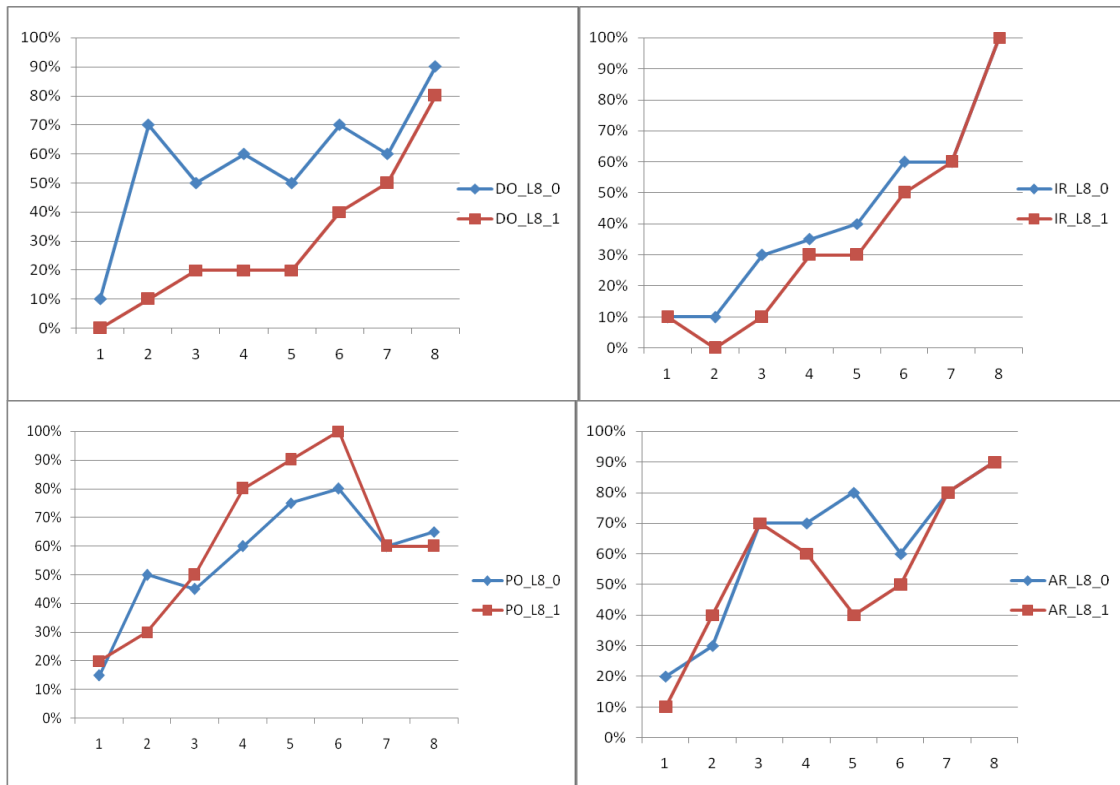


Figure 44 : croissance linéaire du taux d'identification pour quatre attitudes dans lesquelles l'effet du gate est identifié : *doute-incrédulité* [DO] (en haut à gauche), *irritation* [IR] (en haut à droite), *politesse-simple* [PO] (en bas à gauche) et *arrogance-impolitesse* [AR] (en bas à droite) pour les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais.

Lorsqu'on regarde les détails pour l'identification de chaque attitude, on voit d'abord, une différence significative entre deux phrases testées pour *doute-incrédulité* ($F = 14.109$, $p. < .01$). Cette attitude exprimée par la phrase sans accent lexical est

identifiée dès la position du second gate (70%) tandis que la même attitude exprimée par la phrase avec un accent lexical en première more atteint le même taux d'identification seulement en fin de phrase. L'attitude de *politesse-simple* enregistre la baisse du score en position de 7^{ème} gate. Enfin, *arrogance-impolitesse* exprimée par deux phrases montre le taux d'identification presque identique en chaque position du gate sauf en position du 5^{ème} gate.

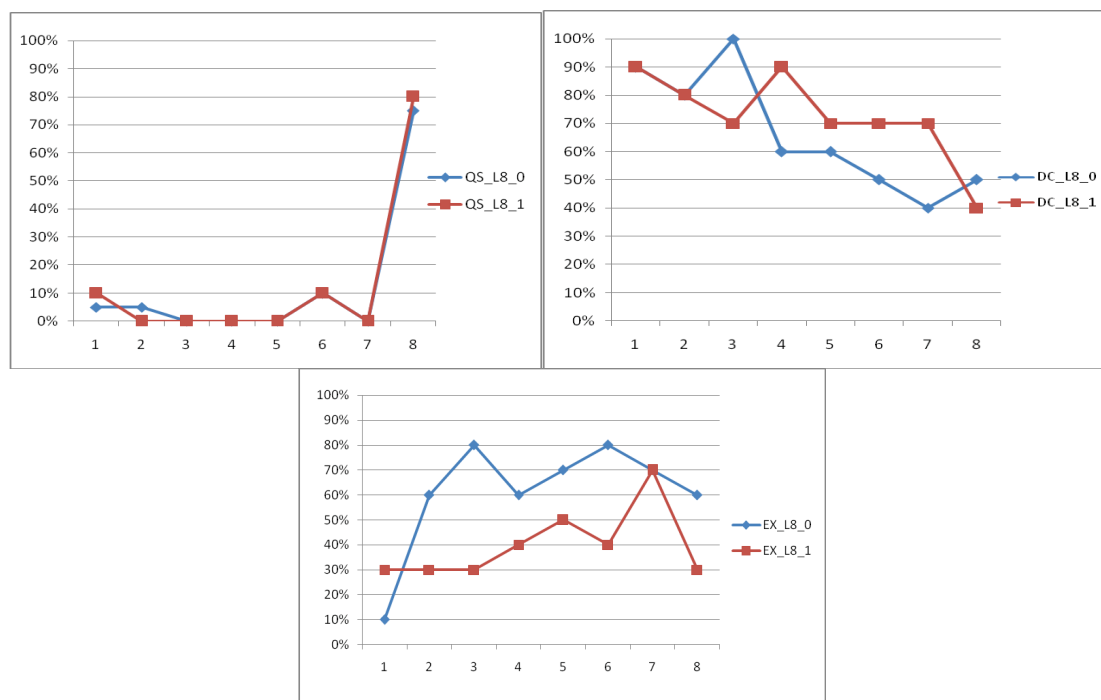


Figure 45 : Le mouvement du taux d'identification en chaque position du gate pour trois autres attitudes dans lesquelles l'effet du gate est identifié : question-simple [QS] (en haut à gauche), déclaration [DC] (en haut à droite) et exclamation de surprise [EX] (en bas) pour les phrases de 8 mores. Les abscisses indiquent l'évolution du gate, les ordonnées le taux de reconnaissance moyenne de 10 auditeurs japonais.

Nous voyons maintenant le mouvement du taux d'identification pour trois autres attitudes dans lesquelles nous avons reconnu l'effet du gate (cf. figure 45). D'abord, *question-simple* est reconnu uniquement en dernière position du gate. Ce résultat avec les phrases de 8 mores confirme le même phénomène avec les phrases de 2 et de 5 mores présenté *supra*.

Le taux de reconnaissance pour *déclaration* montre la diminution du début vers la fin pour deux phrases testées. Une explication possible pour ce phénomène est que les auditeurs ont tendance à choisir *déclaration* lorsqu'ils ne savaient pas identifier d'autres attitudes surtout en position du 1^{ère} gate. Enfin, pour l'attitude d'*exclamation de surprise*, nous avons identifié une différence significative du taux de reconnaissance entre deux phrases testées ($F = 7.462$, $p < .01$). Comme pour l'attitude de *doute-incrédulité*, cette attitude exprimée par la phrase sans accent lexical est mieux identifiée dès la position du second gate (60%) tant dit que la même attitude exprimée par la phrase avec un

accent lexical en première more atteint le même taux d'identification seulement en 7^{ème} position du gate.

Par ailleurs, trois attitudes pour lesquelles nous n'avons pas pu identifier l'effet du gate sont présentées en figure 46. Parmi ces trois attitudes, deux attitudes : *admiration* et *évidence* montrent une différence significative du taux de reconnaissance entre deux phrases testées ($F = 10.523$, $p. < .01$ pour admiration, et $F = 10.523$, $p. < .01$ pour évidence). *Admiration* exprimée avec la phrase de 8 mores sans accent lexical n'est pas du tout reconnue (0%) sauf en position du 1^{ère} et 6^{ème} gate. Afin de comprendre la cause de cette faible reconnaissance pour cette attitude avec la phrase de 8 mores sans accent, nous voyons la matrice de confusions pour cette attitude (cf. tableau 15). Selon le tableau, *admiration* est confondu avec *déclaration* jusqu'à la 4^{ème} position du gate, mais cette attitude est mélangée avec les expressions de politesse (*politesse-simple*, *kyoshuku* et *sincérité-politesse*).

Quant à *évidence*, nous observons un fort écart du degré d'identification entre deux phrases en 2^{ème} à 4^{ème} position du gate. En général, la phrase avec l'accent lexical est mieux reconnue dans ces portions du gate par rapport à la phrase sans accent, mais nous voyons l'augmentation du taux de reconnaissance pour ces deux phrases en fin de phrase (à partir de la position de 6^{ème} gate).

Tableau 15 : réponses des 10 auditeurs japonais à chaque gate, pour le stimulus d'admiration exprimé avec la phrase de 8 mores sans accent lexical. Les chiffres en gras italique indiquent le taux de reconnaissance, et les chiffres en gras désignent un degré de confusion important.

8 mores sans accent	Nombre de gate présentés aux sujets							
	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
<i>Admiration</i>	30%	0%	0%	0%	0%	10%	0%	0%
Arrogance	0%	10%	0%	0%	0%	0%	20%	10%
Autorité	0%	0%	0%	0%	0%	0%	0%	0%
Déclaration	50%	30%	20%	20%	10%	10%	0%	0%
Doute-incrédulité	10%	10%	10%	0%	0%	10%	0%	0%
Evidence	0%	0%	10%	0%	0%	0%	30%	0%
Surprise	0%	10%	0%	10%	10%	0%	0%	0%
Irritation	10%	0%	0%	0%	0%	0%	0%	0%
Kyoshuku	0%	10%	0%	10%	20%	10%	20%	20%
Politesse	0%	10%	40%	40%	30%	40%	10%	50%
Question	0%	20%	20%	10%	20%	0%	0%	0%
Sincérité-Politesse	0%	0%	0%	10%	10%	20%	20%	20%

Pour l'attitude d'autorité, nous n'avons pas identifié l'effet significatif de l'accent lexical pour l'identification de l'attitude ($F = 4.198$, $p. > .01$), mais nous observons tout de même un différent degré du taux d'identification entre les deux phrases en position de

7^{ème} et 8^{ème} position du gate. Le taux d'identification de cette attitude exprimée en phrase sans accent lexical chute en position de septième gate (50% à 0% d'identification), mais elle est reconnue surtout comme irritation (30% de confusion) ou arrogance-impolitesse (30% de confusion) en dernière position du gate. Cela veut dire que les auditeurs perçoivent l'imposition de l'avis du locuteur (*autorité*, *arrogance-impolitesse* et *irritation*) avec le taux de 70% en dernière position du gate des deux phrases.

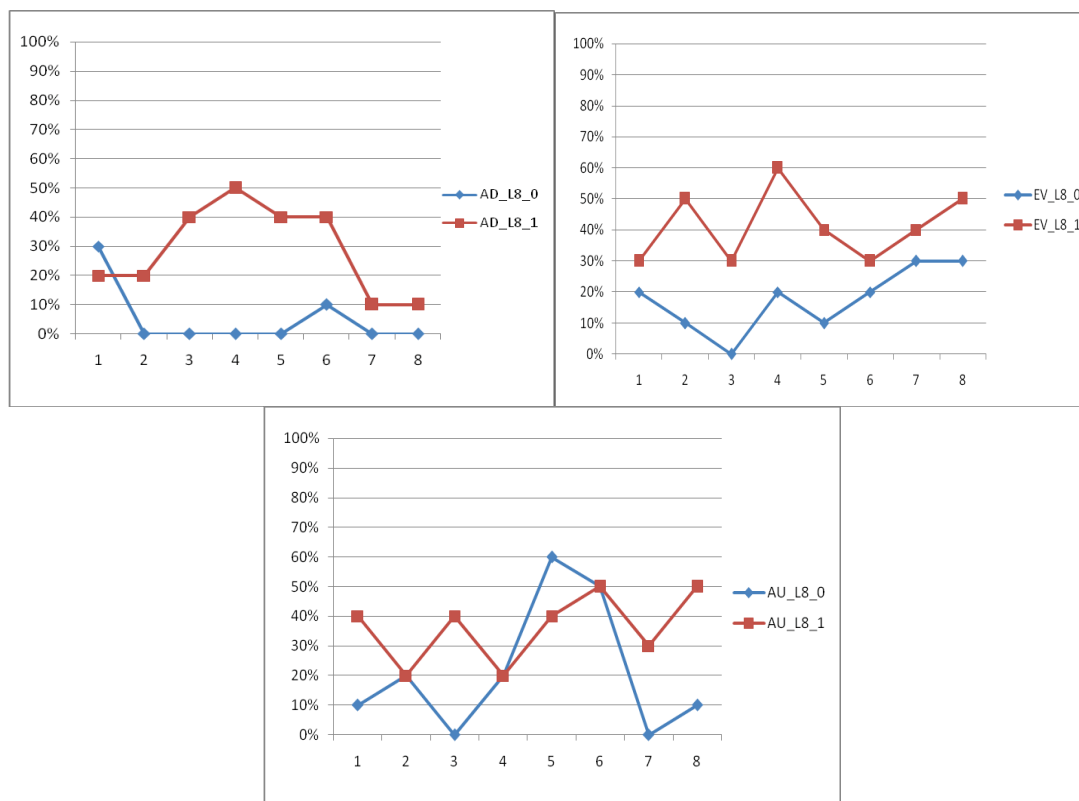


Figure 46 : Trois attitudes pour lesquelles nous n'avons pas pu identifier l'effet du gate : admiration [AD] (en haut à gauche), évidence [EV] (en haut à droite) et autorité [AU] (en bas) pour les phrases de 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais.

En ce qui concerne les expressions culturelles de la politesse (*kyoshuku* et *sincérité-politesse*), le taux de reconnaissance varie au alentour de 10% - 30% pour l'attitude *kyoshuku*, et 10% à 20% pour *sincérité-politesse* sauf trois dernières positions du gate (cf. figure 47). Il est à noter que le taux de reconnaissance pour l'attitude de *kyoshuku* augmente en position de septième gate. Quant à *sincérité-politesse*, le taux d'identification reste faible jusqu'au dernier gate, puis ils reconnaissent mieux en dernière position du gate.

Cependant, nous ne pouvons pas savoir si les auditeurs ont identifié globalement la notion de politesse à partir de ces graphes car ces graphes ne prennent pas en compte la confusion des attitudes perçues. La figure 48 montre la perception globale de la notion de politesse pour deux expressions de politesse. Selon ces graphes, la reconnaissance de

la notion de politesse pour l'expression de *kyoshuku* augmente progressivement du début jusqu'au dernier gate contrairement aux résultats des phrases de 2 et de 5 mores dans lesquelles les auditeurs n'arrivaient pas à identifier cette attitude. Lorsqu'ils écoutent jusqu'au dernier gate (c'est-à-dire la phrase entière) de la phrase de 8 mores, environ 60 % des auditeurs perçoivent *kyoshuku* comme l'expression de politesse. En ce qui concerne *sincérité-politesse*, cette attitude a été bien perçue comme expression de politesse dès la 2^{ème} position du gate avec le taux de 60 %, puis ce taux d'identification croît vers le dernier gate. Ce résultat nous indique que même si les auditeurs reconnaissent mal *sincérité-politesse*, cette expression est perçue tout de suite comme une expression de politesse au début de l'énoncé.

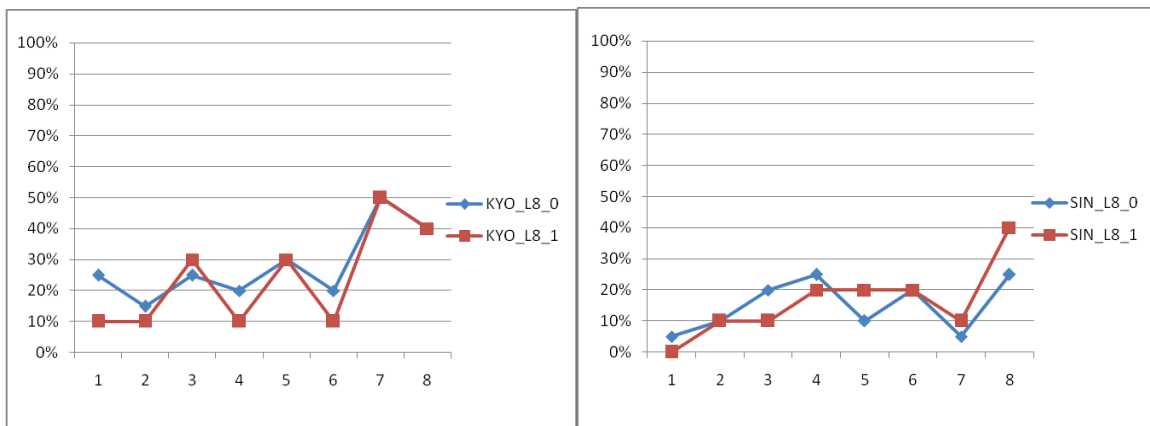


Figure 47 : La reconnaissance pour les expressions culturelles de la politesse pour les phrases de 8 mores. Le graphe à gauche pour le stimulus de *kyoshuku*, et *sincérité-politesse* à droite. La ligne bleue indique la phrase de 8 mores sans accent lexical, et celle en rouge indique la phrase avec un accent sur la 1^{ère} more. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais.

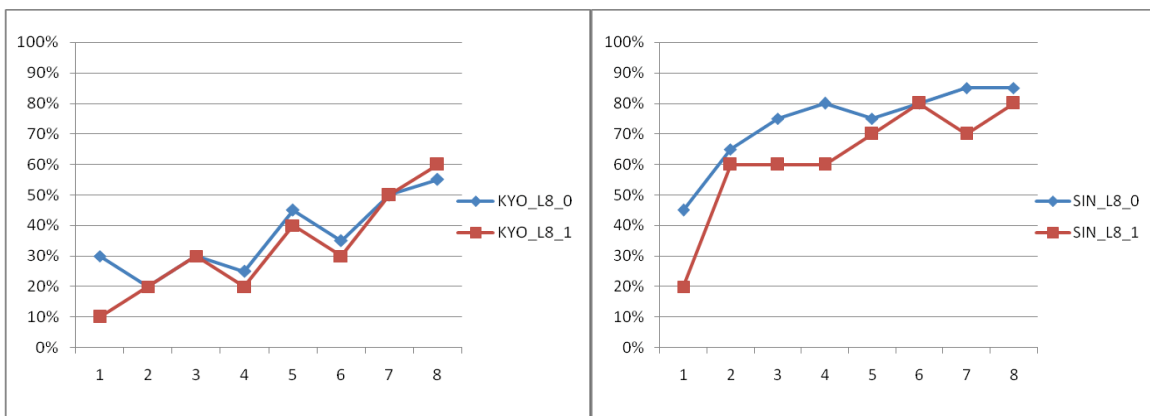


Figure 48 : la reconnaissance globale de la politesse pour les phrases de 8 mores. Le graphe à gauche pour le stimulus de *kyoshuku*, et *sincérité-politesse* à droite. La ligne bleue indique la phrase de 8 mores sans accent lexical, et celle en rouge indique la phrase avec un accent sur la 1^{ère} more. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées montre le taux de reconnaissance moyen des auditeurs japonais.

5.6.3.4. Conclusion

Dans le test perceptif précédent avec les phrases de 2 et de 5 mores, nous avons observé la confusion importante entre l'attitude de *kyoshuku* et *irritation* ou *autorité*. Nous constatons que c'est dû à l'influence du contenu sémantique qui a donné la difficulté aux auditeurs de soumettre une telle attitude dans le contexte donné. C'est pourquoi dans ce second test perceptif, nous avons choisi la phrase complète (GN + GV) qui convient davantage au contexte dans lequel ce type de politesse est pratiqué. Pour ce test perceptif dans le paradigme du gate, 10 Japonais natifs ont participé. Deux phrases composées de 8 mores : celle avec un accent lexical sur la 1^{ère} more et celle qui n'a pas d'accent lexical, sont sélectionnées à partir du corpus validé *supra*. Les phrases sont découpées en chaque position du gate approprié en y ajoutant du bruit blanc.

L'ANOVA (trois facteurs : accent, attitude, gate) sur les résultats montre que l'effet d'accent lexical sur les réponses n'est pas observé. En revanche, l'influence significative du gate a été identifiée. Lorsqu'on observe le taux d'identification de toutes les attitudes confondues, nous voyons la croissance linéaire du taux global de reconnaissance avec l'évolution de la position du gate. Ce résultat nous confirme l'effet du gate pour l'identification des attitudes, autrement dit, plus les auditeurs écoutent le reste d'énoncés, plus ils reconnaissent l'attitude exprimée.

Pourtant, le « simple main effect analysis » indique que l'effet du gate est identifié uniquement sur sept attitudes suivantes : *déclaration*, *question-simple*, *exclamation de surprise*, *doute-incrédulité*, *irritation*, *politesse-simple* et *arrogance-impolitesse*. Parmi ces attitudes qui ont subi l'influence significative du gate, l'attitude du *doute-incrédulité*, *irritation*, *politesse-simple* et *arrogance-impolitesse* montrent la croissance linéaire du taux d'identification avec l'évolution de position du gate.

Parmi trois autres attitudes influencées par la position du gate, *question-simple* est reconnu uniquement en dernière position du gate. Ce résultat avec les phrases de 8 mores confirme le même phénomène avec les phrases de 2 et de 5 mores présenté *supra*. La *déclaration*, quant à elle, est de moins en moins bien reconnue pour les deux phrases testées. Une explication possible pour ce phénomène est que les auditeurs ont tendance à choisir *déclaration* lorsqu'ils ne savent pas identifier d'autres attitudes surtout en position du 1^{ère} gate. Pour l'attitude d'exclamation de surprise, nous avons identifié une différence significative du taux de reconnaissance entre deux phrases testées. Cette attitude exprimée par la phrase sans accent lexical est mieux identifiée dès la position du second gate (60%) tant dit que la même attitude exprimée par la phrase avec un accent lexical en première more atteint le même taux d'identification seulement en 7^{ème} position du gate.

En ce qui concerne les attitudes qui n'ont pas eu d'influence du gate, sans doute, c'est admiration qui montre le plus fort contraste du taux de reconnaissance entre deux phrases testées. Parmi ces deux phrases, *admiration* exprimée avec la phrase de 8 mores sans accent lexical n'est pas identifiée correctement (0%). Cette attitude est reconnue

plutôt comme expression de politesse (*politesse-simple*, *kyoshuku* et *sincérité-politesse*) chez les auditeurs. Pour l'attitude d'autorité, nous observons un différent degré du taux d'identification entre les deux phrases testées en dernière position (7^{ème} et 8^{ème} position) du gate. Le taux d'identification de cette attitude exprimée en phrase sans accent lexical chute en position du 7^{ème} gate (50% à 0% d'identification) en montrant des confusions avec *irritation* (30%) ou *arrogance-impolitesse* (30%) en dernière position du gate. Cela veut dire que les auditeurs perçoivent tout de même l'imposition de l'avis du locuteur (*autorité*, *arrogance-impolitesse* et *irritation*) en dernière position du gate des deux phrases.

Deux expressions culturelles de la politesse (*kyoshuku* et *sincérité-politesse*) qui n'ont pas d'effet du gate, montrent un taux de reconnaissance faible sauf aux trois dernières positions du gate. Cependant, la reconnaissance de la notion de politesse pour l'expression de *kyoshuku* augmente progressivement du début vers la fin de phrase, en registrant le taux d'identification de 60% à la dernière position du gate. Quant à l'attitude de sincérité-politesse, elle est perçue immédiatement comme l'expression de politesse dès la 2^{ème} position du gate avec le score de 60% d'identification. Ce résultat confirme la prédiction rapide de la notion de politesse pour l'expression de *sincérité-politesse* que nous avons identifié *supra*.

5.6.4. Troisième test de gating

5.6.4.1. Introduction

Dans les deux tests perceptifs, l'attitude de *kyoshuku* et admiration (cf. *supra*) ont été mal reconnues par les auditeurs natifs. Nous pensons que la faible identification de ces attitudes est due au biais de l'influence du contenu lexical des phrases utilisées. C'est pourquoi nous avons effectué un 3^{ème} test de perception dans le cadre du gating avec des phrases ayant un contenu lexical qui s'adapte mieux au contexte de la vie quotidienne.

5.6.4.2. Méthodes

a- Sélection de 12 attitudes japonaises

Un ensemble de 12 attitudes japonaises validées *supra* a été utilisé pour le test perceptif.

b- Corpus

Dans cette expérimentation, nous avons sélectionné des phrases qui consistent en trois phrases de 2, 5 et 8 mores (cf. tableau 16). La structure syntaxique des phrases est soit une phrase soit une phrase elliptique réduite à un mot simple, soit une structure simple de type « verbe-objet ».

En ce qui concerne la position d'accent lexical, nous n'avons identifié aucune influence significative dans le test perceptif précédant (cf. *supra*). C'est pourquoi nous

n'avons pas examiné l'influence d'accent lexical sur la réponse des sujets. Pour toutes les phrases, la position de l'accent lexical est alignée sur la première more (Hirata, 1960).

Afin d'exprimer certaines attitudes, la cinquième voyelle [i] de la phrase de 8 mores qui est normalement dévoisée en déclaration peut être voisée. En revanche, la septième voyelle [i] est dévoisée pour toutes les attitudes exprimées.

Les phrases ont été construites afin de n'avoir aucune connotation sémantique particulière dans quelle que soit la région du Japon. Cependant ces phrases nous semblent être utilisées assez fréquemment dans un contexte dans lequel les 12 attitudes sont exprimées. Comme le corpus précédant, chaque phrase est produite avec les douze expressions attitudinales par un enseignant natif de la langue japonaise qui a aussi produit le corpus précédant.

Tableau 16 : *Phrases sélectionnées pour le 3^{ème} test de dévoilement progressif.*

nb de more	Phrase	Traduction
2	Na*ra	Nara
5	Na*ra no machi	La ville de Nara
8	Na*ra e ikimashita	Il est allé à Nara

Un ensemble de 36 énoncés (3 phrases x 12 attitudes x 1 locuteur) est enregistré en numérique à 44,1 kHz, 16bits dans une chambre sourde au LIMSI, à Paris. Le locuteur est debout devant le micro AKG C414B qui est placé à 40 cm de distance par rapport à la bouche du locuteur. Ce micro est relié avec un « USBPre sound device » vers un ordinateur qui se situe en dehors de la chambre sourde.

Ensuite, les 36 énoncés enregistrés ont été découpés à la fin de chaque more ; la fin de l'énoncé, après chaque gate, est remplacé par du bruit blanc. La durée de ce bruit a été fixée de façon à obtenir des stimuli d'une durée constante égale à la durée de la phrase entière de 8 mores.

Ainsi un ensemble de 180 stimuli (15 gates x 3 phrases x 12 attitudes) ont été utilisés pour le test perceptif suivant.

c- Protocole expérimental

Afin de déterminer à quel moment les auditeurs japonais reconnaissent chaque attitude exprimée, les stimuli sont présentés aux auditeurs à partir en commençant par le premier gate jusqu'au dernier gate (c-à-d, écouter la phrase entière).

Le test perceptif est effectué auditivement avec des casques Victor ML-H200 dans une salle calme. Tout le test est effectué sur l'interface développée sous le logiciel « Runtime revolution » avec un ordinateur PC Windows XP professionnel. L'interface du test affiche 12 choix des attitudes avec leurs définitions. Avant le démarrage du test, nous avons distribué une feuille avec la définition pour chaque attitude et des exemples de contextes d'occurrences possibles des attitudes, afin d'éviter de possibles ambiguïtés

de compréhension. Ainsi aucun sujet n'a affirmé avoir éprouvé des difficultés pour comprendre les concepts référés par les étiquettes des attitudes.

Tous les sujets ayant participé à cette expérimentation ont écouté chaque stimulus une fois seulement. Pour chaque stimulus, nous leur avons demandé de choisir l'attitude perçue parmi les 12 possibles. L'ordre de présentation des stimuli était aléatoire et différent pour chaque sujet.

Le test de perception est effectué auprès de 33 Japonais natifs (24 femmes et 9 hommes) parlant le dialecte du Tokyo. L'âge moyen des sujets est de 19 ans. Aucun auditeur n'a fait état de trouble d'audition.

5.6.4.3. Résultats

a- L'effet du gate

Tout d'abord, l'effet du gate sur les résultats de réponses est évalué par l'ANOVA. Les résultats montrent que l'influence significative du gate a été observé sur trois différentes phrases ($F=29.333$ $p<.01$ pour la phrase de 2 mores, $F=22.196$ $p<.01$ pour la phrase de 5 mores et $F=12.201$ $p<.01$ pour la phrase de 8 mores). L'effet significatif du choix des attitudes est également observé pour les trois phrases ($F=6.9283$ $p<.01$ pour la phrase de 2 mores, $F=18.751$ $p<.01$ pour la phrase de 5 mores, $F=24.754$ $p<.01$ pour la phrase de 8 mores). Par contre, l'effet de la longueur de la phrase sur la perception des attitudes n'est pas confirmé ($F=4.4$ $p>.01$). Cela veut dire, comme illustré dans la figure 49, que lorsque les auditeurs écoutent la différente longueur des phrases, leur comportement perceptif dépend des attitudes plutôt que la longueur de la phrase.

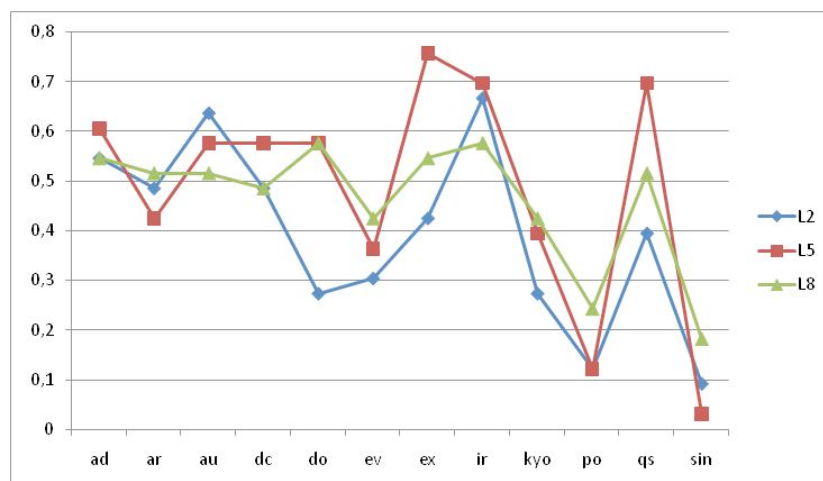


Figure 49 : taux de reconnaissance des attitudes pour trois longueurs différentes (la phrase de 2, 5 et 8 mores respectivement). L'axe des abscisses indique les attitudes. L'axe des ordonnées indique le taux d'identification moyenne de 33 auditeurs japonais.

Selon la figure 50, le graphe de chaque longueur de phrase illustre clairement une corrélation positive entre l'évolution de la position de gate et la croissance du taux global de reconnaissance pour toutes les attitudes confondues. Ce résultat nous suggère que

plus les auditeurs écoutent une partie importante de l'énoncé, mieux ils reconnaissent l'attitude exprimée.

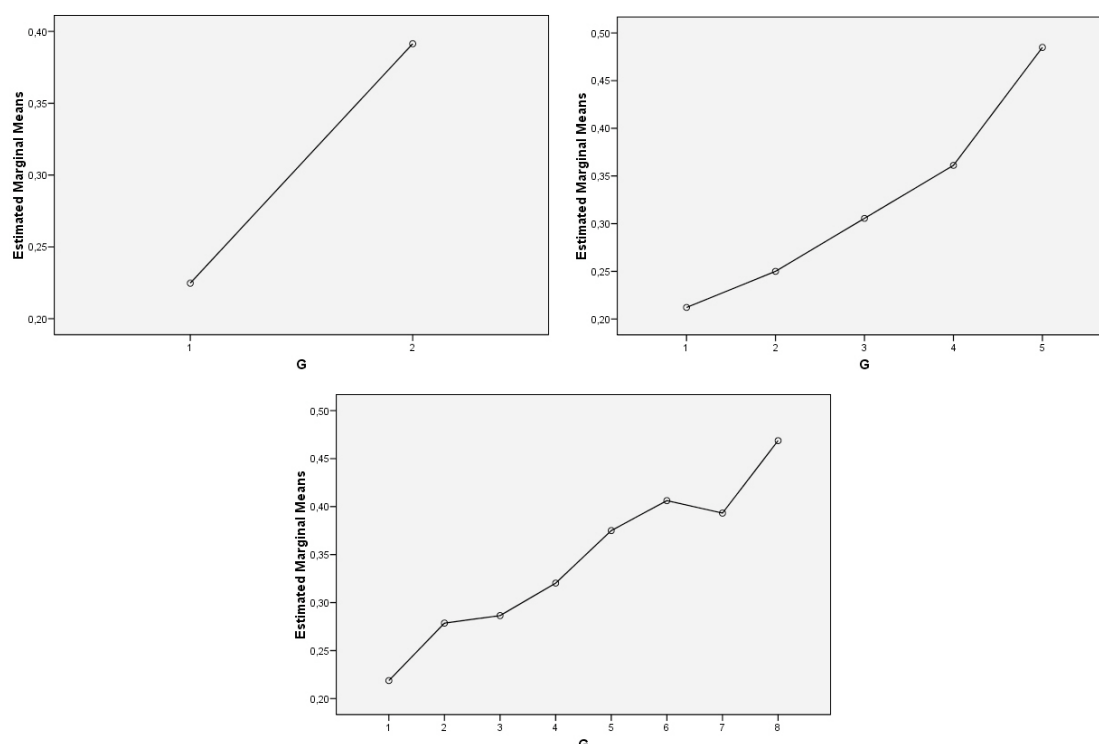


Figure 50 : croissance linéaire de la reconnaissance globale pour toutes les attitudes confondues sur les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais.

Pourtant, cette augmentation du taux d'identification avec l'évolution du gate n'est pas justifiée pour toutes les attitudes. La figure 51 représente l'évolution globalement linéaire de la reconnaissance pour cinq attitudes. Selon ces graphes, l'attitude d'*admiration*, *arrogance-impolitesse*, *déclaration*, *doute-incrédulité* et *évidence* montrent une croissance linéaire du taux d'identification, mais ce phénomène n'est pas confirmé pour les autres cinq attitudes.

Par exemple, le taux d'identification pour l'attitude d'*autorité* et d'*exclamation de surprise* marque environ 50 ou 60 % dès le premier gate contrairement aux autres attitudes (cf. figure 52). Ce résultat indique que les auditeurs reconnaissent probablement ces deux attitudes très tôt (dès le début) de l'énoncé.

En revanche, le taux de reconnaissance pour *question-simple* reste toujours très faible jusqu'à l'avant-dernier gate, puis le taux d'identification monte brutalement juste au dernier gate (cf. figure 53). Ce phénomène nous permet de conclure que la perception du dernier gate contient les informations déterminantes pour reconnaître l'interrogation japonaise. De plus, la perception pour *irritation* nous semble de montrer le pattern similaire que *question-simple* sauf pour la phrase du 8 mores dans laquelle les

informations déterminantes ne sont pas localisées sur le dernier gate, mais plutôt au alentours du sixième gate où le taux de reconnaissance accroît du 30% à 65%.

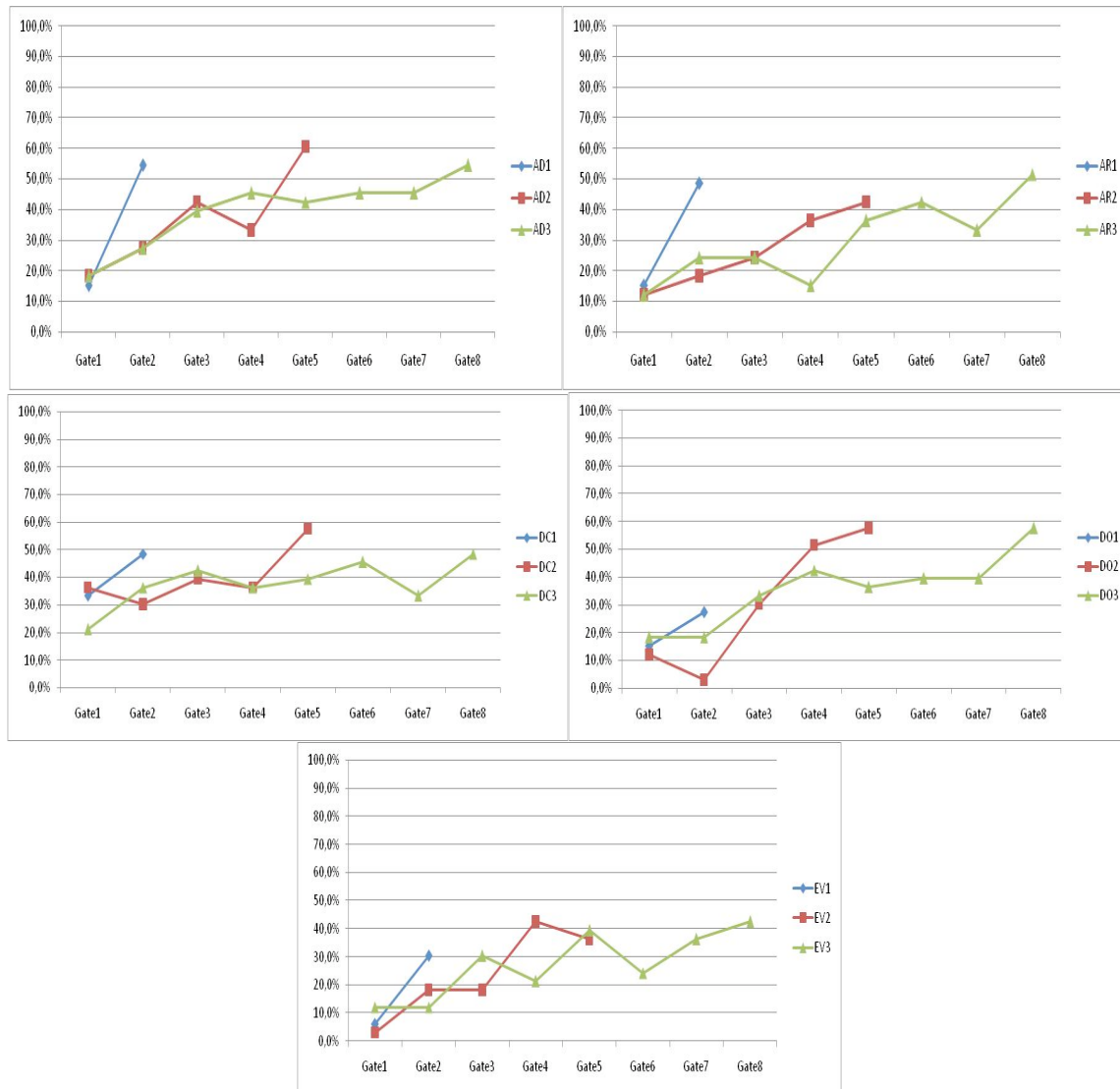


Figure 51 : croissance linéaire de la reconnaissance de cinq attitudes (admiration, arrogance-impolitesse, déclaration, doute-incrédulité et évidence) pour les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais.

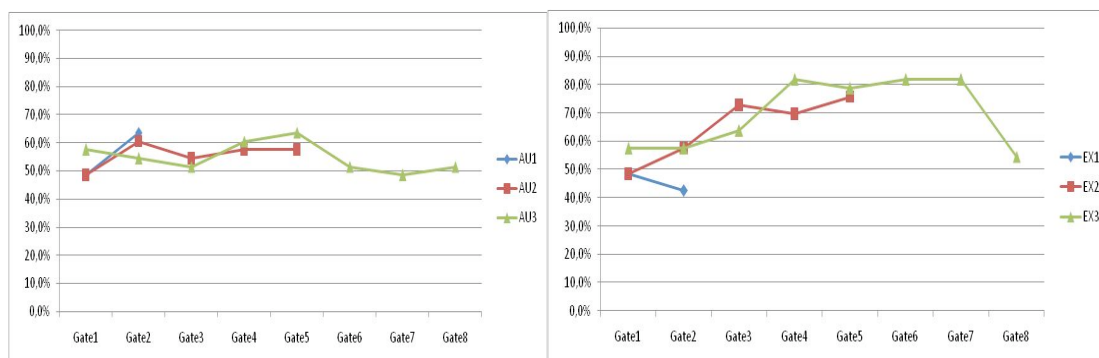


Figure 52 : attitudes perçues dès le premier gate : autorité et exclamation de surprise, pour les phrases de 2, 5 et 8 mores. L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est le taux de reconnaissance moyenne de 33 auditeurs japonais.

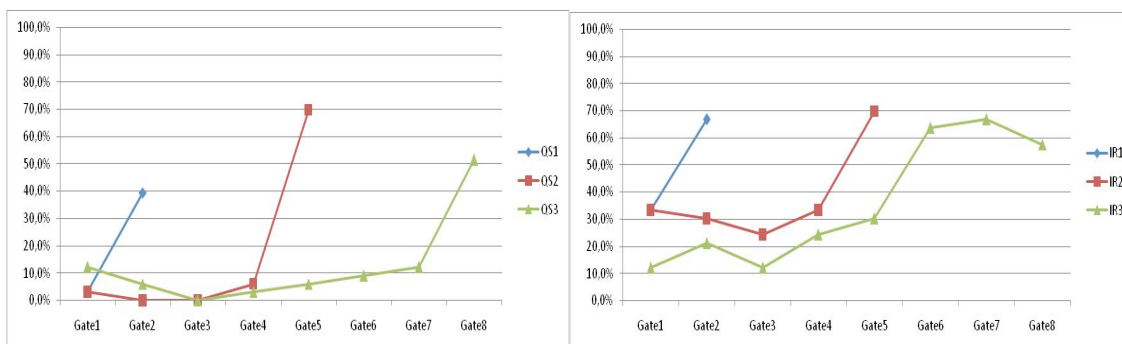


Figure 53 : attitudes reconnues à un gate précis (le dernier, sauf l'irritation de 8 mores): interrogation et irritation, phrases de 2, 5 et 8 mores. Les abscisses indiquent le gate et les ordonnées le taux de reconnaissance moyen des 33 auditeurs japonais.

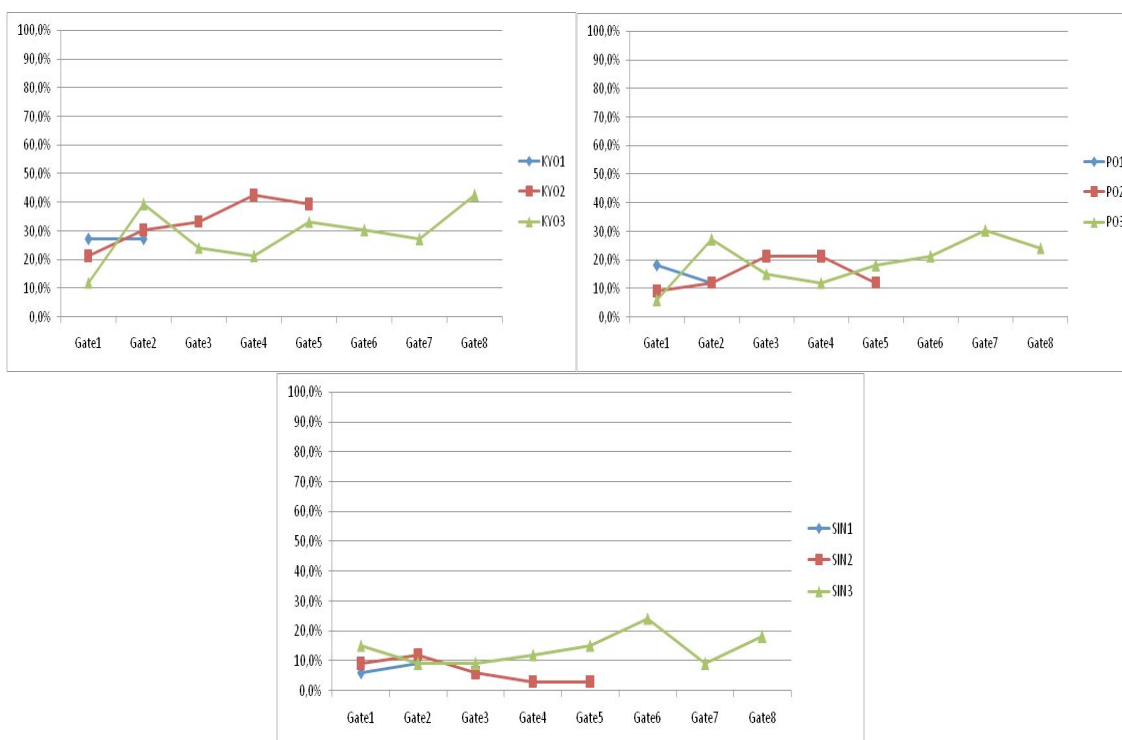


Figure 54 : attitudes de politesse pour les phrases de 2, 5 et 8 mores. En haut à gauche : kyoshuku ; en haut à droite : politesse-simple ; en bas : sincérité-politesse. Les abscisses indiquent le gate et les ordonnées le taux de reconnaissance moyen.

En ce qui concerne les expressions de la politesse (*kyoshuku*, *politesse-simple* et *sincérité-politesse*), le taux de reconnaissance reste presque plat depuis le premier gate jusqu'au dernier gate en montrant une faible reconnaissance (cf. figure 54). Mais nous ne pouvons pas savoir si les auditeurs ont identifié globalement la notion de politesse à partir de ces graphes car ces graphes ne prennent pas en compte la confusion des attitudes perçues. La figure 55 montre la perception globale de la notion de politesse pour ces trois expressions de politesse. Selon ces graphes, la reconnaissance de la notion de politesse pour l'expression de *kyoshuku* augmente progressivement du début jusqu'au dernier gate. Surtout lorsqu'ils écoutent jusqu'au dernier gate (c'est-à-dire la phrase entière) de la phrase de 8 mores, 90 % des auditeurs perçoivent *kyoshuku* comme l'expression de politesse. En ce qui concerne *sincérité-politesse*, cette attitude a été bien perçue comme expression de politesse dès le premier gate avec le taux de 60 % pour la phrase de 8 mores, puis ce taux d'identification croît vers le dernier gate. Ce résultat nous indique que même si les auditeurs reconnaissent mal *sincérité-politesse*, cette expression est perçue immédiatement comme l'expression de politesse dès le début de l'énoncé.

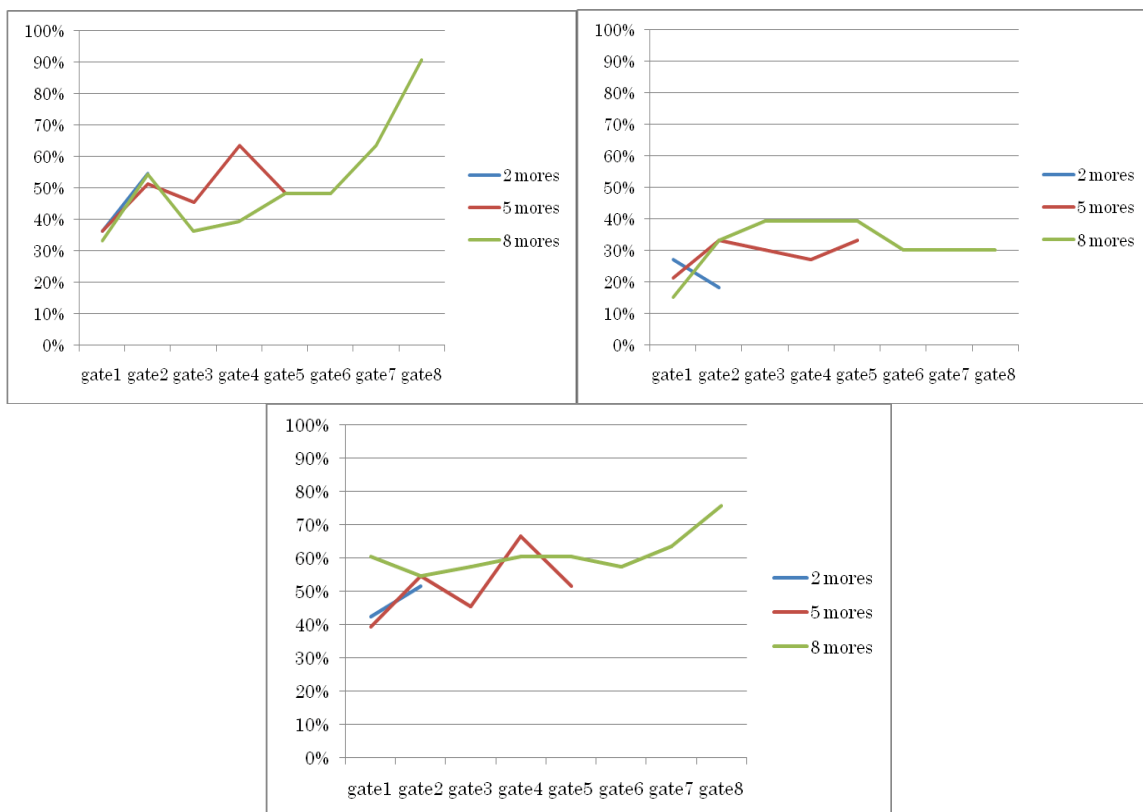


Figure 55 : reconnaissance globale de la politesse pour les phrases de 2, 5 et 8 mores. Le graphe en haut à gauche pour le stimulus de *kyoshuku*, *politesse-simple* en haut à droite et la graphe pour le stimulus de *sincérité-politesse* est en bas). L'axe des abscisses indique l'évolution du gate, l'axe des ordonnées est un taux de reconnaissance moyenne de 33 auditeurs japonais.

Quant à la *politesse-simple*, le taux du degré global de politesse reste aux alentours de 30 à 40%, sans montrer d'identification progressive (cf. figure 55 en haut à droite) comme le graphe de l'identification de cette attitude (cf. figure 54 en haut à droite). La matrice de confusion pour cette attitude montre une confusion importante pour cette attitude avec d'autres expressions de politesse contrairement aux deux autres types de politesse que nous avons mentionnés ci-dessus (cf. tableau 17). Cette attitude a majoritairement confondu avec la simple déclaration. Nous allons confirmer ce point dans le paragraphe suivant.

Tableau 17 : résultats des 33 auditeurs japonais pour chaque gate, pour la politesse exprimée avec la phrase de 8 mores. Les chiffres en gras italique indiquent le taux de reconnaissance, et les chiffres en gras désignent un degré de confusions important.

<i>Phrases de 8 mores</i>	<i>Nombre de gate présentés aux sujets</i>							
	<i>1/8</i>	<i>2/8</i>	<i>3/8</i>	<i>4/8</i>	<i>5/8</i>	<i>6/8</i>	<i>7/8</i>	<i>8/8</i>
Admiration	3%	12%	6%	6%	6%	3%	6%	3%
Arrogance	3%	6%	3%	3%	3%	0%	3%	0%
Autorité	3%	3%	0%	0%	0%	3%	0%	3%
Déclaration	27%	27%	24%	39%	33%	54%	48%	54%
Doute-Incrédulité	6%	9%	6%	0%	0%	3%	3%	0%
Evidence	9%	6%	9%	0%	12%	3%	6%	9%
Surprise	3%	0%	3%	3%	0%	0%	3%	0%
Irritation	15%	3%	6%	6%	0%	0%	0%	0%
Kyoshuku	6%	3%	6%	6%	6%	0%	0,00%	0%
Politesse	6%	27%	15%	12%	18%	21%	30%	24%
Question	15%	0%	3%	3%	6%	3%	0%	0%
Sincérité-Politesse	3%	3%	18%	21%	15%	9%	0%	6%

b- Reconnaissance des attitudes par l'analyse de correspondance.

Dans ce paragraphe, nous voulons observer le degré d'identification de chaque attitude lorsque les auditeurs écoutent la phrase entière, en plus de l'étude de la distribution des attitudes et des principales confusions inter-attitudes.

Afin de confirmer un degré d'identification des attitudes observées dans le paragraphe précédent, nous avons appliqué cet outil statistique aux résultats des énoncés du dernier gate avec tous les longueurs de phrase confondues (c-à-d, aux phrases entières qui contiennent toutes les informations prosodiques). Cette analyse nous permet de présenter la distance entre des objets (ou comportement humain) sous forme des points calculés en espace de deux dimensions. Cela permet également d'observer la distance entre le point du stimulus et le point perceptif d'une attitude donnée.

Selon les résultats obtenus, plus de 80% de variance sont expliquées dans les quatre premières dimensions. La figure 56 (en haut) représente la distribution des points perceptifs et points stimuli pour les deux premières dimensions. Toutes les attitudes sauf *politesse-simple* ont été très bien identifiées, et elles sont indiquées sur le graphe en cercles bleus. En ce qui concerne ces attitudes, les points stimuli se situent très proche de points perceptifs. Cela explique que les auditeurs japonais ont reconnu ces attitudes correctement. Quant à *politesse-simple*, le point du stimulus pour cette attitude est proche du point perceptif de *déclaration* plutôt qu'être près des autres expressions de la politesse. Par contre le point perceptif de *politesse-simple* est près des autres expressions de politesse plutôt que celui de la *déclaration*.

Quant aux attitudes ayant des confusions avec d'autres attitudes, nous voyons que les attitudes confondues se situent près de ces attitudes, et un ensemble de ces attitudes se forment des catégories des attitudes, et elles sont dans les cercles remplis par la couleur violette. Selon la figure 56, quatre catégories sont globalement observées par l'analyse de correspondance.

Les expressions de la politesse (*kyoshuku*, *sincérité-politesse* et *politesse-simple*) et *admiration* se situent en bas à gauche du graphe, et elles se forment une catégorie perceptive. Il est à noter que le phénomène de l'interconnexion entre la politesse japonaise et l'admiration a été déjà observé *supra*, et cette interconnexion est justifiée par un entré lexical du japonais qui comporte le sens de la politesse et de l'admiration (par exemple, « *Keifuku* » est un terme employé au sens de l'admiration et de la politesse).

Les attitudes de *doute-incrédulité*, *exclamation de surprise* et *question-simple* sont localisées en haut à gauche du graphe. Toutes ces attitudes de cette catégorie réclament la « réponse » de la part du locuteur.

Les attitudes d'*autorité* et d'*irritation* sont placées au milieu à droite du graphe. Cette catégorie d'attitudes impose l'opinion du locuteur aux autres.

Les attitudes d'*évidence* et d'*arrogance-impolitesse* se situent au milieu du graphe. Nous avons déjà vu dans le passage de validation (*supra*) que *évidence* a une tendance d'être interprétée comme *arrogance-impolitesse* car *évidence* montre que le locuteur est sûr de lui-même et cette expression de la certitude peut parfois être perçue comme irrespectueux vis-à-vis de l'interlocuteur.

La figure 56 (en bas) représente la distribution des points perceptifs et points stimuli en 3^{ème} et 4^{ème} dimensions. Selon le graphe présenté, *exclamation de surprise* qui est située en haut, et *admiration* localisée en bas à gauche du graphe sont bien distinguées par rapport aux autres attitudes qui sont concentrées à droite du graphe. Cela signifie que les auditeurs ont très bien discriminé ces deux attitudes par rapport aux autres.

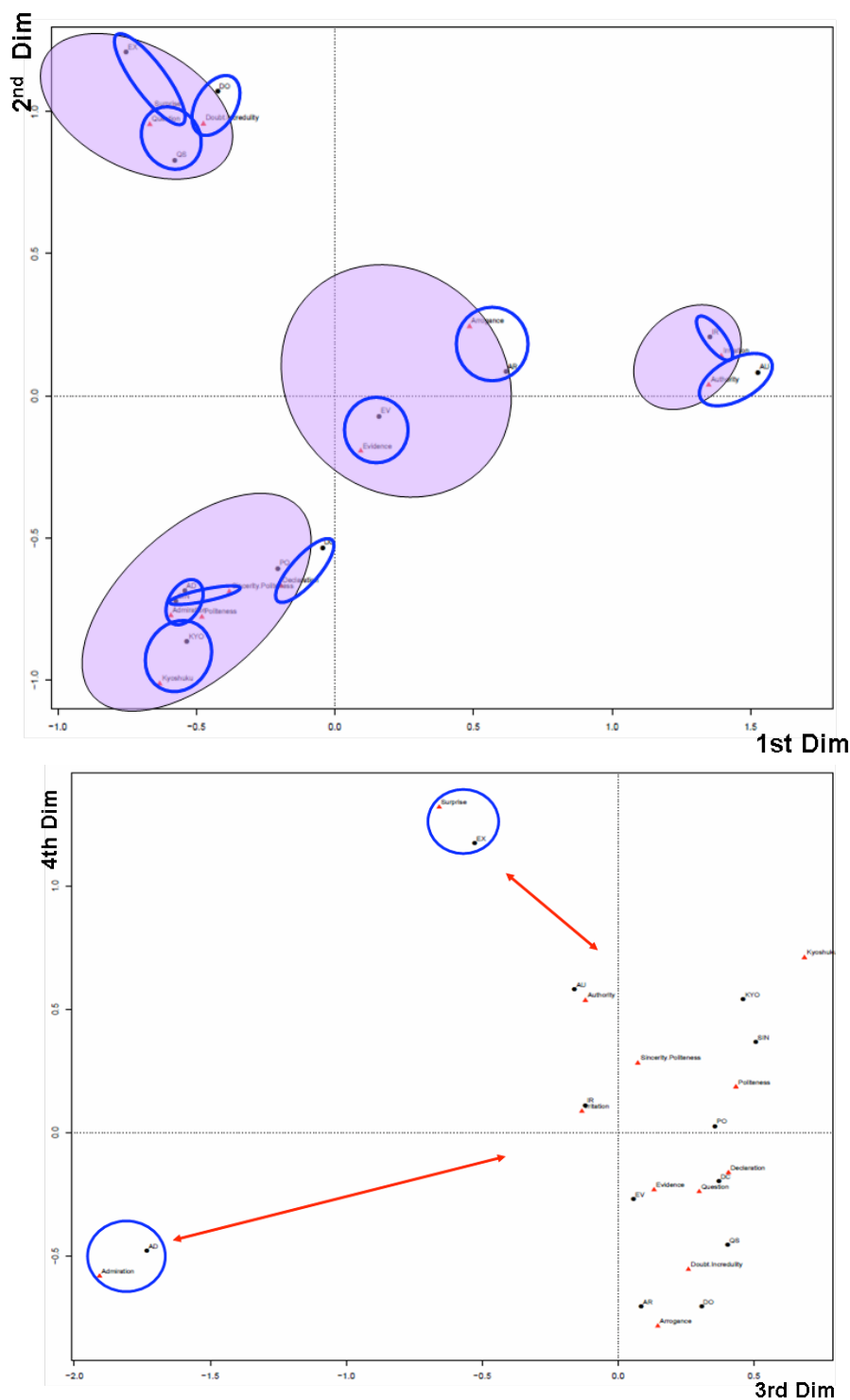


Figure 56 : distribution des points perceptifs et des points stimuli en deux dimensions calculées par l'analyse de correspondance. Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles.

5.6.4.4. Paramètres prosodiques et analyse

Dans ce paragraphe, nous allons analyser le lien entre les caractéristiques acoustiques et les catégories perceptives de chaque position du gate afin de déterminer une (ou des) partie(s) clef(s) pour la perception de chacune des attitudes exprimées.

Pour cela, nous avons extrait les valeurs acoustiques telles que la durée, l'intensité et la fréquence fondamentale de chaque more des phrases utilisées lors du test perceptif d'une part, et d'autre part, la description auditive de la qualité de voix telle que la voix modale, soufflée, rauque, tendue, nasale et dévoisée, pour chaque voyelle de phrase est aussi intégrée dans cette analyse.

5.6.4.5. Extraction des paramètres prosodiques

Les paramètres acoustiques (la pente de F0, la pente en dernière voyelle de F0, l'écart type de F0, la moyenne de F0, allongement de la dernière more, l'écart type de la durée, la durée moyenne, la pente d'intensité, l'écart type de l'intensité et la moyenne d'intensité) de la prosodie sont extraits automatiquement par la méthode présentée *supra*. Toutes ces valeurs extraites sont utilisées comme les entrées pour l'analyse en composantes principales (ACP) et pour la classification hiérarchique.

5.6.4.6. ACP et classification hiérarchique.

En utilisant les paramètres objectifs décrits dans le paragraphe précédent comme les caractéristiques, nous les avons analysé par l'ACP et la classification hiérarchique afin d'identifier quel facteur (ou composant) peut véhiculer un tel comportement des auditeurs.

Les résultats de l'ACP représentent l'identification des paramètres corrélés avec chacune des attitudes exprimées. Ensuite, la classification hiérarchique permet de regrouper les attitudes en catégories partageant un ensemble de caractéristiques similaires.

5.6.4.7. ACP sur les paramètres prosodiques

L'ACP basée sur les paramètres prosodiques montre les caractéristiques distinctives suivantes. Le niveau élevé de l'intensité moyenne, l'écart type de l'intensité fort et l'écart type de F0 fort caractérisent *évidence* et *autorité*, et *irritation* à un moindre degré. L'attitude de *kyoshuku* est exprimée par la grande pente d'intensité et également par la grande pente de F0 sur la dernière voyelle. *Question-simple*, *doute-incrédulité* et *sincérité-politesse* aussi montrent une pente d'intensité assez importante. L'attitude d'*admiration* est caractérisée par la grande pente de F0, la grande pente en dernière voyelle de F0, la durée moyenne plus longue, la forte variabilité de durée par l'indice de l'écart type, l'allongement de la dernière more plus long et F0 moyenne plus élevée. L'attitude d'*irritation* porte les mêmes caractéristiques prosodiques qu'*admiration*, mais à un degré beaucoup moins fort.

Tableau 18 : L'importance des composants sur chaque dimension pour les résultats de l'ACP pour les paramètres acoustiques. Les deux premières dimensions expliquent 73% de la variance.

Importance of components:

Dimension	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2,03	1,79	1,05	0,87	0,71	0,41	0,28	0,19	0,08	0,02
Proportion of Variance	41%	32%	11%	8%	5%	2%	1%	0%	0%	0%
Cumulative Proportion	41%	73%	84%	92%	97%	99%	100%	100%	100%	100%

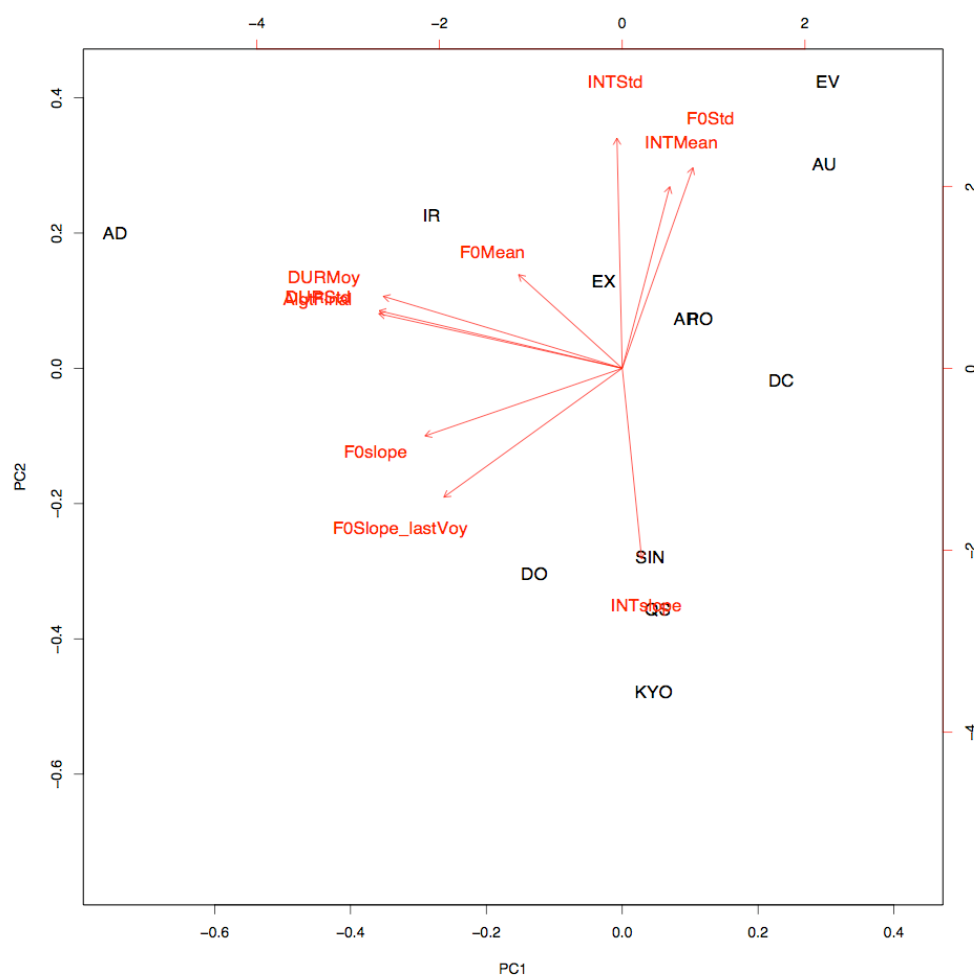


Figure 57 : Deux dimensions principales de l'ACP pour les paramètres prosodiques. Les étiquettes en rouge indiquent les paramètres pris en compte pour l'ACP. La distribution des attitudes qui sont abrégées et indiquée en noir. En ce qui concerne les vecteurs rouges, ils indiquent l'importance de paramètre. Paramètres pris en compte : la pente de F0 (F0slope), la pente en dernière voyelle de F0 (F0SlopeLastVoy), l'écart type de F0 (F0Std), la moyenne de F0 (F0Mean), allongement de la dernière more (DURAlgt), l'écart type de la durée (DURStd), la durée moyenne (DURMoy), la pente d'intensité (INTSlope), l'écart type de l'intensité (INTStd) et la moyenne d'intensité (INTMean).

5.6.4.8. Classification hiérarchique des paramètres prosodiques

Ensuite, nous avons classé les 12 attitudes en plusieurs groupes à partir de la similarité des valeurs de paramètres prosodiques par la classification hiérarchique. Selon la figure 57, l'*autorité* est distinct des autres attitudes par les caractéristiques acoustiques telles que l'intensité moyenne forte, la forte variation de F0 et de l'intensité. Le groupe des quatre attitudes : *doute-incrédulité*, *kyoshuku*, *question-simple* et *sincérité-politesse*, est caractérisé par l'intensité moyenne faible, la faible variation de F0 et de l'intensité d'une part, et d'autre part la grande pente de l'intensité, la grande pente de F0 et la grande pente de F0 en dernière more. Le groupe des attitudes : *exclamation de surprise*, *admiration* et *irritation*, porte une similaire caractéristique de l'écart type de l'intensité. Surtout *irritation* et *admiration* montrent des caractéristiques spécifiques en fin de phrase selon l'ACP (la grande pente en dernière voyelle de F0 et l'allongement de la dernière more plus long). En revanche, le groupe d'*évidence*, *politesse-simple*, *arrogance-impolitesse* et *déclaration*, est caractérisé par la faible pente de F0, la durée moyenne plus courte, l'écart type de durée plus faible et un faible allongement de la dernière more.

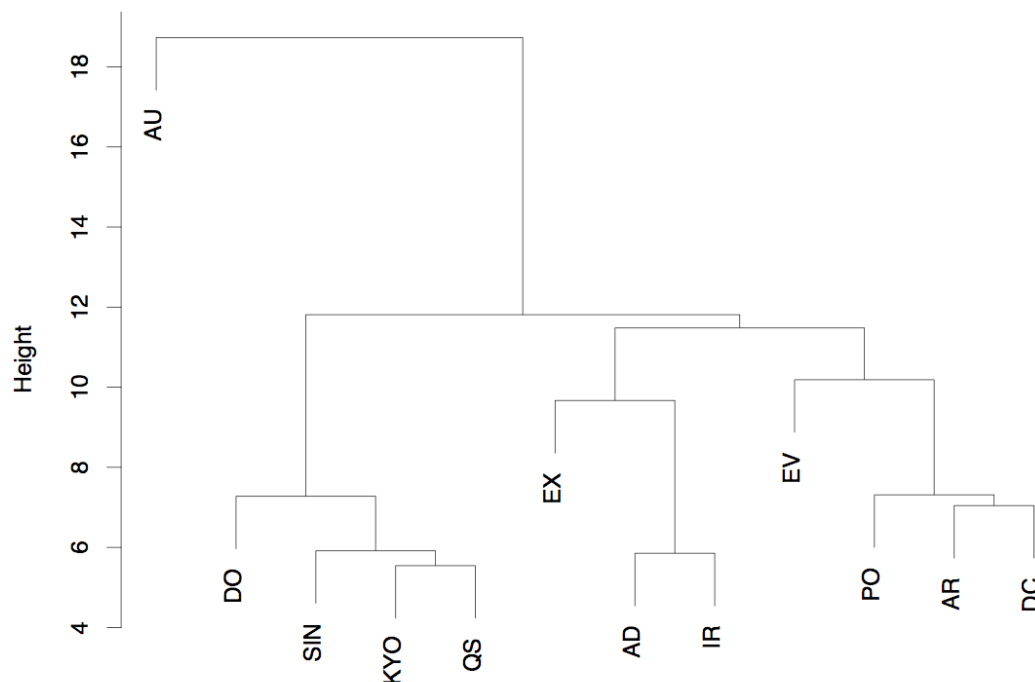


Figure 58 : Catégories des attitudes à partir de la similarité des valeurs de paramètres prosodiques selon la classification hiérarchique (AR).

5.6.4.9. ACP sur les attitudes et la qualité de voix

Les six paramètres de la qualité de voix : la voix modale, soufflée, rauque, tendue, nasale et le dévoisement pour toutes les voyelles ont été évalués par impression auditive par l'auteur, et toutes les valeurs sont normalisées afin de les traiter dans une ACP. Les résultats montrent que la voix soufflée est fortement corrélée avec *sincérité-politesse*,

admiration et *exclamation de surprise*. L'attitude de *kyoshuku* est caractérisée par la voix rauque et tendue. Elle est fréquemment dévoisée aussi. *Irritation* possède la même caractéristique, mais degré plus faible. Quant à *arrogance-impolitesse*, elle est plus nasalisée que les autres attitudes.

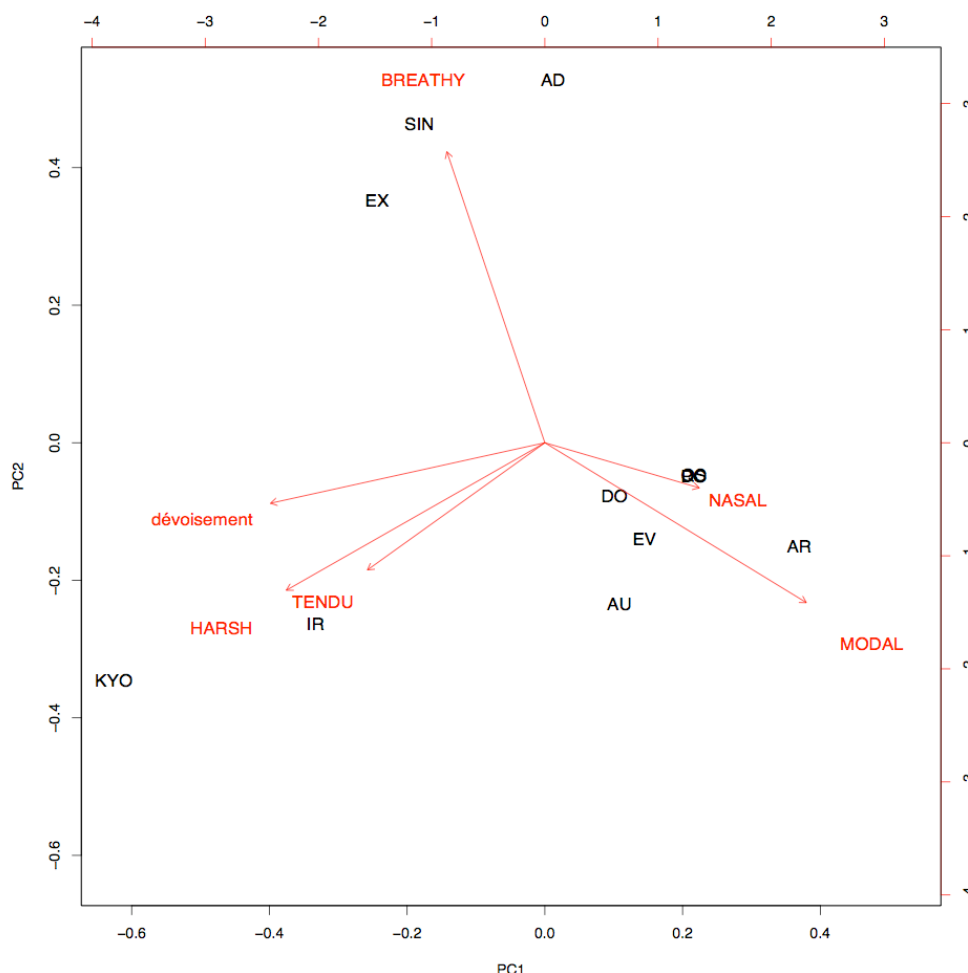


Figure 59 : deux dimensions principales de l'ACP pour les paramètres de qualité de voix. Les étiquettes rouges indiquent les paramètres de l'ACP. La distribution des attitudes qui sont abrégées et indiquée en noir. En ce qui concerne les vecteurs rouges, ils indiquent l'importance de paramètre. *Breathy* (= soufflée), *Harsh* (= rauque).

Tableau 19 : L'importance des composants sur chaque dimension pour les résultats de l'ACP pour les paramètres de la qualité de voix.

Importance of components:

Dimension	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1,67	1,25	0,95	0,79	0,34	0,05
Proportion of Variance	47%	26%	15%	10%	2%	0%
Cumulative Proportion	47%	73%	88%	98%	100%	100%

5.6.4.10. Classification hiérarchique de la qualité de voix

À la suite de l'ACP, la classification hiérarchique permet de regrouper globalement les attitudes en trois groupes d'expressions attitudinales différenciées par leur qualité de voix. Selon la figure 60, *exclamation de surprise*, *admiration* et *sincérité-politesse* se trouvent dans le même groupe qui partage une caractéristique de la voix soufflée. Le second groupe est composé d'*irritation* et de *kyoshuku* et est caractérisé par la voix rauque et tendue selon l'ACP. Les autres attitudes partagent la caractéristique de la voix modale, mais *autorité* est exprimée par la voix plus tendue que les autres attitudes du même groupe, et *arrogance-impolitesse* est exprimée par la voix plus nasale que les autres attitudes.

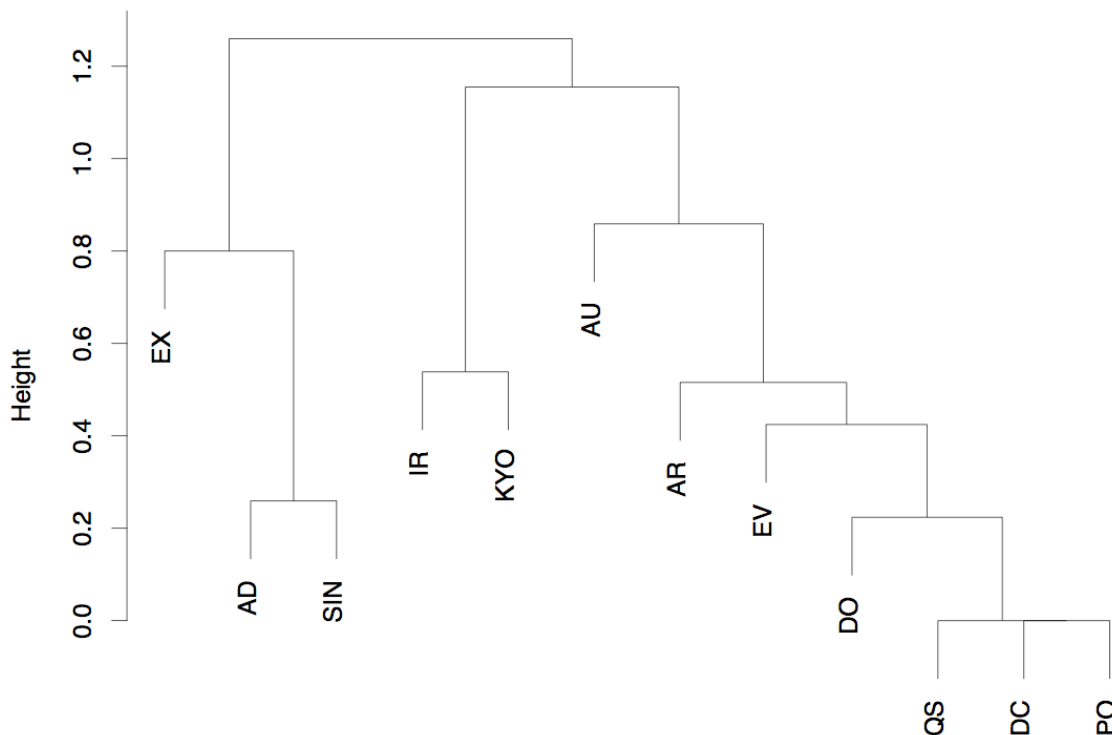


Figure 60 : Catégories des attitudes à partir de la similarité des valeurs de paramètres de la qualité de voix selon la classification hiérarchique.

5.6.4.11. Analyse de la prosodie et de la qualité de voix

Enfin nous avons observé des catégories d'attitudes qui partagent les mêmes caractéristiques prosodiques et de qualité de voix par la classification hiérarchique. L'analyse conjointe de ces résultats nous permet d'extraire certains groupes des expressions attitudinales. D'abord, deux expressions culturelles de politesse (*kyoshuku* et *sincérité-politesse*) se situent dans le même groupe. C'est probablement l'influence des paramètres prosodiques plutôt que le facteur de la qualité de voix qui est nettement distinguée entre les deux attitudes. Le groupe des attitudes : *exclamation de surprise*, *admiration* et *irritation*, porte une caractéristique similaire d'écart type de l'intensité.

Surtout *irritation* et *admiration* possèdent les caractéristiques similaires en fin de phrase. Une autre catégorie des attitudes : *évidence* et *autorité*, est caractérisée par le niveau élevé de l'intensité moyenne, l'écart type de l'intensité fort et l'écart type de F0 fort. Le groupe composé de *politesse-simple*, *déclaration* et *arrogance-impolitesse* partage les mêmes caractéristiques acoustiques, mais *arrogance-impolitesse* est exprimée par la voix plus nasale que les deux autres attitudes. *Doute-incrédulité* et *question-simple* sont tous les deux caractérisés par la grande pente de F0 en dernière more et la grande pente de l'intensité.

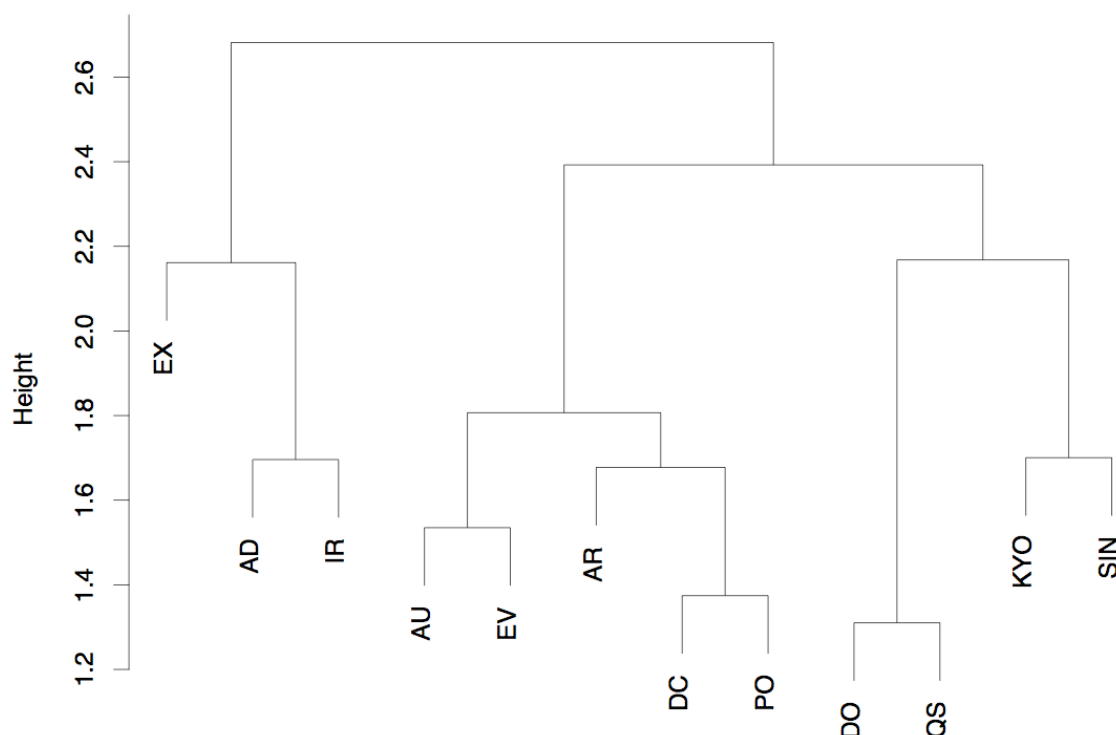


Figure 61 : Catégories des attitudes à partir de la similarité des valeurs de paramètres prosodiques et de la qualité de voix selon la classification hiérarchique.

5.6.4.12. Relations entre prosodie et perception

Selon les résultats du test perceptif, nous avons observé trois façons différentes d'identifier les attitudes présentées.

1. La perception progressive des attitudes (majorité des attitudes)
2. La perception immédiate (*autorité* et *exclamation de surprise*)
3. La perception déterminée sur une partie de l'énoncé (*question-simple* et *irritation*)

Nous considérons que pour les attitudes de la catégorie 1, il n'y a pas de partie déterminante afin d'identifier l'attitude exprimée. Autrement dit, les caractéristiques globales de l'énoncé sont directement liées à l'identification de l'attitude exprimée. Par

contre, la perception immédiate, dès la première more, de l'*autorité* est sans doute due à la voix tendue, le niveau élevé d'intensité et la forte variabilité de l'intensité et de la F0. Quant à *exclamation de surprise* qui est aussi reconnue immédiatement, les auditeurs ont probablement perçu tout de suite la voix tendue et soufflée qui est caractéristique de cette attitude.

En ce qui concerne les attitudes de la catégorie 3, nous ne pouvons pas déterminer de facteurs prosodiques déterminants pour l'identification de ces attitudes à partir de l'ACP ou de la classification hiérarchique que nous avons décrites. La figure 62 illustre le contour de F0 et d'intensité pour *question-simple* en trois phrases de longueurs différentes en japonais. Lors du test perceptif, cette attitude a été reconnue seulement à la fin de la phrase, et la réalisation de cette expression nous indique la montée systématique de F0 et la diminution progressive de l'intensité en fin de phrase. Nous considérons que ces caractéristiques prosodiques ont permis aux sujets japonais d'identifier cette attitude en fin de phrase.

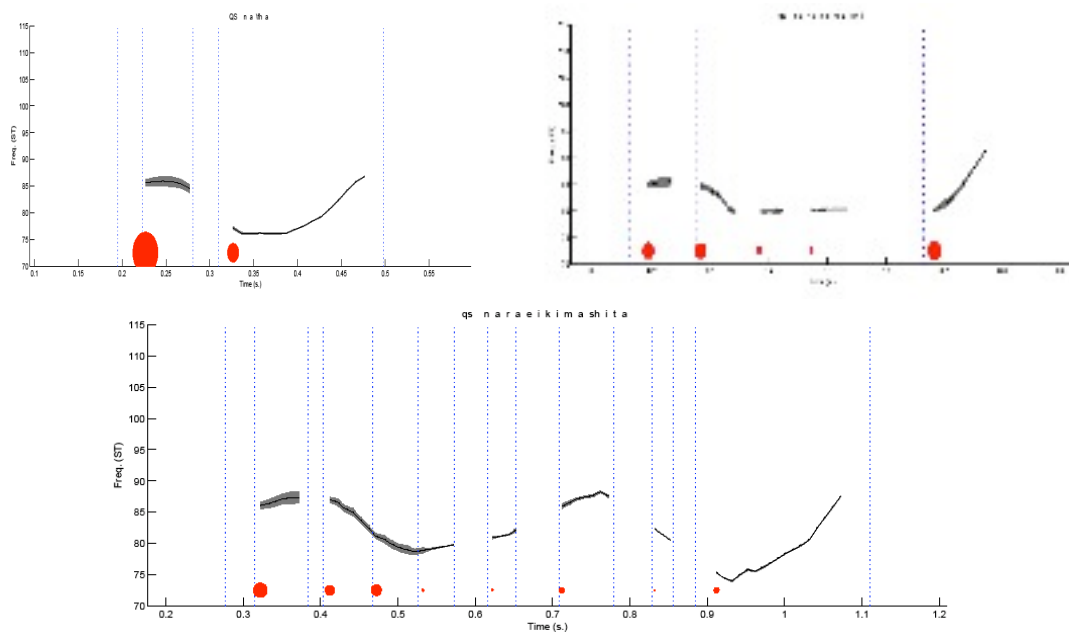


Figure 62 : Contour de F0 et d'intensité pour question-simple japonaise sur la phrase de 2 mores présenté en haut à gauche, celle de 5 mores en haut à droite et 8 mores en bas. La ligne noire indique le contour de F0 présenté en demi ton, l'épaisseur de la ligne représente l'intensité et la taille des points rouges indique également la force d'intensité de l'attaque de chaque point perceptif (P-center). La ligne pointillée en bleue indique la frontière d'un segment.

Quant à *Irritation*, elle a été identifiée en fin de phrase, comme *question-simple* pour les phrases de 2 et de 5 mores lors du test perceptif. La figure 63 représente le contour de F0 et d'intensité pour l'expression d'*irritation* en trois phrases de longueurs différentes en japonais. Pour la phrase de 2 mores, les signaux en 2^{ème} more sont saturés

du fait de la qualité de voix rauque réalisée en second more. Cette partie de l'énoncé est caractérisée également par la voix tendue avec un important allongement. Pour la phrase de 5 mores, la dernière more qui est un endroit déterminant pour l'identification de cette attitude, est caractérisé par l'intensité forte avec la voix rauque et tendue, et aussi par l'allongement important.

En ce qui concerne la phrase de 8 mores, les auditeurs l'ont identifiée aux alentours de la 6^{ème} more lors du test perceptif. Cette partie de l'énoncé est caractérisée acoustiquement par une montée brutale de F0 qui débute à partir de la 4^{ème} more (c'est-à-dire à la frontière syntagmatique) avec une intensité forte, et également par la voix rauque et tendue.

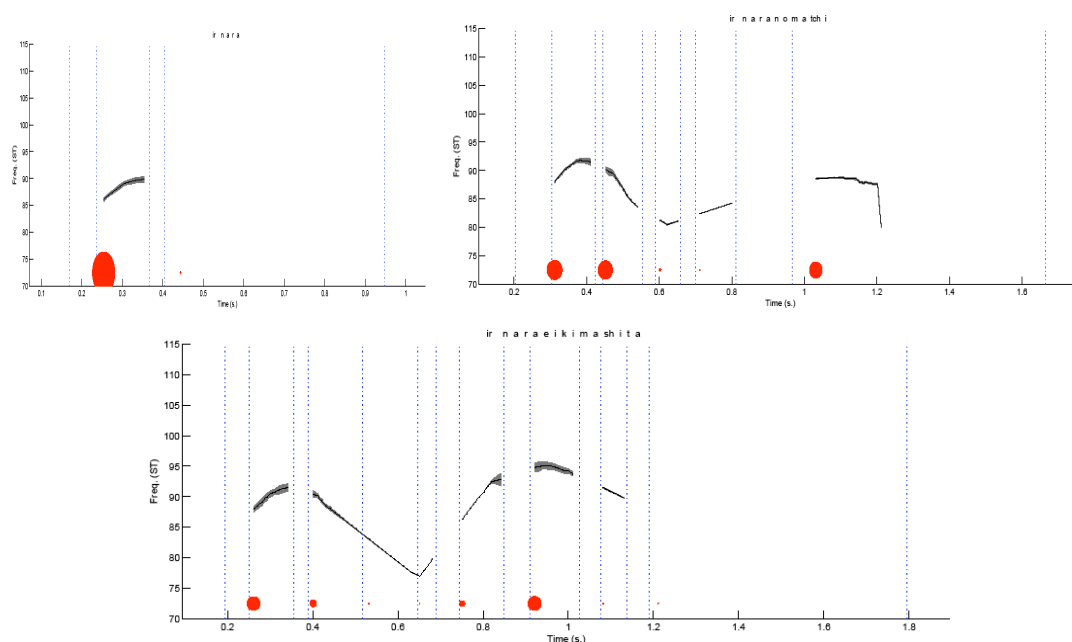


Figure 63 : Contour de F0 et d'intensité pour irritation japonaise sur la phrase de 2 mores présenté en haut à gauche, celle de 5 mores en haut à droite et 8 mores en bas. La ligne noire indique le contour de F0 présenté en demi-ton, l'épaisseur de la ligne représente l'intensité et la taille des points rouges indique également la force d'intensité de l'attaque de chaque point perceptif (P-center). La ligne pointillée en bleue indique la frontière d'un segment.

5.6.4.13. Conclusion

Dans les deux tests perceptifs, l'attitude de *kyoshuku* et d'*admiration* (cf. *supra*) ont été mal reconnues par les auditeurs natifs. Nous pensons que la faible identification de ces attitudes est due au biais de l'influence du contenu lexical des phrases utilisées. C'est pourquoi nous avons effectué 3^{ème} test de perception dans le cadre du gating avec des phrases ayant un contenu lexical mieux adapté au contexte de la vie quotidienne. 33 Japonais natifs ont participé à ce test perceptif de dévoilement progressif. Trois phrases de différentes longueurs sont produites selon 12 attitudes différentes par un enseignant

natif du japonais. Par la suite, ces 36 énoncés enregistrés sont découpés à la fin de chaque more : jusqu'à la dernière more d'une phrase, puis ils sont regroupés par « paquet » de même phrase en ajoutant le bruit blanc.

Les résultats montrent que l'effet de la longueur des phrases sur la perception des attitudes n'est pas confirmé par l'analyse d'ANOVA, mais l'influence significative du gate sur le comportement perceptif des auditeurs est mise en évidence, avec une corrélation positive entre l'évolution de la position du gate et l'augmentation du taux de reconnaissance global de toutes les attitudes. Pourtant, cette identification progressive n'est pas observée pour toutes les attitudes. *Autorité* et *exclamation de surprise* sont toutes les deux reconnues immédiatement dès le début de la phrase. Probablement les auditeurs ont perçu, dès la première more, la voix tendue, le niveau élevé d'intensité et la forte variabilité de l'intensité et de la F0 qui sont corrélés avec l'attitude d'autorité. Quant à *exclamation de surprise*, les auditeurs l'ont probablement perçu tout de suite la voix tendue et soufflée qui est la caractéristique de cette attitude. *A contrario*, *question-simple* est reconnue seulement à la fin de la phrase où la montée de F0 et la diminution progressive de l'intensité sont réalisées. L'*irritation* exprimée avec les phrases de 2 et de 5 mores est identifiée en fin de phrase comme *question-simple*. Cette partie des énoncés est caractérisée par l'intensité forte avec la voix rauque et tendue, et aussi par l'allongement long. En ce qui concerne la même attitude réalisée sur la phrase de 8 more, elle a été identifiée à la position de 6^{ème} gate où la montée brutale de F0 avec l'intensité forte, et surtout avec la qualité de voix rauque et tendue ont été adoptés afin d'exprimer cette attitude. Il est aussi important de noter que le taux de reconnaissance pour les expressions de la politesse reste faible, mais si on regardait la perception globale de la notion de politesse pour ces trois expressions de politesse, la reconnaissance de la notion de politesse pour l'expression de *kyoshuku* augmente progressivement du début jusqu'au dernier gate, et *sincérité-politesse* a aussi été bien perçue comme expression de politesse dès le premier gate. Ce résultat nous indique que même si les auditeurs reconnaissent mal *sincérité-politesse*, cette expression est perçue immédiatement comme une expression de politesse dès le début de l'énoncé par un indice de la forte variation de l'intensité et la voix soufflée.

Quant à *politesse-simple*, les auditeurs ont peu confondu cette attitude avec d'autres expressions de politesse contrairement aux deux autres types de politesse. Selon l'analyse de correspondance, cette attitude a été majoritairement confondue avec *déclaration* plutôt qu'avec d'autres expressions de la politesse. Cette même analyse statistique nous montre que les auditeurs ont bien identifié toutes les attitudes sauf *politesse-simple*.

En outre, quatre catégories perceptives ont été observées selon l'AC. La première catégorie regroupe toutes les attitudes de politesse et *admiration*. La seconde est composée des attitudes qui réclament la réponse du locuteur : *doute-incrédulité*, *exclamation de surprise* et *question-simple*. La troisième consiste en les attitudes qui imposent l'opinion du locuteur aux autres : *autorité* et *irritation*. Enfin la quatrième

catégorie est le groupe des attitudes arrogantes : *évidence* et *arrogance-impolitesse*. Parmi les 12 attitudes, *admiration* et *exclamation de surprise* ont été particulièrement bien perçues.

Puisque nous avons saisi les caractéristiques perceptives et acoustiques de chacune des attitudes japonaises présentées, notre intérêt est maintenant de savoir comment ces valeurs attitudinales des langues validées (japonais, français et anglais) sont perçues chez les auditeurs non natifs. Dans cet objectif, nous allons effectuer un test perceptif croisé avec les attitudes en trois langues que nous avons validées en chapitre 2 et 3 chez les Français et les Américains dans le chapitre suivant.

6. Perception interculturelle des attitudes

6.1. Introduction

Dans les chapitres précédents, nous avons étudié l'expression des attitudes prosodiques de trois langues : le français, l'anglais et le japonais. Nous avons montré la validité de ces attitudes et nous en avons également observé les variations d'identification et les confusions.

Ce chapitre s'attachera à présenter d'abord des travaux effectués sur les expressions affectives au plan interculturel. Ensuite, nous nous attarderons sur une série de tests de perception des attitudes japonaises, anglaises et françaises, effectuées auprès de publics non natifs pour chacune de ces langues : des Français, des Américains et des Japonais passent ainsi les tests qui ne portent pas sur leur propre langue. Ces tests ont pour objectif une comparaison interculturelle des attitudes prosodiques entre ces trois publics de langue et culture différentes.

6.2. Études interculturelles de perception

Nous avons déjà vu dans le chapitre 1 que les attitudes qui nécessitent un contrôle volontaire, sont mises en œuvre spécifiquement par / pour la langue et la culture. Ainsi, les expressions attitudinales sont acquises pendant le développement de l'enfant dans un environnement donné. Il est donc possible que les concepts attitudinaux ou les expressions attitudinales varient d'une culture à l'autre, avec évidemment le risque de créer une mauvaise interprétation de l'intention des messages (Pavlenko 2005). Les études interculturelles sont un moyen particulièrement utile pour découvrir les similarités et les décalages de perception des attitudes.

L'universalité de la perception des affects (notamment des expressions faciales des émotions) est depuis longtemps un thème qui a attiré l'attention des chercheurs (par exemple, Darwin, 1872 ; Ekman, 1972 ; Kramer, 1964 ; Izard 1971 etc.). Cependant, la spécificité culturelle des affects sociaux, autrement dit le « décalage » perceptif des attitudes selon la culture, est rarement traitée dans ce cadre d'étude. En effet leur but prioritaire est d'identifier les émotions universelles plutôt que la différence qui existe entre les langues / cultures (Matsumoto & Assar, 1992). Toutefois ces recherches sur les affects universaux nous ont montré les décalages de perception des affects sociaux dans un contexte multi-langues / cultures. Par exemple, Izard (1971) montre que les Américains et les européens identifient mieux des affects exprimés sur le visage que les

sujets japonais ou africains. Ces travaux dénotent d'une réelle prise de conscience du rôle des attitudes au sein des langues/cultures.

Plus récemment, Fónagy (1991) note dans une description de la confusion entre « joie » et « colère » : *“Six sujets japonais sur trente ont interprété la querelle [hongroise] comme une jubilation”* (p.138). Ainsi, parallèlement au phénomène de reconnaissance/non-reconnaissance des attitudes en langue étrangère, se pose le problème de l'interprétation erronée des attitudes en langue/culture étrangère, due à l'influence de la langue/culture maternelle.

Kramer (1964) est sans doute l'un des premiers à avoir traité de façon systématique les expressions affectives dans le cadre interculturel. Il a effectué des travaux sur la base de phrases portant une connotation de colère en anglais et en japonais, et simulées par des natifs selon cinq expressions affectives (« colère », « amour », « mépris », « indifférence », « chagrin »).

Puis, ayant délexicalisé les phrases en anglais à partir des stimuli naturels, trois tests perceptifs (stimuli naturels vs. délexicalisés en anglais et des stimuli naturels en japonais) auprès de 27 sujets américains n'ayant aucune connaissance de la langue japonaise ont été effectuées. Les résultats montrent que le taux d'identification global pour les stimuli naturels est supérieur à ceux qui ne possèdent pas de contenu lexical (70% contre 61%). Il montre également que les sujets ont mieux identifié les stimuli de leur langue maternelle (c-à-d, en anglais américain) plutôt que les énoncés en japonais (58% du taux d'identification pour les stimuli en japonais). Avec les stimuli naturels, l'attitude de « mépris » est la mieux reconnue (85%), et l'attitude d'« amour » est la moins bien reconnue (56%). Avec les stimuli délexicalisés, « colère » est la mieux reconnue (77%), et « mépris » et « amour » sont les moins bien reconnues (48%). Quant aux expressions en japonais, le « chagrin » japonais est le mieux reconnu par les Américains (90%), mais « amour » et « mépris » ne sont pas bien identifiés (38% pour « amour », et 20% pour « mépris »). Ces résultats indiquent d'une part que l'expression attitudinale de « chagrin » peut être une des expressions partagées à la fois par le japonais et l'anglais, et d'autre part « amour » et « mépris » peuvent être exprimés différemment dans les deux langues.

On notera d'autres travaux, comme ceux de Van Bezooijen et al. (1983), qui ont étudié la perception des expressions affectives dans la langue néerlandaise chez trois groupes des sujets : 48 sujets natifs, 41 Japonais et 40 Taïwanais parlant le chinois. Les sujets non natifs ne possèdent aucune compétence en langue néerlandaise. Une phrase n'ayant aucune connotation particulière a été récitée, à partir d'un texte écrit, par huit locuteurs néerlandais (quatre hommes et quatre femmes) selon 10 expressions affectives : « colère », « honte », « tristesse », « angoisse », « dégoût », « joie », « surprise », « curiosité », « mépris » et « neutre ». Les résultats montrent que les natifs arrivent à reconnaître les expressions exprimées avec un taux de 66%, alors que les sujets non natifs les identifient nettement moins bien (37% pour les Taïwanais, 33% pour les Japonais).

D'un autre point de vue, Scherer et al. (2001) ont effectué un test de perception avec des pseudo-phrases contenant des syllabes pouvant exister en six langues européennes (allemand, français, anglais, néerlandais, italien et espagnol). Ces phrases ont été produites selon cinq expressions affectives allemandes (« colère », « angoisse », « tristesse », « joie » et « neutre ») par quatre locuteurs professionnels allemands (deux hommes et deux femmes). Les sujets de cette expérimentation perceptive sont au nombre de 428 : 70 Allemands, 45 Suisses francophones, 40 Anglais britanniques, 60 Néerlandais, 32 Américains de la côte Ouest, 43 Italiens, 51 Français, 49 Espagnols et 38 Indonésiens. Les résultats montrent que malgré l'absence d'accès lexical dans la tâche liée aux pseudo-phrases, et malgré l'utilisation de syllabes existant dans les six langues européennes (soit tous les sujets concernés sauf les indonésiens), ce sont les sujets allemands qui ont le mieux identifié les expressions affectives allemandes (74% du taux d'identification). A contrario, les sujets indonésiens ont beaucoup moins bien identifié les stimuli, avec un taux de reconnaissance de 52%. Ceci implique que les sujets perçoivent moins bien les expressions affectives de cultures éloignées plutôt que de cultures proches.

Pourtant, on assiste parfois à des situations paradoxales où les scores de perception des deux groupes de sujets – natifs et non-natifs – sont quasi identiques voire inversés –lorsque les non-natifs obtiennent un meilleur score de reconnaissance que les natifs eux-mêmes.

Ainsi, Erickson & Shochi (2006a) ont travaillé sur la perception des expressions affectives japonaises pour des sujets de trois langues/cultures différentes : 20 Japonais, 20 Américains et 9 Coréens. Une phrase sémantiquement neutre a été élicitée selon cinq expressions affectives : « colère », « tristesse », « surprise », « doute » et « joie », par une locutrice japonaise. Les résultats montrent que malgré le manque de compétence en japonais, les auditeurs non natifs sont capables d'identifier les cinq expressions affectives au dessus de 85% du taux de reconnaissance (taux d'identification globale de 92,8% pour les Américains, 89% pour les Coréens). A l'inverse, les auditeurs japonais ont le moins bien reconnu ces expressions de leur langue maternelle (taux d'identification moyen de 83,6%), notamment en ce qui concerne l'identification de « colère » (75%). Les auteurs expliquent ces scores plus faibles des natifs par le fait qu'ils seraient perturbés par le décalage entre contenu sémantique de la phrase et émotions/attitudes exprimées, contrairement aux sujets non natifs, que le contenu lexical ne perturbe pas.

Mejvaldová (2000) a quant à elle travaillé sur la perception/production des expressions attitudinales du tchèque et du français. Elle a montré que les natifs perçoivent généralement mieux les attitudes exprimées (Tchèque vs. Français). Dans ses ouvrages, elle a identifié un potentiel de « faux amis » prosodiques (« homonymie » selon sa propre terminologie : une morphologie prosodique unique correspondant à deux attitudes distinctes) : par exemple l'attitude d'« ennui » en français, bien perçue par les natifs français, mais que les sujets tchèques ont tendance à interpréter comme « triste » ou « furieux » ou la « déclaration » et la « surprise » de deux locuteurs tchèques, perçues comme « joyeux » par les sujets français.

D'un point de vue socioculturel, la question du rôle du genre dans l'expression/perception des affects en langue maternelle et leur influence sur la perception des affects en langue étrangère peut se poser. Erickson et al (2002) ont travaillé sur la perception de l'expression du « sarcasme » en anglais américain, élicité par deux locuteurs américains (un homme et une femme), par des natifs et par des Japonais. Les résultats montrent que le sarcasme américain élicité par l'homme est mal reconnu par les Japonais, alors qu'il l'est nettement mieux lorsqu'il est élicité par l'Américaine. Il est possible d'expliquer cette différence de reconnaissance pour « sarcasme » par les sujets japonais, par le fait que le contour de F0 est montant en fin de phrase pour l'élicitation de l'Américaine, contrairement à la phrase élicité par l'Américain. Ce mouvement du contour de F0 est similaire au « sarcasme » japonais (Hamasaki et al., 1999). Par ailleurs, il est à noter que les Américains perçoivent nettement mieux cette attitude (95% contre 42,5%) que les Japonais : il s'agit ici du « *native effect* ».

Dans une optique différente, Erickson et al. (2003) ont voulu observer le rôle des informations paralinguistiques dans la parole expressive : d'une part les compléments paralinguistiques (désormais nommés CP) tels que « humeur », « politesse », « énergie », « vivacité », « dureté », « intérêt », d'autre part les informations paralinguistiques (désormais IP) tels que « sincérité », « surprise », « habitué », « joie », « réflexion », « sommeil », « indifférence », « admiration », « excitation », « tristesse », « dépression », « soupçon », « sarcasme », « insatisfaction », « doute », « colère », « angoisse », « calme », « irritation », « plaisanterie », « taquinerie », « voilé », « précipitation » et « consternation ». Plus précisément, les auteurs ont analysé la perception de ces informations par des sujets japonais et américains. Ils ont ainsi effectué un test de perception auprès de 20 étudiants japonais (femmes de 18-21 ans) de différentes régions, et 10 étudiants américains (7 hommes et 3 femmes de 22-29 ans). Ces sujets ont écouté un total de 60 stimuli et évalué chacun d'eux sur une échelle de valeur allant de -3 (plus faible) à +3 (plus fort), selon 6 critères (« humeur », « politesse », « énergie », « vivacité », « dureté », « intérêt »). Suite à l'évaluation, les auditeurs devaient sélectionner quelle était l'« information paralinguistique » que le locuteur avait voulu exprimer parmi les 24 items proposés. Les sujets pouvaient choisir deux items au maximum. Les résultats à ce test de perception montrent que les sujets des deux langues étaient capables de percevoir les IP et les CP par le seul indice de la prosodie et sans se référer au contexte dans lequel la phrase était produite. Cependant, ils n'étaient pas tous d'accord quant au choix des IP, cela même parmi les auditeurs natifs d'une même langue. La meilleure identification d'IP par les auditeurs américains était de 46% (irritation), et celle des auditeurs japonais de 67% (surprise). De plus, les résultats nous informent que leurs choix de IP sont corrélés avec l'évaluation du degré de chacun des CP. Par exemple, un stimulus a été classé dans les IP « colère » (48%) ou « surprise » (29%) chez les Japonais, et dans les IP « irritation » (46%) ou « colère » (23%) chez les Américains ; mais ce même stimulus, en ce qui concerne les CP, a été perçu comme « mauvaise humeur », « impoli », « énergique », « dure » et « intéressé » pour les auditeurs des deux langues. Nous voyons ici que le choix

des IP correspond au degré de chaque CP. Cependant, les sujets ont montré une différence culturelle (ou linguistique) au niveau des choix des IP et des valeurs données aux CP. Par exemple l'énoncé monosyllabique "Eh" a été perçu, par les sujets japonais, comme : légèrement « mauvaise humeur » (-0.6), ni poli et ni impoli (0), « énergie » faible (-1.5), faible « vivacité » (-1.4), moins « dure » (-0.2) et peu « intéressé » (-0.7) ; concernant les IP, les sujets japonais ont répondu que le locuteur était en train de réfléchir (44%) ou « calme » (12%).

De l'autre côté, les sujets américains ont perçu ce même énoncé comme : « impoli » (-1.5), « vivacité », plus ou moins « neutre » (-0.2), absence de « dureté » (-1.9).

Les auteurs considèrent que cette différence est due au rôle important joué par le contenu sémantique chez les Japonais, cet accès au contenu sémantique étant absent chez les sujets américains, ces derniers ont par conséquent évalué ces stimuli uniquement à partir de la prosodie.

Parallèlement au test de perception, les auteurs ont mesuré la durée, la moyenne de F0, l'Amplitude Quotient (AQ) et le *pitch-Normalized Amplitude Quotient* (NAQ) de ces six stimuli. Les sujets américains montrent une corrélation négative entre AQ/NAQ et la dureté de la voix. Cela signifie que plus la voix devient soufflée (la valeur haute de NAQ), moins ce stimulus est perçu comme rauque chez les Américains. La voix soufflée entraîne la perception du « doute », « voilé » ou de « tristesse » au niveau des IP. En revanche, aucune valeur significative de cette corrélation négative n'a été identifiée chez les sujets japonais. Ces derniers montrent quant à eux une corrélation positive entre F0 et « humeur », « vivacité » et « intérêt ». Par exemple, l'énoncé "Eh" réalisé avec une F0 très haute et une durée très courte est perçu comme « surprise ». Les auteurs notent qu'une F0 très haute est fortement utilisée par les Japonais pour l'attitude de « surprise ».

Enfin, à la suite de travaux sur la perception croisée chez les auditeurs adultes, il sera intéressant d'envisager l'acquisition des expressions attitudinales chez des enfants. Il n'existe en effet que peu de recherches sur le plan développemental, mais quelques unes de celles effectuées indiquent une acquisition relativement précoces (entre 7 et 11 ans) de certaines attitudes françaises (Clément, 1999). D'autres études préliminaires (Erickson&Shochi, 2006b) ont également abordé la reconnaissance des affects (« colère », « joie », « suspicion », « surprise », « tristesse ») chez des enfants japonais (12 ans) pour ce qui est de leur langue maternelle. Les auteurs ont ainsi montré que la « colère irritée », « élicitée » par un adulte masculin japonais peut être perçue comme « tristesse » ou « joie », et que « surprise » peut être reconnue comme « colère irritée ».

Dans une perspective parallèle et concernant également la perception des attitudes chez les enfants, on trouve des études portant sur des sujets bilingues. Roseberry-McKibbin & Brice (1999), ont ainsi travaillé sur la perception des affects de deux langues (l'espagnol mexicain et l'anglais américain) chez des enfants bilingues (langue maternelle espagnole et langue seconde anglaise). Ils ont enregistré une phrase

interrogative, en soit affectivement neutre, dans les deux langues (espagnol et anglais) produite par des natifs (un homme et une femme pour chaque langue) selon trois « émotions » : « joie », « tristesse » et « colère ». Les stimuli ont ensuite été présentés auprès de neuf sujets mexico-américains (enfants de 5–7 ans). Selon l'ANOVA, la différence de réponses des enfants entre les affects des deux langues est significative, avec un meilleur taux de reconnaissance des affects espagnols par rapport aux anglais. Cependant, les auteurs n'ont pu expliquer si ce résultat était dû à la compétence des enfants testés ou au locuteur américain qui n'aurait pas exprimé suffisamment les affects par rapport aux locuteurs espagnols. Étudier le « bilinguisme » dans le cadre des perceptions croisées nécessite des analyses prudentes.

Finalement, concernant les perceptions attitudinales du point de vue interculturel, il est indispensable de citer des travaux prenant en compte l'influence de deux modalités différentes (audio et visuel) dans la communication en situation réelle.

Dans ses études multimodales, Shigeno (1998) a essayé de trouver des similarités et des différences culturelles dans la reconnaissance de stimuli audio-visuels, entre des Japonais et des Américains.

Cinq phrases en deux langues, comportant des contenus lexicaux relativement neutres, ont été choisies. Ces phrases enregistrées en chambre sourde et élicitées par deux acteurs professionnels (un Japonais et un Américain) selon six émotions fondamentales : « joie », « surprise », « colère », « dégoût », « angoisse », et « tristesse » ont en même temps été filmées de manière à avoir les expressions faciales des locuteurs. Au total, 36 stimuli (six expressions vocales x six expressions faciales) ont été sélectionnés et utilisés lors du test perceptif.

Les sujets, des étudiants japonais et américains du Nord (au nombre de sept et cinq pour chaque groupe de sujets) n'avaient jamais quitté leur pays plus d'un an. Il a été procédé à l'expérimentation perceptive en trois sessions en chambre sourde : audio seul, visuel seul et audio-visuel.

Les résultats en perception des stimuli en audio seul montrent que les sujets ont mieux perçu les émotions exprimées dans leur langue maternelle. En effet, les sujets japonais ont mieux perçu les expressions émotionnelles du locuteur japonais par rapport à celles du locuteur américain, et les sujets américains ont quant à eux mieux reconnu les stimuli du locuteur américain.

Toutefois, les japonais ont eu des difficultés à percevoir l'angoisse, à la fois en audio seul et en visuel seul). Son taux d'identification était en effet nettement inférieur à celui du même stimulus chez les sujets américains.

Par ailleurs, en général, tous les sujets (Japonais et Américains) ont identifié les stimuli grâce aux indices visuels plutôt qu'aux indices auditifs. Les auteurs ont pu conclure que les natifs perçoivent généralement mieux les attitudes exprimées au niveau visuel.

Tous les travaux que nous venons de décrire donnent ainsi quelques pistes pour mieux cerner la perception interculturelle des expressions affectives.

D'abord, quelles que soient les langues traitées, les natifs les perçoivent mieux que les non natifs. Toutefois, les natifs sont plus sensibles aux contenus lexicaux, et s'ils trouvent un décalage entre l'information prosodique d'une attitude et le contenu sémantique de la phrase, il leur arrive de mal identifier à leur tour l'attitude exprimée (Erickson & Shochi 2006a).

Un autre point intéressant est que la distance culturelle est un facteur important pour l'identification des affects. En effet, plus cette distance entre la langue/culture cible et la langue/culture maternelle est courte, et meilleurs seront les scores de reconnaissance.

Par ailleurs, si les auditeurs non-natifs montrent un comportement perceptif différent pour certains affects par rapport aux natifs, il est possible que l'expression d'affect concerné renvoie à une notion différente dans la culture/langue des sujets non natifs.

Pour conclure, la majorité des affects s'avère largement identifiable chez les non natifs, même par l'indice auditif seul, mais il existe également des affects sociaux bien spécifiques à chaque langue/culture.

Revenons maintenant à l'analyse de la perception des attitudes pour les trois langues cibles qui nous intéressent : japonais, anglais et français.

6.3. Attitudes japonaises perçues par des Français et des Américains

6.3.1. Objectif du test perceptif

L'objectif de ces tests est d'appliquer le même paradigme perceptif que celui utilisé auprès des sujets japonais à des sujets anglophones et francophones afin de mesurer le décalage perceptif de comportement entre ces sujets de trois langues différentes.

6.3.2. Composition du Corpus

Nous avons donc utilisé les 12 attitudes validées auprès des auditeurs japonais (*supra*). Il s'agit de : *doute-incrédulité*, *évidence*, *exclamation de surprise*, *autorité*, *irritation*, *arrogance-impolitesse*, *sincérité-politesse*, *admiration*, *kyoshuku*, *politesse-simple*, *déclaration* et *question-simple*.

Tableau 20 : *Corpus des attitudes japonaises : sept phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque.*

n° de more	Phrase	Traduction
1	Me	Les yeux
2	Na* ra	Nara
5	Na* rade neru	Il dort à Nara
8	Na* goyade nomimas [u]	Il boit à Nagoya
8	Nara* shide nomimas [u]	Il boit dans la ville de Nara
8	Matsuri* de nomimas [u]	Il boit à la fête
8	Naniwade nomimas [u]	Il boit à Naniwa

Les définitions des attitudes utilisées ici sont identiques à celles présentées aux sujets japonais, mais traduites dans la langue maternelle du sujet (c-à-d., français ou anglais). Elles sont présentées en annexe. En ce qui concerne l'attitude de *kyoshuku*, il n'existe pas l'entrée lexicale qui correspond à ce type de politesse en français et en anglais. Par conséquent, nous l'avons nommé *kyoshuku* comme en japonais dans la liste des attitudes présentées, en ajoutant une explication complémentaire afin que les sujets français et américains comprennent la notion et le contexte de cette attitude. Toutes ces définitions sont présentées aux auditeurs lors du test perceptif.

Les phrases que nous avons sélectionnées sont celles que nous avons employées pour les sujets japonais (cf. tableau 20).

6.3.3. Validation perceptive

Le test de perception a été effectué auprès d'auditeurs français et américains n'ayant aucune connaissance sur la langue japonaise. L'objectif de cette série du test perceptif est de connaître le décalage de comportement perceptif entre les sujets de ces trois langues.

6.3.3.1. Méthodologie expérimentale

a- Stimuli

Nous avons utilisé exactement les mêmes stimuli que ceux validés sur le japonais. C'est-à-dire sept énoncés différents produits selon 12 attitudes, soit 84 stimuli.

b- Matériel

Le test perceptif est effectué à l'aide de casques Pioneer SE-M570 dans une salle calme. Tout le test se déroule grâce à une interface développée sous le logiciel « Runtime revolution » avec un ordinateur PC Windows XP professionnel. Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 12 attitudes

possibles et les définitions des attitudes présentées. Cette interface a été créée en langue maternelle du sujet³⁰ (français et anglais).

c- Procédure expérimentale

Le test perceptif a été conduit sujet après sujet, individuellement. Lors du test, le sujet est assis en face d'un écran et perçoit un stimulus présenté dans un ordre aléatoire et différent pour chaque sujet. Pendant le test, le sujet doit juger de l'attitude exprimée parmi un choix fermé des 12 attitudes possibles. Le sujet ne peut écouter qu'une seule fois chaque stimulus avant de choisir l'attitude qu'il a perçue afin d'éviter que le sujet s'entraîne au cours du test.

Lors du commencement du test perceptif, on a demandé au sujet de lire la consigne affichée sur l'écran. Ensuite, on a expliqué brièvement la tâche et la consigne, puis la définition des 12 attitudes dans le but d'habituer le sujet à la voix du locuteur (qui se trouve être aussi l'expérimentateur). Enfin à la fin du test perceptif, nous avons interrogé le sujet à propos des attitudes afin de savoir quelles attitudes étaient difficiles à distinguer ou s'il avait perçu d'autres attitudes qui n'étaient pas proposées.

6.3.3.2. Sujets

Les auditeurs sont au nombre de 35 : 15 auditeurs français (5 sujets masculins et 10 sujets féminins) et 20 auditeurs américains (8 sujets masculins et 12 sujets féminins). L'âge moyen est de 25,4 pour les auditeurs français et 24,9 pour les auditeurs américains. Aucun sujet n'a fait état de trouble perceptif.

En ce qui concerne des sujets français, la plupart sont des étudiants et des chercheurs du Gipsa-lab, département de la parole et cognition. Ils habitent tous dans la région grenobloise et ils n'ont aucune connaissance sur la langue japonaise.

Tous les auditeurs américains sont des étudiants à Black Hills State University, Spearfish dans l'état de South Dakota aux Etats-Unis. Ils n'ont aucune connaissance sur la langue et la culture japonaise.

6.3.4. Résultats du test perceptif

6.3.4.1. Les sujets français

a- Répartition des attitudes par rapport au hasard

De la même manière que pour les résultats des sujets japonais, nous avons effectué un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 21).

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

³⁰ Les interfaces utilisées sont présentées en annexe.

Ce premier résultat montre donc que des auditeurs français, lorsqu'ils perçoivent des attitudes japonaises, répondent de manière cohérente. Les variations expressives introduites par les attitudes japonaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs français, même si ces identifications ne correspondent pas forcément au sens de ces attitudes en japonais. Nous analyserons plus loin ces confusions.

Tableau 21 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	120.9	(ddl : 11, p.<0.01)
Arrogance-impolitesse (AR)	117.3	(ddl : 11, p. < 0.01)
Autorité (AU)	556.4	(ddl : 11, p. < 0.01)
Déclaration (DC)	291.9	(ddl : 11, p. < 0.01)
Doute-incrédulité (DO)	120.3	(ddl : 11, p. < 0.01)
Évidence (EV)	112.3	(ddl : 11, p. < 0.01)
Exclamation de surprise (EX)	362.3	(ddl : 11, p. < 0.01)
Irritation (IR)	467.7	(ddl : 11, p. < 0.01)
Kyoshuku (KYO)	273.6	(ddl : 11, p. < 0.01)
Politesse-simple (PO)	135.8	(ddl : 11, p. < 0.01)
Question-simple (QS)	149.3	(ddl : 11, p. < 0.01)
Sincérité-politesse (SIN)	59.0	(ddl : 11, p. < 0.01)

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Pour observer l'effet de la longueur sur le choix des attitudes, nous avons effectué le test de χ^2 sur les résultats moyens par longueur, toutes attitudes confondues. Ce test donne un résultat significativement différent du hasard (avec p. < 0.05, ddl = 33, χ^2 = 62.2).

On peut donc constater que, comme pour les auditeurs japonais, la longueur de l'énoncé influence le choix des attitudes. Mais ce résultat ne nous indique pas à partir de quelle longueur d'énoncé les distributions deviennent significatives. Pour répondre à cette question, nous avons effectué un test de χ^2 sur les résultats moyens entre les longueurs de 1 et 2 mores, de 2 et 5 mores et de 5 et 8 mores, toutes attitudes confondues.

Le test de χ^2 sur les résultats moyens par longueur de 1 more et 2 mores, toutes attitudes confondues, donne un résultat significatif (ddl = 11, χ^2 = 32.2, p. < 0.05).

A contrario, entre 2 et 5 mores, nous n'avons pas identifié d'effet significatif de la longueur (ddl = 11, χ^2 = 10.9, p. > 0.05).

En ce qui concerne la longueur de 5 et de 8 mores, toutes attitudes confondues le test χ^2 donne un effet significatif (ddl = 11, χ^2 = 34.5, p. < 0.05).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe, cependant cet effet a été identifié entre les mots isolés monomoraïques et dimoraïques, et entre les phrases de 5 mores et de 8 mores.

c- Effet de l'accent lexical

Quant à l'influence de la position de l'accent lexical sur les attitudes choisies, nous avons, dans un premier temps, effectué un test de χ^2 sur les résultats moyens des réponses toutes attitudes confondues, et toutes longueurs confondues. L'hypothèse nulle est que l'accent ne fait pas changer les réponses des sujets.

Le test de χ^2 pour ces données n'est pas significatif (ddl = 33, $\chi^2 = 39.9$, $p. > 0.05$), donc on ne peut pas rejeter l'hypothèse nulle : il n'y a donc pas d'effet significatif de la position de l'accent. Afin de soustraire l'influence de la longueur de l'énoncé présente dans ce test, un autre test de χ^2 a été conduit sur les résultats moyens par position de l'accent avec toujours la même longueur de 8 mores, toutes attitudes confondues. Ce test ne donne pas d'effet significatif (ddl = 33, $\chi^2 = 28.8$, $p. > 0.05$). On peut donc rejeter l'effet de la position de l'accent comme facteur d'influence sur le choix des attitudes par des auditeurs français.

d- Taux d'identification des attitudes

Ensuite, nous avons calculé le taux d'identification des attitudes en pourcentage afin d'envisager le degré d'identification des attitudes japonaises auprès des auditeurs français.

La figure 64 présente le taux d'identification des 12 attitudes japonaises pour les 15 auditeurs français.

Les résultats montrent que toutes les attitudes sauf celle de *kyoshuku* ont été au dessus du seuil du hasard (8,4%).

Le taux global d'identification des attitudes japonaises chez les Français étaient en moyenne inférieur à celui des Japonais (35% chez les Français contre 55% chez les Japonais). Autrement dit, les auditeurs français ont moins bien identifié les bonnes réponses que les auditeurs japonais. Ce résultat confirme ce « native effect » observé précédemment dans les travaux d'Erickson et al. (2003), de Mejvaldová (2000), Scherer et al.(2001) et de Shigeno (1998).

Parmi 12 attitudes, *autorité* et *irritation* sont particulièrement bien reconnues chez les auditeurs français. En outre, *déclaration* et *exclamation de surprise* sont aussi bien reconnues en montrant au dessus de 50 % du taux d'identification. Le reste des attitudes ont été reconnu au dessus du seuil du hasard, mais avec confusions importantes.

Les Français ont montré beaucoup de difficultés à percevoir les distinctions entre les attitudes de politesse bien particulières à la société japonaise (*sincérité-politesse* 15%, *kyoshuku* 0%, contre 32% (*sincérité-politesse*) et 25% (*kyoshuku*) de reconnaissance chez

les auditeurs japonais. Ce faible taux de reconnaissance de ces attitudes vient probablement de l'absence de ce concept dans la société française d'une part, et d'autre part de la qualité de voix très particulière exprimant cette politesse japonaise, qui transmet (manifestement) une signification (ou intention) complètement différente pour des francophones.

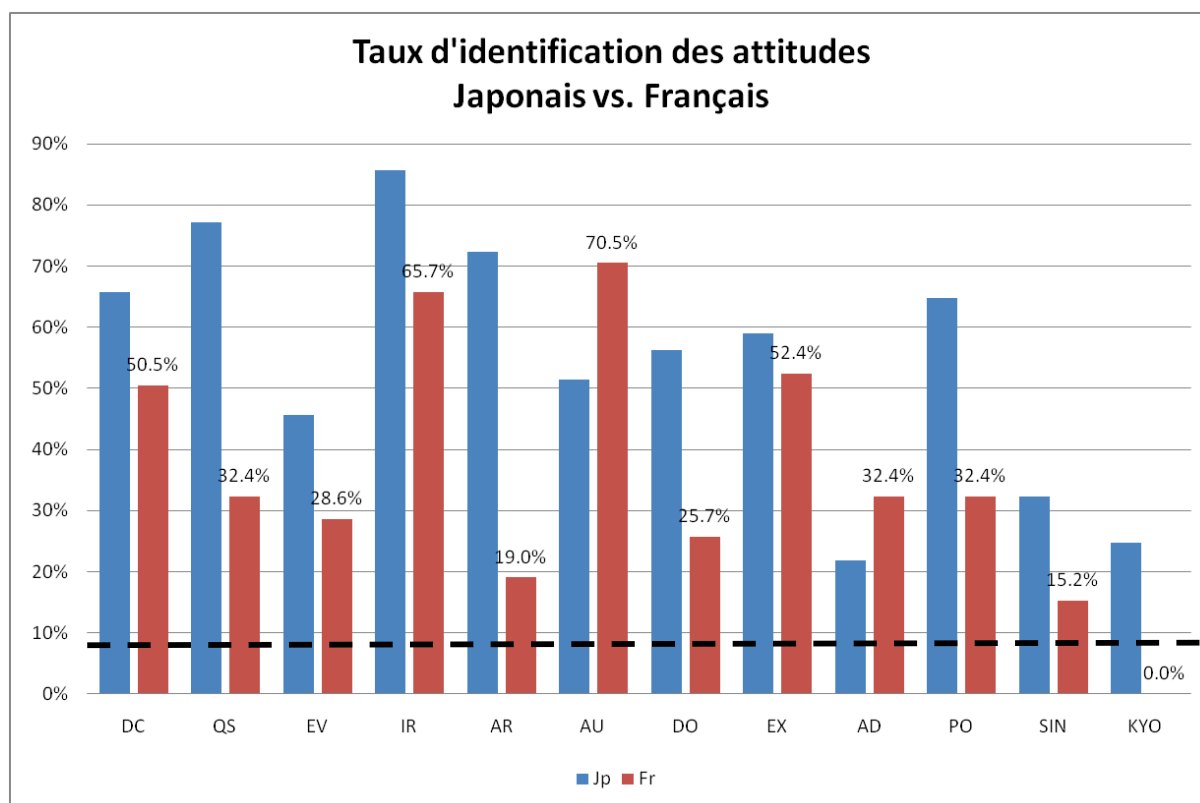


Figure 64 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs français et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO (doute-incrédulité), EV (évidence), EX (exclamation de surprise), IR (irritation), KYO (kyoshuku), PO (politesse-simple), QS (question-simple) et SIN (sincérité-politesse)

En revanche, les attitudes qui imposent l'opinion du locuteur telles qu'*autorité* et *irritation* ont été bien reconnues. Surtout *autorité* qui est mieux reconnue par les Français que par les Japonais (70% contre 51%). Par contre, *arrogance-impolitesse* a montré une faible reconnaissance même si cela a été reconnu (19%) alors que les auditeurs japonais l'ont très bien perçue sans confusion (72%). Ce grand écart de taux d'identification amène à suggérer que la notion du terme « arrogance-impolitesse » peut être interprétée différemment entre les deux langues/cultures lointaines.

On peut également observer un grand écart de taux d'identification de *question-simple* entre les deux groupes de sujets. Contrairement aux sujets japonais qui l'ont bien reconnue (77%), le taux d'identification de cette attitude était nettement plus faible chez

les Français (32%). Nous reviendrons ultérieurement sur ce point dans le paragraphe concernant la confusion des attitudes.

Quant à *admiration*, les Français l'ont reconnue à 32%, alors que les Japonais l'ont reconnue à 22% seulement.

e- **Comportement perceptif des sujets français en l'analyse de correspondance**

Afin d'observer le comportement perceptif des sujets français, nous avons calculé la distance des points cognitifs de 12 attitudes par rapport aux points stimuli selon l'analyse de correspondance comme pour les résultats des Japonais.

Selon ce test statistique, quatre premières dimensions expliquent 85% de la variance.

Maintenant, en regardant la figure 65 où les points stimuli se situent très proche de points perceptifs sauf *question-simple*, *arrogance-impolitesse* et *kyoshuku*. La perception de ces attitudes qui montrent la distance longue par rapport aux points de distribution des stimuli. Cela indique que leur reconnaissance est très faible. Il est important de noter que le comportement perceptif des auditeurs français pour l'attitude de *kyoshuku* se situe près d'*irritation*, et loin de *sincérité-politesse* ou de *politesse-simple*. Cela explique que les sujets français ne reconnaissent pas *kyoshuku* comme expression de politesse, et que c'est interprété comme *irritation*. Un autre point intéressant est que le comportement pour *question-simple* se situe loin d'exclamation de surprise et de *doute-incrédulité*, mais plutôt proche d'*admiration*. Cette attitude montre également une confusion importante avec déclaration en troisième et quatrième dimension de la figure 65. Ce comportement des Français est très différent du celui des Japonais, et cela veut dire que les Français n'entendent pas forcément la phrase interrogative japonais, en montrant la confusion avec la phrase affirmative.

Quant à *arrogance-impolitesse*, elle est localisée dans un endroit isolé, c'est-à-dire, cette attitude est mal reconnue, mais les Français ont du mal à répondre à quoi cette expression ressemble.

Cette figure montre également certaines confusions (ou un comportement perceptif similaire) qui sont présentées par des disques violets (entre *déclaration* et *évidence*, entre *politesse-simple* et *sincérité-politesse* et entre *doute-incrédulité* et *exclamation de surprise*).

Nous allons analyser plus concrètement les phénomènes de confusions entre les attitudes dans le paragraphe suivant.

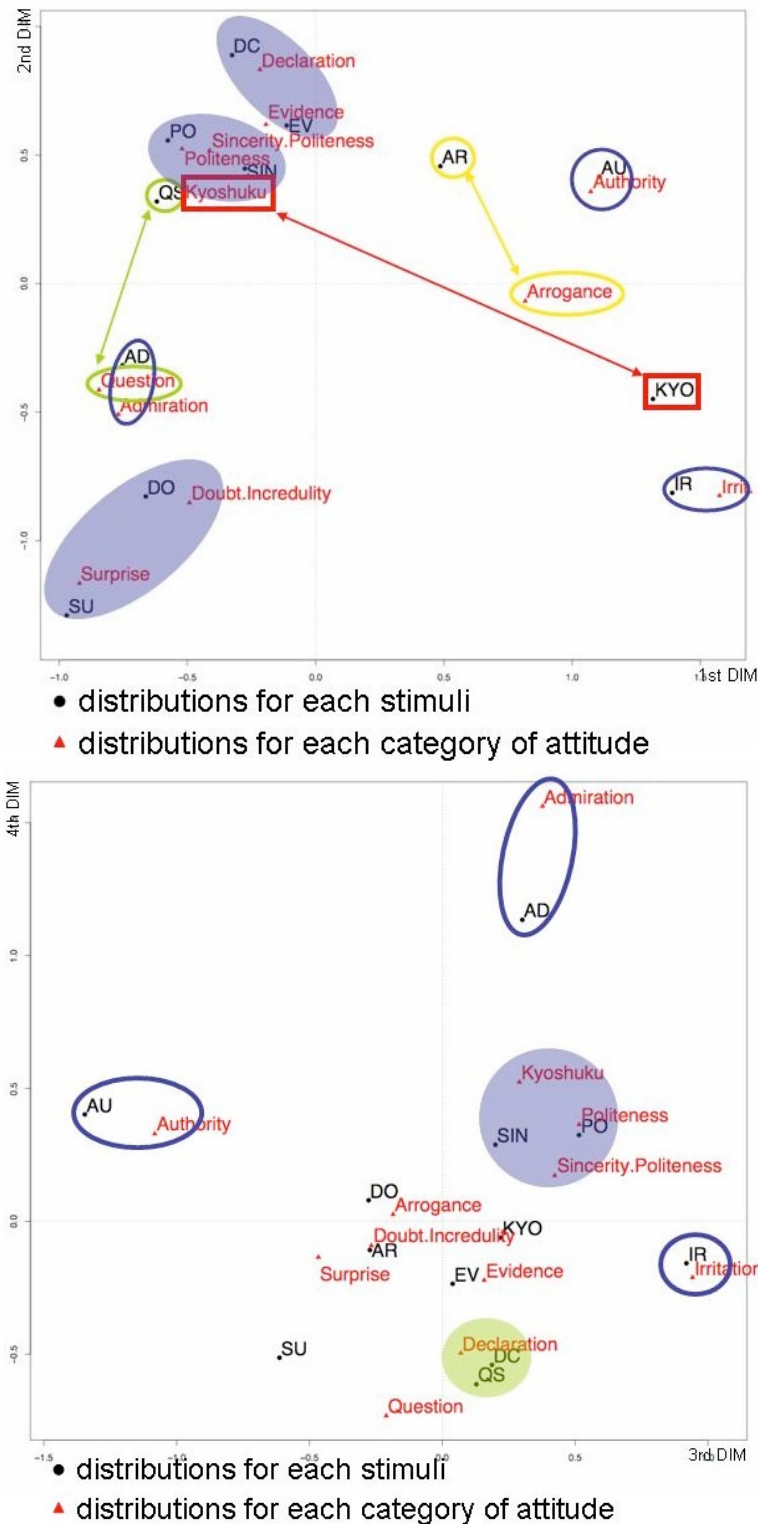


Figure 65 : Comportement perceptif de 15 auditeurs français pour les 12 attitudes japonaises présenté en deux dimensions (figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Signification des étiquettes : les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles.

f- Confusion entre attitudes

En comparant le taux d'identification de chacune des attitudes, l'attitude de *kyoshuku* n'est pas été reconnue au dessous du seuil du hasard. Par la suite, l'analyse de correspondance a démontré la faible reconnaissance de *Kyoshuku*, *question-simple* et *arrogance-impolitesse*.

Maintenant nous allons voir le taux de confusion pour chaque attitude présentée (cf. tableau 22 et figure 66). D'après les résultats présentés, trois attitudes : *admiration*, *irritation* et *autorité* ont été reconnues sans aucune confusion particulière. Notamment *irritation* (65,7%) et *autorité* (70,5%) ont été très bien perçue chez les auditeurs français.

En revanche, deux attitudes de politesse : *kyoshuku* et *sincérité-politesse* ont été mal reconnues. L'attitude de *kyoshuku* n'a pas du tout été identifiée par les auditeurs français. Cette attitude a été reconnue comme *irritation*, *arrogance-impolitesse* ou *autorité*. Cela était attendu, étant donné que cette attitude n'existe pas du tout dans la société française, non plus que la qualité de voix rauque adoptée pour cette expression ne se rapporte en rien à l'expression de la politesse en français. Par ce fait, cette expression japonaise renvoie une notion tout à fait différente chez les français, et cela peut être un des cas typique de « faux-amis » prosodique.

Tableau 22 : Les matrices de confusion des 15 auditeurs français. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : AD (*admiration*), PO (*politesse-simple*), KYO (*kyoshuku*), SIN (*sincérité-politesse*), AR (*arrogance-impolitesse*), AU (*autorité*), IR (*irritation*), DO (*doute-incrédulité*), EX (*exclamation de surprise*), QS (*question-simple*), EV (*évidence*) et DC (*déclaration*).

Attitudes perçues	Attitudes présentées											
	AD	AR	AU	DC	DO	EV	SU	IR	KYO	PO	QS	SIN
Admiration	32%	0%	0%	0%	10%	0%	1%	2%	0%	4%	1%	2%
Arrogance	1%	19%	14%	0%	10%	11%	0%	10%	29%	0%	0%	6%
Autorité	0%	24%	70%	3%	0%	8%	0%	10%	20%	0%	0%	9%
Déclaration	0%	22%	8%	50%	2%	22%	2%	0%	1%	15%	22%	14%
Doute-incrédulité	7%	3%	1%	1%	26%	3%	22%	5%	5%	1%	4%	8%
Evidence	10%	16%	2%	19%	2%	29%	0%	2%	3%	8%	15%	6%
Surprise	16%	0%	0%	3%	27%	3%	52%	2%	0%	6%	1%	2%
Irritation	0%	8%	4%	0%	0%	0%	0%	66%	42%	0%	0%	1%
Kyoshuku	8%	2%	1%	2%	8%	4%	0%	0%	0%	12%	4%	17%
Politesse	11%	4%	0%	6%	1%	5%	1%	2%	0%	32%	12%	19%
Question	3%	0%	0%	3%	12%	3%	22%	0%	0%	3%	32%	2%
Sincérité-Politesse	11%	3%	0%	13%	4%	13%	0%	3%	1%	19%	9%	15%

Contrairement à *kyoshuku*, l'attitude de *sincérité-politesse* a été confondue avec *kyoshuku* ou *politesse-simple*. Cela explique que l'expression de *sincérité-politesse* a été perçue une expression de politesse malgré le faible taux de reconnaissance.

Sept autres attitudes ont été reconnues, mais avec des confusions. *Politesse-simple* a été confondue avec *sincérité-politesse*. Les auditeurs français n'ont pas confondu cette attitude avec celle de *kyoshuku*.

L'attitude d'*arrogance-impolitesse* est mal reconnue par les auditeurs français (19%) ; et cela correspond au résultat tiré par l'analyse de correspondance que nous avons observé dans le paragraphe précédent. Ce taux est juste à peine au-dessus de seuil de reconnaissance. Cette attitude est confondue avec *déclaration* et *autorité*. Cela nous indique que les auditeurs français ont tendance à confondre *arrogance-impolitesse* japonaise avec une simple déclaration ou une sorte de l'autorité.

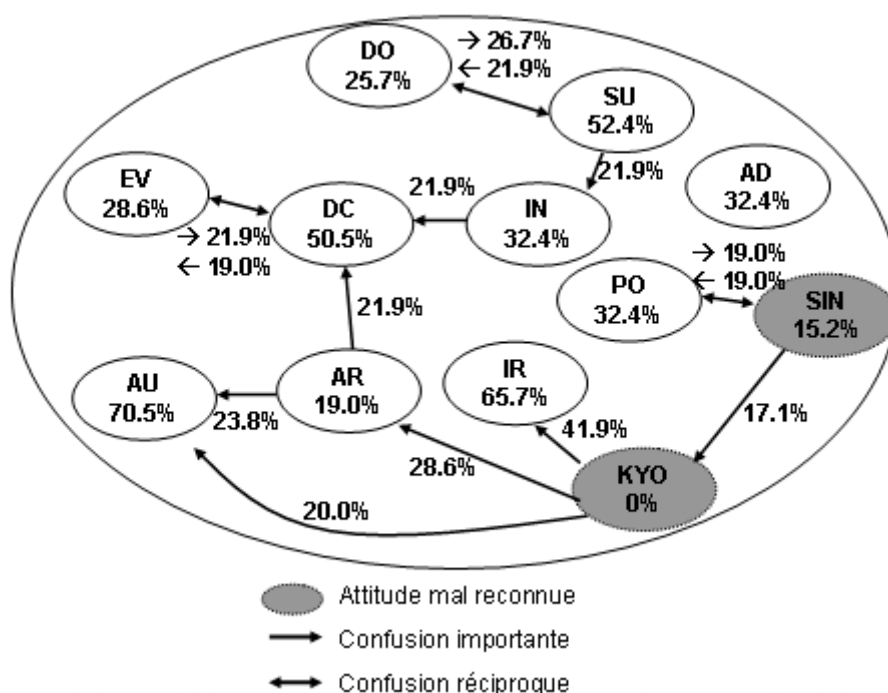


Figure 66 : Graphe de Confusion pour 15 auditeurs français (niveau 0): pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO (doute-incrédulité), EV (évidence), EX (exclamation de surprise), IR (irritation), KYO (kyoshuku), PO (politesse-simple), QS (question-simple) et SIN (sincérité-politesse)

L'attitude de *doute-incrédulité* et *exclamation de surprise* montrent une confusion réciproque. *Exclamation de surprise* a été confondue également avec une *question-simple*. Il est à noter que le degré de cette confusion est différent entre ces attitudes. Le taux d'identification d'*exclamation de surprise* était le double (52,4%) de celui des attitudes confondues (*doute-incrédulité* 21,9%, *question-simple* 21,9%). En revanche, le taux d'identification de *doute-incrédulité* (25,7%) a été inférieur à celui d'*exclamation de surprise* (26,7%). *Déclaration* et *évidence* avaient aussi une confusion réciproque.

Une autre confusion que nous avons identifiée entre *question-simple* et *déclaration* est également observée dans cette analyse. Cette confusion nous semble tout à fait étonnante car *question-simple* en français qui a été identifiée auprès des natifs (Aubergé et al., 2002) montre des caractéristiques de F0 similaires à celles du japonais, c'est-à-dire la montée de F0 systématique à la fin de phrase. La seule différence de la question-simple en deux langues peut être le degré d'intensité en fin de phase. La question japonaise montre une diminution progressive de l'intensité vers la fin, tant dit que la question en français maintient le même degré d'intensité jusqu'à la fin de phrase (Aubergé et al., 2002). Tous ces résultats montrent la possibilité pour les Français de percevoir une question japonaise comme une simple déclaration.

6.3.4.2. Les sujets américains

a- Répartition des attitudes par rapport au hasard

Voici un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 23).

Tableau 23 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	94,7	(ddl : 11, p. < 0.01)
Arrogance-impolitesse (AR)	113,0	(ddl : 11, p. < 0.01)
Autorité (AU)	382,7	(ddl : 11, p. < 0.01)
Déclaration (DC)	227,5	(ddl : 11, p. < 0.01)
Doute-incrédulité (DO)	164,8	(ddl : 11, p. < 0.01)
Évidence (EV)	127,1	(ddl : 11, p. < 0.01)
Exclamation de surprise (EX)	307,8	(ddl : 11, p. < 0.01)
Irritation (IR)	506,8	(ddl : 11, p. < 0.01)
Kyoshuku (KYO)	521,4	(ddl : 11, p. < 0.01)
Politesse-simple (PO)	80,5	(ddl : 11, p. < 0.01)
Question-simple (QS)	180,4	(ddl : 11, p. < 0.01)
Sincérité-politesse (SIN)	30,4	(ddl : 11, p. < 0.01)

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Ce résultat montre donc que des auditeurs américains lorsqu'ils perçoivent des attitudes japonaises, répondent de manière cohérente. Les variations expressives introduites par les attitudes japonaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs américains, même si ces identifications ne correspondent pas forcément au sens de ces attitudes en japonais.

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Nous avons effectué un test de χ^2 sur les résultats moyens entre les longueurs de 1 et 2 mores, de 2 et 5 mores et de 5 et 8 mores, toutes attitudes confondues.

Le test de χ^2 sur les résultat moyens par longueur de 1 more et 2 mores, toutes attitudes confondues, ne donne pas de résultat significatif (ddl = 11, $\chi^2 = 17.76$, $p. < 0.05$).

A contrario, entre 2 et 5 mores, nous avons identifié un effet significatif de la longueur (ddl = 11, $\chi^2 = 35.88$, $p. < 0.05$).

En ce qui concerne la longueur de 5 et de 8 mores, toutes attitudes confondues le test χ^2 ne donne pas d'effet significatif (ddl = 11, $\chi^2 = 12.90$, $p. > 0.05$).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe, cependant cet effet a été identifié uniquement entre les phrases de 2 mores et de 5 mores, ce qui est opposé aux résultats obtenus avec les auditeurs japonais.

c- Effet de l'accent lexical

Quant à l'influence de la position de l'accent lexical sur les attitudes choisies, nous avons effectué un test de χ^2 sur les résultats moyens par position de l'accent avec toujours la même longueur de 8 mores, toutes attitudes confondues. Ce test ne donne pas d'effet significatif (ddl = 33, $\chi^2 = 40.75$, $p. > 0.05$). On peut donc rejeter l'effet de la position de l'accent comme facteur d'influence sur le choix des attitudes par des auditeurs américains.

d- Taux d'identification des attitudes

D'après le taux d'identification des attitudes chez les Américains comparant celle des Japonais (Figure 67), le taux global d'identification des attitudes japonaises chez les Américains étaient nettement inférieur à celui des Japonais et des Français (25% chez les Américains contre 35% chez les Français, 55% chez les Japonais). Autrement dit, les auditeurs américains ont le moins bien identifié les attitudes exprimées. Ce résultat confirme, cette fois aussi, que les natifs perçoivent mieux que les auditeurs non-natifs.

Les résultats montrent que le taux d'identification pour toutes les attitudes sauf les expressions culturelles de la politesse japonais (*kyoshuku* et *sincérité-politesse*) était au dessus du seuil du hasard. Cela signifie que ces dix attitudes ont été identifiées correctement.

Comme les résultats des auditeurs français, les auditeurs américains ont très bien identifié autorité et irritation japonaise. Trois autres attitudes : *exclamation de surprise*, *question-simple* et *déclaration*, sont aussi reconnues au dessus de 30%, mais ce taux d'identification est nettement plus faible que celui des auditeurs français.

Selon la même figure, les auditeurs américains n'ont pas réussi à identifier deux attitudes culturelles : *kyoshuku* et *sincérité-politesse* (*kyoshuku* 0.7%, *sincérité-politesse* 7.9% vs. *arrogance-impolitesse* 15%). Ce faible taux de reconnaissance des Américains est partagée avec les Français, et ce concept de politesse est probablement absent dans la société occidentale, notamment pour la perception de *arrogance-impolitesse* (15% chez

les Américains, 19% chez les Français contre 72% chez les Japonais) même si cette attitude a été reconnue au dessus du seuil du hasard.

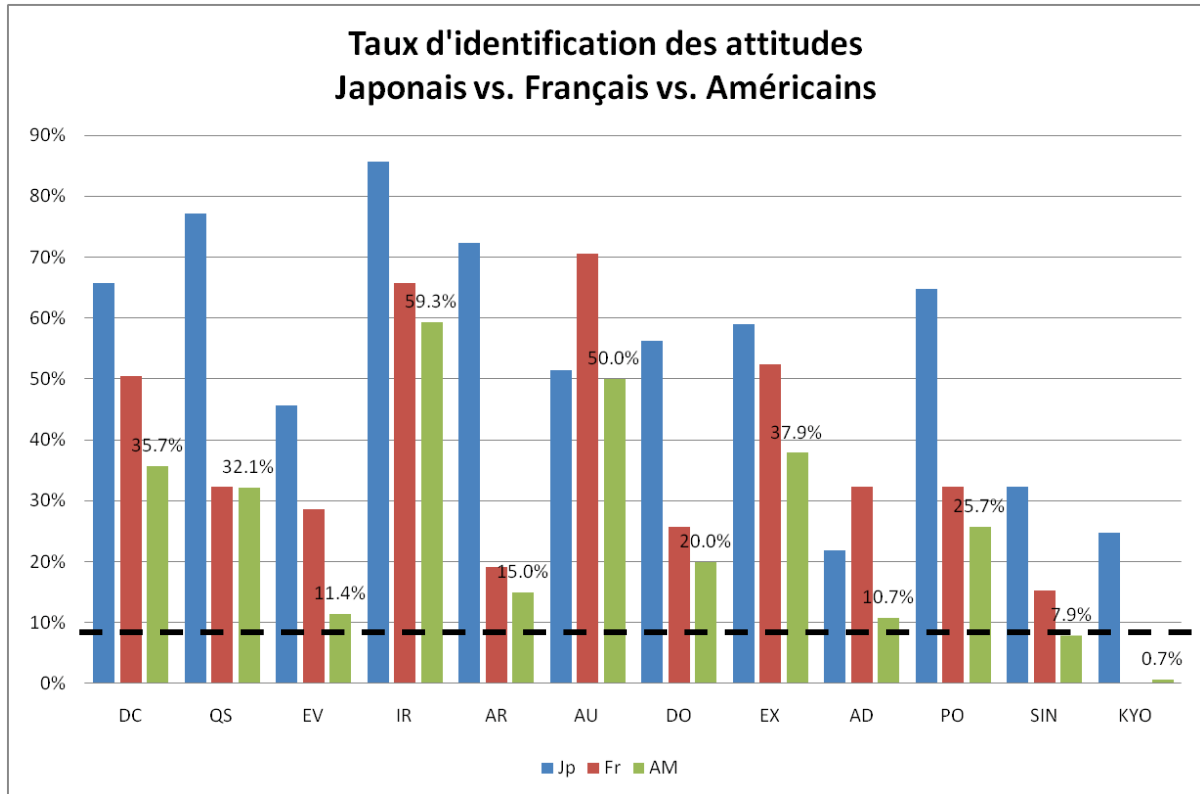


Figure 67 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs américains, français et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : AD (admiration), PO (politesse-simple), KYO (kyoshuku), SIN (sincérité-politesse), AR (arrogance-impolitesse), AU (autorité), IR (irritation), DO (doute-incrédulité), EX (exclamation de surprise), QS (question-simple), EV (évidence) et DC (déclaration)

En ce qui concerne *admiration*, *doute-incrédulité* et *évidence* ayant été reconnues chez les Français, ces attitudes ont été reconnues avec un taux d'identification nettement plus faible par les sujets américains. Ce taux d'identification explique que la réalisation des expressions d'*admiration*, de *doute-incrédulité* et d'*évidence* en anglais peut être différente de celles du Japonais.

Du même que les résultats des Français, les attitudes d'imposition de l'avis du locuteur telles qu'*autorité* et *irritation* ont bien été reconnues. Ces deux attitudes peuvent être considérées comme attitudes potentiellement partagées à la fois au niveau conceptuel et aussi au niveau expressif dans la culture américaine et japonaise.

Nous pouvons également observer un faible taux d'identification de *question-simple* par rapport aux Japonais (32 % chez les Américains contre 77 % chez les Japonais). Ce contraste a été déjà confirmé entre Français vs. Japonais.

e- Comportement perceptif des sujets américains en analyse de correspondance

A la suite de l'observation de l'effet de la longueur et de l'accent lexical, et de la comparaison du taux d'identification de chacune des attitudes, nous allons maintenant évaluer le degré d'identification par l'analyse de correspondance. Plus concrètement nous avons calculé la distance cognitive de 12 attitudes à partir de résultats du test perceptif afin d'observer le comportement perceptif des sujets américains par rapport aux stimuli.

Selon ce test statistique, les quatre premières dimensions expliquent 90% de la variance.

Selon la figure 68, il est clair que le comportement perceptif des auditeurs américains pour l'attitude de *kyoshuku* ressemble au comportement d'*irritation*, et leur réaction perceptive est nettement différente de *sincérité-politesse* ou de *politesse-simple*. Cela implique que les sujets américains ne reconnaissent pas *kyoshuku* comme expression de politesse, et que c'est interprété comme *irritation*.

Par ailleurs, le point perceptif d'*arrogance-impolitesse* est loin du point stimulus de cette attitude, et plutôt proche d'*autorité*. Le comportement pour *doute-incrédulité* et *exclamation de surprise* se situe proche, et nous les mettons dans la même catégorie. De même, les trois expressions de politesse (*kyoshuku*, *sincérité-politesse* et *politesse-simple*) sont proches, et elles forment une catégorie de politesse en dimensions 3 et 4, cependant, ces trois expressions sont dispersées dans la figure en dimensions 1 et 2.

La perception de *politesse-simple*, *autorité* et *irritation* est bonne selon cette figure.

f- Confusion entre attitudes

En comparant le taux d'identification de chacune des attitudes, deux attitudes (*sincérité-politesse* et *kyoshuku*) n'ont pas été reconnues. Selon l'analyse de correspondance, les auditeurs américains ont montré un comportement similaire entre *kyoshuku* et *irritation*. Par contre, la perception pour l'expression de *sincérité-politesse* reste dans la catégorie des expressions de politesse. Quant à *arrogance-impolitesse*, leur comportement perceptif pour cette attitude était similaire d'*autorité*. Par ailleurs, *admiration*, *doute-incrédulité* et *évidence* ont été mal reconnues en comparant aux auditeurs japonais.

Pour observer plus de détails de cette faible reconnaissance et de la tendance de confusions entre les attitudes, nous allons analyser ces résultats à travers des matrices de confusion des données (cf tableau 24).

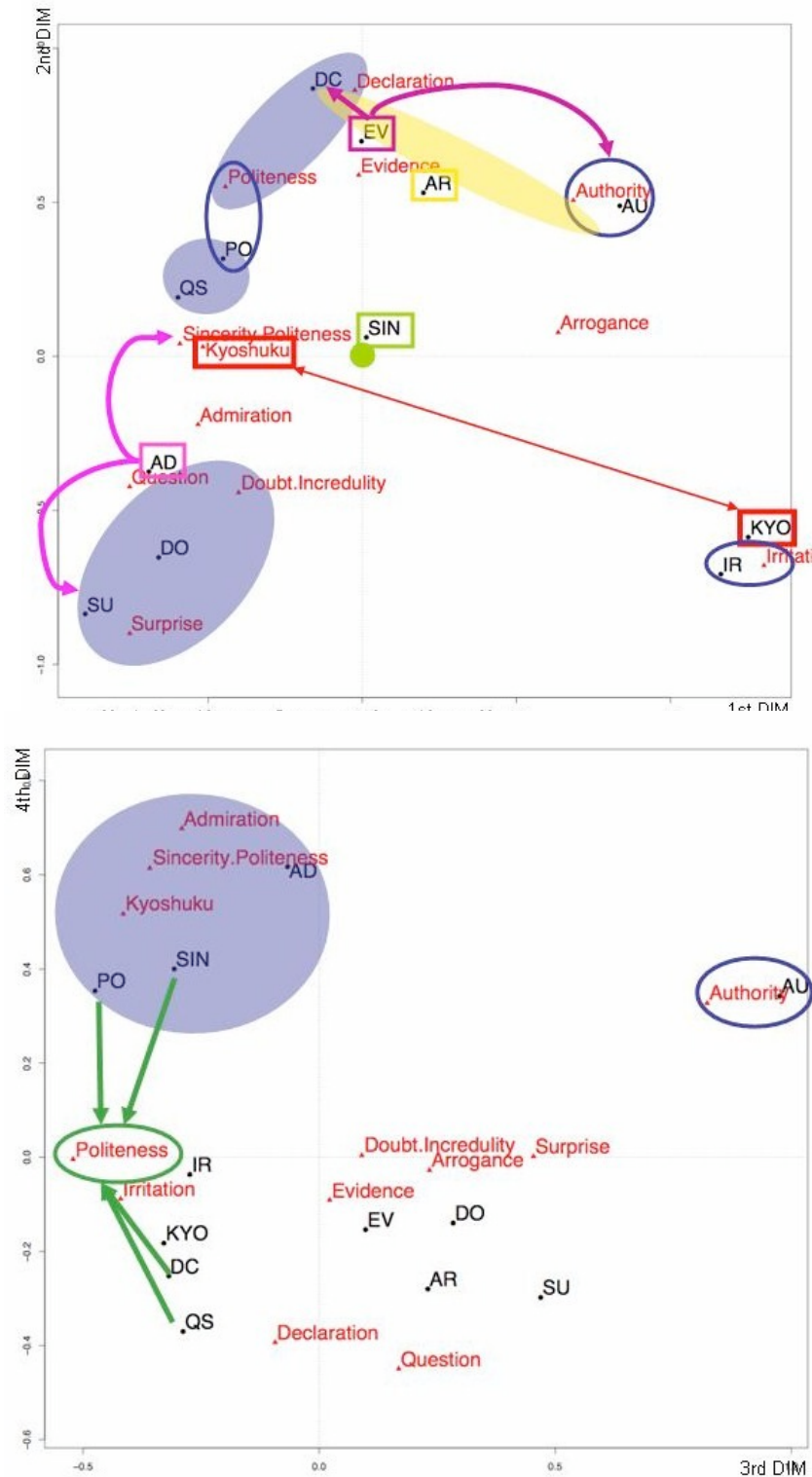


Figure 68 : Comportement perceptif de 20 auditeurs américains pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4. Les étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles.

Tableau 24 : Matrices de confusion des 20 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : AD(admiration), PO(politesse-simple), KYO (kyoshuku), SIN(sincérité-politesse), AR(arrogance-impolitesse), AU (autorité), IR (irritation), DO(doute-incrédulité), EX (exclamation de surprise), QS(question-simple), EV (évidence) et DC (déclaration).

Attitudes perçues	Attitudes présentées											
	AD	AR	AU	DC	DO	EV	SU	IR	KY	PO	QS	SIN
Admiration	11%	0%	0%	1%	4%	3%	4%	2%	0%	7%	2%	7%
Arrogance	1%	15%	19%	6%	7%	6%	1%	11%	18%	4%	2%	6%
Autorité	3%	14%	50%	6%	1%	17%	1%	11%	8%	4%	1%	9%
Déclaration	2%	26%	9%	36%	1%	29%	1%	3%	3%	9%	14%	6%
Doute-incrédulité	9%	6%	1%	3%	20%	3%	12%	7%	1%	5%	4%	11%
Evidence	4%	15%	6%	9%	1%	11%	2%	0%	4%	7%	6%	7%
Surprise	25%	2%	1%	2%	21%	2%	38%	6%	2%	2%	1%	1%
Irritation	1%	6%	12%	1%	2%	3%	0%	59%	59%	2%	1%	15%
Kyoshuku	10%	2%	0%	4%	5%	1%	2%	0%	1%	9%	6%	12%
Politesse	9%	4%	1%	26%	3%	14%	4%	0%	1%	26%	23%	14%
Question	8%	6%	1%	1%	29%	6%	32%	0%	4%	9%	32%	4%
Sincérité-Politesse	18%	1%	1%	5%	6%	4%	3%	1%	0%	15%	6%	8%

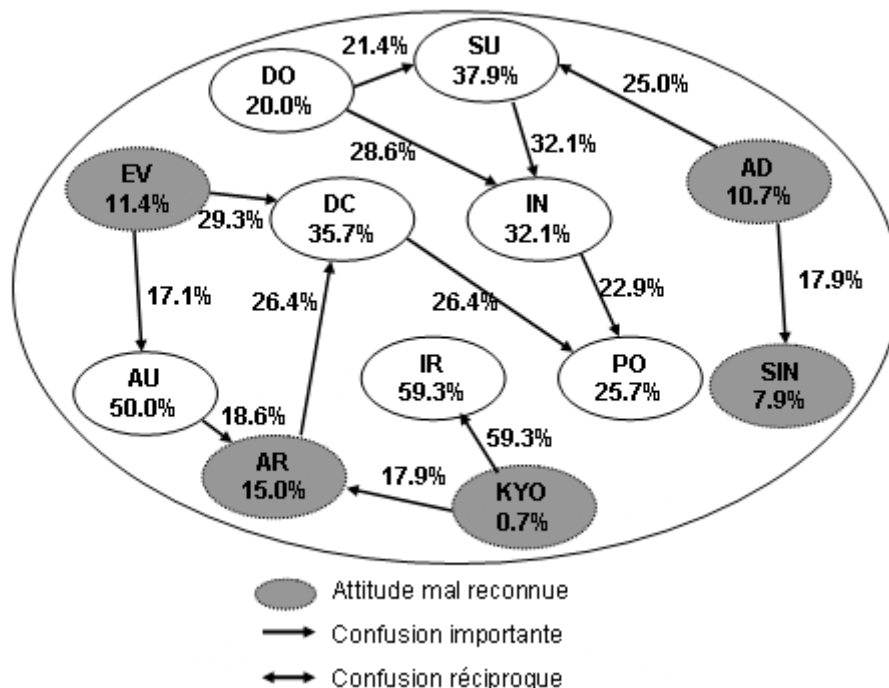


Figure 69 : Graphe de Confusion pour 20 auditeurs américains (niveau 0) : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Abréviations : AD (admiration), AR (arrogance-impolitesse), AU (autorité), DC (déclaration), DO(doute-incrédulité),

EV(évidence), *EX* (exclamation de surprise), *IR* (irritation), *KYO* (*kyoshuku*), *PO* (*politesse-simple*), *QS* (*question-simple*) et *SIN* (*sincérité-politesse*)

La figure 69 représente le graphe de confusion schématique pour les auditeurs américains. D'après ce graphe, deux attitudes : *irritation* et *politesse-simple* ont été reconnues sans aucune confusion particulière. Cependant, le taux d'identification de *politesse-simple* reste très faible (25.7%). Le taux de reconnaissance pour *irritation* est très élevé, et ce phénomène est observé en effet chez tous les groupes que nous avons testés. Alors, cela nous laisse supposer que cette attitude est probablement partagée entre les trois cultures distinctes (japonais, français et américain)

Les auditeurs américains ont eu du mal à reconnaître les deux attitudes culturelles de politesse : *kyoshuku* et *sincérité-politesse* comme les résultats des auditeurs français. L'attitude de *kyoshuku* a été reconnue comme *arrogance-impolitesse* (17,9%), mais surtout comme *irritation* (59,3%) confirmant les résultats de l'AC. Ce similaire comportement entre Français et Américains pour *kyoshuku* implique que cette attitude peut être absente dans la société occidentale, et que une telle qualité de voix n'est pas liée avec l'expression de la politesse ni en anglais et ni en français. Par ce fait, cette expression japonaise renvoie une notion d'énervement, d'arrogant ou d'autoritaire chez les occidentaux, et cette attitude peut être aussi un des cas typique de « faux-amis » prosodique chez les Américains.

Quant à l'attitude de *sincérité-politesse*, cette attitude n'a pas été reconnue, mais cela n'a montré aucune tendance de confusion particulière (voir la distribution de réponse pour cette attitude sur le tableau 24). Cela explique peut-être que cette expression exprimée par la voix soufflée n'est pas liée avec les 12 attitudes présentées, et que cette expression est affectée avec la séduction ou l'intimité en anglais américain par cette qualité de voix (Pavlenko, 2002). Pour le moment, cette attitude reste comme attitude inconnue chez les auditeurs américains.

Enfin, l'attitude d'*arrogance-impolitesse* a été confondue avec *déclaration* et *autorité*. Rappelons que les auditeurs français ont montré aussi leur faible reconnaissance (19%), par contraste du résultat des sujets japonais (72.4%). Nous pouvons constater par ces résultats qu'*arrogance-impolitesse* peut être aussi un des « faux-amis » chez les occidentaux.

En ce qui concerne *admiration* et *évidence*, ces deux attitudes qui ont été reconnues chez les auditeurs japonais et français, ont été mal identifiées par les auditeurs américains. *Admiration* a eu des confusions importantes avec *sincérité-politesse* et *exclamation de surprise*, et *évidence* avec *autorité* et *déclaration*.

Cinq autres attitudes ont été reconnues (*déclaration*, *question-simple*, *autorité*, *doute-incrédulité* et *exclamation de surprise*) avec confusion.

D'abord, deux attitudes *déclaration* et *question-simple* ont été tous les deux confondue avec *politesse-simple*. Curieusement le taux d'identification de *question-*

simple chez les non-natifs (Américains, 32,1 % et Français, 32,4 %) est nettement plus faible par rapport à celui des natif (Japonais, 77,1 %). Un autre point important à noter est que les sujets américains n'ont pas montré la confusion entre *déclaration* et *question-simple* contrairement aux auditeurs français. Cela veut donc dire que ce phénomène est un cas particulier chez les Français.

Doute-incrédulité et *exclamation de surprise* ont montré une confusion importante avec *question-simple*. *Doute-incrédulité* a montré également une autre confusion avec *exclamation de surprise*, pourtant les sujets américains n'ont pas confondu entre *doute-incrédulité* et *exclamation de surprise* d'une manière réciproque comme chez les auditeurs français.

Quant à l'attitude d'*autorité*, les auditeurs américains ont relativement bien identifié cette attitude (50 %), mais cette attitude a montré une légère confusion avec *arrogance-impolitesse*.

6.3.5. Conclusion

L'objectif global de ce chapitre était de découvrir des attitudes prosodiques partagées, mal décodées ou inconnue d'une culture à une autre en manipulant une série des attitudes de trois langues distinctes. Il s'agit de la langue japonaise, anglaise et française.

L'objectif de cette section de la thèse était de mesurer perceptivement le « décalage » de comportement des sujets francophones et anglophones par rapport aux celui des Japonais. Selon l'hypothèse sur laquelle ce travail est basé (Aubergé, 2002), les attitudes semblent construites pour et par la langue qui les manipule, ce qui a pour conséquence que des attitudes peuvent ne pas exister d'une langue à l'autre. Dans cette perspective socioculturelle des attitudes du locuteur, les morphologies prosodiques propres aux attitudes d'une langue donnée peuvent être mal décodées ou ambiguë dans la langue de non-natifs. Le japonais, le français et l'anglais américain présentent des caractéristiques de distance culturelles intéressantes, en particulier en ce qui concerne la politesse qui existe en langue occidentale (anglais et français) et qui existe sous plusieurs catégories en japonais.

Pour l'objectif ci-dessus, nous avons utilisé 12 attitudes représentatives de la langue japonaise qui ont été validées auprès des 15 auditeurs japonais dans le chapitre 3. Avec ce corpus, nous avons effectué le même test perceptif avec ces 12 attitudes japonaises auprès des sujets américains et français

Une série du test khi deux relève les points suivants. D'abord, la distribution des attitudes est différente du hasard même pour les auditeurs français et américains qui n'ont pas du tout la compétence de la langue japonaise. Second point important est que la position de l'accent lexical du japonais qui est réalisé par une forte amplitude de fréquence fondamentale n'a pas donné d'influence significative pour le comportement perceptif des auditeurs de trois groupes. Ce résultat est bien intéressant surtout pour les

auditeurs français car les contours de F0 sont nettement différents selon la position des accents lexicaux, cela n'a pas influencé le choix des attitudes parmi les auditeurs français qui ne possèdent pas d'accent lexical et pour qui la variation de F0 est reliée à la prosodie linguistique ou affective (Aubergé et al., 1997). En revanche, ce test statistique montre un effet significatif de longueur des énoncés pour les Japonais, ainsi que pour les Français et les Américains. Concernant les sujets français, nous avons observé un effet significatif de la longueur entre les mots isolés monomoraïques et dimoraïques, et entre les phrases de 5 mores et de 8 mores. Quant aux sujets américains, l'influence de la longueur a été observée uniquement entre les phrases de 2 mores et de 5 mores. Les Japonais aussi ont eu une influence significative de la longueur entre 2 à 5 mores, mais aussi entre 5 à 8 mores qu'on trouve un effet de la longueur sur la distribution des attitudes.

Dans un premier temps les analyses perceptives menées sur des auditeurs japonais valident globalement la qualité du corpus, alors qu'un test perceptif sur ces mêmes attitudes, appliqué à des auditeurs de langue maternelle française et anglaise non exposés au japonais sont très largement en dessous de ce score, certaines attitudes culturelles sont pas identifiées, confirmant ainsi notre hypothèse initiale.

Nous avons déjà observé dans le chapitre 3 que les auditeurs japonais comportent différemment entre trois attitudes (*question-simple*, *exclamation de surprise* et *deux*) et les autres attitudes selon l'analyse de correspondance. Les sujets japonais distinguent clairement ces trois attitudes par rapport aux autres selon l'intention de locuteur qui réclame (ou qui ne réclame pas) la réaction de l'interlocuteur, liée avec l'intonation montante (ou descendante) en fin de phrase. Nous observons également des attitudes de politesse (*kyoshuku*, *sincérité-politesse*, *politesse-simple*) se regroupent les uns et les autres, et quatre attitudes (*autorité*, *arrogance-impolitesse* et *irritation*) se situent perceptivement très proche.

A contrario, selon le même test statistique, le comportement perceptif des auditeurs français pour l'attitude de *kyoshuku* se situe près d'*arrogance-impolitesse* et de *autorité*, et loin de *sincérité-politesse* ou de *politesse-simple*. Quant au comportement perceptif des auditeurs américains, leur comportement pour l'attitude de *kyoshuku* ressemble au comportement pour *irritation*, et leur réaction perceptive est nettement différente de *sincérité-politesse* ou de *politesse-simple*. Cela implique que les sujets américains et français ne reconnaissent pas *kyoshuku* comme expression de politesse, et que c'est interprété comme *irritation*, *arrogance-impolitesse* ou *autorité*.

Un autre point intéressant est que le comportement des sujets français pour *question-simple* se situe loin de *exclamation de surprise* et de *doute-incrédulité*, mais plutôt proche de *sincérité-politesse* et *politesse-simple*. Ce comportement des Français est très différent du celui des Japonais et des Américains, et cela veut dire que les seuls Français n'entendent pas forcément la phrase interrogative japonais, en montrant la confusion avec la phrase affirmative.

Selon l'analyse de confusion des attitudes perçues, l'attitude de *kyoshuku* a été reconnue parmi les sujets japonais ayant une confusion réciproque avec *sincérité-politesse*. Ces deux attitudes sont socialement reliées par la notion d'humilité qu'elles impliquent fortement face à un supérieur sociale. C'est l'utilisation pragmatique qui permet de choisir entre ces deux catégories.

En revanche, cette attitude n'a pas été identifiée du tout (0 % d'identification pour les sujets français et 0,7% pour les sujets américains) chez les auditeurs occidentaux.

Les auditeurs français l'ont perçu comme arrogance ou autorité, et les sujets américains, quant à eux, ils l'ont reconnu comme arrogance ou irritation. Ce résultat confirme le comportement perceptif pour ces attitudes décrites par l'analyse de correspondance.

Si cette confusion n'est pas observée globalement pour les résultats des sujets japonais (*irritation*, *autorité* et *arrogance* ne sont pas confondues avec *kyoshuku*), il est très important de noter qu'elle est systématique pour nos auditeurs occidentaux, et ce similaire comportement entre Français et Américains pour *kyoshuku* implique qu'une telle qualité de voix rauque n'est pas liée avec l'expression de la politesse en anglais et en français. Pour cette raison que cette expression japonaise renvoie une notion d'énervement, d'arrogant ou d'autoritaire chez les occidentaux, et cette attitude peut donc être un des cas typique de « faux-amis » prosodique chez les Américains et les Français.

Un autre « faux-amis » potentiel que nous avons observé est l'attitude d'*arrogance-impolitesse*. Cette attitude a été reconnue chez les auditeurs japonais avec un taux de reconnaissance très élevé (72,4 %) n'ayant aucune confusion particulière.

En revanche, les auditeurs américains ne l'ont pas reconnu, et ils l'ont perçu comme simple déclaration. Les auditeurs français ont montré aussi leur faible reconnaissance (19%) ayant des confusions avec *autorité* et *déclaration*.

A propos de l'attitude de *sincérité-politesse*, cette attitude a été reconnue chez les sujets japonais ayant des confusions à l'intérieur des expressions de politesse (par exemple, *politesse-simple* et *kyoshuku*). Les auditeurs français ne l'ont pas reconnu, mais perçu comme expression de politesse. Cependant les américains, quant à eux, n'ont pas reconnue cette attitude, mais cela n'a montré aucune tendance de confusion particulière. Il est à noter que cette attitude est produite avec la qualité de voix soufflée, et cette qualité de voix largement utilisée en anglais américaine pour exprimer l'intimité sexuelle (Pavlenko, 2005). Cela explique pourquoi la sincérité-politesse en japonais n'affecte avec aucune des 12 attitudes présentées chez les Américains, et cette attitude reste comme attitude inconnue, selon les résultats du test, chez les auditeurs américains.

Enfin *politesse-simple* a été reconnue parmi les trois groupes d'auditeurs. Les seuls auditeurs français ont montré une confusion importante, mais cette confusion reste dans le cadre d'une expression de politesse. Cette attitude de courtoisie nous semble être potentiellement une des attitudes universelles.

En dehors des expressions liées avec la politesse, irritation a été reconnue très bien chez les auditeurs de trois groupes sans aucune confusion importante. Cette attitude est aussi partagée parmi les trois groupes d'auditeurs autant au niveau conceptuel qu'au niveau d'expressivité.

Un autre point important à noter que les seuls sujets français font une confusion importante entre *déclaration* et *question-simple*, ce qui n'est pas le cas des japonais et des Américains. Cette confusion n'est pas à notre connaissance mentionnée dans la littérature. Dans les travaux précédents, la question-simple en français montre la caractéristique de F0 similaire que celle du japonais, c'est-à-dire la montée de F0 systématique à la fin de phrase (Aubergé et al., 2002). La question-simple en japonais est aussi caractérisée par la montée de F0 en fin de phrase. La seule différence de la question-simple en deux langues peut être le degré d'intensité en fin de phase. La question japonaise montre une diminution progressive de l'intensité vers la fin, tant dit que la question en français maintient le même degré d'intensité jusqu'à la fin de phrase (Aubergé et al., 2002). Tous ces résultats impliquent l'explication de cette perturbation pour les Français de percevoir, parfois, une question japonaise comme une simple déclaration.

6.4. Attitudes anglaises perçues par des Français, des Américains et des Japonais

6.4.1. Objectif du test perceptif

L'objectif de ce test est d'évaluer les comportements perceptifs des sujets non britanniques face à des attitudes anglaises britanniques. Il nous semble particulièrement intéressant de soumettre à un même test perceptif trois groupes culturels différents : des Français riches d'une langue proximité géographique et culturelle, mais dont la langue n'a pas les mêmes racines ; des Américains, qui eux partagent la même langue mais dont la culture est singularisée et enfin des Japonais dont ni la langue ni la culture ne se rapprochent et pour qui l'anglais britannique représente le statut de langue étrangère.

6.4.2. Choix des attitudes de l'anglais

Nous avons cité directement un ensemble de 11 attitudes représentatives de l'anglais (voir Diaferia, 2002 pour les détails d'élaboration du corpus). Il s'agit de : *doubt-incredulity* (doute-incrédulité), *evidence* (évidence), *surprise* (exclamation de surprise), *Command-authority* (autorité), *irritation*, *scorn-aloofness* (mépris) *Sarcastic irony* (ironie sarcastique), *Seduction* (séduction), *Politeness* (politesse), *declaration* (déclaration) et interrogation (question-simple). Certaines de ces attitudes sont spécifiques ou particulières à la culture occidentale, notamment celle de séduction de la part des

hommes qui peuvent être difficile à éliciter en japonais. De plus, l'ironie sarcastique, la politesse ou le mépris peuvent aussi être les attitudes qui possèdent les aspects culturels forts.

Concernant la définition de chacune des attitudes sélectionnées en langue maternelle des sujets, cf. *supra*.

6.4.3. Composition du Corpus

Les phrases que nous avons sélectionnées sont aussi identiques que celles que nous avons employées pour les sujets anglais.

Tableau 25 : *Corpus des attitudes anglaises : 12 phrases de longueur variée avec les différentes positions de l'accent lexical marquées par un astérisque.*

Syllabe	Phrase	Structure
2	Ma [*] ry	mot isolé
2	Beware [*]	mot isolé
2	Men [*] work [*]	N + V
2	Man [*] counts [*]	N + V
5	Li [*] beralizes	mot isolé
5	Unpre [*] cedented	mot isolé
5	Interroga [*] tion	mot isolé
5	Man [*] commu [*] nicates	N + V
5	A man [*] ma [*] nages	Det + N + V
5	A new [*] man [*] ar [*] gued	Det + Adj + N
5	A new [*] hu [*] man counts [*]	Det + Adj + N
5	Tele [*] phony works [*]	N + V

6.4.4. Méthodologie expérimentale

6.4.4.1. Stimuli

Nous avons ainsi choisi exactement les mêmes phrases qui ont été utilisées lors du test de validation.

12 énoncés différents produits selon 11 attitudes, soit un nombre total de 132 stimuli ont été utilisés pour les tests de perception suivants.

6.4.4.2. Matériel

Le test perceptif a été effectué sur l'interface du logiciel « Revolution ». Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 11 attitudes possibles et les définitions des attitudes présentées. Cette interface a été créée en langue maternelle du sujet³¹ (japonais, français et anglais).

³¹ Les interfaces en d'autres langues sont présentées en annexe.

6.4.4.3. Procédure

Le test perceptif a été conduit sujet après sujet, individuellement. Lors du test, le sujet est assis en face d'un ordinateur ; et l'ordre de présentation des stimuli est aléatoire et différent pour chaque sujet. Pendant le test, le sujet doit choisir quelle attitude et a perçue dans un choix fermé des 11 attitudes possibles. Le sujet ne peut écouter qu'une seule fois chaque stimulus avant de choisir l'attitude qu'il a perçu afin d'éviter que le sujet s'entraîne au cours du test.

Lors du commencement du test perceptif, on a demandé au sujet de lire la consigne affichée sur l'écran. Ensuite, on a expliqué brièvement la tâche et la consigne, puis la définition des 11 attitudes dans le but d'habituer le sujet à la voix du locuteur. Enfin à la fin du test perceptif, nous avons interrogé le sujet à propos des attitudes afin de savoir quelles attitudes étaient difficiles à distinguer ou s'il avait perçu d'autres attitudes qui n'étaient pas proposées.

6.4.4.4. Sujets

Un nombre total des auditeurs qui ont participé à ce test perceptif est de 84 :

34 auditeurs japonais (14 sujets masculins et 20 sujets féminins)

30 auditeurs français (14 sujets masculins et 16 sujets féminins)

20 auditeurs américains (9 sujets masculins et 11 sujets féminins).

L'âge moyen est de 19,8 pour les auditeurs japonais, 22,5 pour les auditeurs français et 25,1 pour les auditeurs américains. Aucun sujet n'a réclamé le trouble d'écouter lors du test.

En ce qui concerne les auditeurs japonais, ils sont tous Japonais natifs venant de régions variées. Ils habitent tous au Japon pour toujours, sans jamais séjourner à l'étranger. Ils ont appris la langue anglaise, mais leur niveau de l'anglais n'est pas suffisant pour dire ce qu'ils veulent exprimer.

Concernant les sujets français, la plupart sont des étudiants à l'université Stendhal Grenoble III. Ils habitent tous dans la région grenobloise et ils connaissent plus ou moins l'anglais, mais leur niveau de l'anglais n'est pas suffisant pour communiquer aisément en anglais.

Tous les auditeurs américains sont des étudiants à Black Hills State University, Spearfish dans l'état de South Dakota aux Etats-Unis. Ils n'ont jamais séjourné en Europe avant la passation du test.

6.4.5. Résultats du test perceptif

6.4.5.1. Les sujets américains

a- Répartition des attitudes par rapport au hasard

A la suite de l'analyse des résultats des Anglais que nous avons observés *supra*, il est maintenant intéressant de regarder les résultats des sujets américains qui partagent la même langue avec les Anglais britanniques, mais qui possèdent une culture différente.

De la même manière que pour les résultats des sujets anglais, nous avons effectué un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 26).

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Nous avons donc confirmé que des auditeurs américains lorsqu'ils perçoivent des attitudes anglaises, répondent de manière cohérente.

Tableau 26 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2	
Command-authority (CO)	354.8	(ddl : 10, p.<0.01)
Declaration (DC)	620.6	(ddl : 10, p.< 0.01)
Doubt-incredulity (DO)	340.4	(ddl : 10, p.< 0.01)
Evidence (EV)	251.9	(ddl : 10, p.< 0.01)
Interrogation (IN)	346.9	(ddl : 10, p.< 0.01)
Irritation (IR)	460.4	(ddl : 10, p.< 0.01)
Politeness (PO)	166.5	(ddl : 10, p.< 0.01)
Sarcastic irony (SI)	129.8	(ddl : 10, p.< 0.01)
Scorn-alloofness (AL)	131.3	(ddl : 10, p.< 0.01)
Seduction (SE)	676.5	(ddl : 10, p.< 0.01)
Surprise (SU)	349.4	(ddl : 10, p.< 0.01)

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Pour observer l'effet de la longueur sur le choix des attitudes, nous avons effectué le test de χ^2 sur les résultats moyens par longueur de 2 syllabes et 5 syllabes, toutes attitudes confondues, ne donne pas de résultat significatif (ddl=10, $\chi^2=9,53$ p. <0.05). Ce résultat démontre donc que l'effet de la longueur de l'énoncé n'existe pas sur le choix des attitudes chez les auditeurs américains.

c- Taux d'identification des attitudes

Ensuite, nous avons calculé le taux d'identification des attitudes en pourcentage afin de comparer le degré d'identification des attitudes anglaises chez les auditeurs américains par rapport aux sujets anglais. Cela nous a permis de regrouper les attitudes. La figure 70 présente l'histogramme du taux d'identification des 11 attitudes anglaises pour les 20 auditeurs américains.

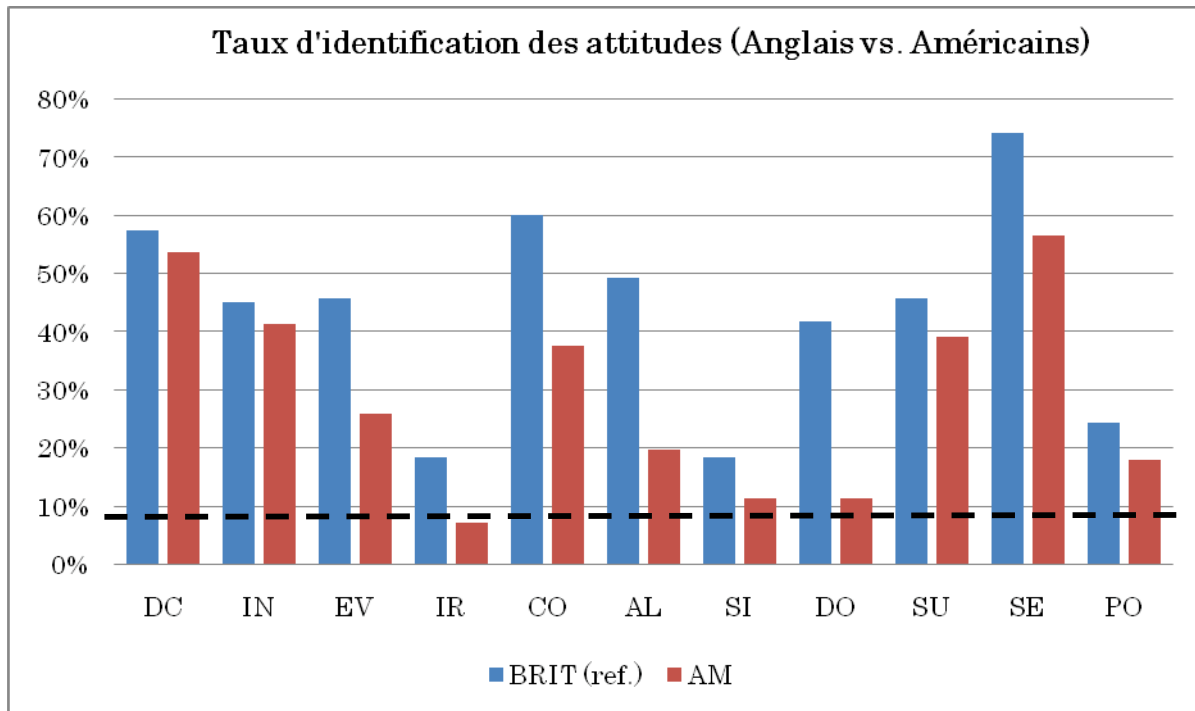


Figure 70 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais et américains. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness).

Les résultats montrent que les sujets natifs perçoivent mieux globalement les 11 attitudes comparant aux auditeurs américains. Rappelons que ce « native effect » a été déjà observé dans les résultats des attitudes japonaises (cf. *supra*), et ce résultat avec les attitudes anglaises confirme de nouveau ce phénomène.

Les auditeurs américains ont généralement reconnues toutes les attitudes sauf *irritation* en montrant le taux de reconnaissance assez faible pour *politeness* et *sarcastic irony*. Comme nous avons vu dans Diaferia (2002), que même les Anglais natifs ont montré leur faible reconnaissance pour l'attitude d'*irritation*, *politeness* et *sarcastic irony*, nous pouvons alors facilement comprendre pourquoi les Américains n'ont pas réussi à percevoir ces trois attitudes. Par contre, la faible reconnaissance de l'attitude de *doubt-incrédulity*, contrairement aux sujets anglais qui l'ont très bien perçu, est un

comportement intéressant. Nous allons analyser ce curieux comportement au plus tard dans le paragraphe dans lequel nous traitons des confusions des attitudes.

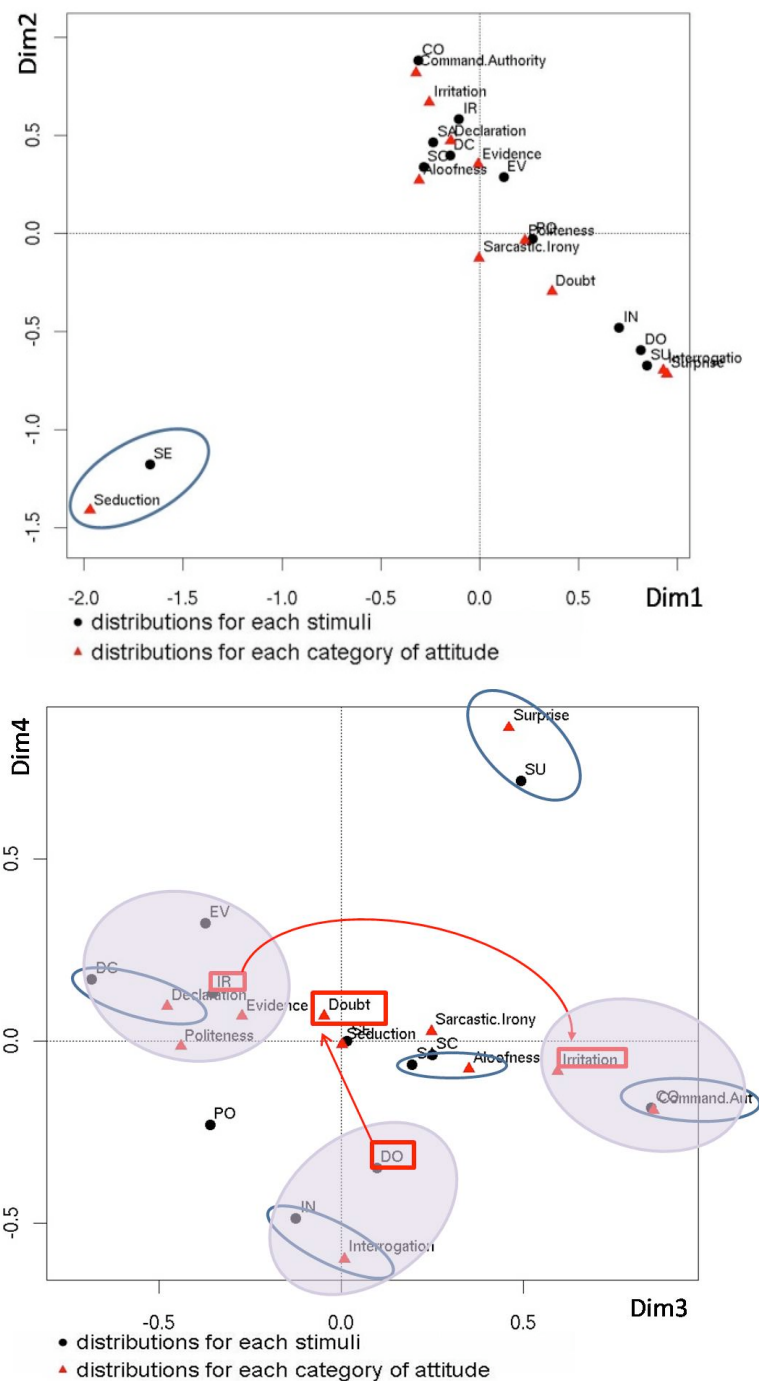


Figure 71 : Comportement perceptif de 20 auditeurs américains pour les 11 attitudes anglaises (la figure du haut présente les dimensions 1 et 2 ; celle du bas les dimensions 3 et 4). Les étiquettes rouges correspondent aux distributions des stimuli, les étiquettes noires indiquent les distributions des catégories d'attitudes. Les cercles bleus indiquent la bonne perception des attitudes, les disques violets indiquent les confusions des attitudes qui sont à l'intérieur.

Un autre point important est que les Américains ont très bien saisi l'attitude de séduction anglaise comme chez les natifs.

d- Comportement perceptif des sujets américains en l'analyse de correspondance

Afin d'observer le comportement perceptif des sujets américains, nous avons calculé la distance des points cognitifs de 12 attitudes par rapport aux points stimuli selon l'analyse de correspondance comme dans pour les résultats des sujets anglais.

Selon ce test statistique, les quatre premières dimensions expliquent 91% de la variance.

Maintenant, voyons figure 71 en haut qui représente leur comportement perceptif en deux dimensions (1^{ère} et 2^{ème} Dimension). Selon le graphe présenté, presque tous les points stimuli, et points perceptifs sauf *seduction* se situent à droite du graphe. Cela signifie que l'attitude de *seduction* a été particulièrement bien discriminée par rapport aux autres attitudes. L'attitude d'*exclamation de surprise* a été également bien discriminée selon la figure 71 en bas. En revanche, le même graphe montre que les points du stimulus pour *irritation* et *doubt-incredulity* sont éloignés de leurs points perceptifs. Cela signifie que ces deux attitudes anglaises ont été mal identifiées par les Américains.

Par ailleurs, cette figure montre trois catégories des attitudes que les auditeurs ont perçues d'une manière similaire. Les attitudes qui imposent l'opinion du locuteur : *irritation* et *command-authority* se forment une catégorie localisée à droite du graphe. *Interrogation* et *doubt-incredulity* se situent plutôt en bas du graphe, et ce group d'attitudes réclame la réaction de la part d'interlocuteur. Enfin la troisième catégorie est composée de *declaration*, *evidence*, et *politeness*.

Nous allons approfondir les analyses de ces phénomènes de confusions entre les attitudes dans le paragraphe suivant.

e- Confusion entre attitudes

Dans les paragraphes précédents, la comparaison du taux d'identification des attitudes, ainsi que la distribution des points perceptifs de ces attitudes selon l'analyse de correspondance ont montré que cinq attitudes : *declaration*, *interrogation*, *command-authority*, *exclamation de surprise* et *seduction*, ont été bien reconnues. Nous nous concentrons donc sur six autres attitudes qui ont montré des confusions importantes.

Comme nous avons mentionné la catégorie des attitudes d'*evidence*, *declaration* et *politeness*, la matrice de confusion (cf. tableau 27) montre également que les sujets américains ont confondu *evidence* avec *declaration* comme les Anglais, mais ils l'ont confondu également avec *politeness*. Concernant l'attitude de *politeness*, cela aussi a été perçu comme attitude qui manque d'intention particulière, c'est-à-dire déclaration ou interrogation. Il est à noter que les Anglais ont perçu cette attitude sans confusion particulière, mais avec le taux d'identification assez faible (24,2%)

Tableau 27 : Matrices de confusion des 20 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC(declaration), IN(interrogation), SU(surprise), DO(doubt-incredulity), EV(evidence), CO(command-authority), IR(irritation), PO(politeness), AL(scorn-aloofness), SE(seduction) et SI(sarcastic irony).

Attitudes perçues	Attitudes présentées										
	CO	DC	DO	EV	IN	IR	PO	SI	SC	SE	SU
Command Authority	38%	2%	2%	6%	1%	8%	3%	10%	7%	3%	2%
Declaration	18%	54%	3%	24%	12%	47%	23%	23%	18%	10%	2%
Doubt	3%	8%	11%	6%	8%	2%	8%	4%	3%	4%	12%
Evidence	8%	11%	7%	26%	7%	16%	12%	13%	13%	4%	2%
Interrogation	0%	1%	40%	0%	41%	2%	20%	3%	2%	1%	19%
Irritation	20%	4%	1%	3%	1%	7%	3%	14%	12%	2%	3%
Politeness	3%	12%	10%	21%	12%	6%	18%	4%	3%	5%	6%
Sarcastic Irony	3%	2%	8%	2%	5%	5%	7%	11%	17%	8%	11%
Scorn Aloofness	7%	2%	3%	3%	2%	5%	3%	15%	20%	6%	3%
Seduction	1%	4%	0%	0%	1%	0%	2%	3%	5%	57%	0%
Surprise	1%	1%	14%	9%	9%	2%	2%	0%	0%	1%	39%

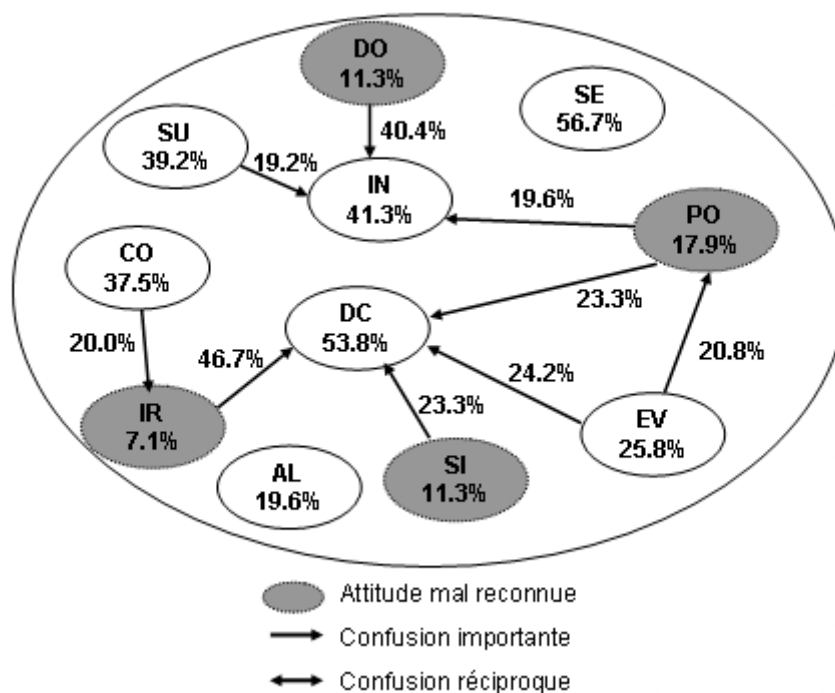


Figure 72 : Graphe de Confusion pour 20 auditeurs américains : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes: DC (declaration), IN (interrogation), SU (surprise), DO (doubt-incredulity), EV (evidence), CO (command-authority), IR (irritation), PO (politeness), AL (scorn-aloofness), SE (seduction) et SI (sarcastic irony).

En ce qui concerne *irritation* et *sarcastic irony*, ces attitudes qui n'ont pas été reconnues au dessus du seuil de hasard (9%) ont été perçues chez les Américains comme une simple déclaration. Puisque les sujets anglais ont reconnu ces attitudes avec le taux d'identification à peine au dessus de seuil du hasard, les résultats des Américains pour ces attitudes semblent liés avec la production de ces deux attitudes qui doivent être peu variées (ou peu intenses).

Quant à la catégorie des attitudes « interrogatives » confirmées par l'analyse de correspondance, *doubt-incredulity* a été perçu comme *interrogation*. Bien que les auditeurs anglais l'aient aussi confondu avec *interrogation*, leur taux d'identification pour *doubt-incredulity* est nettement supérieur à celui des Américains (41,7% pour les Anglais contre 11,3% pour les Américains). Ce décalage du différent taux d'identification pour l'attitude de doute entre deux groupes concernés, implique que la réalisation de *doubt-incredulity* peut être différente chez les Américains, et le doute anglais leur semble une attitude qui n'affirme aucune intention particulière.

Par conséquent, ils perçoivent mal le doute exprimé en anglais britannique, et ils reconnaissent ce doute comme simple interrogation systématiquement.

6.4.5.2. Les sujets français

A la suite des résultats des Américains qui partagent la même langue que les Anglais britanniques, il s'agit cette fois des résultats des auditeurs français qui sont européens qui habitent géographiquement juste à côté des Anglais, et ils partagent beaucoup de points communs culturels par rapport aux Américains.

a- Répartition des attitudes par rapport au hasard

Voici un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 28).

Tableau 28 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2
Command-authority (CO)	694.3 (ddl : 10, p.<0.01)
Declaration (DC)	573.1 (ddl : 10, p.< 0.01)
Doubt-incredulity (DO)	463.0 (ddl : 10, p.< 0.01)
Evidence (EV)	605.1 (ddl : 10, p.< 0.01)
Interrogation (IN)	645.9 (ddl : 10, p.< 0.01)
Irritation (IR)	809.4 (ddl : 10, p.< 0.01)
Politeness (PO)	313.4 (ddl : 10, p.< 0.01)
Sarcastic irony (SI)	253.6 (ddl : 10, p.< 0.01)
Scorn-aloofness (AL)	333.6 (ddl : 10, p.< 0.01)
Seduction (SE)	1099.3 (ddl : 10, p.< 0.01)
Surprise (SU)	361.9 (ddl : 10, p.< 0.01)

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Cela explique donc que les variations expressives introduites par les attitudes anglaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs français.

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Nous avons effectué un test de χ^2 sur les résultats moyens entre les longueurs de 2 et 5 syllabes, toutes attitudes confondues.

Le résultat donne un effet significatif (ddl = 10, $\chi^2 = 73.42$, $p. < 0.05$).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe entre les phrases de 2 syllabes et de 5 syllabes, ce qui est opposé aux résultats obtenus avec les auditeurs américains.

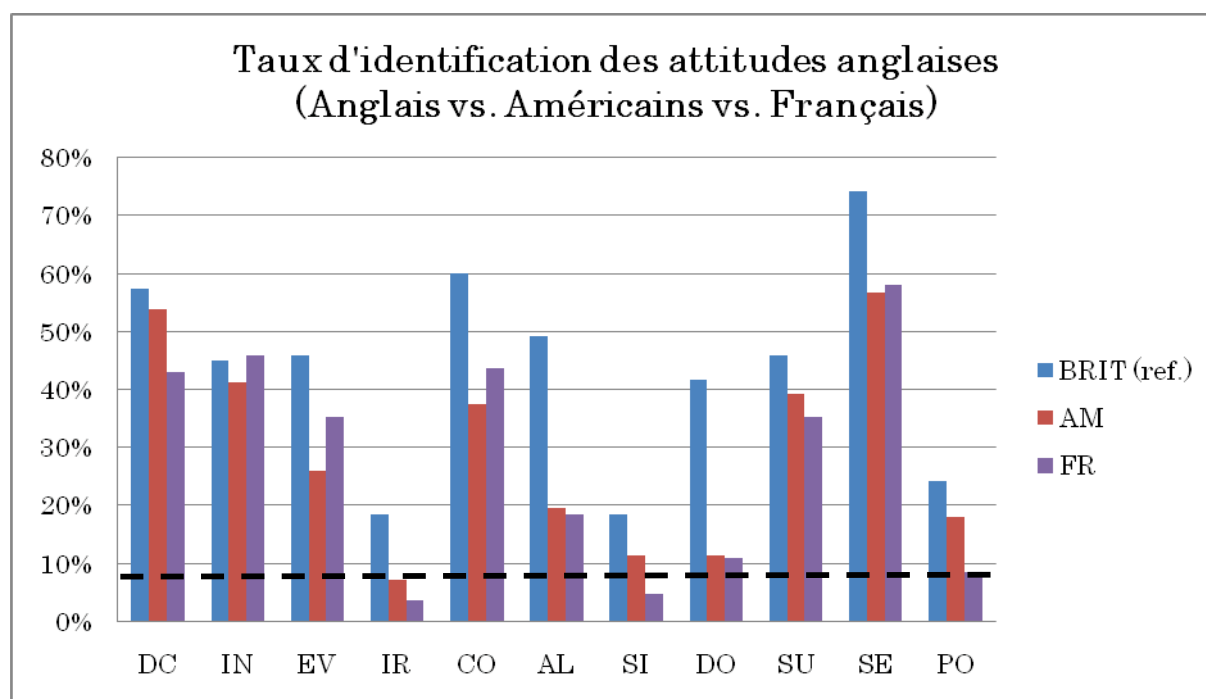


Figure 73 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais, américains et français. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-aloofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness).

a- Taux d'identification des attitudes

D'après le taux d'identification des attitudes en pourcentage chez les Français comparé à celui des anglophones (cf. figure 73), le taux global d'identification des

auditeurs français reste inférieur aux natifs sauf *interrogation* qui marque plus ou moins le même score que les natifs.

Lorsqu'on compare le taux d'identification de chacune des attitudes chez trois groupes d'auditeurs, les auditeurs français n'ont pas réussi à reconnaître *irritation* et *sarcastic irony* comme les sujets américains. De plus, ils n'arrivaient pas à identifier *politeness* (score au dessous du seuil du hasard).

Parmi ces 11 attitudes, nous constatons que quatre attitudes telles que *scorn-aloofness*, *sarcastic irony*, *seduction* et *politeness* sont classées comme les attitudes culturelles. D'abord la politesse de chaque langue est étroitement liée avec la culture de chacun. Deuxièmement, la morphologie de l'attitude de sarcasme par exemple, peut apparaître différemment en japonais (Hamasaki et al., 1999).

En prenant compte de ces aspects culturels, si nous regardons le comportement des Français pour ces quatre attitudes culturelles, il est facile de remarquer que seule l'attitude de séduction est très bien reconnue autant chez les auditeurs américains que chez les français. Par contre, la perception de l'attitude de *scorn-aloofness* (mépris) entre les sujets natifs et les non natifs montre un décalage important. Les auditeurs non natifs montrent un taux de perception nettement plus faible que celui des auditeurs anglais.

Quant à *politeness*, les sujets américains ont mieux perçu cette attitude culturelle par rapport aux sujets français. Cela implique peut-être que l'influence de la langue est plus importante que la distance géographique de ces deux pays pour la perception de la politesse anglaise.

En ce qui concerne *sarcastic irony*, le taux de reconnaissance de tous les groupes incluant le groupe des sujets anglais est faible. Il est donc possible que ce soit l'influence du locuteur au niveau de la production.

b- Comportement perceptif des sujets français en analyse de correspondance

Ensuite, nous avons analysé le comportement perceptif des sujets français par rapport aux stimuli par l'analyse de correspondance.

Tout d'abord, les quatre premières dimensions extraites par cette analyse statistique expliquent 94% de la variance.

La figure 74 en haut représente leur comportement perceptif en deux dimensions (Dimension 1 et 2). Selon le graphe présenté, nous voyons que deux attitudes : *command-authority* et *seduction*, sont localisées dans une position isolée par rapport aux autres attitudes. Cela signifie que ces deux attitudes ont été très bien discriminées. Par contre, le point perceptif d'*irritation* est décalé par rapport au point de stimulus pour l'attitude correspondante. C'est-à-dire, *irritation* a été mal reconnue chez les auditeurs français.

Par ailleurs, cette figure montre globalement deux catégories d'attitudes que les auditeurs ont perçues d'une manière similaire. Le point perceptif pour *declaration*, *evidence*, *irritation*, *politeness*, *sarcastic irony* et *scorn-aloofness* s'est répartie sur la partie gauche du graphe, et l'autre catégorie est composée des attitudes qui réclament la réaction de l'interlocuteur telles que *exclamation of surprise*, *interrogation* et *doubt-incredulity*.

La figure 74 en bas illustre la mauvaise reconnaissance d'*irritation* et de *sarcastic irony* chez les Français d'une part, et la bonne reconnaissance d'*exclamation of surprise*, *seduction*, *command-authority* et *interrogation* d'autre part.

Dans le paragraphe suivant, nous traitons plus de détails sur ce mécanisme de confusion entre les attitudes.

a- Confusion entre attitudes

En comparant le taux d'identification de chacune des attitudes, les auditeurs français n'ont pas réussi à identifier trois attitudes (*irritation*, *politeness* et *sarcastic irony*). Bien que leur taux d'identification pour *doubt-incredulity* et *scorn-aloofness* fût au dessus de hasard, leur degré de reconnaissance pour ces deux attitudes était nettement plus faible que les natifs.

Tableau 29 : Matrices de confusion des 30 auditeurs français. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (*declaration*), IN (*interrogation*), EV (*evidence*), IR (*irritation*), CO (*command-authority*), AL (*scorn-aloofness*), SI (*sarcastic irony*), DO (*doubt-incredulity*), SU (*surprise*), SE (*seduction*) et PO (*politeness*).

Attitudes perçues	Attitudes présentées										
	CO	DC	DO	EV	IN	IR	PO	SI	SC	SE	SU
Command Authority	44%	3%	0%	6%	1%	4%	2%	7%	2%	1%	5%
Declaration	15%	43%	11%	31%	10%	48%	33%	27%	24%	12%	5%
Doubt	0%	3%	11%	1%	12%	1%	8%	2%	2%	2%	5%
Evidence	12%	11%	7%	35%	8%	24%	15%	16%	15%	9%	13%
Interrogation	0%	0%	38%	2%	46%	0%	13%	0%	0%	0%	12%
Irritation	18%	10%	1%	6%	1%	4%	7%	10%	2%	1%	6%
Politeness	1%	6%	8%	10%	5%	5%	8%	5%	8%	6%	2%
Sarcastic Irony	2%	3%	6%	2%	4%	4%	2%	5%	6%	6%	11%
Scorn Aloofness	7%	11%	1%	1%	2%	6%	7%	13%	18%	4%	5%
Seduction	1%	9%	2%	0%	3%	3%	1%	14%	23%	58%	1%
Surprise	0%	0%	15%	4%	9%	1%	4%	1%	1%	1%	35%

En revanche, quatre attitudes (*seduction*, *command-authority*, *interrogation* et *exclamation of surprise*) ont été bien discriminées selon l'analyse de correspondance. Nous avons également identifié deux catégories du comportement des auditeurs pour les attitudes. La première catégorie des attitudes est composée des attitudes « interrogative » : *exclamation of surprise*, *interrogation* et *doubt-incredulity*, et la

deuxième consiste en *declaration*, *evidence*, *irritation*, *politeness*, *sarcastic irony* et *scorn-aloofness*.

Voyons maintenant le taux de confusion pour chaque attitude présenté (cf. tableau 29).

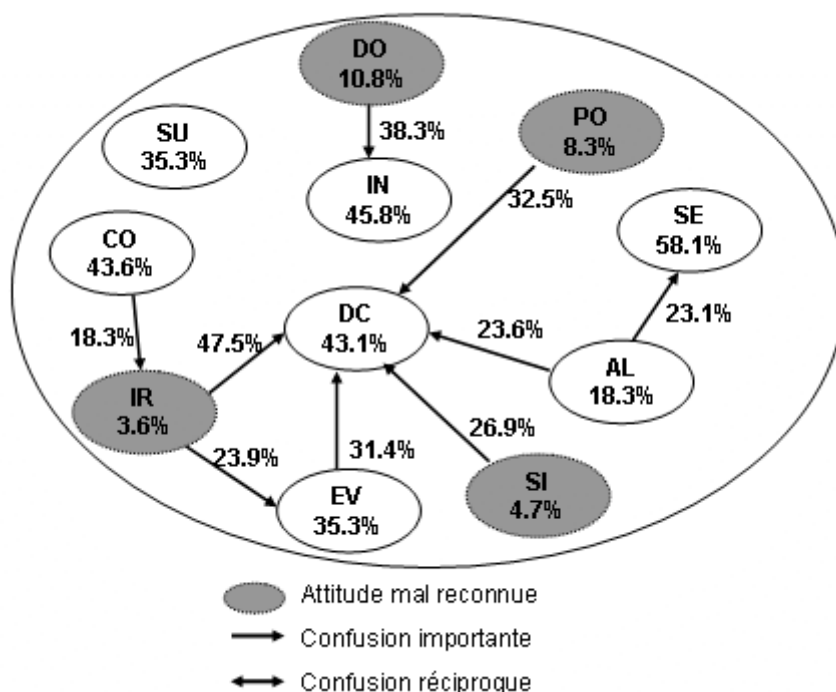


Figure 75 : Graphe de Confusion pour 30 auditeurs français : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes: DC(*declaration*), IN(*interrogation*), EV(*evidence*), IR(*irritation*), CO(*command-authority*), AL(*scorn-aloofness*), SI(*sarcastic irony*), DO(*doubt-incredulity*), SU(*surprise*), SE(*seduction*) et PO(*politeness*).

Les résultats tirés de ces matrices de confusions confirment que trois attitudes : *interrogation*, *exclamation of surprise* et *seduction* ont été reconnues sans aucune confusion particulière. Quant à *command-authority*, elle a été bien perçue aussi en montrant une légère confusion avec *irritation*. Nous ajoutons également *declaration* dans ce groupe des attitudes bien perçues.

Evidence a été confondue avec *declaration*. Contrairement aux auditeurs américains, les Français n'ont pas confondu cette attitude avec *politeness*.

En ce qui concerne *scorn-aloofness*, il a été confondu avec *declaration* et *seduction*. Cette confusion avec séduction est intéressante car ces deux attitudes ne ressemblent pas du tout au niveau cognitif. Ce phénomène n'a pas été observé chez les auditeurs américains.

Quatre attitudes ayant été mal reconnues : *doubt-incredulity*, *irritation*, *politeness* et *sarcastic irony* ont été perçus généralement comme attitude sans intention particulière.

Politeness et *sarcastic irony* ont été perçues comme *declaration*, et *doubt-incredulity* a été perçue comme interrogation.

Seule *irritation* a été perçue comme *evidence* ou *declaration*.

Un autre point important à noter que les auditeurs français n'ont pas montré de confusion entre *declaration* et *interrogation*. Cela implique donc que cette confusion est bien particulière chez les Français pour seulement la déclaration japonaise.

6.4.5.3. Les sujets japonais

A la suite des résultats des occidentaux (Américains et Français), il s'agit cette fois des résultats des auditeurs japonais qui sont éloignées de l'anglais géographiquement et culturellement.

a- Répartition des attitudes par rapport au hasard

Voici un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 30).

Tableau 30 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2	
Command-authority (CO)	585.7	(ddl : 10, p.<0.01)
Declaration (DC)	452.3	(ddl : 10, p.< 0.01)
Doubt-incredulity (DO)	669.5	(ddl : 10, p.< 0.01)
Evidence (EV)	479.1	(ddl : 10, p.< 0.01)
Interrogation (IN)	1150.7	(ddl : 10, p.< 0.01)
Irritation (IR)	819.4	(ddl : 10, p.< 0.01)
Politeness (PO)	382.5	(ddl : 10, p.< 0.01)
Sarcastic irony (SI)	298.9	(ddl : 10, p.< 0.01)
Scorn-aloofness (AL)	358.3	(ddl : 10, p.< 0.01)
Seduction (SE)	285.7	(ddl : 10, p.< 0.01)
Surprise (SU)	847.3	(ddl : 10, p.< 0.01)

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Cela explique donc que les variations expressives introduites par les attitudes anglaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs japonais.

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Nous avons effectué un test de χ^2 sur les résultats moyens entre les longueurs de 2 et 5 syllabes, toutes attitudes confondues.

Le résultat donne un effet significatif (ddl = 10, $\chi^2 = 80.67$, $p. < 0.05$).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe entre les phrases de 2 syllabes et de 5 syllabes, ce qui est opposé aux résultats obtenus avec les auditeurs américains.

c- Taux d'identification des attitudes

Selon la figure 76 qui représente le taux d'identification des attitudes en pourcentage chez les Japonais comparant avec celle des natifs, le taux global d'identification des auditeurs japonais reste inférieur aux natifs sauf *interrogation* et *exclamation of surprise*.

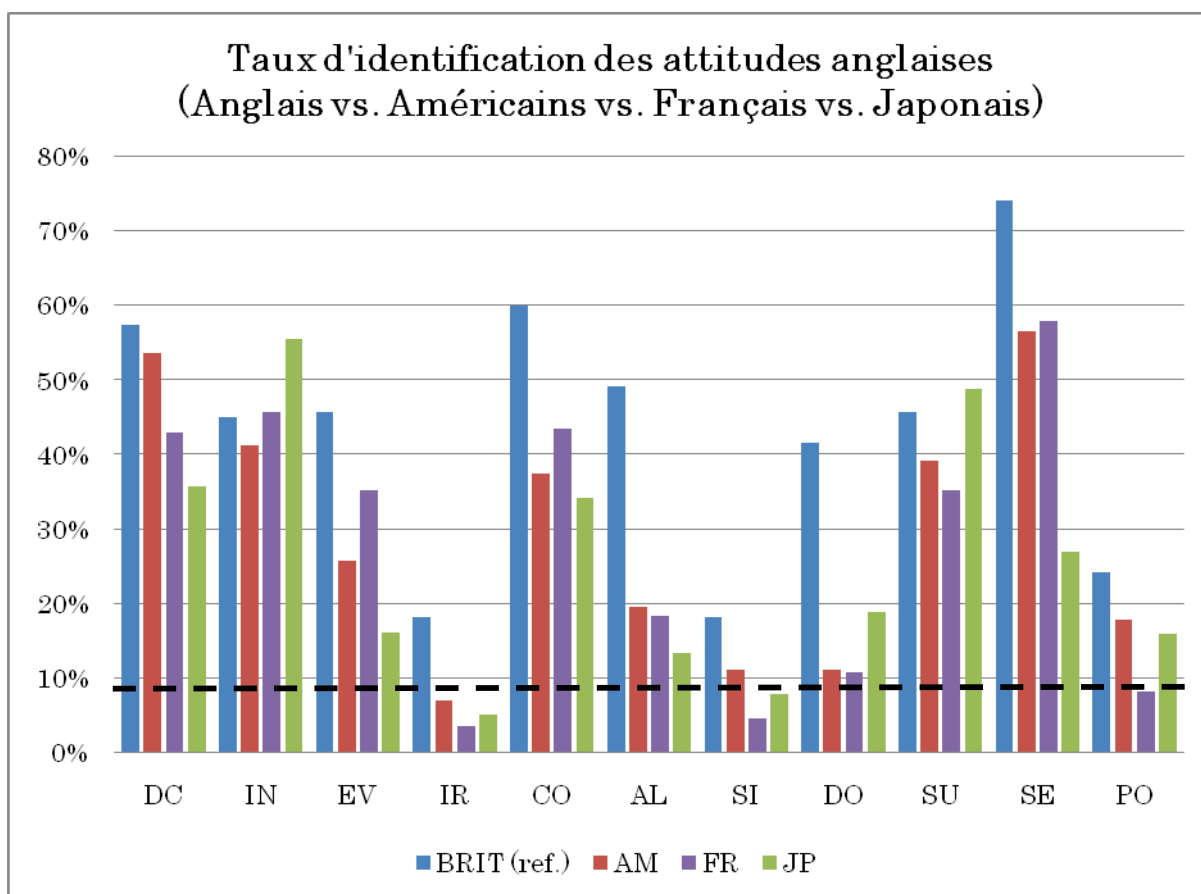


Figure 76 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais, américains, français et japonais. La ligne pointillée indique le seuil du hasard (9%). Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-alooftness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness).

Comme les résultats des sujets occidentaux, les sujets japonais n'ont pas réussi à reconnaître deux attitudes : *irritation* et *sarcastic irony*. Concernant le faible taux

d'identification d'*irritation*, *sarcastic irony* est expliqué par le faible taux de validité (cf. le taux d'identification des natifs). En ce qui concerne *politeness*, les auditeurs japonais l'ont mieux reconnue que les sujets français, mais le taux de reconnaissance reste faible. Par contre le taux d'identification de *scorn-aloofness* des sujets de quatre groupes testés indique clairement que les seuls natifs qui marquent le taux d'identification largement supérieur par rapport aux sujets non natifs, et celui des autres groupes reste juste au dessus du bar du seuil.

Nous mentionnons également deux autres points importants. D'abord, les auditeurs japonais n'ont pas aussi bien reconnu l'attitude d'*evidence* que les sujets occidentaux. Le second point est que *doubt-incredulity* a été mieux reconnu chez les Japonais par rapport aux sujets occidentaux. Rappelons que cette attitude a été une des attitudes plus difficiles à percevoir chez les auditeurs américains et français.

Enfin, l'attitude de séduction qui ne semble peut-être pas exister chez les hommes japonais a été reconnue correctement chez les auditeurs japonais, mais avec le taux d'identification largement au dessous de celui des autres groupes.

d- Comportement perceptif des sujets japonais en analyse de correspondance

Dans ce paragraphe, nous observons la distribution des points perceptifs pour les sujets japonais par rapport aux stimuli en deux dimensions calculées par l'analyse de correspondance.

Tout d'abord, les quatre premières dimensions extraites par cette analyse statistique expliquent 96% de la variance.

La figure 77 en haut représente leur comportement perceptif en deux dimensions (1^{ère} et 2^{ème} Dimension). Selon le graphe présenté, nous voyons que deux attitudes : *exclamation of surprise* et *seduction*, sont localisées dans une position isolée par rapport aux autres attitudes. Cela signifie que ces deux attitudes ont été très bien discriminées. Par contre, le point perceptif d'*irritation* et *sarcastic irony* est décalé par rapport au point de stimulus pour chacune des attitudes correspondantes. C'est-à-dire, ces deux attitudes ont été mal reconnues chez les auditeurs japonais comme deux autres groupes des sujets non natifs (les Américains et les Français).

Par ailleurs, aucune catégorie d'attitudes n'est identifiée visuellement, et la plupart des attitudes se confondent avec déclaration. Ce phénomène nous suggère que les sujets japonais ont eu du mal à identifier la subtilité de variété des attitudes que le locuteur anglais a exprimées.

La figure 77 en bas illustre de nouveau la mauvaise reconnaissance d'*irritation* d'une part, et la bonne reconnaissance d'*exclamation of surprise*, *seduction*, *command-authority* et *interrogation* d'autre part.

Dans le paragraphe suivant, nous analysons des détails de matrices de confusions entre les attitudes.

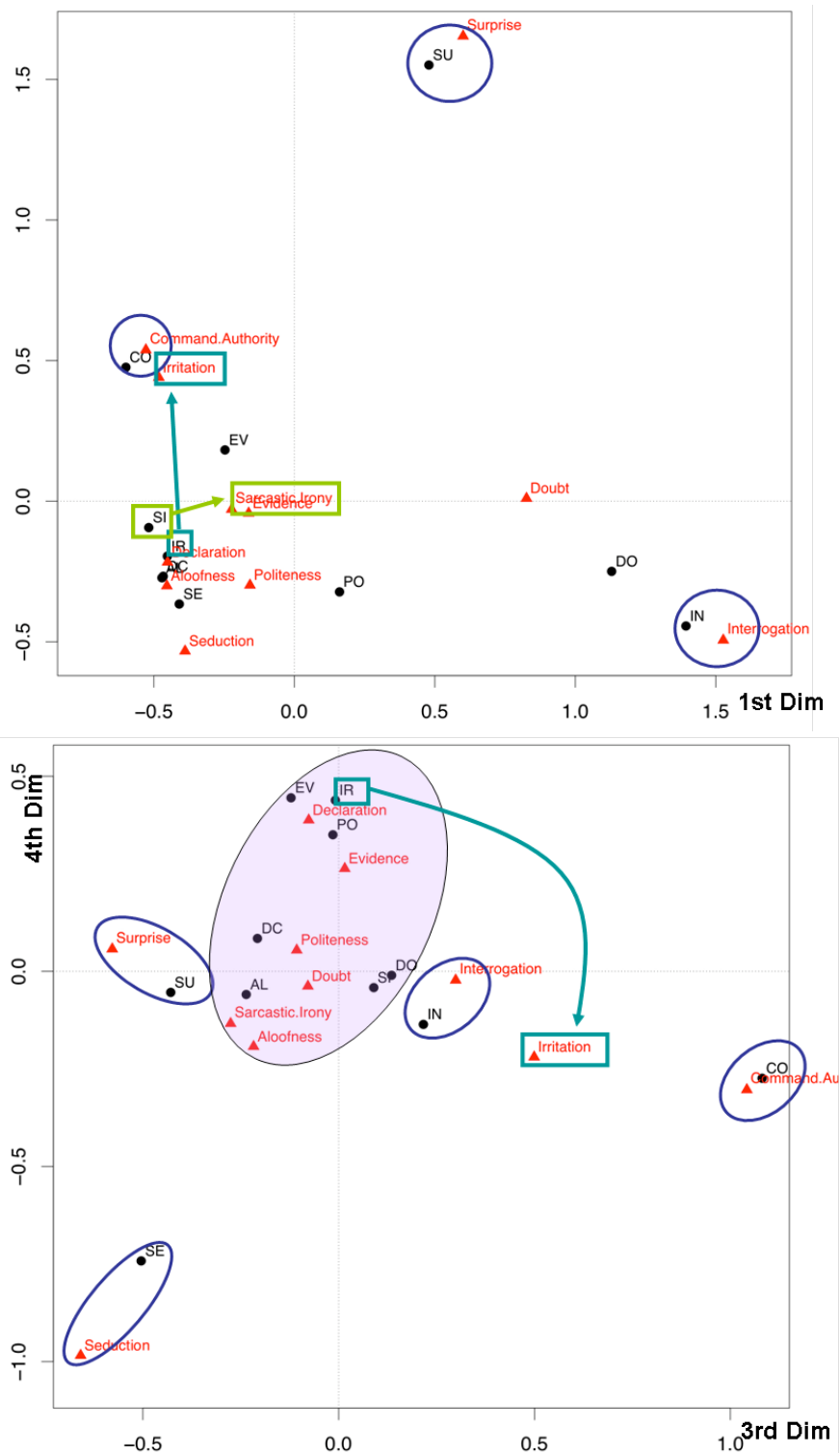


Figure 77 : Comportement perceptif de 34 auditeurs japonais pour les 11 attitudes anglaises (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les étiquettes rouges correspondent aux distributions des stimuli, les étiquettes noires indiquent les distributions des catégories d'attitudes. Les cercles bleus indiquent la bonne perception des attitudes, les disques violets indiquent les confusions d'attitudes qui sont à l'intérieur.

e- Confusion entre attitudes

A la suite de la comparaison du taux d'identification des attitudes chez les auditeurs japonais par rapport aux sujets occidentaux, ainsi par l'analyse de correspondance pour ces données, les auditeurs japonais n'ont pas réussi à identifier deux attitudes (*irritation* et *sarcastic irony*) comme les autres groupes des auditeurs non natifs.

Concernant l'attitude de *politeness*, les auditeurs japonais l'ont mieux reconnue que les sujets français, mais le taux de reconnaissance reste faible. Leur taux d'identification pour *doubt-incredulity* et *scorn-aloofness* était au dessus de hasard, mais leur degré de reconnaissance pour ces deux attitudes était nettement plus faible que les natifs.

L'attitude de *seduction* a été bien identifiée chez les Japonais comme les autres groupes, mais leur taux d'identification était largement au dessous par rapports aux sujets occidentaux. Du même, trois attitudes (*command-authority*, *interrogation* et *exclamation of surprise*) ont été bien discriminées selon l'analyse de correspondance. Nous n'avons identifié aucune catégorie des attitudes similaires. Regardons maintenant le taux de confusion pour chaque attitude présenté (cf. tableau 31).

Ces matrices de confusions confirment que trois attitudes : (*declaration*, *interrogation*, *surprise* et *seduction*) ont été reconnues sans aucune confusion particulière comme dans les résultats des auditeurs français. Toute fois, le taux d'identification de *seduction* reste très faible (27.0%) comparant à celui des Français (58,1%).

Command-authority était une des attitudes très bien reconnues selon l'analyse de correspondance, mais nous voyons dans ce tableau qu'elle a été confondue avec *irritation*. Nous nous rappelons que cette confusion est partagée chez tous les sujets non-natifs, mais les natifs n'ont pas montré cette tendance. Quant à *doubt-incredulity*, elle a été reconnue avec un taux d'identification très faible (18,9%), et perçu comme simple interrogation en plus part des stimuli.

Ensuite, quatre attitudes (cercles remplis de couleur grise en figure 78) ont été mal reconnues. Il s'agit d'*evidence*, *politeness*, *irritation*, *scorn-aloofness* et *sarcastic irony* qui ont été perçu plupart des cas comme attitude sans intention particulière.

C'est-à-dire, toutes ces attitudes sauf *politeness* ont été perçues tous comme simple déclaration, et *politeness*, quant à elle, a été perçu comme une déclaration ou une simple interrogation.

Tableau 31 : Matrices de confusion pour 34 auditeurs japonais. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (declaration), IN (interrogation), EV (evidence), IR (irritation), CO (command-authority), AL (scorn-alooofness), SI (sarcastic irony), DO (doubt-incredulity), SU (surprise), SE (seduction) et PO (politeness).

Attitudes perçues	Attitudes présentées										
	CO	DC	DO	EV	IN	IR	PO	SA	SC	SE	SU
Command Authority	34%	2%	2%	5%	1%	7%	1%	9%	3%	4%	7%
Declaration	16%	36%	6%	37%	2%	48%	31%	30%	32%	15%	2%
Doubt	0%	3%	19%	2%	15%	3%	8%	3%	4%	4%	12%
Evidence	7%	6%	7%	16%	3%	7%	10%	8%	6%	6%	3%
Interrogation	0%	1%	43%	2%	56%	1%	18%	0%	0%	0%	4%
Irritation	26%	8%	1%	5%	0%	5%	3%	14%	8%	5%	12%
Politeness	7%	12%	8%	12%	9%	16%	16%	9%	12%	17%	2%
Sarcastic Irony	4%	8%	5%	5%	2%	5%	7%	8%	12%	11%	7%
Scorn Aloofness	4%	16%	1%	2%	3%	4%	4%	13%	13%	10%	1%
Seduction	0%	7%	2%	1%	4%	2%	0%	6%	10%	27%	0%
Surprise	2%	1%	7%	12%	5%	1%	1%	0%	0%	1%	49%

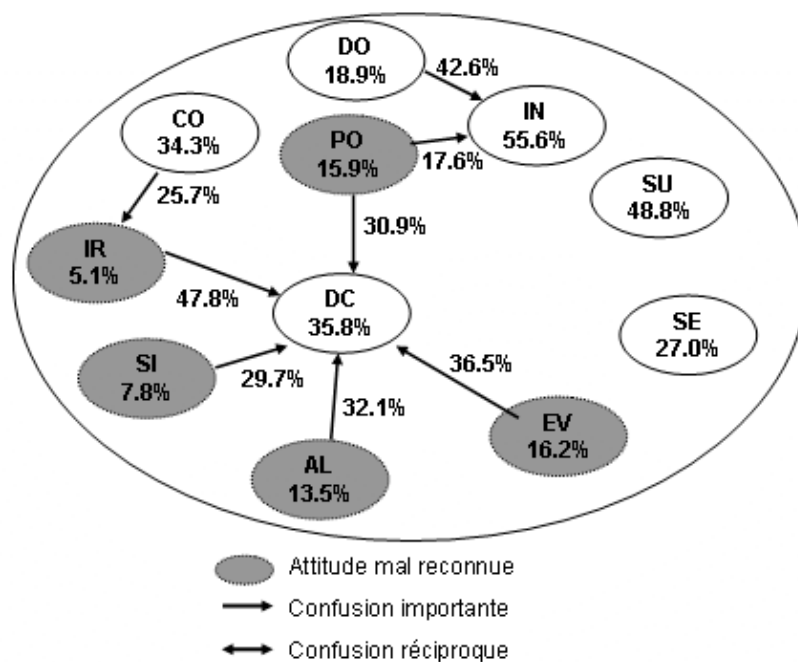


Figure 78 : Graphe de Confusion pour 34 auditeurs japonais : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : DC(declaration), IN(interrogation), EV(evidence), IR(irritation), CO(command-authority), AL(scorn-alooofness), SI(sarcastic irony), DO(doubt-incredulity), SU(surprise), SE(seduction) et PO(politeness).

6.4.6. Conclusion

L'objectif de ce test perceptif avec les attitudes anglaises pour les sujets non natifs est d'évaluer des comportements perceptifs des auditeurs de différents pays pour les attitudes anglaises britanniques qui ont été validé dans les travaux de Diaferia (2002) décrit *supra*. Puisque les attitudes sont acquies dans la société, et elles reflètent donc un aspect culturel fort, il est particulièrement intéressant de voir un décalage au niveau perceptif des Anglais avec les sujets français qui partage beaucoup de points communs culturels avec les Anglais, et avec les sujets américains qui partage la même langue, et enfin avec les sujets japonais qui sont éloignés de cette langue culturellement, et qui reçoivent des informations uniquement par les médias.

En manipulant le corpus des 11 attitudes anglaises britanniques ayant été validé par les Anglais britanniques, nous avons effectué une série du test de perception auprès des auditeurs américains, français et japonais afin de mesurer la distance perceptive entre les attitudes et de voir le décalage de comportement entre ces groupes de sujets.

Une série du test khi deux relève les points suivants. D'abord, la distribution des attitudes est différente du hasard pour tous les auditeurs qui ont participé à ce test. Second point important est qu'il y a un effet significatif de longueur des énoncés pour les Japonais, et les Français, mais cette influence était absente chez les anglophones (les Anglais et les Américains).

Dans un premier temps nous avons récapitulé la validation d'un ensemble des 11 attitudes auprès des Anglais britanniques dans les travaux de Diaferia (2002), puis la comparaison du taux d'identification des attitudes anglaises chez quatre groupes des sujets testés montre que les natifs ont perçu généralement mieux que les sujets non-natifs. Ce résultat correspond au résultat obtenu par les résultats de 12 attitudes japonaises, et cela confirme le « native effect » mentionné dans les travaux d'Erickson (2003), Scherer et al. (2001), Shigeno (1998).

Contrairement aux natifs, les auditeurs non-natifs ont du mal à reconnaître trois attitudes : *irritation*, *sarcastic irony* et *politeness*. L'origine de ce faible taux d'identification est expliquée par le faible taux de validité des auditeurs anglais (cf. le taux d'identification des natifs).

En ce qui concerne l'attitude de *scorn-aloofness*, le taux d'identification pour cette attitude chez les auditeurs non natifs est nettement plus faible que celui des natifs (19,6% pour les Américains, 18,3% pour les Français, 13,5% pour les Japonais). Ce résultat indique clairement que les seuls natifs qui marquent le taux d'identification largement au dessus du seuil du hasard (49,2%), et celui des autres groupes reste à peine au dessus du bar du seuil.

Un autre point important à noter que les auditeurs japonais ne reconnaissent pas bien l'attitude d'*evidence* que les sujets occidentaux. En revanche, *doubt-incredulity* est mieux reconnu chez les Japonais par rapport aux sujets occidentaux.

Enfin, l'attitude de *seduction* qui nous semble une des attitudes culturelles est reconnue correctement auprès de tous les auditeurs non natifs, mais le taux d'identification de cette attitude pour les auditeurs japonais reste largement au dessous de celui des autres groupes (27,0% pour les Japonais contre 74,2% pour les Anglais, 56,7% pour les Américains et 58,1% pour les Français).

Ensuite, l'analyse de correspondance montre que tous les auditeurs ayant participé à cette expérimentation ont généralement du mal à percevoir *sarcastic irony* et *irritation*. Ce faible taux d'identification peut être dû au fait que la réalisation de ces expressions est peu variable, et que c'est probablement très susceptible à percevoir même chez les natifs. Bien que les auditeurs natifs aient très bien discriminé *scorn-alooofness* par rapport aux autres attitudes, cette attitude est souvent mélangée avec d'autres attitudes chez les non natifs. De plus, la mauvaise reconnaissance de *doubt-incredulity* est observée chez les auditeurs américains et français. Le point perceptif de ces deux groupes d'auditeurs pour cette attitude se trouve à côté d'*interrogation*.

Par ailleurs, nous avons identifié deux autres catégories du comportement des attitudes chez les sujets américains : Une catégorie des attitudes qui imposent l'opinion du locuteur (*irritation* et *command-authority*) et l'autre est composée de *declaration*, *evidence*, et *politeness*.

En ce qui concerne les auditeurs japonais, nous n'avons identifié aucune catégorie des attitudes, et la plupart des attitudes se confondent avec déclaration. Ce phénomène nous suggère que les sujets japonais ont eu du mal à identifier la subtilité de variété des attitudes que le locuteur anglais a exprimées.

Par la suite, l'analyse des matrices de confusions montre un peu plus de détails sur le degré de confusions entre les attitudes perçues.

Les auditeurs anglais reconnaissent quatre attitudes : *declaration*, *command-authority*, *politeness* et *seduction*, sans montrer de confusion particulière. Parmi ces quatre attitudes, les auditeurs non natifs reconnaissent *declaration* et *seduction* aussi bien que les natifs. Toute fois, le taux d'identification de *seduction* chez les auditeurs japonais reste très faible comparant aux sujets occidentaux. Une explication possible pour ce phénomène est que l'attitude de séduction reflète un aspect culturel très fort et que cette expression chez les hommes japonais peut être absente ou tout à fait différente par rapport à la notion de cette attitude au sein de la culture occidentale.

Les auditeurs non natifs reconnaissent également *interrogation* et *exclamation of surprise* sans confusion importante. En ce qui concerne ces deux attitudes, ce sont les auditeurs japonais qui identifient le mieux parmi quatre groupes testés. Rappelons que lors du test perceptif avec les attitudes japonaises, *exclamation de surprise* a été parmi les attitudes très bien perçues chez les auditeurs n'ayant aucune connaissance de cette langue. Nous considérons que les auditeurs japonais reçoivent le moins l'influence du contenu sémantique parmi tous les sujets participants pour cette expérimentation car la plupart des sujets japonais réclamaient, après le test, la difficulté de compréhension sur

ceux qui ont été dit lors du test perceptif. Ce degré d'influence lexicale a probablement donné ce résultat. Il est à noter également que la confusion entre *interrogation* et *declaration* chez les auditeurs français ayant eu une confusion entre ces deux attitudes lors du test perceptif avec les attitudes japonaises est absente.

Conformément les résultats d'analyse de correspondance, *command-authority* est également généralement bien perçu chez les auditeurs non natifs. Cependant, ils montrent une légère confusion cette attitude avec *irritation* contrairement aux sujets natifs qui n'ont pas cette confusion.

A contrario, nous avons identifié par l'analyse de correspondance que tous les auditeurs ayant participé à cette expérimentation ont généralement du mal à percevoir *sarcastic irony* et *irritation*, et nous avons conclu que ce faible taux d'identification peut être du au fait que la réalisation de ces expressions est peu variable, et que c'est probablement très susceptible à percevoir même chez les natifs. *Irritation* est confondue avec *declaration* et *evidence* chez les auditeurs anglais et français. Les auditeurs américain et japonais la reconnaissent comme simple déclaration. Quant à *sarcastic irony*, cette attitude est confondue avec irritation dans les résultats des sujets anglais. Cela prouve que les sujets anglais considèrent cette attitude comme une attitude d'imposition. Par contre, les auditeurs non natifs perçoivent cette attitude comme déclaration simple.

Quant à *Politeness*, contrairement aux sujets anglais qui reconnaissent cette attitude sans confusion particulière malgré son faible taux d'identification, cette attitude culturelle est reconnue chez les auditeurs non natifs comme déclaration ou interrogation. Autrement-dit, les sujets natifs ne reconnaissent aucune intention particulière de la part du locuteur pour l'expression de cette attitude.

Une autre attitude culturelle : *scorn-alooofness*, est reconnue en montrant une confusion avec *sarcastic irony* parmi les auditeurs anglais. En revanche, les auditeurs non natifs ont du mal à reconnaître cette attitude, mais avec le taux d'identification assez faible. Cette attitude a généralement une confusion avec déclaration ou interrogation, c'est-à-dire, les attitudes n'ayant aucune intention particulière. Seuls sujets français montrent aussi une confusion entre cette attitude et *seduction*. Cette confusion avec séduction est intéressante car ces deux attitudes ne ressemblent pas du tout au niveau cognitif.

Concernant *doubt-incredulity*, cette attitude est confondue avec *interrogation* chez les Anglais. Cette confusion est partagée avec les sujets non natifs, mais leur taux d'identification pour cette attitude est nettement supérieur à celui des Américains. Ce décalage du différent taux d'identification pour l'attitude de doute entre les natifs vs. les non natifs, implique que la réalisation de *doubt-incredulity* peut être différente chez les non natifs, et le doute anglais leur semble une attitude qui n'affirme aucune intention particulière. Par conséquent, ils perçoivent mal le doute exprimé par l'Anglais, et ils reconnaissent ce doute comme simple interrogation systématiquement.

Enfin l'attitude d'*evidence* est confondue avec *simple déclaration* parmi tous les quatre groupes testés. Seuls auditeurs américains confondent cette attitude avec *politeness*.

6.5. Attitudes françaises perçues par des Américains et des Japonais

6.5.1. Objectif du test perceptif

L'objectif de ce test est d'évaluer les comportements perceptifs des auditeurs de différents pays pour les attitudes françaises qui ont été validées *supra*. Ici, notre intérêt est d'identifier un décalage du comportement perceptif entre les sujets natifs (Français), les Américains qui partagent la culture occidentale, et les sujets japonais qui sont éloignés de cette langue culturellement, et qui n'ont aucune d'idée à propos de la langue française.

6.5.2. Choix des attitudes du français

Nous avons décrit les 12 attitudes représentatives que nous avons retenues pour la validation des attitudes françaises *supra*. Il s'agit de : *doute-incrédulité*, *évidence*, *exclamation de surprise*, *autorité*, *irritation*, *mépris*, ironie sarcastique, *séduction*, *politesse*, *déclaration* et *question-simple*. Comme les attitudes anglaises que nous avons sélectionnées, certaines de ces attitudes sont spécifiques ou particulières à la culture occidentale : *séduction* qui est probablement difficile à éliciter en japonais, *ironie sarcastique*, *politesse* et *mépris* qui peuvent aussi être les attitudes qui possèdent les aspects culturels forts.

Concernant la définition de chacune des attitudes sélectionnées en langue maternelle des sujets, voir Annexe.

Tableau 32 : *Corpus des attitudes français : quatre phrases de longueur variée.*

Syllabe	Phrase	Structure
1	L'eau	mot isolé
2	L'eau bout	N + V
5	L'auto dérapait	N + V
7	Les gars coupaient des rondins	N + V + C

6.5.3. Composition du Corpus

Les phrases que nous avons sélectionnées sont aussi identiques à celles que nous avons employées pour les sujets français.

6.5.4. Méthodologie expérimentale

6.5.4.1. Stimuli

Nous avons ainsi choisi exactement les mêmes phrases qui ont été utilisées lors du test de validation.

4 énoncés différents produits selon 12 attitudes, soit un nombre total de 48 stimuli ont été utilisés pour les tests de perception suivants.

6.5.4.2. Matériel

Le test perceptif a été effectué sur l'interface du logiciel « Revolution ». Cette interface comporte, de façon permanente, la consigne du test perceptif, le choix fermé des 12 attitudes possibles et les définitions des attitudes présentées. Cette interface a été créée en langue maternelle du sujet³² (japonais et anglais).

6.5.4.3. Procédure

Le test perceptif a été conduit sujet après sujet, individuellement. Lors du test, le sujet est assis en face d'un ordinateur ; et l'ordre de présentation des stimuli est aléatoire et différent pour chaque sujet. Pendant le test, le sujet doit choisir quelle attitude et a perçue dans un choix fermé des 12 attitudes possibles. Le sujet ne peut écouter qu'une seule fois chaque stimulus avant de choisir l'attitude qu'il a perçu afin d'éviter que le sujet s'entraîne au cours du test.

Lors du commencement du test perceptif, on a demandé au sujet de lire la consigne affichée sur l'écran. Ensuite, on a expliqué brièvement la tâche et la consigne, puis la définition des 12 attitudes en langue maternelle du sujet afin d'assurer la compréhension de chacune des attitudes utilisées. Enfin à la fin du test perceptif, nous avons interrogé le sujet à propos des attitudes afin de savoir quelles attitudes étaient difficiles à distinguer ou s'il avait perçu d'autres attitudes qui n'étaient pas proposées.

6.5.4.4. Sujets

Un nombre total des auditeurs qui ont participé à ce test perceptif est de 59 :

21 auditeurs américains (9 sujets masculins et 12 sujets féminins)

38 auditeurs japonais (15 sujets masculins et 23 sujets féminins).

L'âge moyen est de 28,8 pour les auditeurs américains, 20,3 pour les auditeurs japonais. Aucun sujet n'a réclamé le trouble d'écouter lors du test.

En ce qui concerne les auditeurs japonais, ils sont tous Japonais natifs venant de régions variées. Ils habitent tous au Japon pour toujours, sans jamais séjourner à l'étranger. Ils n'ont aucune connaissance de la langue française.

³² Les interfaces en d'autres langues sont présentées en annexe.

Tous les auditeurs américains sont des étudiants à Black Hills State University, Spearfish dans l'état de South Dakota aux Etats-Unis. Ils n'ont jamais séjourné en Europe avant la passation du test, et ils n'ont aucune connaissance sur la langue française comme les sujets japonais.

6.5.5. Résultats du test perceptif

6.5.5.1. Les sujets américains

Afin de comparer le comportement perceptif des sujets américains par rapport aux natifs, voyons d'abord les résultats extraits par la série des tests statistiques.

a- Répartition des attitudes par rapport au hasard

De la même manière que pour les résultats des sujets français, nous avons effectué un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 33).

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Nous avons donc confirmé que des auditeurs américains lorsqu'ils perçoivent des attitudes anglaises, répondent de manière cohérente.

Tableau 33 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	62.3	(ddl : 11, p.<0.01)
Autorité (AU)	173.0	(ddl : 11, p.<0.01)
Déclaration (DC)	179.0	(ddl : 11, p.< 0.01)
Doute-incrédulité (DO)	224.3	(ddl : 11, p.< 0.01)
Evidence (EV)	20.4	(ddl : 11, p.< 0.01)
Exclamation de surprise (EX)	325.3	(ddl : 11, p.< 0.01)
Irritation (IR)	120.6	(ddl : 11, p.< 0.01)
Mépris (ME)	39.4	(ddl : 11, p.< 0.01)
Politesse (PO)	118.6	(ddl : 11, p.< 0.01)
Question-simple (QS)	190.3	(ddl : 11, p.< 0.01)
Ironie-sarcastique (SA)	34.3	(ddl : 11, p.< 0.01)
Séduction (SE)	78.0	(ddl : 11, p.< 0.01)

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Pour observer l'effet de la longueur sur le choix des attitudes, nous avons effectué le test de χ^2 sur les résultats moyens entre les longueurs de 1 et 2 syllabes, de 2 et 5 syllabes et de 5 et 7 syllabes, toutes attitudes confondues.

Le test de χ^2 sur les résultats moyens par longueur de 1 syllabe et 2 syllabes, toutes attitudes confondues, ne donne pas d'effet significatif (ddl = 11, $\chi^2 = 17.26$, $p. > 0.05$).

Concernant la longueur de 2 et 5 syllabes, toutes attitudes confondues, nous n'avons pas identifié d'effet significatif non plus (ddl = 11, $\chi^2 = 9.68$, $p. > 0.05$).

Enfin pour la longueur de 5 et de 7 syllabes, toutes attitudes confondues, le test χ^2 ne donne pas non plus d'effet significatif (ddl = 11, $\chi^2 = 9.86$, $p. > 0.05$).

Ce résultat démontre donc que l'effet de la longueur de l'énoncé n'existe pas sur le choix des attitudes chez les auditeurs américains.

c- Taux d'identification des attitudes

Ensuite, nous avons calculé le taux d'identification des attitudes en pourcentage afin de comparer le degré d'identification des attitudes anglaises chez les auditeurs américains par rapport aux sujets français.

La figure 79 présente l'histogramme du taux d'identification des 12 attitudes françaises pour les 21 auditeurs américains.

Les auditeurs américains ont reconnu toutes les attitudes au dessus du seuil de hasard. Pourtant, les résultats montrent que les natifs reconnaissent globalement mieux ces 12 attitudes françaises comparant aux sujets américains.

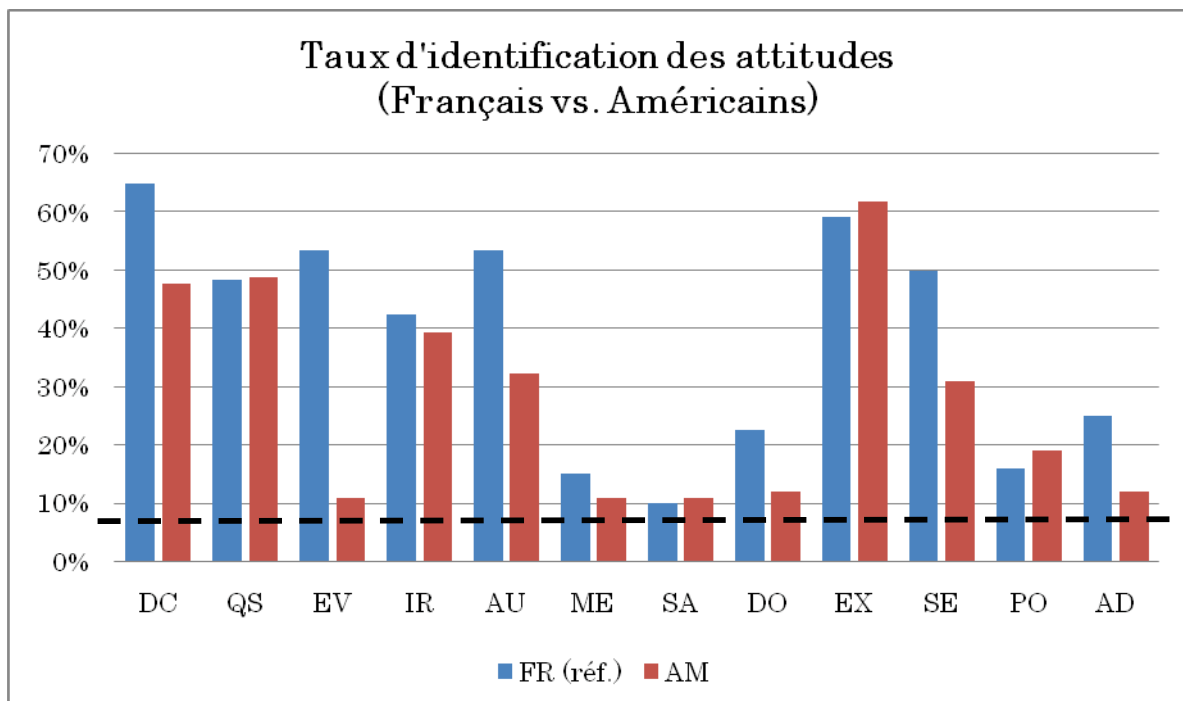


Figure 79 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs anglais et américains. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO

(*doute-incrédulité*), *EX* (*exclamation de surprise*), *SE* (*séduction*), *PO* (*politesse*) et *AD* (*admiration*).

Lorsqu'on regarde le taux de chacune des attitudes, les sujets américains ont aussi bien reconnu que les natifs pour *question-simple*, *irritation* et *exclamation de surprise*.

En ce qui concerne *mépris*, *ironie sarcastique* et *politesse*, le degré de leur reconnaissance reste faible comme les résultats des natifs. Cette faible reconnaissance est probablement due à la faible expressivité de la part de la locutrice.

En revanche, cinq attitudes : *évidence*, *autorité*, *doute-incrédulité*, *séduction* et *admiration*, ont été nettement moins bien reconnues chez les américains par rapport aux natifs. Surtout l'attitude d'*évidence* a été à peine au dessus du seuil de hasard chez les américains, alors que les Français l'ont très bien reconnue.

Nous nous rappelons que généralement les Américains reconnaissent mal l'attitude d'*évidence* de trois langues cibles (japonais, anglais et français), et leur expression d'*évidence* en anglais américain peut être très différente par rapport à ces trois langues.

d- Comportement perceptif des sujets américains en l'analyse de correspondance

Afin d'observer le comportement perceptif des sujets américains, nous avons calculé la distance des points cognitifs de 12 attitudes par rapport aux points stimuli selon l'analyse de correspondance comme pour les résultats des sujets français.

Selon ce test statistique, les quatre premières dimensions expliquent 90% de la variance.

Voyons sur la figure 80 en haut qui représente leur comportement perceptif en deux dimensions (1^{ère} et 2^{ème} Dimension). Selon le graphe présenté, le point de stimuli pour trois attitudes : *exclamation de surprise*, *séduction* et *déclaration*, se situe près de leur point perceptif (voir les attitudes dans le cercle en couleur bleu). La figure 80 (en bas) confirme également la bonne reconnaissance de ces attitudes et *question simple*. Cela signifie que ces quatre attitudes ont été bien reconnues chez les auditeurs américains. Nous pouvons identifier également les attitudes mal reconnues par l'indication de la longue distance entre le point de stimulus et le point perceptif dans le graphe de gauche. Il s'agit, par exemple, de *doute-incrédulité* et d'*admiration* qui sont toutes les deux mal reconnues chez les Américains, et leurs points perceptifs se situent plutôt à côté du point de stimulus pour *question-simple*. Autrement-dit, les sujets américains confondent ces deux attitudes avec *question-simple*.

En ce qui concerne l'attitude de *mépris* et de *politesse*, leurs points de stimulus se situent près de *déclaration*. Cela signifie que les Américains considèrent la notion de deux attitudes : *politesse* et *mépris* français comme une attitude similaire de *déclaration*.

Quant aux deux attitudes d'imposition de l'avis du locuteur : *autorité* et *irritation*, le point perceptif de chacune des attitudes se trouvent le point de stimulus de l'autre attitude. Cela explique que ces attitudes sont confondues simultanément.

Nous allons analyser plus concrètement les phénomènes de confusions entre les attitudes dans le paragraphe suivant.

e- Confusion entre attitudes

Dans le paragraphe précédent, nous avons identifié la bonne reconnaissance de *déclaration*, *question-simple*, *séduction* et *exclamation de surprise*. En ce qui concerne *irritation*, le taux d'identification montre que cette attitude a été bien reconnue, mais cela a été confondu avec *autorité* selon l'analyse de correspondance.

Dans ce paragraphe, nous allons observer des phénomènes de confusions parmi les attitudes mal reconnues, mais surtout les cinq attitudes : *évidence*, *doute-incrédulité* et *admiration*, qui ont été nettement moins bien reconnues chez les américains par rapport aux natifs.

Selon le tableau 34, les sujets américains ont montré une importante confusion pour l'expression d'*évidence* avec *autorité*. Cette confusion a déjà été observée pour l'*évidence* japonaise, mais l'*évidence* anglaise a été mélangée plutôt avec la politesse.

Le même tableau montre que *doute-incrédulité* ayant été mal reconnue est reconnue plutôt comme *question-simple*. Du même, *admiration* est mélangé avec *exclamation de surprise* mais surtout avec *question-simple*. Ce comportement perceptif de ces deux attitudes justifie les résultats de l'analyse de correspondance.

Tableau 34 : Matrices de confusion pour 21 auditeurs américains. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (*déclaration*), QS (*question-simple*), EV (*évidence*), IR (*irritation*), AU (*autorité*), ME (*mépris*), SA (*ironie-sarcastique*), DO (*doute-incrédulité*), EX (*exclamation de surprise*), SE (*séduction*), PO (*politesse*) et AD (*admiration*).

Attitudes perçues	Attitudes présentées											
	AD	AU	DC	DO	EV	EX	IR	ME	PO	QS	SA	SE
Admiration	12%	0%	2%	2%	1%	2%	1%	1%	10%	0%	5%	4%
Autorité	2%	32%	4%	0%	26%	1%	25%	8%	2%	5%	19%	1%
Déclaration	1%	1%	48%	1%	14%	0%	7%	19%	35%	15%	20%	7%
Doute-incrédulité	8%	1%	6%	12%	12%	5%	1%	6%	5%	2%	2%	10%
Evidence	6%	5%	7%	0%	11%	1%	7%	14%	14%	5%	11%	7%
Surprise	20%	2%	2%	24%	7%	62%	2%	2%	2%	0%	1%	2%
Irritation	4%	45%	8%	1%	7%	2%	39%	1%	1%	4%	8%	0%
Mépris	1%	7%	6%	0%	10%	1%	7%	11%	7%	4%	12%	2%
Politesse	13%	0%	13%	4%	1%	4%	1%	19%	19%	8%	5%	14%
Question	25%	1%	4%	49%	7%	15%	4%	4%	2%	49%	6%	17%
Ironie-sarcastique	4%	5%	0%	7%	4%	5%	4%	8%	0%	4%	11%	5%
Séduction	4%	0%	0%	0%	0%	1%	1%	6%	2%	5%	0%	31%

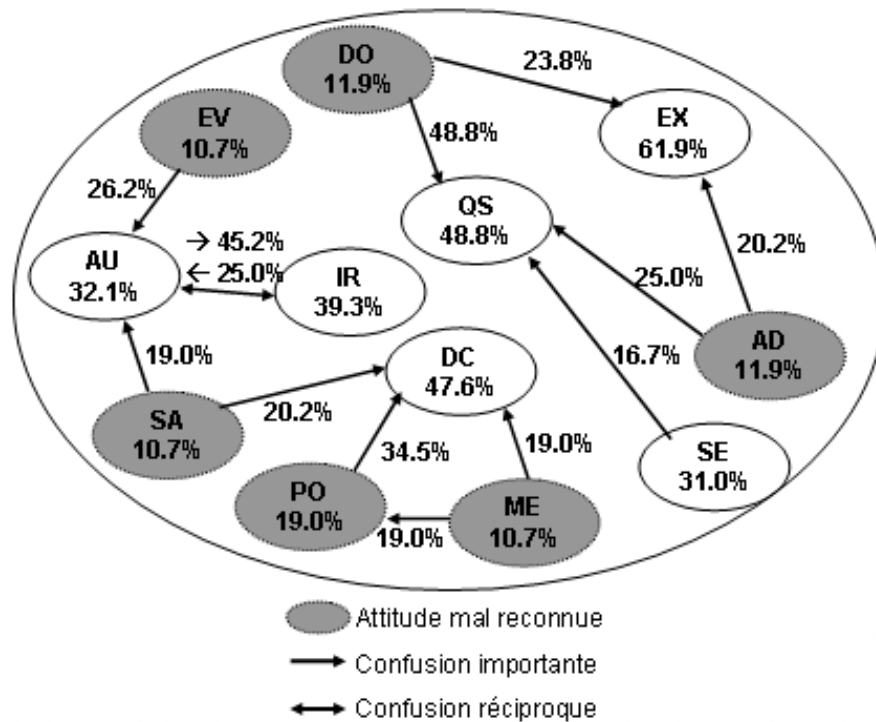


Figure 81 : Graphe de Confusion pour 21 auditeurs américains : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Abréviations : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

Nous voyons maintenant les attitudes généralement mal reconnues comme les résultats des sujets français. L'attitude de *mépris* montre une confusion importante avec *politesse* et *déclaration*. Ensuite, *ironie sarcastique*, quant à elle, a été confondue avec *autorité* et *déclaration*. Enfin, *politesse* a été mélangée avec *déclaration* comme le comportement des sujets français. Leur comportement pour ces attitudes nous indique qu'ils ont la tendance à percevoir ces attitudes comme une attitude sans intention particulière.

Il est à noter également que l'attitude de *séduction* est nettement moins bien reconnue chez les auditeurs américains par rapport aux français, mais cela n'a pas montré de confusion particulière.

6.5.5.2. Les sujets japonais

A la suite des résultats des occidentaux (Américains et Français), il s'agit cette fois des résultats des auditeurs japonais qui sont éloignées de langues occidentales géographiquement et culturellement.

a- Répartition des attitudes par rapport au hasard

Voici un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 35).

Tableau 35 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	72.7	(ddl : 11, p.<0.01)
Autorité (AU)	518.7	(ddl : 11, p.< 0.01)
Déclaration (DC)	495.8	(ddl : 11, p.< 0.01)
Doute-incrédulité (DO)	388.2	(ddl : 11, p.< 0.01)
Evidence (EV)	84.1	(ddl : 11, p.< 0.01)
Exclamation de surprise (EX)	604.6	(ddl : 11, p.< 0.01)
Irritation (IR)	252.8	(ddl : 11, p.< 0.01)
Mépris (ME)	316.5	(ddl : 11, p.< 0.01)
Politesse (PO)	274.2	(ddl : 11, p.< 0.01)
Question-simple (QS)	577.2	(ddl : 11, p.< 0.01)
Ironie-sarcastique (SA)	115.9	(ddl : 11, p.< 0.01)
Séduction (SE)	369.5	(ddl : 11, p.< 0.01)

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Cela explique donc que les variations expressives introduites par les attitudes françaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs japonais.

b- Effet de la longueur des énoncés

Répartition des différentes longueurs et des attitudes par rapport au hasard

Ici, nous avons effectué le test de χ^2 sur les résultats moyens entre les longueurs de 1 et 2 syllabes, de 2 et 5 syllabes et de 5 et 7 syllabes, toutes attitudes confondues.

Le test de χ^2 sur les résultats moyens par longueur de 1 syllabe et 2 syllabes, toutes attitudes confondues, ne donne pas d'effet significatif (ddl = 11, χ^2 = 15.00, p. > 0.05).

A contrario, pour la longueur de 2 et 5 syllabes, toutes attitudes confondues, nous avons identifié un effet significatif de la longueur (ddl = 11, χ^2 = 46.29, p. < 0.05).

En ce qui concerne la longueur de 5 et de 7 syllabes, toutes attitudes confondues, le test χ^2 ne donne pas d'effet significatif (ddl = 11, χ^2 = 6.11, p. > 0.05).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe entre les phrases de 2 syllabes et de 5 syllabes, ce qui est opposé aux résultats obtenus avec les auditeurs américains.

c- Taux d'identification des attitudes

La figure 82 représente la comparaison du taux d'identification des attitudes en pourcentage chez les Français vs. les Américains vs. les Japonais.

Tout d'abord, les auditeurs japonais ont reconnu toutes les attitudes sauf *mépris* et *ironie sarcastique* au dessus du seuil de hasard. Les deux attitudes : *mépris* et *ironie sarcastique*, n'ont pas été reconnues seulement chez les auditeurs japonais, mais le degré de reconnaissance pour ces attitudes chez les Français et les Américains reste aussi faible, et c'est probablement dû à la faible expressivité de la locutrice.

Comparant avec celle des natifs, le taux global d'identification des auditeurs japonais reste inférieur aux natifs sauf *interrogation* et *doute-incrédulité*. En ce qui concerne *exclamation de surprise* et *politesse*, le taux d'identification de ces attitudes chez les Japonais et les Français est presque identique.

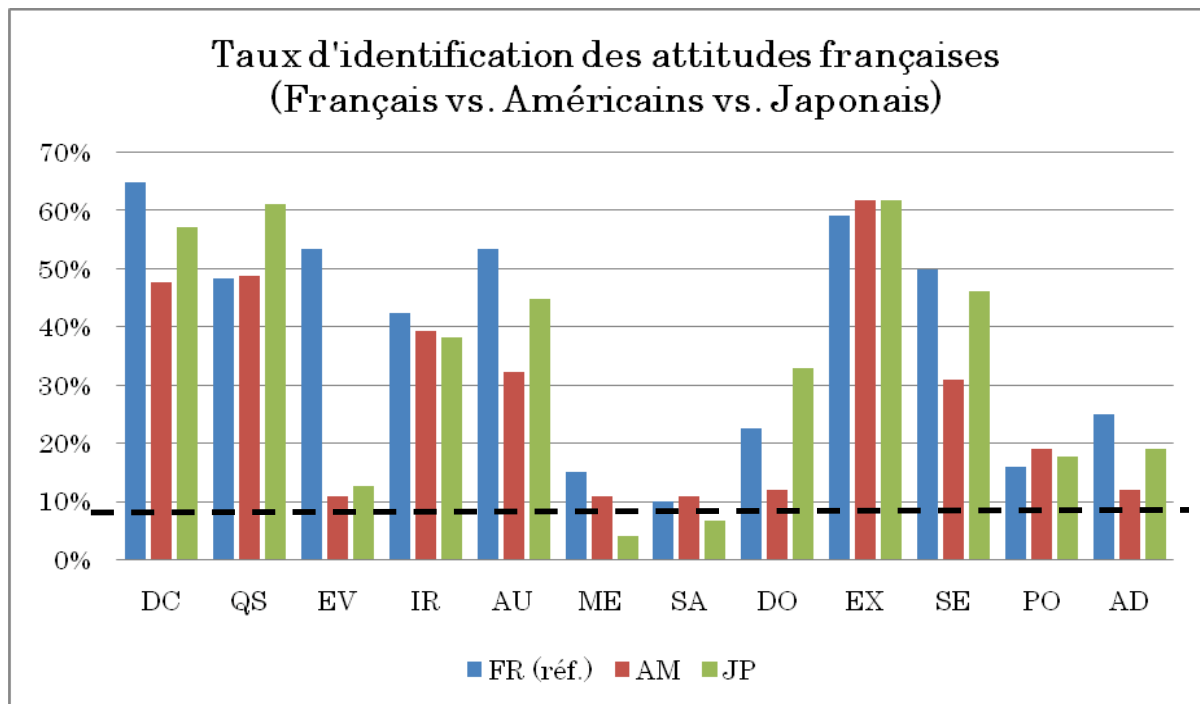


Figure 82 : Taux d'identification en pourcentage pour chaque attitude, comparaison entre auditeurs français, américains et japonais. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

Lorsqu'on regarde le taux de chacune des attitudes, les sujets japonais ont très bien reconnue *déclaration*, *question-simple*, *irritation* et *exclamation de surprise* comme les deux groupes précédents. Ils reconnaissent bien également *autorité*, *séduction* et *politesse*. Il est important à noter que les sujets japonais ont nettement mieux reconnu, comparant aux résultats des américains, les attitudes suivantes : *déclaration*, *question-*

simple, *autorité*, *séduction* et *admiration* même si leur taux de reconnaissance reste inférieur par rapport aux natifs. Leur comportement pour l'attitude de séduction est intéressant car la séduction anglaise a été plutôt mal perçue comparant aux sujets occidentaux, mais en ce qui concerne la séduction française, les sujets japonais reconnaissent mieux que les Américains.

Nous mentionnons également un autre point important. Les auditeurs japonais reconnaissent *évidence* à peine au dessus du hasard comme les sujets américains, alors que les Français l'ont très bien reconnue. Nous allons décrire plus tard le comportement perceptif des Japonais pour cette attitude, mais *a priori*, cette attitude française peut être un potentiel faux-amis entre ces trois groupes des sujets.

d- Comportement perceptif des sujets japonais en analyse de correspondance

Comme les résultats des sujets occidentaux, la distribution des points stimuli et des points perceptifs pour les résultats des sujets japonais est présentée en figure 83 selon l'analyse de correspondance.

Selon ce test statistique, quatre premières dimensions comportent 89% de variances.

La figure 83 (haut) illustre le comportement perceptif des sujets japonais en deux dimensions (Dimension 1 et 2). Selon le graphe présenté, le point de stimuli pour trois attitudes : *exclamation de surprise*, *séduction* et *question-simple*, se situe près de leur point perceptif (voir les attitudes dans le cercle en couleur bleu). La figure 83 (bas) confirme également le même phénomène pour ces attitudes. Cela signifie que ces trois attitudes ont été bien reconnues chez les auditeurs japonais.

Nous pouvons identifier également les attitudes mal reconnues par l'indication de la longue distance entre le point de stimulus et le point perceptif dans le graphe du haut. Il s'agit d'*évidence*, *ironie sarcastique* et *mépris*. Le point de stimulus pour *ironie sarcastique* se trouve dans le cercle de confusion avec *autorité* et *irritation* (voir le cercle rempli par la couleur violette qui se situe à gauche du graphe en 1^{ère} et 2^{ème} dimension). Rappelons que la confusion entre les attitudes d'imposition de l'avis du locuteur : *autorité* et *irritation* était présente dans les résultats des sujets américains. Les sujets japonais ont confondu également entre *question-simple* et *doute-incrédulité*. Nous avons déjà identifié cette confusion chez les sujets occidentaux, mais comparant à eux, les auditeurs japonais ont bien reconnu la *doute-incrédulité* française.

Voyons dans le paragraphe suivant comment les auditeurs japonais ont confondu pour les attitudes mal reconnues afin de comprendre les phénomènes de confusions entre les attitudes.

a- Confusion entre attitudes

Le tableau 36 représente les matrices de confusion des 12 attitudes françaises chez 38 auditeurs japonais, et la figure 84 illustre la direction et la catégorie des confusions parmi 12 attitudes présentées.

Ces matrices de confusion confirment que *ironie sarcastique* qui n'a pas été reconnue chez les sujets japonais a été reconnue comme *autorité* ou *irritation* comme nous l'avons déjà observé lors de l'analyse de correspondance. Ce résultat indique que même si les sujets japonais n'arrivent pas à reconnaître précisément *ironie sarcastique* de la part du locuteur, ils reconnaissent parfaitement l'imposition de l'opinion de la part du locuteur.

Une autre attitude n'ayant pas été reconnue : *mépris*, a, en fait, été reconnue comme déclaration. Autrement dit, les sujets japonais n'ont pas réussi à identifier l'intention du locuteur.

En revanche, trois attitudes : *déclaration*, *question-simple* et *exclamation de surprise*, ayant été très bien reconnues selon l'analyse de correspondance, montrent aucune confusion particulière. Une autre attitude : *séduction* qui est bien identifiée (cf. *supra*) a été mélangée avec *question-simple*.

L'attitude de *doute-incrédulité* a été confondue également avec *question-simple*. Cette confusion a été déjà identifiée dans le paragraphe précédent par l'analyse de correspondance. Bien que cette confusion soit partagée avec des sujets français et américains, le degré d'identification de cette attitude chez les Japonais reste nettement supérieur par rapport aux deux autres groupes d'auditeurs (cf. *supra*).

Tableau 36 : Matrices de confusion des 38 auditeurs japonais. Les valeurs présentées en gras indiquent le taux d'identification. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

<i>Attitudes perçues</i>	<i>Attitudes présentées</i>											
	<i>AD</i>	<i>AU</i>	<i>DC</i>	<i>DO</i>	<i>EV</i>	<i>EX</i>	<i>IR</i>	<i>ME</i>	<i>PO</i>	<i>QS</i>	<i>SA</i>	<i>SE</i>
Admiration	19%	2%	0%	1%	3%	3%	2%	5%	3%	0%	5%	1%
Autorité	1%	45%	1%	1%	12%	3%	19%	2%	1%	1%	20%	0%
Déclaration	9%	0%	57%	1%	14%	0%	6%	45%	42%	7%	8%	8%
Doute-incrédulité	15%	0%	1%	33%	3%	13%	0%	3%	6%	12%	4%	11%
Evidence	6%	1%	6%	1%	13%	0%	11%	14%	4%	5%	9%	1%
Surprise	18%	7%	0%	11%	10%	62%	3%	3%	1%	1%	11%	0%
Irritation	7%	40%	3%	3%	23%	9%	38%	2%	3%	5%	25%	0%
Mépris	5%	3%	9%	2%	9%	1%	14%	4%	7%	2%	10%	2%
Politesse	3%	0%	5%	1%	3%	0%	0%	14%	18%	2%	1%	9%
Question	10%	0%	3%	41%	3%	9%	1%	2%	0%	59%	2%	22%
Ironie-sarcastique	4%	3%	9%	4%	8%	1%	5%	4%	5%	2%	7%	1%
Séduction	5%	0%	6%	0%	1%	0%	0%	1%	10%	4%	0%	46%

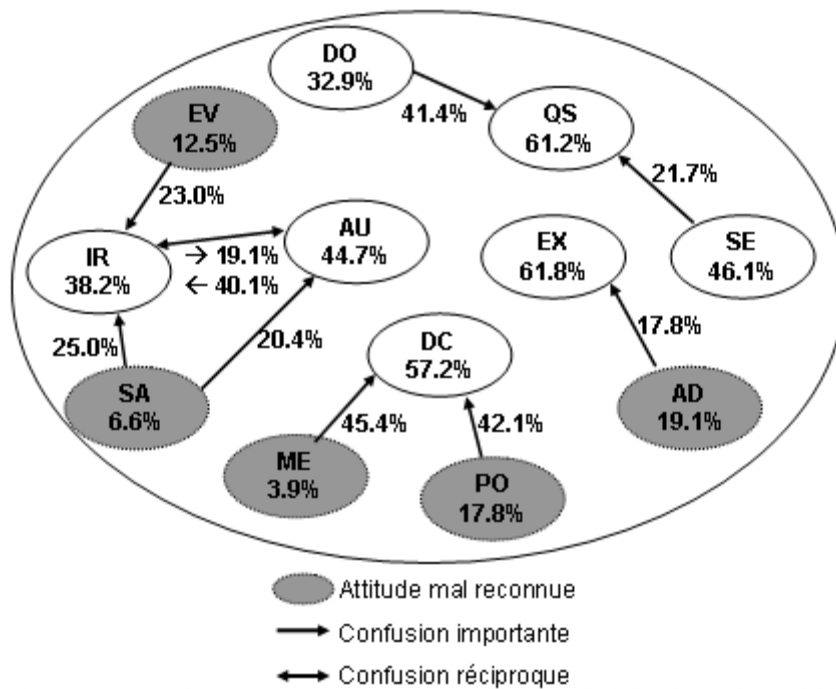


Figure 84 : Graphe de Confusion pour 38 auditeurs japonais : pourcentage indiqué en dehors de cercle signifie le degré de confusion en valeur relative. Autres pourcentages écrits au dessous d'étiquette de chaque attitude représentent le taux de reconnaissance des attitudes en valeur relative. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AU (autorité), ME (mépris), SA (ironie-sarcastique), DO (doute-incrédulité), EX (exclamation de surprise), SE (séduction), PO (politesse) et AD (admiration).

Conformément aux résultats de l'analyse de correspondance, nous pouvons identifier une confusion réciproque entre *autorité* et *irritation* dans ces matrices de confusion.

Quant à *évidence*, cette attitude a montré une confusion importante avec *irritation*. Ce comportement des Japonais ressemble à celui des sujets américains qui ont confondu cette attitude avec *autorité*, mais ce type de confusion est totalement absent chez les natifs qui, eux, l'ont identifié sans confusion particulière.

La politesse française a été mal reconnue chez les sujets japonais (17,8%) comme les auditeurs occidentaux en montrant une confusion importante avec *simple déclaration*. c'est probablement dû à la faible expressivité de politesse de la part de la locutrice.

L'attitude d'admiration qui est faiblement reconnue chez les Japonais comme d'autres sujets français et américains, montre une confusion particulière avec exclamation de surprise.

6.5.5.3. Conclusion

L'objectif de ce test est d'évaluer les comportements perceptifs des sujets non francophones face à des attitudes françaises. Il nous semble particulièrement intéressant de soumettre à un même test perceptif trois groupes culturels différents : des Américains, qui partagent la culture occidentale et des Japonais dont ni la langue ni la culture ne se rapprochent et pour qui le français représente le statut de langue étrangère inconnue.

Notre hypothèse initiale est qu'à travers des résultats obtenus de ce test, le taux d'identification inférieure chez les non natifs par rapport aux Français d'une part, et surtout un décalage du comportement perceptif entre les natifs et les sujets non natifs d'autre part.

En manipulant le corpus des 12 attitudes françaises ayant été validé par les natifs *supra*, nous avons effectué une série du test de perception auprès des auditeurs américains et japonais afin de mesurer la distance perceptive entre les attitudes et de voir le décalage de comportement entre ces groupes de sujets.

Une série du test khi deux relève les points suivants. D'abord, la distribution des attitudes est différente du hasard pour tous les auditeurs qui ont participé à ce test. Second point important est qu'il y a un effet significatif de longueur des énoncés pour les français et japonais, mais cet effet est absent chez les auditeurs américains. En effet, pour les Américains, nous n'avons pas identifié d'effet significatif de la longueur sur la distribution des attitudes entre toutes les longueurs de phrases choisies (c'est-à-dire entre les mots isolés monosyllabiques et dissyllabiques, entre les phrases de 2 et de 5 syllabes, entre les phrases de 5 et de 7 syllabes). Les sujets japonais, par contre, ont eu une influence significative de la longueur sur la distribution des attitudes entre les phrases de 2 et de 5 syllabes.

Dans un premier temps nous avons comparé le degré du taux d'identification des attitudes françaises chez trois groupes des sujets testés. Le résultat montre que les natifs perçoivent généralement mieux les attitudes appropriées par rapport aux sujets non-natifs. Ce résultat confirme l'hypothèse initiale que nous avons postulée.

Le taux de reconnaissance des non natifs pour toutes les attitudes françaises sauf mépris et ironie sarcastique était tous au dessus du seuil du hasard, autrement dit, les auditeurs américains et japonais ont reconnu plupart des attitudes testées significativement. Plus particulièrement, les sujets non natifs ont très bien reconnue *déclaration*, *question-simple*, *irritation* et *exclamation de surprise* comme chez les natifs selon l'analyse de correspondance.

En ce qui concerne *autorité* et *séduction*, les natifs et les sujets japonais identifie bien, mais le degré d'identification des sujets américains pour ces attitudes est nettement plus faible comparant à deux autres groupes.

Ici, le comportement des Japonais pour *séduction* est intéressant car la séduction anglaise a été plutôt mal perçue chez les Japonais comparant aux sujets occidentaux,

mais en ce qui concerne la séduction française, les sujets japonais reconnaissent mieux que les Américains.

Par contraste, *évidence*, *mépris*, *ironie sarcastique*, *politesse* et *admiration* sont mal reconnues chez les sujets non natifs. Notamment, *ironie sarcastique* et *mépris* sont les deux attitudes que les auditeurs sentaient autant de difficulté que les sujets natifs. Ce faible taux d'identification est justifié par l'analyse de correspondance, et c'est peut-être du au fait que la réalisation de ces expressions est peu variable, et que c'est probablement très difficile à percevoir même chez les natifs.

Il est à noter que *doute-incrédulité* a été mal perçu chez les Américains, mais les sujets japonais l'ont très bien reconnu (même plus que les natifs). Rappelons que les Japonais avaient mieux identifié le doute anglais par rapport aux autres sujets non natifs (cf. *supra*), et ces résultats nous suggèrent que les Japonais ont plus de facilité à percevoir le doute en trois langues malgré la distance culturelle.

Nous mentionnons également un autre point important. Les auditeurs non natifs reconnaissent *évidence* à peine au dessus du hasard, alors que les Français l'ont très bien reconnue. Cette attitude française peut être, a priori, un potentiel « faux-amis » entre ces trois groupes de sujets.

Selon l'analyse de correspondance, quelques catégories de comportements perceptifs similaires entre *irritation* et *autorité*, et également entre *doute-incrédulité* et *question-simple* sont observées chez tous les participants de ce test.

Conformément à ces résultats de l'analyse de correspondance, nous pouvons identifier une confusion réciproque entre *autorité* et *irritation* chez les non natifs selon l'analyse de degré de confusion. De même, la confusion entre *doute-incrédulité* et *question-simple* est aussi partagée parmi tous les auditeurs ayant participé à ce test perceptif.

Une légère confusion entre *question-simple* et *déclaration* qui a lieu chez les sujets français n'est pas identifiée chez les sujets non natifs.

L'attitude d'*ironie sarcastique* qui est particulièrement mal reconnue parmi tous les groupes, a été confondue souvent avec *mépris*, *irritation* ou *autorité* pour tous les groupes. Ce résultat indique que même si les sujets n'arrivent pas précisément *ironie sarcastique* de la part du locuteur, ils reconnaissent parfaitement l'imposition de l'opinion de la part du locuteur.

Une autre attitude ayant été mal reconnue : *mépris*, a été confondue majoritairement avec *déclaration*. Autrement dit, tous les sujets de tous les groupes n'ont pas facilement réussi à identifier l'intention du locuteur.

En revanche, trois attitudes : *déclaration*, *question-simple* et *exclamation de surprise*, ayant été très bien reconnues les non natifs selon l'analyse de correspondance, montrent aucune confusion particulière. Une autre attitude : *séduction* qui est bien identifiée parmi les trois groupes a été mélangée avec *question-simple*.

Quant à *évidence*, cette attitude a montré une confusion importante avec *irritation* pour les Japonais et *autorité* pour les Américains, alors que ce type de confusion est totalement absent chez les natifs qui, eux, l'ont identifié sans confusion particulière. Ce résultat nous permet de conclure qu'*a priori*, cette attitude française nous semble un des « faux-amis ».

7. Apprentissage des attitudes japonaises

7.1. Problématique

Dans le chapitre 4, nous avons effectué une série de tests de perception avec les 12 attitudes japonaises auprès des sujets provenant de cultures différentes (Américains et Français), et d'après l'analyse des données obtenues, nous avons identifié les attitudes communes qui sont reconnues parmi ces trois groupes d'auditeurs, les « faux-amis » dû à la spécificité de ces attitudes culturelles et enfin les attitudes inconnues qui sont bien identifiées chez les natifs, mais étaient ininterprétable pour les auditeurs d'une différente culture.

En prenant en compte tous ces phénomènes perceptifs chez les auditeurs n'ayant aucune connaissance de la langue et de la culture cible (japonaise), nous nous penchons, dans ce chapitre, une évolution (ou plutôt un changement) du comportement perceptif des auditeurs de la même langue en observant des comportements des auditeurs ayant acquis certaines compétences de la langue (et culture) japonaise.

Plus précisément, nous avons effectué le même test de perception avec les attitudes japonaises auprès d'auditeurs français en cours d'apprentissage du japonais. Il s'agit, cette fois, de voir l'effet de l'apprentissage sur la perception des attitudes japonaises chez les Français natifs.

7.2. Revue de littérature sur l'apprentissage des attitudes

Malheureusement il n'y a que peu de recherches qui traitent de l'effet d'apprentissage des valeurs attitudinales, mais nous présentons ici quelques travaux qui se concentrent sur ces domaines d'études.

Voyons d'abord les travaux sur la comparaison interculturelle de la prosodie des attitudes entre le français et le tchèque (Mejvaldová, 2000) que nous venons de présenter brièvement *supra*. Dans ses travaux, 15 phrases de modalités différentes (affirmatif, interrogatif) ont été produites par 7 locuteurs (5 locuteurs tchèques et 2 locuteurs français, tous bilingues) en deux langues différentes selon 8 expressions affectives (« peur », « colère », « tristesse », « ennui », « surprise », « admiration », « joie » et « neutralité »). Ensuite ces stimuli ont été évalués perceptivement auprès de 4 groupes des sujets suivants (des Français ayant une bonne connaissance de la langue tchèque, des Français n'ayant pas de connaissance sur la langue tchèque, des Tchèques ayant

acquis une bonne connaissance de la langue française et des Tchèques qui ne connaissent pas le français).

Les résultats des tests perceptifs ontrent les trois points suivants. Les auditeurs ayant une bonne connaissance de la langue avaient une facilité d'identification des attitudes transmises par la prosodie. Les auditeurs ont mieux perçu les attitudes de leur langue maternelle. En revanche, la réalisation des énoncés par un locuteur natif n'a pas donné un effet particulier pour l'identification des attitudes.

L'auteur a identifié un potentiel « faux amis » prosodique (« homonymie » selon sa propre terminologie : une morphologie prosodique liée avec deux attitudes différentes entre les deux langues. Il s'agit de l'attitude d'« ennui » produite par un locuteur français qui a été bien perçue par les auditeurs français, mais les sujets tchèques avaient tendance à l'interpréter comme « triste » ou « furieux ». De même que la « déclaration » et la « surprise » produites par deux locuteurs tchèques ont été perçues comme « joyeux » par les auditeurs français.

Abelin & Allwood (2000, 2002), quant à eux, travaillent sur l'identification de huit expressions affectives suédoises : colère, angoisse, joie, tristesse, surprise, dégoût, dominance, timidité, par 35 sujets suédois d'une part, et d'autre part 58 sujet non natifs (23 Espagnols, 23 Finnois, 12 Anglais) parlant la langue suédoise comme leur seconde langue. La compétence de la langue suédoise chez les participants non natifs varie du très haut niveau au niveau débutant. Lors du test perceptif, les sujets doivent répondre librement à écrit après la présentation de chaque stimulus. Ensuite, les auteurs ont regroupé les réponses similaires afin d'attribuer une étiquette commune pour chaque catégorie créée. En cours de ce processus, les réponses pour deux expressions (dégoût et timidité) ont été exclues du sujet d'analyse suivant dû au faible taux de reconnaissance. Les résultats indiquent que les auditeurs suédois (qui sont natifs) montrent un taux d'identification de 73% contre 59% (Anglais), 52% (Finnois) et 42% (Espagnols) locuteurs du suédois comme L2. Lorsqu'on regarde le degré de confusion entre les affects perçus, la colère et la tristesse sont les expressions qui sont les mieux identifiées, en revanche, la joie suédoise a tendance à être perçue comme la tristesse chez les Espagnols (35%) et les Finnois (17%). Il est à noter que la tristesse est mieux identifiée chez les non natifs que les natifs qui perçoivent parfois cette expression comme de la déception. La colère est aussi identifiée mieux chez les Espagnols et les Finnois par rapport aux sujets anglais et suédois qui l'interprètent parfois comme de la domination.

Matsumoto & Sadanobu (2001) s'intéressent à l'interprétation de la voix pressée chez 61 apprenants chinois de japonais. Ils avaient le niveau moyen ou avancé. Dans cet optique, deux phrases : “Umainaa” (Il est doué pour cela) et “Nomura-san, saanee” (M. Nomura, je ne le connais pas) exprimées avec une voix pressée, ont préalablement été enregistrées dans une chambre sourde.

La première phrase avec la voix pressée est perçue chez les Japonais comme de l'admiration (Il est doué de cela et j'ai de l'admiration.). Pour la deuxième phrase comme

de la sincérité (Il a réfléchi sérieusement à propos de M. Nomura, mais il ne le connaît pas). Les sujets ont choisi une des attitudes qui correspond à chaque stimulus parmi cinq choix d'attitudes. Après cette expérimentation perceptive, les tendances suivantes ont été observées. D'abord, les apprenants chinois étaient capables d'identifier une différence de qualité de voix (par exemple, la voix pressée semble être une articulation inhabituelle). Cependant, seuls 37,7% des sujets ont fait une interprétation appropriée de la première phrase, et 49,2% pour la deuxième phrase. Cela signifie que les apprenants chinois ont tendance à interpréter la voix pressée comme une attitude négative (59% des sujets ont interprété la première phrase comme de l'ironie ou de l'indifférence, et pour la deuxième phrase, 26,2% des sujets l'ont interprété comme du dégoût).

Nakagawa et al. (2007) a effectué un test perceptif auprès des 15 apprenants français débutants de langue japonaise avec six phrases n'ayant pas de connotation particulière, prononcées avec des expressions affectives variées. Ces énoncés sont produits par deux locutrices japonaises lors d'une conversation spontanée au téléphone dans une salle relativement calme au sein de l'ATR, Japon.

Avant le test perceptif, 10 codeurs naïfs ont attribué une étiquette à chaque expression qu'ils viennent d'entendre en choix libre. C'est ainsi qu'ils ont attribué une étiquette pour laquelle plus de la moitié des codeurs se sont mis d'accord, à chaque expression. Ainsi, 22 stimuli répétés trois fois en ordre aléatoire sont présentés aux 15 japonais, 15 français naïfs et 15 français débutants en choix fermé de trois réponses en mettant deux fausses réponses.

Selon les résultats, les sujets japonais ont identifié 13 sur 23 expressions avec un taux d'identification supérieur à deux fois du seuil de hasard (66,7%). Si on se concentre sur ces 13 énoncés correctement perçus chez les natifs, Français naïfs ont identifié seulement 4 sur 13 expressions. Parmi les expressions qui n'ont pas été identifiées, le mécontentement est perçu majoritairement comme la compassion. Contrairement aux Français naïfs, les sujets français du niveau débutant ont identifié 8 sur 13 expressions, et ils concluent que l'effet d'apprentissage est identifié sur certaines attitudes (embarras, sincérité, sympathie), cependant, l'hésitation, le mécontentement, politesse (avec le sentiment d'embarras) restent les attitudes difficiles à identifier malgré l'apprentissage du japonais. Ce résultat nous indique deux points suivants. D'abord, selon le niveau d'apprentissage de la langue cible, l'on améliore l'identification des attitudes. Cependant, certaines attitudes attachées à la culture de la langue cible ont tendance à être mal perçues malgré leur compétence de langue cible.

7.3. Objectif du test perceptif

Notre objectif de ce chapitre est donc de vérifier l'effet de l'apprentissage de la langue japonaise sur la perception des attitudes. C'est-à-dire, ici, de comparer les résultats des apprenants français du japonais par rapport aux Français naïfs d'une part, et de comparer ces résultats avec ceux des Japonais natifs d'autre part.

7.4. Sujets

Nous avons sélectionné 32 sujets français (6 femmes et 26 hommes) ayant déjà acquis certaines connaissances de la langue et de la culture japonaises à l'université en France.

L'âge moyen des sujets est de 21,5 ans.

28 sujets choisis sont des étudiants ayant pour spécialité « anglais – japonais » en Langues Etrangères Appliquées (LEA). Ils sont 2^{ème} et 3^{ème} années. Il y a aussi 4 sujets d'autres spécialités que les langues (par exemple, mathématiques etc.). Ils apprennent le japonais comme option. Ils sont en 4^{ème} ou 5^{ème} années en LANGues pour Spécialistes d'Autres Disciplines (LANSAD).

Le nombre d'heures d'apprentissage du japonais varie de 142 heures (dont 71 heures de cours d'oral) à 480 heures (dont 140 heures de cours d'oral). La moyenne du nombre d'heures d'apprentissage du japonais est de 339 heures (dont 106 heures en moyenne de cours d'oral)

Parmi ces 32 sujets, seuls 8 ont déjà séjourné au Japon. La durée de leur séjour au Japon varie de deux semaines à une année.

7.5. Procédure

7.5.1. Deux entretiens

En complément du test perceptif, nous avons effectué deux entretiens (l'un avant le test perceptif et l'autre après). Le premier entretien, avant le test de perception, a pour but d'obtenir d'une part des informations personnelles (âge, lieu de résidence, etc.) et d'autre part, des informations ayant trait à la langue japonaise, à d'éventuels séjours linguistiques et des informations portant sur le contexte et l'entourage japonais.

En ce qui concerne les sujets n'ayant peu de connaissance en japonais et qui ne maîtrisent pas encore suffisamment la langue cible, l'interview a eu lieu en langue maternelle. Cependant, pour ce qui est des sujets en cours d'apprentissage du japonais et

qui maîtrisent bien cette langue, nous leur avons posé des questions en japonais afin d'avoir une idée sur leurs compétences dans cette langue. Dans le cas où des sujets n'arriveraient pas à comprendre, nous avons reposé les mêmes questions en langue maternelle. De même, si la difficulté se situait au niveau de la capacité à répondre en japonais, nous leur avons aussi donné la possibilité de s'exprimer en langue maternelle.

L'entretien ayant lieu après le test de perception permet d'obtenir des réactions, des réflexes ou encore des explications sur le choix des attitudes repérées pendant le test. Pour cette interview également, les questions ont été prioritairement posées en japonais puis si nécessaire, en langue maternelle. Il est à noter que les questions posées demandent des réponses particulièrement concrètes et relativement compliquées ; tous les sujets ont donc répondu en langue maternelle.

7.5.2. Catégorisation des sujets selon le niveau de langue

7.5.2.1. Objectif

Ces entretiens avec les sujets ne sont évidemment pas suffisants pour faire état de leurs compétences en langue japonaise. Ici encore, il faut dissocier trois types de public : les sujets n'ayant aucune connaissance du japonais (degré 0 de compétence) ; les sujets en cours d'apprentissage du japonais (niveau1); et enfin les sujets pour qui le japonais est langue seconde (ils possèdent cette langue, niveau2). Le cas des apprenants de japonais langue étrangère est celui qui nécessite une évaluation de leurs compétences.

7.5.2.2. Méthodes d'évaluation de la compétence en langue japonaise

Il existe en fait, de nombreuses méthodes d'évaluation des compétences en japonais langue étrangère. Parallèlement aux méthodes habituelles d'évaluation, on retiendra deux nouvelles méthodes.

La première méthode, dite « OPI » (Oral Proficiency Interview) vise à évaluer le niveau des apprenants de japonais langue étrangère dans les compétences orales. Elle nécessite que l'examineur ait reçu une formation spécifique d'un an et ait réussi les épreuves nécessaires à l'obtention du diplôme professionnel.

La seconde méthode « SPOT » (Simple Performance-Oriented Test), a été inventée au sein de l'université de Tsukuba au Japon (Kobayashi et al., 1996). Différemment d'« OPI », « SPOT » propose d'évaluer globalement les compétences orales et écrites. La particularité de ce test porte sur sa courte durée (dix minutes), et sa simplicité. Il est largement utilisé comme test de regroupement des apprenants par niveau, dans la plupart des établissements proposant l'apprentissage du japonais au Japon, aux Etats-Unis et en Australie.

7.5.2.3. Qu'est-ce que « SPOT » ?

Selon la méthode « SPOT », les apprenants écoutent un certain nombre de phrases les unes après les autres. Les apprenants suivent avec les yeux ces mêmes phrases écrites incomplètes : il faut remplir des blancs à l'aide de l'enregistrement perçu et des compétences en langue japonaise. « SPOT » est actuellement proposé sur deux niveaux de difficulté. Selon N. Kobayashi, le didacticien qui est à l'origine de ce test, « SPOT » n'est ni un test de discrimination phonétique ni un test qui évalue uniquement la compréhension orale. De récentes recherches sur la comparaison entre « SPOT » et d'autres types d'évaluation ont montré que « SPOT » n'est pas fait pour mesurer le niveau de compétence de points discrets, mais vise plutôt à évaluer globalement les compétences générales en langue cible.

Par ailleurs, Iwasaki (2002) a montré une forte corrélation entre les résultats obtenus selon « SPOT » et selon d'autres méthodes plus classiques ou même celle de l'« OPI ». Ce qui signifie plus ou moins que les résultats obtenus selon « SPOT » sont autant le reflet des compétences globales qu'orales en particulier.

Nous avons retenu la méthode « SPOT » pour évaluer les compétences des sujets testés en langue cible pour plusieurs raisons. D'abord parce que ce test peut être pratiqué par tout examinateur, ensuite parce que l'examineur n'a pas à créer lui-même ce test car il dépend des consignes et des sujets établis par l'université de Tsukuba. Enfin parce que les résultats obtenus par « SPOT » donnent une idée générale des compétences linguistiques des sujets testés. Ces résultats permettent donc de faire apparaître clairement et simplement des catégories d'apprenants, regroupés par niveau.

7.5.2.4. Protocole du test « SPOT »

Afin d'évaluer la compétence globale des sujets, nous avons ainsi effectué les deux tests « SPOT » A et B (« SPOT » A est plus difficile que « SPOT » B) pour 32 sujets français sélectionnés.

Tous les sujets ont passé les deux tests de « SPOT » l'un après l'autre. Ils ont passé d'abord le test B qui consiste en 60 phrases, ensuite le test A composé de 65 phrases. Chaque test prend environ sept minutes sans compter la durée d'un essai de 10 phrases et de l'explication du test. La durée totale de ces tests est d'environ 20 minutes. Entre les passations de ces deux tests, les sujets ont eu 5 minutes de pause.

Ces tests ont été effectués en général dans une salle de cours pendant le cours de japonais avec l'aide des enseignants. Pour ceux qui ont été absents lors du test, ils l'ont passé individuellement dans une salle calme.

7.5.2.5. Résultats du test « SPOT »

Afin de catégoriser ces 32 sujets en sous-groupes selon leur niveau de japonais, nous les avons répartis en deux catégories en accord avec les critères de classement utilisés à l'université de Tsukuba. C'est-à-dire que les sujets qui ont obtenu au-delà de

60% de taux de réussite comme moyenne des deux tests (« SPOT » A et B) sont classés dans la catégorie du niveau 2 qui est grosso modo équivalente du niveau intermédiaire³³. Quant aux sujets ayant obtenu une note en dessous de 60% de taux de réussite de deux tests « SPOT », ils sont classés dans la catégorie de niveau 1 qui est équivalente au niveau débutant.

C'est ainsi que nous avons établi deux catégories selon les niveaux acquis en langue japonaise : 16 sujets dans la catégorie du niveau 1 et 16 sujets dans la catégorie du niveau 2.

Selon les résultats obtenus, nous avons observé un large écart parmi ces 32 sujets testés. Pourtant, nous n'avons pas pu identifier de corrélation entre la compétence en japonais mesurée par « SPOT » et le nombre d'heures d'apprentissage de chaque sujet. Par contre, ceux qui ont déjà séjourné au Japon ont tous eu un score supérieur à 60% de taux de réussite en moyenne des deux tests. Afin de vérifier que ces résultats correspondent réellement à leur compétence en japonais, nous avons consulté le résultat universitaire (partiels oraux et écrits) de chaque sujet avec l'aide des enseignants responsables, et leurs résultats scolaires ont montré une forte corrélation avec les résultats tirés du test « SPOT ».

³³ Les termes tels que « débutant », « intermédiaire » portent un sens différent selon les classifications utilisées par les universités ou autres centres d'enseignement des langues dans chaque pays. C'est pourquoi nous avons utilisé les termes « niveau 1 » et « niveau 2 ». Néanmoins, le « niveau 1 » et le « niveau 2 » correspondent aux niveaux débutant et intermédiaire de la catégorisation utilisée à l'université de Tsukuba.

Tableau 37 : Les informations linguistiques des 32 sujets à partir de la 4^{ème} colonne : la moyenne de taux de réussite des deux tests « SPOT », la catégorie de niveau du japonais à laquelle ils appartiennent, leur âge, le nombre d'heures d'enseignement qu'ils ont suivi, le nombre d'heures de pratique orale, le nombre d'heure d'enseignement d'une autre langue étrangère, l'éventualité d'un séjour au Japon et leur possibilité de rencontre avec la langue japonaise dans leur vie quotidienne.

sujet	âge	genre	Score (SPOT)	catégorie	japonais	cours oral	autre spécialité	séjour au Japon	rencontre avec des Japonais
JS	23	M	88%	Niveau2	372h	132h	anglais 246h	Oui (5 mois)	amis (en japonais)
RM	23	M	87%	Niveau2	432h	120h	anglais 360h	Oui (2 mois)	amis (en japonais 50%)
JE	22	M	87%	Niveau2	336h	136h	anglais 336h	Oui (1 mois)	amis (en japonais)
VC	21	M	85%	Niveau2	336h	136h	anglais 336h	Oui (2 mois)	amis (en japonais)
HT	23	M	84%	Niveau2	322h	130h	anglais 276h	Oui (1 année)	amis (en japonais)
RA	22	M	78%	Niveau2	360h	132h	anglais 360h	Oui (2 mois)	amis (en japonais)
EK	21	M	78%	Niveau2	324h	84h	anglais 246h	Non	amis correspondance
JP	20	M	78%	Niveau2	432h	120h	anglais 360h	Non	amis (en japonais)
CR	18	F	72%	Niveau2	256h	92h	littérature fr	Non	amis (en japonais)
LS	20	M	70%	Niveau2	276h	84h	anglais 246h	Non	amis (en japonais 50%)
AA	20	M	70%	Niveau2	432h	120h	anglais 360h	Non	amis (en japonais)
LR	22	F	69%	Niveau2	432h	120h	anglais 360h	Non	uniquement dans la classe
EA	22	M	66%	Niveau2	276h	84h	anglais 246h	Oui (2 semaines)	amis (en japonais 50%)
GN	19	M	66%	Niveau2	276h	84h	anglais 246h	Non	amis (en japonais 50%)
TK	21	M	66%	Niveau2	432h	120h	anglais 360h	Non	amis (en japonais 50%)
FN	23	M	64%	Niveau2	432h	120h	anglais 360h	Non	uniquement dans la classe
PM	20	F	59%	Niveau1	432h	120h	anglais 360h	Non	uniquement dans la classe
GIB	20	M	58%	Niveau1	422h	122h	anglais 288h	Non	petite amie (en français)
FP	22	M	54%	Niveau1	276h	84h	anglais 246h	Non	uniquement dans la classe
AL	23	M	49%	Niveau1	276h	84h	anglais 246h	Non	uniquement dans la classe
RB	22	M	49%	Niveau1	276h	84h	anglais 246h	Non	amis (en français)
CB	27	M	45%	Niveau1	236h	94h	chimie	Oui (2mois)	petite amie (en français)
LV	21	F	45%	Niveau1	480h	140h	anglais 480h	Non	uniquement dans la classe
PP	22	M	41%	Niveau1	432h	120h	anglais 360h	Non	uniquement dans la classe
RBM	22	M	40%	Niveau1	432h	120h	anglais 360h	Non	uniquement dans la classe
BM	21	M	38%	Niveau1	432h	120h	anglais 360h	Non	uniquement dans la classe
LB	19	F	38%	Niveau1	276h	84h	anglais 246h	Non	amis (en français)
TNK	19	M	36%	Niveau1	276h	84h	anglais 246h	Non	uniquement dans la classe
GB	19	M	34%	Niveau1	276h	84h	anglais 246h	Non	uniquement dans la classe
VL	25	F	30%	Niveau1	176h	88h	mathématique	Non	uniquement dans la classe
DH	24	M	28%	Niveau1	142h	71h	technique	Non	amis (en français)
JC	22	M	26%	Niveau1	276h	84h	anglais 246h	Non	uniquement dans la classe

7.6. Les sujets français de niveau 1

7.6.1. Répartition des attitudes par rapport au hasard

Afin d'analyser précisément les résultats de cette première expérience, nous avons effectué une série de tests de χ^2 afin de comparer la distribution des résultats par rapport à une distribution théorique due au hasard.

Tableau 38 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues).

Attitude	χ^2	
Admiration (AD)	93,5	(ddl : 11, p.<.001)
Arrogance-impolitesse (AR)	104,4	(ddl : 11, p.<.001)
Autorité (AU)	413,0	(ddl : 11, p.<.001)
Déclaration (DC)	417,5	(ddl : 11, p.<.001)
Doute-incrédulité (DO)	186,3	(ddl : 11, p.<.001)
Évidence (EV)	111,9	(ddl : 11, p.<.001)
Exclamation de surprise (EX)	445,8	(ddl : 11, p.<.001)
Irritation (IR)	434,6	(ddl : 11, p.<.001)
Kyoshuku (KYO)	195,9	(ddl : 11, p.<.001)
Politesse-simple (PO)	196,8	(ddl : 11, p.<.001)
Question simple (QS)	199,6	(ddl : 11, p.<.001)
Sincérité-politesse (SIN)	77,4	(ddl : 11, p.<.001)

Ce premier résultat très important montre que les réponses correspondent à des distributions significativement différentes du hasard (cf. tableau 38).

Nous utiliserons ce critère à défaut de pouvoir utiliser un autre test de χ^2 , les conditions d'application n'étant pas remplies pour effectuer des comparaisons plus précises que celles du tableau 38.

Ce premier résultat montre donc que des apprenants français du niveau débutant (niveau 1) lorsqu'ils perçoivent des attitudes japonaises, répondent de manière cohérente. Les variations expressives introduites par les attitudes japonaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs français de ce group, même si ces identifications ne correspondent pas forcément au sens de ces attitudes en japonais. Nous analyserons maintenant le taux d'identification de chacune des attitudes.

7.6.2. Répartition des différentes longueurs et des attitudes par rapport au hasard

Pour observer l'effet de la longueur sur le choix des attitudes, nous avons effectué le test de χ^2 sur les résultats moyens par longueur, toutes attitudes confondues. Ce test donne un résultat significativement différent du hasard (avec $p < 0.05$ ddl = 33, $\chi^2 = 123.4$).

On peut donc constater que, comme pour les auditeurs japonais, la longueur de l'énoncé influence le choix des attitudes. Mais ce résultat ne nous indique pas à partir de quelle longueur d'énoncé les distributions deviennent significatives. Pour répondre à cette question, nous avons effectué un test de χ^2 sur les résultats moyens entre les longueurs de 1 et 2 mores, de 2 et 5 mores et de 5 et 8 mores, toutes attitudes confondues.

Le test de χ^2 sur les résultats moyens par longueur de 1 more et 2 mores, toutes attitudes confondues, donne un résultat significatif (ddl = 11, $\chi^2 = 19.9$ p. < 0.05).

Entre 2 et 5 mores, nous avons également identifié un effet significatif de la longueur (ddl = 11, $\chi^2 = 73.5$, p. < 0.05).

En ce qui concerne la longueur de 5 et de 8 mores, toutes attitudes confondues, le test χ^2 donne un effet significatif (ddl = 11, $\chi^2 = 23.1$, p. < 0.05).

Ces résultats démontrent que l'effet de la longueur de l'énoncé existe entre toutes les longueurs des phrases choisies (1, 2, 5 et 8 mores)

7.6.3. Effet de l'accent lexical

Quant à l'influence de la position de l'accent lexical sur les attitudes choisies, nous avons, dans un premier temps, effectué un test de χ^2 sur les résultats moyens des réponses toutes attitudes confondues, et toutes longueurs confondues. L'hypothèse nulle est que l'accent ne fait pas changer les réponses des sujets.

Le test de χ^2 pour ces données est significatif (ddl = 33, $\chi^2 = 60.4$, p. < 0.01), mais ces données contiennent forcément l'influence de la longueur de l'énoncé. C'est pourquoi un autre test de χ^2 a été conduit sur les résultats moyens pour chaque position d'accent avec uniquement les phrases de 8 mores, toutes attitudes confondues. Ce test ne donne pas d'effet significatif (ddl = 33, $\chi^2 = 18.3$, p. < 0.05). On peut donc rejeter l'effet de la position de l'accent comme facteur d'influence sur le choix des attitudes par des auditeurs français de niveau 1.

7.6.4. Taux d'identification des attitudes

La figure 85 présente l'histogramme du taux d'identification des 12 attitudes japonaises pour les 16 auditeurs français de niveau 1 (niveau débutant) en comparant à celui des français naïfs et des natifs.

Les résultats montrent que toutes les attitudes sauf l'attitude de *kyoshuku* ont été au dessus du seuil du hasard (8,4%).

A leur niveau de la compétence du japonais, le taux global d'identification des attitudes japonaises est presque égal à celui des sujets français n'ayant aucune connaissance du japonais (35% chez les Français naïfs contre 37,5% chez les apprenants français du niveau 1). Ce score est nettement inférieur à celui des sujets japonais (55%).

À partir de la figure 85, selon le degré de reconnaissance, nous avons constitué les trois groupes d'attitudes suivants :

Attitudes très bien reconnues : il s'agit d'attitudes ayant reçu un taux d'identification sans confusion importante avec une autre attitude. Il s'agit d'*admiration*, *autorité*, *irritation*, *exclamation de surprise* et *déclaration*.

Attitude reconnues, mais avec des confusions importantes : *doute-incrédulité*, *évidence*, *politesse-simple*, *question simple*, *arrogance-impolitesse* et *sincérité-politesse*. Ces attitudes ont reçu un taux d'identification au dessus du seuil du hasard, mais sont aussi confondues de manière importante avec une ou plusieurs autres attitudes.

Attitude qui n'a pas été reconnue : *kyoshuku*. Cette attitude n'est pas reconnue par les sujets français de niveau 1.

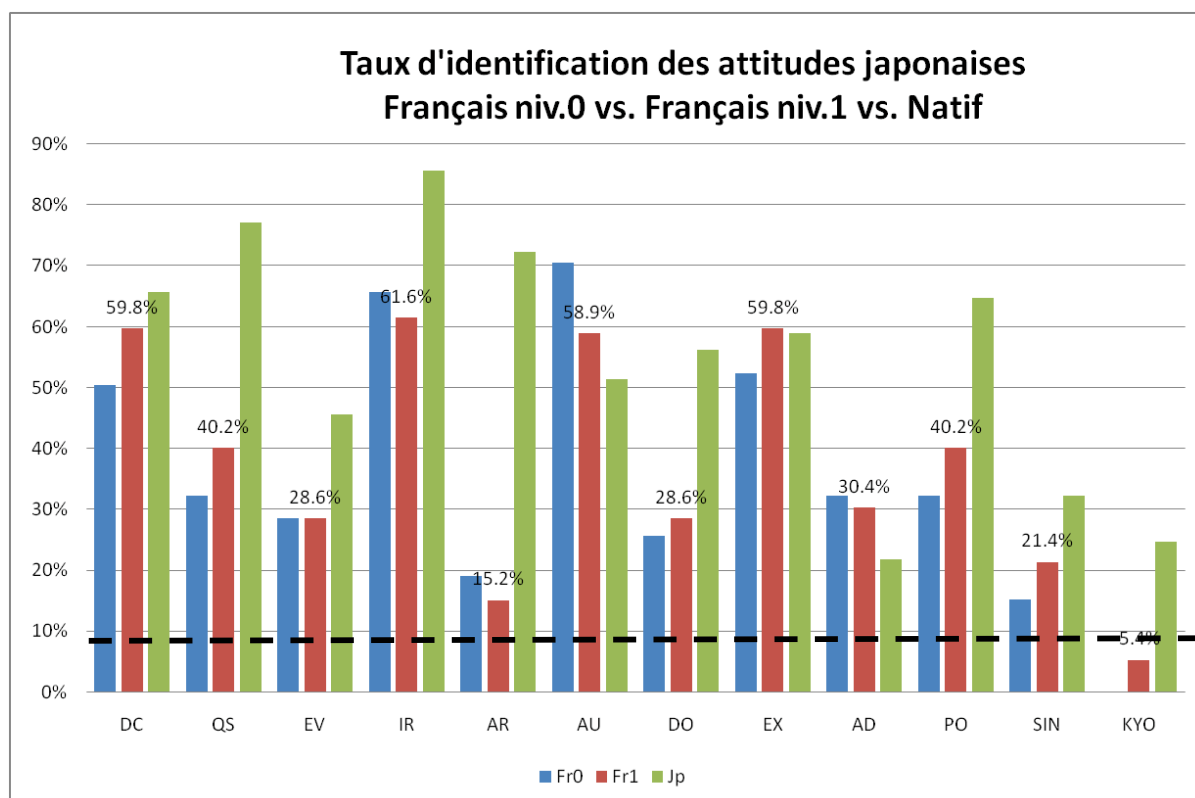


Figure 85 : Taux d'identification de chaque attitude chez les 16 auditeurs français de niveau 1 vs. Français naïfs vs. natifs. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse), SIN (sincérité-politesse) et KYO (kyoshuku).

Bien que *arrogance-impolitesse* soit reconnue, le taux de reconnaissance reste très faible (15,2%), et les deux attitudes le moins bien perçues chez les apprenants français du niveau 1 sont *kyoshuku* et *arrogance-impolitesse*. Il est à noter que ce faible taux de reconnaissance pour ces deux attitudes a déjà été observé dans les résultats obtenus

auprès des Français naïfs. De ce fait, nous n'avons pas pu identifier l'effet d'apprentissage sur ces deux attitudes.

Contrairement à ces deux attitudes de politesse mal reconnues, l'attitude de *sincérité-politesse* a été reconnue, mais elle montre aussi un taux de reconnaissance assez faible comme les résultats de Français naïfs (21.4% contre 15.2%).

Les auditeurs de niveau 1 ont très bien reconnu les attitudes d'*autorité*, d'*irritation* comme les Français naïfs. *Autorité* surtout est mieux reconnue par les Français du niveau 1 que par les Japonais (58% contre 51%).

On peut également observer un grand écart de taux d'identification de *question-simple* entre les sujets japonais et les sujets français du niveau1. Contrairement aux sujets japonais qui l'ont bien reconnue (77%), le taux d'identification de cette attitude était nettement plus faible chez les auditeurs français du niveau1 (40.2%). Nous reviendrons ultérieurement sur ce point dans le paragraphe concernant la confusion des attitudes.

Quant à *admiration*, les auditeurs français de niveau1 l'ont reconnue à 30.4%. Ce taux d'identification est supérieur à celui des auditeurs japonais (22%). Cette tendance avait déjà émergée dans les résultats des Français naïfs (32%)

7.6.5. Comportement perceptif des sujets français niveau 1 en analyse de correspondance

Afin d'observer le comportement perceptif des apprenants français du niveau débutant, nous avons calculé la distance des points cognitifs de 12 attitudes par rapport aux points stimuli selon l'analyse de correspondance comme pour dans les chapitres précédents.

Selon l'AC, quatre premières dimensions expliquent 79% de variances.

Maintenant, voyons la figure 86. Dans cette figure, les points stimuli se situent très proche de points perceptifs sauf *arrogance-impolitesse* et *kyoshuku*. La perception de ces attitudes montre la distance longue par rapport aux points de distribution des stimuli. Cela indique que leur reconnaissance est très faible. Il est à noter également que le comportement perceptif des apprenants débutants pour l'attitude de *kyoshuku* se situe près d'*irritation*, et loin de *sincérité-politesse* ou de *politesse-simple* comme les résultats des Français naïfs, et cela explique qu'à ce niveau de compétence de la langue cible, les auditeurs ne reconnaissent pas *kyoshuku* comme expression de politesse, et que c'est interprété comme *irritation*.

Quant à *arrogance-impolitesse*, le point de stimulus pour cette attitude est localisé dans un endroit isolé, c'est-à-dire, cette attitude est mal reconnue, mais les Français du niveau débutants ont du mal à répondre à quoi cette expression ressemble.

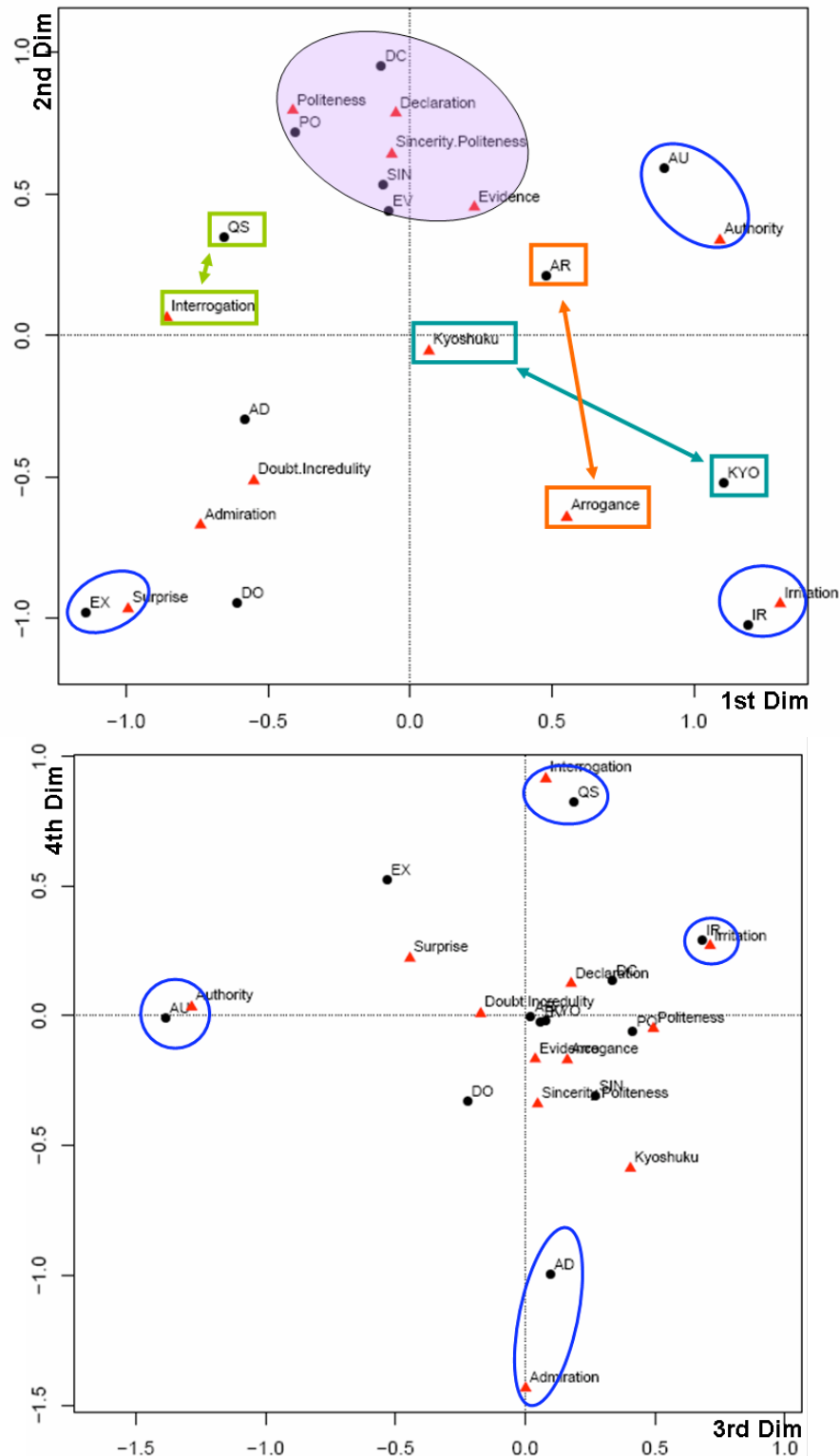


Figure 86 : Comportement perceptif de 16 auditeurs français du niveau 1 (débutant) pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les étiquettes avec le triangle rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles.

En ce qui concerne *question-simple*, la distance entre le point de stimulus et point perceptif n'est pas aussi longue que les deux attitudes citées. Nous nous rappelons que les auditeurs français du niveau 0 montraient une confusion importante avec *déclaration*, mais cette confusion n'est pas observée ici selon l'analyse de correspondance.

Nous avons identifié également une catégorie dans laquelle des attitudes sont confondues. Il s'agit de *déclaration*, *évidence*, *politesse-simple* et *sincérité-politesse*. Ces attitudes confondues partagent la forme affirmative et les expressions de politesse.

Nous allons analyser plus concrètement les phénomènes de confusions entre les attitudes dans le paragraphe suivant.

7.6.6. Confusion entre attitudes

En comparant le taux d'identification de chacune des attitudes, l'attitude de *kyoshuku* n'a pas été reconnue au dessous du seuil du hasard. Par la suite, l'analyse de correspondance a démontré la faible reconnaissance de *Kyoshuku*, et *arrogance-impolitesse* d'une part, et la catégorie du similaire comportement pour les six attitudes : *déclaration*, *évidence*, *politesse-simple* et *sincérité-politesse* d'autre part.

Tableau 39 : Matrice de confusion des 16 auditeurs français, niveau 1. Signification des étiquettes : DC (*déclaration*), QS (*question-simple*), EV (*évidence*), IR (*irritation*), AR (*arrogance-impolitesse*), AU (*autorité*), DO (*doute-incrédulité*), EX (*exclamation de surprise*), AD (*admiration*), PO (*politesse-simple*), SIN (*sincérité-politesse*) et KYO (*kyoshuku*).

Attitudes perçues	Attitudes présentées											
	AD	AR	AU	DC	DO	EV	SU	IR	KYO	PO	QS	SIN
Admiration	30%	0%	0%	0%	15%	0%	3%	1%	1%	1%	1%	3%
Arrogance	3%	15%	0%	1%	21%	7%	0%	12%	26%	0%	1%	1%
Autorité	0%	13%	59%	2%	0%	4%	0%	7%	21%	1%	2%	3%
Déclaration	7%	26%	15%	60%	0%	21%	3%	1%	3%	20%	19%	18%
Doute-incrédulité	4%	5%	2%	2%	29%	10%	19%	4%	2%	5%	7%	7%
Evidence	6%	19%	10%	10%	1%	29%	0%	4%	8%	6%	3%	14%
Surprise	15%	0%	1%	0%	27%	5%	60%	6%	0%	6%	5%	2%
Irritation	3%	12%	2%	1%	2%	0%	1%	62%	33%	0%	1%	4%
Kyoshuku	8%	2%	0%	1%	1%	1%	1%	4%	5%	4%	1%	9%
Politesse	9%	2%	2%	14%	0%	12%	0%	0%	0%	40%	15%	16%
Question	5%	2%	0%	4%	5%	4%	14%	0%	0%	4%	40%	3%
Sincérité-Politesse	9%	5%	10%	6%	0%	6%	0%	0%	2%	13%	5%	21%

Maintenant nous allons voir le taux de confusion pour chaque attitude présentée (cf. tableau 39 et figure 87). D'après les résultats présentés, cinq attitudes : *déclaration*, *exclamation de surprise*, *irritation* et *autorité* sont les attitudes bien reconnues chez les sujets du niveau 1. Surtout *irritation* (61,6%) et *autorité* (58,9%) ont été très bien perçues sans aucune confusion particulière chez les auditeurs du niveau 1.

L'attitude d'*admiration* est reconnue sans avoir une confusion particulière, mais le taux d'identification reste assez faible (30,4%). Cela explique que les auditeurs hésitent avant de se décider pour le choisir, mais pourtant, il n'y avait pas d'autre choix qui correspondait mieux qu'*admiration*.

Concernant les attitudes qui se trouvent dans la même catégorie perceptive : *doute-incrédulité* et *exclamation de surprise* ont montré la confusion réciproque chez les auditeurs comme chez les Français de niveau 0. Cependant les sujets de niveau 1 ont également montré une confusion importante entre *doute-incrédulité* et *arrogance-impolitesse* alors que les Français de niveau 0 n'ont pas eu cette confusion.

Les trois attitudes : *évidence*, *politesse-simple* et *sincérité-politesse* ont été confondu avec l'attitude qui n'affirme aucune intention particulière (déclaration) sans avoir de confusion particulière entre les expressions de la politesse. Rappelons que les sujets de niveau 0 ont confondu les attitudes de politesse, notamment *sincérité-politesse*, avec d'autres expressions de la politesse. Il est possible que les sujets de niveau 1 aient été influencés par le contenu linguistique de phrases qui sont neutres.

Question-simple, elle, a été confondue, comme les Français naïfs avec *déclaration*. Sur ce point, les sujets de niveau 1 ont réclamé, lors de l'entretien, leur difficulté de la percevoir sans particule interrogative « ka » en fin de phrase qui marque l'interrogation. Nous pouvons considérer que les Français qui ne connaissent pas la langue japonaise ont tendance à confondre ces deux modalités, puis lors du début d'apprentissage de cette langue, les Français repère cette « question simple » grâce à un indice de morphème fonctionnel marquant de l'interrogation.

En ce qui concerne les attitudes mal identifiées, *arrogance-impolitesse* a été confondu avec *déclaration* ou *évidence*. Quant à *kyoshuku*, elle a été reconnue comme *irritation*, *arrogance* ou *autorité*. Il est à noter que les Français de niveau 0 ont montré la même tendance de perception pour l'attitude de *kyoshuku*. Nous pouvons donc suggérer que l'effet d'apprentissage de langue cible n'a pas été identifié pour ces attitudes culturelles chez les apprenants débutants.

La première cause d'un tel comportement pour *arrogance-impolitesse* et *kyoshuku* peut être un manque de didactique de la prosodie qui vise l'apprentissage de la prosodie pertinente des politesses. La didactique de japonais langue étrangère insiste l'importance de la forme lexicale afin d'exprimer des degrés de politesse (par exemple, forme polie, forme honorifique ou forme humble etc.) dans certains contextes spécifiques et culturels. Néanmoins la didactique de la prosodie réalisée avec laquelle cette forme grammaticale est employée nous semble d'être encore négligée. Par conséquent, les apprenants du japonais ne semblent pas être sensibles à percevoir la subtilité de degré de politesse uniquement exprimé par la prosodie.

La seconde raison qu'on peut proposer est l'effet du contenu sémantique. Puisque les phrases que nous avons sélectionnées sont le plus neutre possible au niveau sémantique, certains auditeurs ont senti la difficulté d'identifier les attitudes qui

expriment la subtilité de degré de politesse. Souvent leurs choix sont figés par le contenu linguistique exprimé.

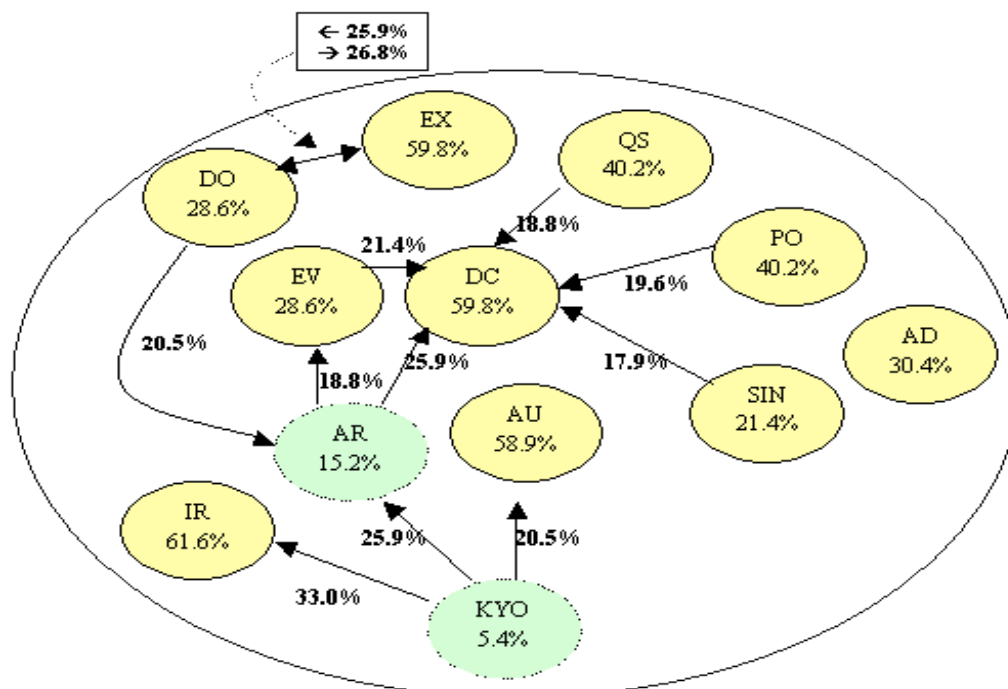


Figure 87 : Graphe de confusions chez les auditeurs français du niveau 1, le chiffre encadré indique le taux de confusion. L'autre chiffre en dessous de l'étiquette d'attitude représente le taux d'identification. Les attitudes qui sont dans les cercles verts sont les attitudes mal reconnues. Le chiffre qui se trouve à côté de la flèche indique le taux de confusion importante.

7.7. Les sujets français de niveau 2

7.7.1. Répartition des attitudes par rapport au hasard

De la même manière que pour les résultats des sujets du niveau 1, nous avons effectué un test de χ^2 afin de comparer la distribution des résultats par rapport au hasard (cf. tableau 40).

Les résultats du test de χ^2 nous montrent que la distribution de toutes les attitudes est significativement différente du hasard.

Ce premier résultat montre donc que des auditeurs français du niveau 2 lorsqu'ils perçoivent des attitudes japonaises, répondent de manière cohérente. Les variations expressives introduites par les attitudes japonaises semblent donc renvoyer aux mêmes significations pour les différents auditeurs du niveau 2, même si ces identifications ne correspondent pas forcément au sens de ces attitudes en japonais. Nous analyserons plus loin ces confusions.

Tableau 40 : Résultats du test de χ^2 sur les résultats moyens des réponses par attitude (toutes longueurs et toutes positions d'accent lexical confondues) par rapport au hasard.

Attitude	χ^2	
Admiration (AD)	58,79	(ddl : 11, p.<.001)
Arrogance-impolitesse (AR)	141,07	(ddl : 11, p.<.001)
Autorité (AU)	228,71	(ddl : 11, p.<.001)
Déclaration (DC)	363,93	(ddl : 11, p.<.001)
Doute-incrédulité (DO)	267,29	(ddl : 11, p.<.001)
Évidence (EV)	90,29	(ddl : 11, p.<.001)
Exclamation de surprise (EX)	476,86	(ddl : 11, p.<.001)
Irritation (IR)	657,93	(ddl : 11, p.<.001)
Kyoshuku (KYO)	192,50	(ddl : 11, p.<.001)
Politesse-simple (PO)	206,00	(ddl : 11, p.<.001)
Question simple (QS)	311,00	(ddl : 11, p.<.001)
Sincérité-politesse (SIN)	41,64	(ddl : 11, p.<.001)

7.7.2. Effet de la longueur des énoncés

Pour observer l'effet de la longueur pour le choix des attitudes, nous avons effectué un test de χ^2 sur les résultats moyens, par longueur, toutes attitudes confondues. Il donne un résultat significativement différent du hasard (ddl = 33, χ^2 = 119.9 p. < 0.01).

On peut donc constater que la longueur de l'énoncé influence le choix de l'attitude. Mais ce résultat ne nous permet pas de savoir s'il existe des différences d'identification en fonction de la longueur des énoncés. Pour répondre à cette question, nous avons effectué le test de χ^2 sur les résultats moyens en les longueurs de 1 et 2 mores, de 2 et 5 mores et 5 et 8 mores, toutes attitudes confondues.

Les résultats montrent que l'effet de la longueur entre les énoncés constitués de 1 et de 2 mores donne un résultat significatif (ddl = 11, χ^2 = 29.2, p. < 0.01). Par contre, entre 2 et 5 mores, on n'a pas observé un effet de la longueur (ddl = 11, χ^2 = 11.5, p. > 0.01).

En ce qui concerne les longueurs de 5 et de 8 mores, nous avons observé un effet significatif de longueur avec toutes attitudes confondues (ddl = 11, χ^2 = 32.9, p. < 0.01).

Ces résultats montrent que l'effet de la longueur de l'énoncé existe, cependant il n'influence pas sur les réponses données aux mots composés de 2 mores et de 5 mores.

7.7.3. Effet de l'accent lexical

Quant à l'influence de la position de l'accent lexical sur les attitudes choisies, nous avons mené un test de χ^2 sur les résultats moyens des réponses toutes attitudes confondues, et toutes longueurs confondues dans un premier temps, pour chaque position d'accent lexical. L'hypothèse nulle est que l'accent ne fait pas changer les réponses des sujets.

Le test de χ^2 pour ces données est significatif (avec $p. < 0.01$, $ddl = 33$, $\chi^2 = 59.9$), mais ces données contiennent forcément l'influence de la longueur de l'énoncé. C'est pourquoi un autre test de χ^2 a été conduit sur les résultats moyens pour chaque position d'accent avec une uniquement les phrases de 8 mores. Ce test mené sur toutes les attitudes ne montre pas d'effet significatif ($ddl = 11$, $\chi^2 = 26.6$, $p. > 0.01$). Ces résultats soutiennent l'hypothèse nulle. On peut donc rejeter l'effet de la position d'accent comme un facteur ayant une influence sur le choix des attitudes.

7.7.4. Taux d'identification des attitudes

Tout d'abord, le taux global d'identification des attitudes japonaises chez les auditeurs français de niveau 2 était identique à celui des auditeurs français de niveau 1 (38% pour deux groupes de différent niveau d'apprentissage). Ce taux global d'identification des attitudes japonaises est presque identique à celui des Français de niveau 0 (38% chez les Français de niveau 1 et 2 contre 35% chez les Français de niveau 0), mais il est nettement inférieur à celui des auditeurs japonais (55%).

Cependant, il est difficile d'identifier un effet d'apprentissage avec ce résultat seul. Nous avons donc calculé le taux d'identification des attitudes en pourcentage afin d'envisager le degré d'identification des attitudes japonaises auprès des auditeurs français. Cela nous a permis de regrouper les attitudes.

La figure 88 présente le taux d'identification des attitudes parmi les 16 auditeurs français du niveau 2 (niveau intermédiaire du japonais).

Selon cette figure, trois attitudes de la politesse : *kyoshuku*, *sincérité-politesse* et *arrogance-impolitesse* ont été mal reconnues. Notamment, le taux d'identification de *kyoshuku* était en dessous du seuil du hasard, donc cette attitude n'a pas reconnu. Nous voyons par ces résultats que les attitudes de la politesse japonaise sont encore très difficiles à identifier malgré leur niveau de japonais.

En revanche, comme les Français du niveau 0 et du niveau 1, les sujets du niveau 2 ont très bien identifié *autorité* et *irritation* sans confusion particulière. Nous pouvons considérer qu'en générale, les auditeurs français peuvent identifier ces deux attitudes sans avoir le rapport avec l'apprentissage de langue.

Les sujets du niveau 2 ont identifié *question-simple* avec un taux d'identification nettement mieux par rapport aux Français naïfs et débutants. De même, ils reconnaissent nettement mieux l'attitude de *doute-incrédulité* par rapport aux Français naïfs ou les apprenants débutants. Cette amélioration du taux de reconnaissance pour ces deux attitudes implique que l'effet d'apprentissage existe sur la perception de ces attitudes. Il est possible qu'à partir du niveau intermédiaire, les apprenants puissent identifier ces attitudes par des indices prosodiques comme c'est le cas chez les Japonais natifs.

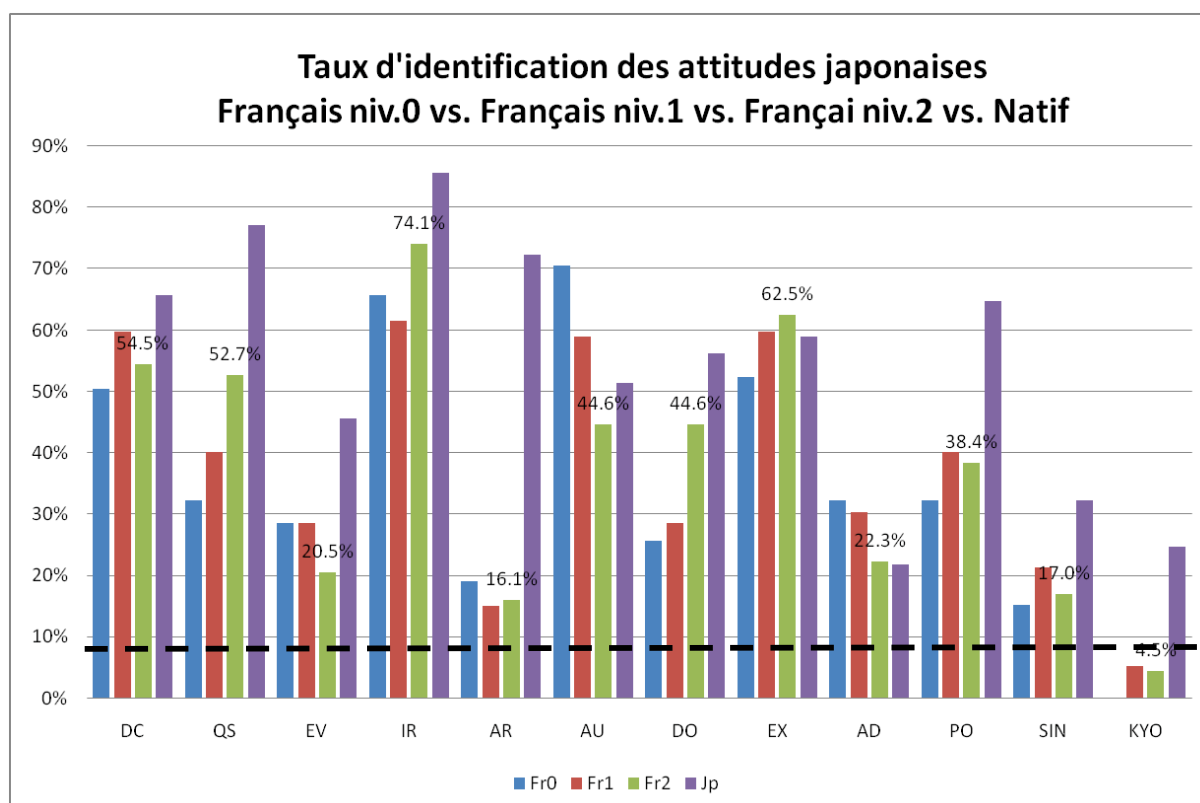


Figure 88 : Comparaison du taux d'identification des attitudes japonaises chez les Français du niveau 0 vs. Français du niveau 1 vs. Français du niveau 2 vs. natifs. La ligne pointillée indique le seuil du hasard (8,4%). Signification des étiquettes : DC (déclaration), QS (question simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse), SIN (sincérité-politesse) et KYO (kyoshuku).

7.7.5. Analyse des correspondances pour les sujets français de niveau 2

Selon l'analyse de correspondance, quatre premières dimensions expliquent 83% de variances.

La figure 89 illustre la distribution des points de stimulus et des points perceptifs pour les sujets du niveau 2. Dans la figure 89 (haut), les points stimuli se situent très proche de points perceptifs sauf *arrogance-impolitesse* et *kyoshuku* comme les sujets du niveau 0 et 1. La perception de ces attitudes montre la distance longue par rapport aux points de distribution des stimuli. Cela indique que même s'ils parlent bien en japonais sans avoir le problème de communication, et malgré leur connaissance sur la culture japonaise, leur reconnaissance pour ces attitudes reste très faible. Il est important de noter que le point du stimulus pour *kyoshuku* se situe juste à côté d'*irritation* et d'*autorité*. Par contre, le point perceptif pour cette attitude se trouve dans la catégorie des expressions de politesse comme pour les résultats des apprenants débutants et naïfs. Cela nous conduit à conclure que les auditeurs français reconnaissent globalement la notion de *kyoshuku* sur le plan cognitif même si ils ne reconnaissent pas cette attitude

sur le plan expressive. Ce résultat nous permet de conclure qu'à ce niveau de compétence de la langue cible, les auditeurs ne reconnaissent toujours pas l'expression de *kyoshuku* comme expression de la politesse, et que c'est interprété comme *irritation*.

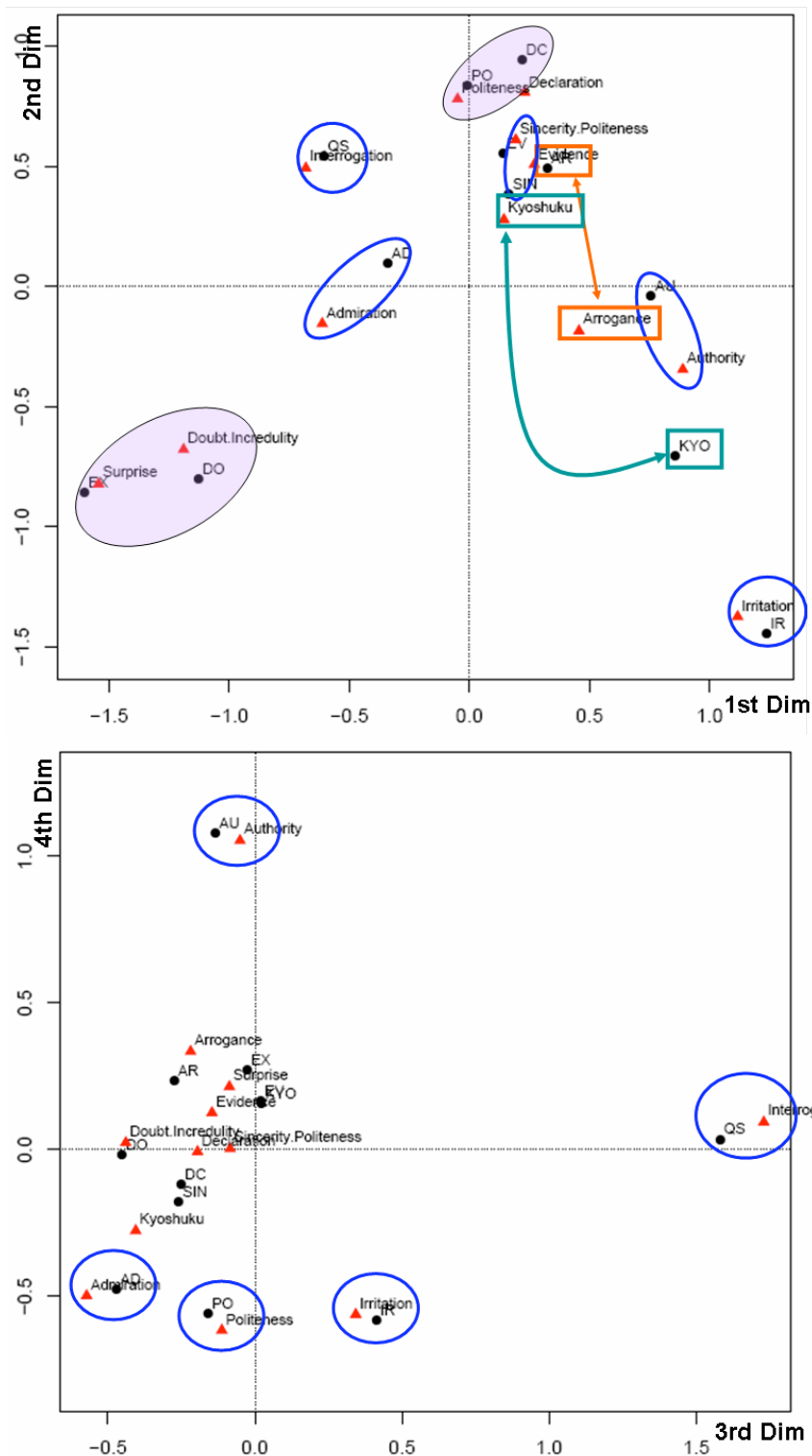


Figure 89 : Comportement perceptif de 16 auditeurs français du niveau 2 (intermédiaire) pour les 12 attitudes japonaises présenté en deux dimensions (la figure du haut présente les dimensions 1 et 2. Celle du bas les dimensions 3 et 4). Les

étiquettes en rouge correspondent aux distributions de chaque stimulus, les étiquettes abrégées en noir indiquent les distributions de chaque catégorie d'attitude. Les cercles bleus indiquent la bonne perception des attitudes, puis les cercles remplis par la couleur violette signifient des catégories (ou confusions) des attitudes qui sont dans ces cercles.

Par ailleurs, l'attitude d'*admiration*, *politesse*, *irritation*, *autorité* et *question-simple* ont été bien identifiées par l'indice de la courte distance entre le point de stimulus et le point perceptif correspondant (voir figure 89 bas). Notamment *irritation*, *autorité* et *question-simple* sont localisés dans une zone bien isolée par rapport aux autres, et cela indique que ce sont les attitudes particulièrement bien discriminées.

Cette figure montre également certaines confusions (ou similaire comportement perceptif) qui sont présentées en cercle violet (entre *politesse-simple* et *déclaration* et entre *doute-incrédulité* et *exclamation de surprise*).

Nous allons analyser plus concrètement les phénomènes de confusions entre les attitudes dans le paragraphe suivant.

7.7.6. Analyse des confusions

Nous venons de voir la comparaison du taux d'identification de chacune des attitudes, ainsi que le degré de perception et des catégories du similaire comportement selon l'analyse de correspondance. Maintenant, nous allons nous pencher à l'analyse des confusions.

Comme nous avons vu dans le paragraphe précédent, les quatre attitudes : *autorité*, *irritation* et *question-simple* sont reconnues sans confusion particulière. *Sincérité-politesse* et *admiration* sont également reconnues sans confusion particulière, mais le taux de reconnaissance des deux attitudes reste faible (22,3% pour *admiration* et 17% pour *sincérité-politesse*). Ces deux attitudes sont reconnues difficilement, mais pourtant les auditeurs du niveau 2 ne font pas d'assimilation avec d'autre attitude.

Par contre, les quatre attitudes : *déclaration*, *doute-incrédulité*, *exclamation de surprise* et *politesse-simple* ont été confondues d'une façon importante avec des autres attitudes.

Tout d'abord, *déclaration* qui a un bon taux de reconnaissance montre une légère confusion avec *sincérité-politesse*. Ensuite, comme on a déjà vu dans l'analyse de correspondance, deux attitudes *doute-incrédulité* et *exclamation de surprise* ont montré la confusion réciproque. Nous vous rappelons que les résultats des auditeurs français du niveau 0 et du niveau 1 ont montré aussi cette confusion réciproque. Mais le taux d'identification de l'attitude de *doute-incrédulité* chez les auditeurs de niveau 2 était nettement supérieur à celui des auditeurs du niveau 0 et du niveau 1 (44.6% pour les auditeurs de niveau 2 contre 25.7% pour les sujets de niveau 0, 28.6% pour les sujets de niveau 1).

Tableau 41 : Matrices de confusion des 16 français niveau 2. Signification des étiquettes : DC (déclaration), QS (question-simple), EV (évidence), IR (irritation), AR (arrogance-impolitesse), AU (autorité), DO (doute-incrédulité), EX (exclamation de surprise), AD (admiration), PO (politesse-simple), SIN (sincérité-politesse) et KYO (kyoshuku).

Attitudes perçues	Attitudes présentées											
	AD	AR	AU	DC	DO	EV	SU	IR	KYO	PO	QS	SIN
Admiration	22%	2%	0%	0%	12%	1%	6%	1%	1%	4%	1%	8%
Arrogance	4%	16%	13%	0%	8%	13%	1%	7%	23%	1%	0%	12%
Autorité	1%	8%	45%	1%	0%	5%	1%	13%	19%	0%	1%	4%
Déclaration	6%	34%	13%	54%	0%	24%	0%	1%	3%	25%	10%	14%
Doute-incrédulité	7%	3%	0%	1%	45%	4%	19%	0%	2%	2%	4%	3%
Evidence	6%	17%	8%	10%	1%	21%	0%	2%	6%	9%	4%	10%
Surprise	11%	0%	0%	0%	24%	1%	63%	1%	1%	2%	8%	1%
Irritation	4%	2%	6%	0%	6%	0%	2%	74%	36%	0%	0%	4%
Kyoshuku	13%	4%	3%	3%	1%	3%	0%	1%	4%	4%	1%	10%
Politesse	16%	4%	2%	11%	1%	9%	0%	0%	0%	38%	10%	14%
Question	1%	2%	1%	2%	3%	10%	9%	0%	0%	4%	53%	3%
Sincérité-Politesse	9%	8%	11%	19%	0%	10%	0%	0%	5%	12%	9%	17%

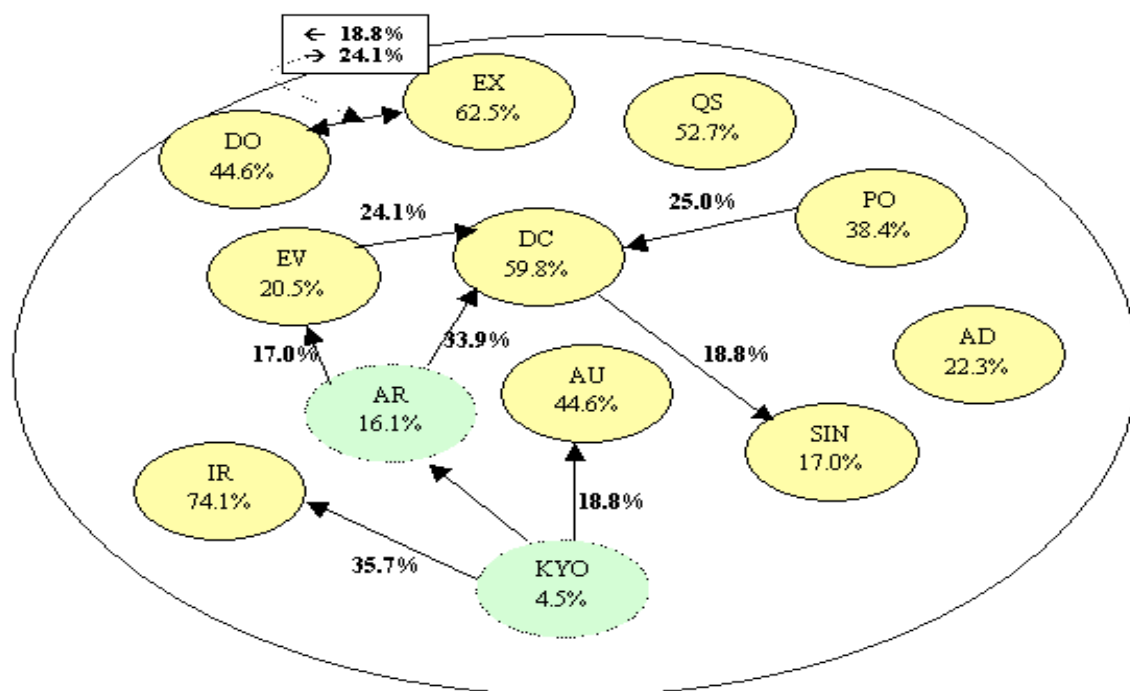


Figure 90 : Graphe de confusions chez les 16 apprenants français du niveau 2. Le chiffre encadré indique le taux de confusion en pourcentage. L'autre chiffre en dessous de l'étiquette d'attitude représente le taux d'identification. Les attitudes qui sont dans les cercles verts sont les attitudes mal reconnues. Le chiffre qui se trouve à côté de la flèche indique le taux de confusion importante.

L'attitude d'*évidence* et *politesse-simple* ont tous les deux été confondue avec la simple déclaration. Selon l'interview effectuée, les sujets du niveau 2 ont répondu qu'ils ont choisi *déclaration* lorsque les stimuli perçus ne correspondaient à rien parmi les 12 attitudes proposées.

Deux attitudes mal reconnues : *arrogance-impolitesse* et *kyoshuku*, n'ont pas été reconnues chez les auditeurs du niveau 2 comme chez les auditeurs du niveau 0 et niveau 1. L'attitude d'*arrogance-impolitesse* a montré une confusion importante, de la même manière que les résultats des sujets de niveau 1, avec *déclaration* et *évidence*.

En ce qui concerne l'attitude de *kyoshuku*, les auditeurs de niveau 2, eux aussi, ont perçu cette attitude comme *irritation*, *arrogance-impolitesse* ou *autorité*. Ce résultat explique la difficulté de percevoir cette expression de la politesse même pour ceux qui ont séjourné au Japon. Selon ces résultats obtenus, l'effet d'apprentissage de langue japonaise était totalement absent pour la bonne identification de cette attitude même chez les auditeurs français qui maîtrisent bien la langue japonaise.

7.8. Conclusion

Après deux tests perceptifs effectués auprès d'auditeurs japonais et d'auditeurs français non exposés au japonais dans le chapitre 3 et 4, nous nous sommes concentrés, dans ce dernier chapitre de la thèse, sur la perception relative des attitudes japonaises par d'apprenants français du japonais au niveau 1 (débutant) vs. des apprenants français au niveau 2 (intermédiaire) afin d'évaluer l'effet de l'apprentissage de la langue japonaise sur la perception des attitudes en comparant les résultats des apprenants français du japonais à ceux de Japonais natifs et de Français non exposés au japonais.

Pour l'objectif ci-dessus, nous avons utilisé 12 attitudes représentatives de la langue japonaise qui ont été validées auprès des 15 auditeurs japonais *supra*. Avec ce corpus, nous avons effectué le même test perceptif avec ces 12 attitudes japonaises auprès des apprenants français du japonais.

Dans un premier temps, nous avons vérifié que la distribution des réponses était différente du hasard pour deux groupes de sujets.

Ensuite, un test de χ^2 montre que les résultats des deux groupes n'ont pas montré d'effet significatif de la position d'accent lexical, pour qui la variation de F0 est reliée à la prosodie linguistique ou affective.

Un autre test de χ^2 montre un effet significatif de la longueur des énoncés pour les deux groupes de sujets. Concernant les sujets du niveau 1, nous avons observé un effet significatif de la longueur sur la distribution des attitudes entre toutes les longueurs de phrases choisies (c'est-à-dire entre les mots isolés monosyllabiques et dissyllabiques, entre les phrases de 2 et de 5 mores, entre les phrases de 5 et de 8 mores). Les sujets du

niveau 2, par contre, n'ont pas eu d'influence significative de la longueur entre les phrases de 2 et de 5 mores, mais c'est plutôt entre les mots isolés monosyllabiques et dissyllabiques et également entre 5 et 8 mores qu'on trouve un effet de la longueur sur la distribution des attitudes.

Selon l'analyse de correspondance, deux expressions portant un rapport avec la politesse, *arrogance-impolitesse* et *kyoshuku* ont été mal reconnues chez tous les apprenants français du japonais (niveau 1 et niveau 2).

Par la suite, l'analyse de confusion a dévoilé que tous les apprenants français testés ont perçu l'attitude de *kyoshuku* comme *irritation*, *arrogance-impolitesse* ou *autorité* comme les Français non exposés au japonais.

De même, tous les apprenants français testés ont confondu l'attitude d'*arrogance-impolitesse*, d'une manière importante, avec *évidence* et *déclaration*. Nous nous rappelons que cette attitude a été mal reconnue chez les Français naïfs avec le taux de reconnaissance en étant très faible (19%).

Une autre attitude qui a un rapport avec la politesse japonaise : *sincérité-politesse*, est généralement mal reconnue chez tous les auditeurs français (c'est-à-dire, le niveau 0, 1 et 2). Avec la compétence du niveau intermédiaire, les sujets ne montrent aucune confusion particulière pour la perception de cette attitude. Cela signifie qu'ils n'arrivent pas à bien identifier cette attitude, mais qu'ils ne trouvent pas d'attitude correspondante parmi les choix proposés.

En ce qui concerne ces trois attitudes culturelles (*kyoshuku*, *arrogance-impolitesse* et *sincérité-politesse*), l'effet d'apprentissage était donc absent.

En revanche, sur certaines attitudes, nous avons observé les effets d'apprentissage de la langue japonaise.

Les auditeurs français étant débutant dans l'apprentissage (niveau 1) ont reconnu *question-simple*, mais ils l'ont confondu avec *déclaration*. Ce résultat ressemble à celui des Français de niveau 0. Par contre, les auditeurs français ayant acquis une bonne connaissance de la langue japonaise (niveau 2) n'ont pas montré cette confusion comme dans les résultats des Japonais natifs, et nous y voyons l'effet d'apprentissage.

Bien que tout les apprenants français du japonais, comme les Français non exposés au japonais, aient montré une confusion réciproque entre *doute-incrédulité* et *exclamation de surprise*, mais l'attitude de *doute-incrédulité* a été nettement mieux identifiée chez les sujets de niveau 2 que chez les sujets de niveau 0 et de niveau 1 (44,6% pour les auditeurs de niveau 2 contre 25,7% pour les sujets de niveau 0, 28,6% pour les sujets de niveau 1). Nous y voyons, ici aussi, l'effet d'apprentissage.

8. Conclusions et Perspectives

L'objectif global de cette thèse était d'envisager les caractéristiques prosodiques d'attitudes stéréotypiques de trois langues : le japonais, le français et l'anglais ; et d'identifier de potentiels « décalages » de perception des attitudes parmi trois groupes de sujets ne partageant ni la même langue, ni la même culture.

À la suite des recherches développées sur les attitudes prosodiques (Aubergé et al., 1997 ; Grépillat, 1996 ; Morlec, 1997 pour les attitudes françaises, Diaferia, 2002 pour l'anglais britannique), nous avons continué dans le même cadre de recherches pour y intégrer les attitudes japonaises. Les Japonais provenant d'une culture distincte, peuvent posséder des attitudes n'étant pas conventionnelles dans la culture occidentale.

Après avoir effectué un large travail bibliographique sur les attitudes japonaises, et notamment sur les travaux extrêmement intéressants que l'on peut trouver sur l'expression de la politesse en japonais (Sadanobu, 2004a ; Ofuka, 2002 ; Ito, 2003 etc.), nous avons élaboré 12 attitudes représentatives de la langue japonaise, puis les avons validées auprès de 15 auditeurs japonais parlant le dialecte de Tokyo.

Dans un premier temps, les analyses perceptives menées sur des auditeurs japonais valident globalement la qualité du corpus. Le taux d'identification de ces 12 attitudes étant supérieur à celui du hasard.

L'analyse objective des indices prosodiques et de la qualité de voix pour les attitudes exprimées nous a donné les informations suivantes.

D'abord, l'*exclamation de surprise* et l'*irritation* sont caractérisées par le niveau élevé de l'intensité moyenne et de F0 moyenne, et par une voix tendue. Une forte pente de F0 sur la phrase ou sur la dernière voyelle, un grand écart-type de F0, un allongement important de la dernière more, un grand écart-type de durée moraïque et une durée moyenne plus longue sont fortement corrélés avec *doute-incrédulité*. Le *kyoshuku* et la *sincérité-politesse* portent des caractéristiques prosodiques similaires, mais la qualité de voix de ces deux attitudes sont nettement distinctes. *Kyoshuku* est caractérisé par une voix rauque et tendue, tandis que *sincérité-politesse* est exprimée par une voix soufflée. L'attitude d'*admiration* est aussi exprimée par la voix soufflée. Quant à l'*arrogance-impolitesse*, elle montre plus de variation d'intensité et elle est plus nasalisée que les autres attitudes. Enfin, la *déclaration* et la *politesse-simple* sont réalisées avec des caractéristiques acoustiques similaires et avec une voix modale.

Du point de vue perceptif, l'analyse de confusion des attitudes perçues montre que les sept attitudes : *arrogance-impolitesse*, *déclaration*, *doute-incrédulité*, *politesse-simple*, *exclamation de surprise*, *irritation* et *question-simple* n'ont pas eu de confusion

particulière. Par contre, les cinq attitudes : *autorité*, *évidence*, *kyoshuku*, *sincérité-politesse* et *admiration*, ont été confondues avec d'autres attitudes.

Si nous prêtons attention aux expressions de politesse, l'attitude de *kyoshuku* a été reconnue par les sujets japonais avec des confusions réciproques avec la *sincérité-politesse*. Ces deux attitudes sont exprimées, dans notre corpus, de manière prosodique similaire, mais distinguées par des qualités de voix différentes (la voix rauque pour *kyoshuku*, et la voix soufflée pour *sincérité-politesse*). Elles sont socialement reliées par la notion d'humilité qu'elles impliquent fortement face à un supérieur social. C'est l'utilisation pragmatique qui permet de choisir entre ces deux catégories.

À propos de l'attitude de *sincérité-politesse*, elle a été reconnue chez les sujets japonais mais confondue à l'intérieur des expressions de politesse (par exemple, *politesse-simple* et *kyoshuku*).

Ensuite, la perception multimodale des attitudes japonaises a été examinée. Généralement la condition audio-visuelle a montré un meilleur score de reconnaissance, et des spécificités de comportement pour les modalités audio et visuelle seules ont été observées. Les attitudes ont été regroupées au sein de catégories perceptives de haut-niveau pour le premier locuteur : les expressions de politesse, les attitudes « interrogatives » et les expressions d'imposition de l'avis du locuteur. Les attitudes de *kyoshuku* et de *surprise* sont particulièrement bien reconnues grâce aux informations visuelles.

L'analyse objective basée sur des indices statiques tels que les caractéristiques acoustiques et visuelles nous donnent les informations suivantes :

- *doute-incrédulité* est caractérisée par un registre de F0 plus large et une F0 maximale très élevée et également par la montée de l'intensité vers la fin de phrase.
- *Exclamation de surprise* possède plus ou moins les mêmes caractéristiques, mais un registre de F0 moins grand en comparaison de *doute-incrédulité*.
- *L'autorité* possède un niveau de F0 assez bas, le débit est plus rapide (la durée de la première more est plus longue que celle de la dernière more), et surtout cette attitude est produite avec plus d'intensité par rapport à *déclaration*.
- *Sincérité-politesse* est élicitée aussi avec un niveau de F0 assez bas, et un débit plus rapide (la durée de la première more est plus longue que celle de la dernière more) par rapport à *doute-incrédulité* ou *exclamation de surprise*, mais contrairement à *autorité*, l'intensité moyenne de l'énoncé n'est pas forte.
- *Kyoshuku* est aussi caractérisée par un bas niveau de F0, un débit plus rapide, et une intensité qui s'accroît progressivement.
- *Arrogance-impolitesse* est prononcée aussi avec un débit plus rapide, mais l'intensité est faible et il n'y a aucune caractéristique importante de la F0.

- *Admiration* est caractérisée par un allongement particulièrement long à la dernière more par rapport aux autres attitudes. En ce qui concerne l'évidence, nous avons identifié une pente de la courbe de F0 plus abrupte (un large écart de la valeur de F0 entre la première more et la dernière more), et cette attitude est prononcée avec plus d'intensité que la *déclaration*.
- Enfin, *déclaration* est marquée par une courbe de F0 plus plate et un débit rapide, *question-simple* se caractérise par la montée de l'intensité vers la fin de phrase.

Selon le test de perception en condition visuelle, l'*irritation* et le *kyoshuku* ont des caractéristiques FACS particulièrement marquées. Cette analyse identifie également la catégorie des attitudes « interrogatives » (c-à-d, *interrogation*, *exclamation de surprise* et *doute-incrédulité*) par un mouvement de l'AU1+2 (une activation musculaire dans la portion intérieure et extérieure des sourcils). Elle montre aussi une autre catégorie qui est composée de *politesse-simple*, *admiration*, *sincérité-politesse* et *arrogance-impolitesse* aux alentours d'AU1 (le mouvement du muscle dans la portion intérieure des sourcils).

Parallèlement aux études multimodales des expressions attitudinales du japonais, nous avons examiné la capacité de prédiction des attitudes par les auditeurs avant l'audition de l'énoncé complet, afin d'identifier à quel moment (point d'unicité), les auditeurs sont capables de reconnaître l'attitude du locuteur. Au total, trois tests ont été effectués dans cette thèse, mais dans les deux premiers tests perceptifs, l'attitude de *kyoshuku* et *admiration* ont été mal reconnues, même lorsque les auditeurs écoutaient la phrase complète. Nous pensons que la faible identification de ces attitudes est due à l'influence du contenu lexical des phrases utilisées. C'est pourquoi nous avons effectué un troisième test de gating avec des phrases ayant un contenu lexical mieux adapté au contexte de la vie quotidienne. Les résultats d'ANOVA montrent que l'effet de la longueur des phrases sur la perception des attitudes n'est pas significatif, mais montrent une influence significative du gate.

Autorité et *exclamation de surprise* sont toutes les deux reconnues immédiatement dès le début de la phrase. Probablement les auditeurs ont perçu, dès la première more, la voix tendue et le niveau élevé d'intensité. Ces facteurs acoustiques ont sans doute permis aux auditeurs d'identifier cette attitude. Quant à l'*exclamation de surprise*, les auditeurs ont probablement perçu tout de suite la voix tendue et soufflée qui est la caractéristique de cette attitude. *A contrario*, la *question-simple* est reconnue seulement à la fin de la phrase où la montée de F0 et la diminution progressive de l'intensité sont réalisées. L'*irritation* exprimée sur les phrases de 2 et de 5 mores est identifiée en fin de phrase comme la *question-simple*. Cette partie des énoncés est caractérisée par une intensité forte avec une voix rauque et tendue, et aussi par un allongement important. En ce qui concerne la même attitude réalisée sur la phrase de 8 mores, elle a été identifiée au sixième gate, où la montée brutale de F0 avec une intensité forte, et surtout avec la qualité de voix rauque et tendue ont été adoptés afin

d'exprimer cette attitude. Il est aussi important de noter que le taux de reconnaissance pour les expressions de la politesse reste plat et faible, mais si on regardait la perception globale de la notion de politesse pour ces trois expressions de politesse, la reconnaissance de la notion de politesse pour l'expression de *kyoshuku* augmente progressivement du début jusqu'au dernier gate, et *sincérité-politesse* a aussi été bien perçue comme expression de politesse dès le premier gate. Ce résultat nous indique que même si les auditeurs reconnaissent mal *sincérité-politesse*, cette expression est perçue immédiatement comme l'expression de politesse dès le début de l'énoncé par un indice de la forte variation d'intensité et la voix soufflée.

Quant à la *politesse-simple*, les auditeurs ont peu confondu cette attitude avec d'autres expressions de politesse, contrairement aux deux autres types de politesse. Selon l'analyse des correspondances, cette attitude a été majoritairement confondue avec la *déclaration* plutôt qu'avec d'autres expressions de la politesse. Cette même analyse statistique nous montre que les auditeurs ont bien identifié toutes les attitudes sauf la *politesse-simple*.

Dans le chapitre des études interculturelles, nous avons effectué des tests de perception croisés en manipulant les corpus validés d'attitudes en trois langues. L'objectif général était de cerner les attitudes partagées en trois langues et d'identifier le « décalage » perceptif pour ces attitudes chez les auditeurs de langue / culture différente. Après les tests perceptifs croisés, nous avons identifié les phénomènes suivants.

D'abord, la série des tests khi deux montre que la distribution des attitudes choisies de tous les groupes de sujets était différente du hasard. Pour les attitudes japonaises, aucune influence de l'accent lexical n'a été identifiée sur la distribution des réponses pour les trois groupes de sujets. Cependant, un effet de la longueur de phrase a été observé. Pour les attitudes anglaises et françaises, les sujets américains n'ont pas montré de sensibilité, sur leur comportement perceptif, à la longueur de phrase comme les sujets anglais qui n'ont pas non plus montré d'effet de la longueur pour la perception des attitudes anglaises. A contrario, les sujets français et japonais ont montré un comportement perceptif différent selon la longueur de phrase pour les attitudes anglaises et françaises. Par ailleurs, les auditeurs natifs reconnaissent correctement toutes les attitudes présentées au-dessus du seuil de hasard.

Le second point important est que deux attitudes : *exclamation de surprise* et *déclaration*, dans trois langues, sont reconnues parmi les trois groupes de sujets (Japonais, Français et Américains). Quelle que soit la langue qui exprime ces attitudes, tous les sujets partagent ces notions et l'expressivité de ces attitudes.

Par ailleurs, lorsqu'on analyse le comportement perceptif des trois groupes de sujets pour les attitudes exprimées dans chacune des trois langues, on identifie, en plus les deux attitudes citées en dessus, des expressions d'attitudes dans une langue qui sont

reconnues de manière similaire par les trois groupes d'auditeurs, même si ces attitudes exprimées en d'autres langues ne sont pas bien reconnues.

(a) Pour les attitudes japonaises, nous avons identifié un comportement similaire parmi les trois groupes des sujets pour la *politesse-simple*, l'*autorité* et l'*irritation*. (b) Pour les attitudes anglaises, seulement une attitude (sans compter l'*exclamation de surprise* et la *déclaration*) : l'*interrogation*, montre un comportement similaire chez les quatre groupes des sujets. (c) En ce qui concerne les attitudes françaises, les sujets des trois groupes montrent un comportement similaire pour la *question-simple* et l'*irritation*.

Nous avons également identifié un comportement différent des trois groupes de sujets (ou quatre pour les attitudes anglaises) pour chacune des langues testées.

(1) **D'abord pour les attitudes japonaises**, nous avons identifié un décalage de comportement parmi les trois groupes pour quatre attitudes : *kyoshuku*, *sincérité*, *arrogance-impolitesse* et *question-simple*.

Kyoshuku : Les sujets japonais montrent une confusion entre *kyoshuku* et *sincérité-politesse* pour cette attitude culturelle. Les sujets occidentaux ne reconnaissent pas du tout cette attitude. Les sujets français l'interprètent comme de l'*arrogance-impolitesse* ou de l'*autorité*. Quant aux sujets américains, ils la perçoivent comme de l'*arrogance-impolitesse* ou de l'*irritation*.

Sincérité-politesse : Les sujets japonais et français montrent une confusion entre cette attitude et *kyoshuku* ou *politesse-simple*. Cela montre bien que les Japonais et les Français reconnaissent cette attitude comme une expression de politesse. En revanche, les sujets américains ne reconnaissent pas cette attitude, mais il n'y a pas d'autres attitudes qui correspondent à ce qui vient d'être exprimé. Par conséquent, cette attitude reste inconnue – incomprise – pour les sujets américains.

Arrogance-impolitesse : Les sujets japonais reconnaissent très bien cette attitude en ne montrant aucune confusion particulière. En revanche, les sujets occidentaux la reconnaissent mal. Les français montrent de multiples confusions avec *déclaration* et *autorité*. Cela montre que les sujets français perçoivent plus ou moins la notion d'imposition du locuteur. Inversement, les auditeurs américains montrent une confusion uniquement entre cette attitude et *déclaration*.

Question-simple : Les sujets japonais reconnaissent très bien cette attitude en ne montrant aucune confusion particulière. Les sujets américains montrent une confusion entre cette attitude et *politesse-simple*. Cette confusion est probablement due à leur perception de degré de politesse entre le niveau neutre et la politesse dans une situation formelle. En revanche, les sujets français montrent une confusion importante avec la *déclaration*. Ce type de confusion n'est pas liée au degré de politesse, mais plutôt à la confusion affirmative vs. interrogative. L'analyse acoustique montre que la *question-simple* en japonais se caractérise par la montée de F0 à la fin de phrase, comme en

français, mais ce qui différencie la *question-simple* des deux langues est l'intensité en fin de phrase. La question-simple en japonais montre une diminution progressive de l'intensité vers la fin, alors que celle en français maintient l'intensité jusqu'à la fin de phrase.

Parmi ces quatre attitudes que les trois groupes perçoivent différemment, on peut considérer que l'attitude de *kyoshuku* est un cas typique de « faux-ami » prosodique. L'attitude de *sincérité-politesse* pour les Américains peut être vue comme une attitude inconnue du fait de l'absence d'expressivité correspondante. La voix soufflée qui exprime cette attitude peut être reliée plutôt à l'intimité sexuelle en anglais (Pavlenko, 2005). En ce qui concerne l'attitude d'*arrogance-impolitesse*, elle peut être aussi un des cas de « faux-ami » chez les Américains car contrairement à l'intention arrogante du locuteur japonais, les Américains perçoivent cette expression comme une *déclaration*, ce qui est le degré « neutre » de la politesse (c'est-à-dire, ni poli, ni impoli). Quant aux Français, ils perçoivent globalement le sens d'imposition de l'avis du locuteur par cette expression même s'ils n'identifient pas exactement l'intention d'arrogance du locuteur.

(2) **Ensuite, pour les attitudes anglaises britanniques**, nous avons identifié trois attitudes qui montrent un décalage du comportement perceptif parmi les quatre groupes : *séduction*, *doute-incrédulité* et *évidence*.

Séduction : aucun groupe ne montre de confusion particulière pour cette attitude, mais les sujets japonais l'identifient mal (27%), contrairement aux sujets occidentaux qui la reconnaissent nettement mieux (72,2% pour les sujets anglais, 56,7% pour les sujets américains et 58,1% pour les sujets français).

Doute-incrédulité : les quatre groupes des sujets ont confondu cette attitude avec une simple interrogation, sans aucune intention du locuteur. Cependant, les sujets anglais seuls identifient bien cette attitude (41,7%), tandis que les sujets non natifs ressentent une difficulté à l'identifier (11,3% pour les sujets américains, 10,8% pour les sujets français et 18,9% pour les sujets japonais), et ils n'identifient aucune intention du locuteur par cette expression anglaise.

Evidence : les quatre groupes de sujets ont confondu cette attitude avec une simple déclaration, sauf les sujets américains qui la confondent avec la *politesse*. Cette attitude est généralement mieux perçue chez les Européens (c'est-à-dire, les Anglais et les Français : 45,8% d'identification pour les sujets anglais et 35,3% pour les sujets français) plutôt que les non Européens (c'est-à-dire, les Américains et les Japonais : 25,8% d'identification pour les sujets américains et 16,2% pour les sujets japonais). Notamment les sujets japonais montrent le score le plus faible des quatre groupes et ils n'identifient aucune intention du locuteur (ils répondent *déclaration*) dans l'expression anglaise.

Parmi ces trois attitudes, il nous semble que la séduction anglaise peut renvoyer à un sens autre que ceux présents dans les 12 choix proposés chez les Japonais,

contrairement aux sujets occidentaux qui la reconnaissent correctement. L'origine de la faible identification de cette attitude chez les Japonais est peut-être l'absence de traduction exacte du terme « séduction » en japonais (par exemple, la séduction est un terme synonyme de « tentation » ou « incitation » selon le Progressive English-Japanese Dictionary, 2006). De même, l'attitude de *doute-incrédulité* anglaise peut être un cas potentiel de « faux-ami » chez les non natifs. Par contre, la notion et l'expressivité d'*évidence* anglaise sont peut-être partagées par une culture géographiquement plus proche plutôt que par une culture qui partage la même langue. Les Japonais qui ne partagent ni la langue, ni la culture européenne montrent le score d'identification le plus faible.

(3) **Enfin, pour les attitudes françaises**, nous avons identifié deux attitudes qui montrent un décalage de comportement perceptif parmi les trois groupes : *séduction* et *évidence*.

Séduction : cette attitude est très bien identifiée par les natifs, sans confusion particulière (50% d'identification). Les sujets non natifs reconnaissent aussi cette attitude (31% pour les sujets américains et 46,1% pour les sujets japonais), mais on observe tout de même une confusion avec la *question-simple*.

Evidence : l'*évidence* française est très bien identifiée chez les natifs avec un taux d'identification de 53,3%. Or les sujets non natifs reconnaissent mal cette attitude (10,7% d'identification pour les sujets américains et 12,5% pour les sujets japonais) en montrant une confusion avec les attitudes de dominance du locuteur (l'*autorité* pour les Américains et l'*irritation* pour les Japonais).

Le comportement perceptif des sujets non natifs pour l'expression de la séduction en français nous indique la possibilité qu'ils perçoivent parfois cette attitude comme une attitude qui ne montre aucune intention particulière. Quant à l'*évidence*, les sujets non-natifs ont tendance à percevoir cette attitude comme une attitude de dominance, contrairement aux natifs qui l'identifient correctement. Il est donc aussi possible que cette attitude soit aussi un cas de « faux-ami » prosodique.

Dans le dernier chapitre de thèse, nous avons examiné l'influence de l'apprentissage de la langue cible sur le comportement perceptif des expressions attitudinales. Nous avons effectué un test perceptif avec le corpus des attitudes japonaises validées précédemment, auprès de 32 apprenants français du japonais.

Les résultats montrent que pour les trois attitudes culturelles de politesse (*kyoshuku*, *arrogance-impolitesse* et *sincérité-politesse*), l'effet d'apprentissage de la langue est absent. Surtout l'attitude de *kyoshuku* est toujours perçue comme de l'*irritation*, de l'*arrogance-impolitesse* ou de l'*autorité*.

En revanche, sur certaines attitudes, nous avons observé les effets de l'apprentissage de la langue japonaise. Il s'agit de *question-simple* et de *doute-incrédulité*.

Les auditeurs français débutant leur apprentissage (niveau débutant) ont reconnu la *question-simple*, mais ils l'ont encore confondue avec *déclaration*. Par contre, les auditeurs français ayant acquis une bonne connaissance de la langue japonaise (niveau intermédiaire) n'ont pas montré cette confusion, comme des Japonais natifs. De même, *doute-incrédulité* a été nettement mieux identifiée chez les sujets de niveau intermédiaire que chez les sujets naïfs et débutants. Nous y voyons pour ces deux attitudes, l'effet de l'apprentissage de la langue.

Perspectives :

Les études préliminaires des attitudes prosodiques et multimodales que nous avons présentées dans cette thèse ont identifié certaines similarités et certaines différences parmi les attitudes prosodiques des trois langues et parmi les groupes de sujets selon leurs degrés d'acquisition de la langue cible. Nous avons également éclairci les caractéristiques acoustiques et décrit l'expression faciale des attitudes japonaises de notre corpus. Cependant, le nombre des sujets participant à chaque test perceptif n'est sans doute pas suffisant pour conclure sur les tendances observées. De plus, afin de chercher des attitudes universelles, nous devons ajouter davantage de langues dans ce paradigme d'études interculturelles. Ainsi, il est également nécessaire d'étudier le comportement d'auditeurs chinois ou coréens, qui partagent plus de points communs culturels avec les occidentaux, afin de voir s'ils peuvent identifier correctement les attitudes prosodiques spécifiques à la culture japonaise qui sont classées comme de « faux-amis » chez les occidentaux.

Dans cette thèse, nous avons analysé les caractéristiques acoustiques des attitudes japonaises, mais les caractéristiques acoustiques des attitudes anglaises et françaises nécessitent une analyse complémentaire pour apporter davantage d'explications au comportement perceptif de ces deux groupes de sujets.

De plus, nous n'avons pas analysé l'influence du sexe sur la production et la perception des attitudes. Or comme ce facteur participe à la personnalité, il doit pouvoir influencer le comportement verbal et non verbal (Ohara, 1992 ; Van Bezooijen, 1995).

Il est aussi intéressant d'effectuer des recherches sur l'acquisition des attitudes chez les enfants afin d'identifier l'âge d'acquisition des attitudes culturelles et des attitudes communes à d'autres langues.

9. Bibliographie

- ABE, I. (1998). Intonation in Japanese. *Intonation systems: a survey of twenty Languages*, Cambridge University Press, 368-375.
- ABELIN, Å. (2004). Cross-cultural multimodal interpretation of emotional expressions - an experimental study of Spanish and Swedish, *Proc. of Speech Prosody 2004*, Nara, Japan, 647-650.
- ABELIN, Å., & ALLWOOD, J. (2000). Cross linguistic interpretation of emotional prosody. *ISCA workshop on Speech and Emotion*, Newcastle, Northern Ireland, 110-113.
- ABELIN, Å., & ALLWOOD, J. (2002). *Cross linguistic interpretation of emotional prosody*. Gothenburg papers in theoretical linguistics, Göteborg, Sweden.
- AJZEN, I., & FISHBEIN, M. (1980). *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs, NJ, Prentice-Hall, Inc.
- d'ALESSANDRO, C. (2006). Voice source parameters and prosodic analysis. *Method in empirical prosody research*, Walter de Gruyter, Berlin/New York, 63-87.
- ALLERTON, D.J., & CRUTTENDEN, A. (1978). Syntactic, illocutionary, thematic and attitudinal factors in the intonation of adverbials. *Journal of Pragmatics*, **2**, 155-188.
- ALLPORT, G.W. (1935). Attitude, extrait d'un article In *Handbook of Social Psychology*, ed. Murchison, Clark University Press, Worcester.
- ANAN, F. (1980). 「イントネーション干渉の研究（日仏対照）」『日本音声学会会報』190号：9-13. [Les interférence intonatives du japonais et du français. *The Bulletin of the Phonetic Society of Japan*, No.190, 9-13.
- ARNOLD, M. B. (1960). *Emotion and personality*, New York, Columbia University Press.
- AUBERGÉ V. (2002). A Gestalt Morphology of Prosody Directed by Functions: the Example of a Step by Step Model Developed at ICP. *Proc. Speech Prosody 2002*, Aix-en Provence, France, 151-154.
- AUBERGÉ, V., & CATHIARD, M. (2003). Can we hear smile ? *Speech Communication*, **40**, 87-97.
- AUBERGÉ, V., GRÉPILLAT, T., & RILLIARD, A. (1997). Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. *Proceedings of EuroSpeech'97*, Rhodes, Greece, **2**, 871-877.

- AUBERGE, V., AUDIBERT, N., & RILLIARD, A. (2004). Acoustic morphology of expressive speech: what about contours? *Proceedings of Speech Prosody 2004*, Nara, Japan, 201-204.
- AVERILL, J. R. (1980). A Constructivist View of Emotion. *Emotion : Theory, Research and Experience : Vol. I. Theories of Emotion*, ed. Plutchik, R., Kellerman, H., 305-339. New York Academic Press. (réédité en 1990 ed. Halberstadt, A. G., & Ellyson, S. L.)
- AYUSAWA, T. (1992). 「日本語の疑問文の韻律的特徴」 『日本語の韻律に見られる母語の干渉 (2) 』 重点領域研究「日本語音声」 D 1 班研究成果報告書, 1-20. [Les caractéristiques prosodiques sur des phrases interrogatives en japonais. *les interférences de la langue maternelle sur la prosodie japonaise (2)*, rapport du résultat des recherches sur la phonétique japonaise, groupe D1, 1-20.]
- AYUSAWA, T. (2001). Acquisition de l'intonation interrogative japonaise par des apprenants français. Études de linguistique contrastive japonais-langue étrangère IX Le japonais et le français phénomènes prosodiques et communication non verbale, Institut National de Recherche sur la Langue Japonaise, 61-85.
- BANSE, R., & SCHERER, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70,(3), 614-636.
- BARKHUYSEN, P., KRAHMER, E., & SWERTS, M. (2007). Cross-modal perception of emotional speech, *Proc. of International Congress of Phonetic Sciences (ICPhS ' 07)*, Saarbrücken, Germany.
- BEIER, E. G., & ZAUTRA, A. J. (1972). Identification of vocal communication of emotions across cultures. *Journal of Consulting and Clinical Psychology*, 39, 166.
- BENZECRI, J.P. (1973). *L'analyse des données*. Paris, Bordas.
- BROWN P., & LEVINSON S. (1986). *Politeness. Some universals in language use*, Cambridge, UK, Cambridge University Press.
- CALBRIS, G., & MONTREDON, J. (1981). *Oh là là. Expression intonative et mimique*. Paris: CLE international.
- CALBRIS, G., & PORCHER, L. (1989). *Geste et communication*. Paris : Hatier-Crédif.
- CALBRIS, G., & MONTREDON, J. (1986). *Des gestes et des mots pour le dire*. Paris: CLE international.
- CALLAMAND, M. (1973). *L'intonation expressive*. Collection le français dans le monde, B.E.L.C. Librairies Hachette et Larousse.
- CALLAMAND, D. (1987). *Grammaire vivante du français : français langue étrangère*, Paris, Larousse, Diffusion CLE International.

- CAMPBELL, N. (1997). 「プラグマティック・イントネーション」 『文法と音声』 音声文法研究会編 くろしお出版, 55-74. [Pragmatic intonation – a function approach to explaining prosody. *Speech and Grammar*, Kurosio, 55-74.]
- CAMPBELL, N. (1999). 「韻律解釈における基本単位」 『文法と音声Ⅱ』 音声文法研究会編 くろしお出版, 73-91. [Basic prosodic elements – building – blocks of a syntax of speech. *Speech and Grammar II*, Kurosio, 73-91.]
- CAMPBELL, N. (2000). Databases of Emotional Speech. *Proc. ISCA 2000*, Northern Ireland, 34-38.
- CAMPBELL, N. (2002). 「音声合成の観点から見た言語音声の特徴」 『言語』 大修館書店, 52-61. [Les caractéristiques phonétiques de la parole du point de vue de la parole synthétisée. *Langage*, Taishukan, 52-61.]
- 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, 109-118.
- CAMPBELL, N. & MOKHTARI, P. (2003). Voice Quality: the 4th Prosodic Dimension. *15th International Congress of Phonetic Sciences*, Barcelone, Espagne, 2417-2420.
- CAMPBELL, N., & ERICKSON, D. (2004). What do people hear? A study of the perception of non-verbal affective information in conversational speech. *Journal of the Phonetic Society of Japan*, 8.1, 9-28.
- CHEVEIGNE, A., & KAWAHARA, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111, 1917-1930.
- CLÉMENT, J. (1999). Structure des représentations prosodiques. Développement normal et pathologique du traitement de la prosodie, Thèse doctorale, Université de Paris V.
- CORNELIUS, R. R. (2000). Theoretical Approaches to Emotion. *ISCA Workshop on Speech and Emotions*, Newcastle, Irlande du Nord, 3-10.
- COTTON, S., & GROJEAN, F. (1984). The gating paradigm : a comparison of successive and individual presentation formats, *Perception and Psychophysics*, 35, 1, 41-48.
- DAIJIRIN (2006). (ed.) Matsumura, A., Sanseido, Japan.
- DAMASIO, A.R. (1994). Descartes' error. Emotion, reason, and the human brain. New-York: G.P. Putnam.
- DANES, F. (1994). Involvement with language and in language. *Journal of Pragmatics*, 22, 251-164.
- DARWIN, C. (1872). The expression of the emotions in Man and Animals, Londres, John Murray (rééd.), Chicago, University of Chicago Press, 1965.

- DELATTRE, P. (1966). Les dix intonations de base du français. *The French Review*, 40 (3), 1-14.
- DIAFERIA, M.L. (2002). *Les Attitudes de l'Anglais : Premiers Indices Prosodiques*. Mémoire de DEA en Science Cognitives, Institut National Polytechnique de Grenoble.
- DOI, T. (1976). 『甘えの構造』 弘文堂. [*jeu d'indulgence*, Kobundo.]
- EKMAN, P. (1999). Facial Expressions. In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion*. Sussex, U.K.: John Wiley & Sons, Ltd, 301-320.
- DROMEY, C., SILVEIRA, J., & SANDOR, P. (2005). Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication*, vol. 47, (3), 351-359.
- DUPOUX, E. (1989). *Identification des mots parlés. Détection de phonèmes et unité prélexical*, Thèse de troisième cycle, Ecole des hautes études en sciences sociales, Paris, France.
- EKMAN, P. (1972). Universals and cultural differences in facial expression of emotion. In J.R. Cole (Eds.), *Nebraska Symposium on Motivation* (pp.207-283). Lincoln: University of Nebraska Press.
- EKMAN, P., FRIESEN, W. C., & HAGER, J. C. (2002). *Facial Action Coding System*. The Manual on CD ROM., Research Nexus division of Network Information Research Corporation.
- ELFENBEIN, H.A., & AMBADY, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, vol.128, (2), 203-235.
- ERICKSON, D., & MAEKAWA, K. (2001). Perception of American English emotion by Japanese listeners, *Proc. Spring Meeting of Acoustical Society of Japan*, 333-334.
- ERICKSON, D., HAYASHI, S., HOSOE, Y., SUZUKI, M. UENO, Y. & MAEKAWA, K. (2002). Perception of American English sarcasm by Japanese listeners. *Proc. of the 2002 Spring Meeting of the Acoustical Society of Japan*, Yokohama, Japan, 277-278.
- ERICKSON, D., OHASHI, S., MAKITA, S., KAJIMOTO, N., & MOKHTARI, P. (2003). Perception of naturally-spoken expressive speech by American English and Japanese listeners. *CREST International Workshop on Expressive Speech Processing*, 31-36.
- ERICKSON, D., & SHOCHI, T. (2006a). 「性差, および母語が感情音声に与える影響—日本語, 韓国語, 英語母語話者を対象として—」 『音声文法V』 くろしお出版, 31-46. [Some gender and Cultural differences in perception of affective expressions. *Speech and Grammar V*, Kurosio, 31-46.]

- ERICKSON, D. & SHOCHI, T. (2006b). Perception of verbal affect by Japanese children. Transactions on Technical Committee of Psychological and Physiological Acoustics, The Acoustical Society of Japan, Vol. **36**. **5**, 433-438.
- ERICKSON, D., SHOCHI, T., MENEZES, C., KAWAHARA, H., & SAKAKIBARA, K. (2007). A comparison of perception by Japanese and American listeners to some non-F0 cues to emotional speech. *Proc. of the international symposium "Japanese Spoken Language Education and its Basic Materials"*, Kobe, Japan.
- ERICKSON, D., SHOCHI, T., MENEZES, C., KAWAHARA, H., & SAKAKIBARA, K. (2008). Some non-F0 cues to emotional speech : An experiment with morphing. *Speech prosody 2008*, Campinas, Brazil.
- ERICKSON, D., SHOCHI, T., MENEZES, C., GUO, X., & DANG, J. (submitted). Some preliminary observations about articulatory, acoustic, and perceptual characteristics of laugh and smile speech in comparison with sad and neutral speech. *Phonetica*.
- FÓNAGY, Y. (1991). *La Vive Voix*, Paris, Payot.
- FUJISAKI, H. (1989). 「日本語の音調の分析とモデル化－語アクセント・統語構造・談話構造と音調との関係－」 『講座日本語と日本語教育 2 : 日本語の音声・音韻 (上)』 明治書院, 266-297. [Modélisation et analyse sur la tonalité en japonais – à travers la relation entre l'accent lexicale, la structure syntaxique, paradigme du discours et la tonalité. *Japanese and Japanese teaching vol. 2 : Japanese phonetics and phonology*, Meiji Syoin, 266-297.]
- FUJISAKI, H. (2004). Information, prosody, and modeling – with emphasis on tonal features of speech. *Proc. of Speech Prosody 2004*, Nara, Japan.
- FUJISAKI, H., & SUDO, H. (1971). 「日本語単語アクセントの基本周波数パターンとその生成機構のモデル」 『日本音響学会誌』 27-9, 445-453. [A model for the generation of fundamental frequency contours of Japanese word accent. *Journal of acoustical society of Japan*, 27-9, 445-453.]
- GARNIER, M., HENRICH, N., CASTELLENGO, M., SOTIROPOULOS, D., & DUBOIS, D. (2007). Characterisation of voice quality in Western lyrical singing: from teachers' judgements to acoustic descriptions. *Journal of Interdisciplinary Music Studies*, vol.1 (2), 62-91.
- GRANSTRÖM, B., & HOUSE, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, **46**, 473–484.
- GRÉPILLAT, T. (1996). *Perçoit-on, par l'intonation, l'attitude d'un locuteur avant la fin de l'énoncé ?* T.E.R. de Maîtrise d'Anglais, Université Stendhal Grenoble III.
- GREENBURG, Y., TSUZAKI, M., KATO, H., & SAGISAKA, Y. (2005). Communicative speech synthesis using constituent word attributes. *Proc. Inter Speech 2005*, Lisbon, Portugal, 517-520.

- GROSJEAN, F. (1985). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 38 (4), 299-310.
- HAMASAKI, T., MITSUMOTO, H., OTAWA, H., TAMURA, S., & YANAGIDA, M. (1999). Estimation of perceptual impression for ironic utterances from their prosody. *Joho Shori Gakkai Kenkyu Hokoku*, vol.99, n°64, 1-8.
- HATTORI, (1955). 「音韻論から見た国語の研究」 『日本の言語学 第二巻 音韻』 柴田武他編 大修館書店 : 364-403. [Les recherches sur la langue japonaise au point de vue phonologique. *Linguistiques au Japon*, vol.2, T. Shibata (ed.), Taishukansyoten, 364-403.]
- HIRATA, T. (1960). 『全国アクセント辞典』 平田輝男編 東京堂出版. [Le dictionnaire de l'accent japonais. ed. Hirata, T., Tokyodo shuppan.]
- HIROSE, K., TAKAHASHI, N., FUJISAKI, H., & OHNO, S. (1994). 「音声の基本周波数パターンにおける話者の意図・感情の表現」 『電子情報通信学会技術報告』 Hc94-41, 33-40. [Representation of intention and emotion of speakers with fundamental Frequency Contours of Speech. *Technical Report of IEICE*, HC94-41, 33-40.]
- HIROSE, K., KAWANAMI, H., & IHARA, N. (1997). Analysis of intonation in emotional speech. *Proc. ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications*, Athens, Greece, 185-188.
- HIRST, D., DI CRISTO, A., & NISHINUMA, Y. (2001). A cross-language approach to prosodic parameters of French. *Etudes de linguistique contrastive japonais-langues étrangères : le japonais et le français*, Institut National de Recherche sur la Langue Japonaise, 7-19.
- IIDA, A., IGA, S., & YASUMURA, M. (1997a). 「韻律にみられる感情表現について」 『第54 回情報処理学会全国大会予稿集』 2 : 227-228. [Study of emotion in relation to intonation. *Proc. 54th Annual Convention, Information Processing Society of Japan*, 2, 227-228.]
- IIDA, A., IGA, S., & YASUMURA, M. (1997b). 「音声にみられる感情表現について : 聴取実験結果の分析」 『情報処理学会研究報告』 1996-SLP-015, 113-118. [Study of emotion in speech: Findings from perceptual experiments. *Proc. SIG Note of IPSJ*, 1996 – SLP – 015, 113-118.]
- IIDA, A., IGA, S., HIGUCHI, F., CAMPBELL, N., & YASUMURA, M. (1998). Acoustic nature and perceptual testing of a corpus of emotional speech. *proc. ICSLP 98*, Sydney, Australia, 1559-1592.
- ITO, M. (2002). Japanese Politeness and Suprasegmentals – a Study based on Natural Speech Materials. *Proc. Speech Prosody 2002*, Aix-en-provence, France, 415-418.
- ITO, M. (2003). The Contribution of Voice Quality to Politeness in Japanese. *VOQUAL'03*, Genève, 157-162.

- IWASAKI, N. (2002). 「日本語能力簡易試験(SPOT)の得点とACTFL口頭能力測定(OPF)のレベルの関係について」 『日本語教育』 114, 日本語教育学会: 100-106. [Corrélation entre le score du SPOT et le niveau du OPI, *Journal of Teaching Japanese as Foreign Language*, 114, 100-106.]
- IZARD, C. E. (1971). *The face of emotion*. Appleton Century Crofts, New York.
- IZARD, C. E. (1977). The Emotions and Emotion Concept in Personality and Culture Reseach. *Handbook of Modern Personality Theory*, ed. Cattell, R.B., Breger, R.M., Washington D.C., Hemisphere.
- JAMES, W. (1902). *La théorie de l'émotion*, Paris, Alcan.
- KAWANO, T. (1993). 「プロソディーと丁寧表現－東京・大阪・名古屋の方言差を考慮して－」 『日本音声学会全国大会研究発表論集』, 29-34. [La prosodie et la politesse, dans trois régions: Tokyo, Osaka et Nagoya. *The Phonetic Society of Japan 1993 Annual Convention*, 29-34.]
- KINDAICHI, H. (1956). 『日本語』 岩波新書. [*La langue japonaise*, Iwanami shinsho.]
- KINDAICHI, H. (1957). 「日本語のアクセント」 『アクセント』 徳川宗賢編 有精堂, 18-36. [L'accent japonais. *L'accent*, ed. Tokugawa, M., Yusedo, 18-36.]
- KITAHARA, Y., TAKEDA, S., ICHIKAWA, A., & TOHKURA, Y. (1987). 「音声言語認知における韻律の役割」 『電子情報通信学会論文誌』 D, Vol.70-D, No.11, 2095-2101. [Role of prosody in cognitive process of spoken language. *Journal of IECE*, D, Vol. J70-D, No.11, 2095-2101.]
- KITAHARA, Y., & TOHKURA, Y. (1989). 「音声の韻律情報と感情表現」 『電子情報通信学会』 SP88-158, 27-32. [Prosodic information and expression of emotions in speech. *Technical report of IEICE*, SP88-158, 27-32.]
- KITAHARA, Y., & TOHKURA, Y. (1992). Prosodic control to express emotions for man-machine speech interaction. *IEICE Trans. Fundamentals*, Vol. E75-A, No. 2, 155-163.
- KO, M. (1993). 「丁寧表現に見られる日本語音声の韻律的特徴」 『日本音声学会全国大会研究発表論集』, 35-40. [Caractéristiques prosodiques dans les expressions de la politesse. *The Phonetic Society of Japan 1993 Annual Convention*, 35-40.]
- KOBAYASHI, N., FORD-TAMBA, J., & YAMAMOTO, H. (1996). 「日本語能力の新しい測定法 [SPOT] 」 『世界の日本語教育』 6, 201-218. [SPOT: A New Method of Testing Japanese Language Proficiency. *Journal of Japanese-Language Education around the Globe*, vol.6, 201-218.]
- KORI, S. (1989). 「強調とイントネーション講座」 『日本語と日本語教育. 第2巻日本語の音声・音韻 (上) 』 明治書院, 316-342. [Intensity and intonation. *Japanese and Japanese teaching vol. 2 : Japanese phonetics and phonology*. Meiji Syoin, 316-342.]

- KORI, S. (1992). 「日本語文音調の研究課題」 『日本語音声』 国際シンポジウム日本語音声の研究と日本語教育. 文部省科学研究重点領域研究. 平成4 年11月, 217-226. [Recherches sur les contours de f0 de la phrase japonaise. *Japanese phonetics*, symposium international de la phonétique japonaise et de l'enseignement du japonais en 1992, 217-226.]
- KORI, S. (1997). 「日本語イントネーション型と機能」 『アクセント・イントネーション・リズムとポーズ』 日本語音声[2] 杉藤美代子監修 国広哲弥・廣瀬肇・河野守夫編 三省堂, 169-202. [Japanese intonation - forms and functions -. *Japanese phonetics* [2], ed. Sugito, M., Sanseido, 169-202.]
- KORI, S. (2001). 「フランス語のイントネーション」 『日本語と外国語との対照研究VI 日本語とフランス語 - 音声と非言語行動-』 国立国語研究所: 21-39. [l'intonation du français. Études de linguistique contrastive japonais-langue étrangère IX Le japonais et le français phénomènes prosodiques et communication non verbale, Institut National de Recherche sur la Langue Japonaise, 21-39.]
- KORI, S. & OHKI, M. (1984). 「フランス語の中立的発話の音調」 『視聴覚外国語教育研究』 第7号, 大阪外国語大学. [F0 contour of French neutral utterances, *Foreign Language Teaching*, No.7, Osaka University of Foreign Studies.]
- KRAMER, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, 68, 390-396.
- KUBOZONO, H. (1997). Syllable and Accent in Japanese: Evidence from Loanword Accentuation. *The Bulletin of the Phonetic Society of Japan*, No.211, 71-82.
- LÉON, P.R. (2000). Phonétisme et prononciation du français, Nathan. Paris.
- LACHERET-DUJOUR, A. & BEAUGENDRE, F., (1999). *La prosodie du français*, CNRS, Paris.
- LADEFOGED, P. & MADDIESON, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.
- LANGE, C. (1895). *Les émotions*, trad. G. Dumas, Paris, Alcan.
- LAPIERE, R.T. (1934). Attitudes vs. Actions. *Social Forces*, 13, 230-237.
- LAVER, J. (1980). *The phonetic description of voice quality*, Cambridge University Press, New York.
- LOVEDAY, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates to English and Japanese politeness formulae. *Language and Speech*, Vol. 24, Part 1, 71-89.
- MAEKAWA, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, 635-638.

- MAEKAWA, K. (1999). Contribution of lexical and prosodic factors to the perception of politeness. *Proc. 14th ISPhS*, San Francisco, 1573-1575.
- MAEKAWA, K. (2004). Production and Perception of 'Paralinguistic' Information. *Proc. of Speech Prosody*, Nara, Japan.
- MAEKAWA, K., & KAGOMIYA, T. (2000). Influence of paralinguistic information on segmental articulation. *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP2000)*, Beijing, 2, 349-352.
- MATSUMOTO, D., & ASSAR, M. (1992). The effects of language on judgments of universal facial expressions of emotion. *Journal of Nonverbal behavior*, 16, 85-99.
- MATSUMOTO, E., & SADANOBU, T. (2001). 『日本語の韻律におけるりきみと日本語学習者の理解度』 ed. Department of Japanese Studies, The Chinese University of Hong Kong and Society of Japanese Language Education, *Quality Japanese Studies and Japanese Language Education in Kanji-Using Areas in the New Century*, Hong Kong Himawari Publishing Co., 455-461.
- MAZZICONACCI, S., & HERMES, D.J. (2000). Expression of emotion and attitude through temporal speech variations. *Proc. of International Conference of Spoken Language Processing*, vol.2, 373-378.
- McCLUSKEY, K., ALBAS, D., NIEMI, R., CUEVAS, C., & FERRER, C. (1975). Cross-cultural differences in the perception of the emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Developmental Psychology*, 11, 551-555.
- McCLUSKEY, K., & ALBAS, D. (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents and adults. *International Journal of Psychology*, 16, 119-132.
- MEJVALDOVÁ, J. (2000). Expressions prosodiques de certaines attitudes en tchèque et en français: étude comparative, Université Paris 7 – Denis Diderot, Paris, France.
- MEJVALDOVÁ, J., & HORÁK, P. (2002). Synonymie et homonymie attitudinale en tchèque et en français. *Proc. Speech Prosody 2002*, Aix-en-provence, France, 491-494.
- MENEZES, C., MAEKAWA, K., & KAWAHARA, H. (2006). Perception of voice quality in paralinguistic information types. *Proceedings of the 20th General meeting of the Phonetic Society of Japan, Special issue of the 80th Anniversary*. Tokyo, Japan, 153-158.
- MIZUTANI, O., & MIZUTANI N. (1979). *Aural Comprehension Practice in Japanese*. The Japan Times.

- MIZUTANI, O., & AYUSAWA, T. (1991). 「韻律の研究と教育をめぐって」 『シンポジウム 日本語音声教育』 凡人社, 95-100. [Recherches sur la prosodie et son application dans l'enseignement. *Japanese teaching education symposium*, Bonjinsha, 95-100.]
- MOKHTARI, P., & CAMPBELL, N. (2003). Automatic measurement of pressed/breathy phonation at acoustic centres of reliability in continuous speech, in Sp. Issue on Speech Information, *Proc. IEICE Transactions on Information and Systems*, (E-86-D), 574-582.
- MORIYAMA, T. (1999). 「モダリティとイントネーション」 『言語』 Vol.28., 大修館書店, 74-79. [Modality and intonation. *Language*, Vol. 28, Taisyukansyoten, 74-79.]
- MORLEC, Y. (1997). *Génération multiparamétrique de la prosodie du français par apprentissage automatique*, Thèse de l'Institut National Polytechnique de Grenoble, Grenoble, France.
- MORLEC, Y., BAILLY, G., & AUBERGÉ, V. (1999). Training an application-dependent prosodic model: corpus, model and evaluation. *European Conference on Speech Communication and Technology*, Budapest, Hungary, 1643-1646.
- MORLEC, Y., BAILLY, G., & AUBERGÉ, V. (2001). Generating prosodic attitudes in French: Data, model and evaluation. *Speech Communication*, **33** (4), 357-371.
- MORONI, V. (1997). Enquête sur les attitudes du français: définition et interprétation. Master thesis, Univ. Grenoble 3, France.
- MOZZICONACCI, S. (2002). Prosody and Emotions. *Proc. Speech Prosody 2002*, Aix-en-provence, France, 1-9.
- NAKAGAWA, A., TABUCHI, S., & SHOCHI, T. (2007). A cross-linguistic and cultural comparison of perception for Japanese communication. – A contrastive study between Japanese and French, *Actes du 9ème symposium sur l'enseignement du japonais en France*, 16-17.
- NAKAMICHI, A., JOGAN, A., USAMI, M., & ERICKSON, D. (2003). Perception by native and non-native listeners of vocal emotion in a bilingual movie. *Bulletin of Gifu City Women's College*, **52**, 87-91.
- NISHINUMA, Y. (1997). 「アメリカ人・フランス人日本語学習者のアクセント聞き取り・母語干渉による知覚のゆがみ」 『21世紀の日本語音声教育にむけて』 新プロ「日本語」研究班3 「音声言語の韻律的特徴に関する実験的研究」 平成8年度研究報告書, 5-12. [Perception of Japanese Accent by American and French Learners: First Language Interference in Perception. *For Japanese pronunciation teaching in 21 century*, Shinpuro Japanese, group 3 experimental research about prosodic features of human language research report of 1997, 5-12.]
- NISHINUMA, Y. & HIRST, D. (2001). 「フランス人の日本語発話に見る韻律の干渉」 『日本語と外国語との対照研究VI 日本語とフランス語 ―音声と非言語行動―』 国立国語

- 研究所, 41–59. [Interférence prosodique en japonais chez les apprenants français. *Études de linguistique contrastive japonais-loangue étrangère IX Le japonais et le français phénomènes prosodiques et communication non verbale*, Institut National de Recherche sur la Langue Japonaise, 41-59.]
- O'CONNER, J.D. (1973). *Phonetics*, Penguin Books, Middlespar exemple,
- OFUKA, E., MCKEOWN, J.D., WATERMAN, M.G., ROACH, P.J. (2000). Prosodic cue for rated politeness in Japanese speech. *Speech Communication*, 32, 199-217.
- OHALA, J.J. (1984). An ethlogical percpective on common cross-language utilization of F0 of voice. *Phonetica*, 41, 1-16.
- OHALA, J.J. (1996). Ethological Theory and the Expression of Emotion in the Voice. *Proc. International Conference on Spoken Language Processing*, Philadelphia, 3, 1812-1815.
- OHARA, Y. (1992). Gender – dependent pitch levels in Japanese and English: A comparative study in Japanese and English. *Locating Power : Proc. of the second Berkeley Women and Language Conference*, ed. Kira, H., Bucholtz, M., Moonwomon, B., Berkeley, Berkeley Women and Language Group, 467-477.
- OHARA, Y. (2001). Finding one's voice in Japanese : A study of the pitch levels of L2 users. *Multilingualism, Second Language Learning, and Gender.*, ed. Pavlenko, A., Brackledge, A., Pillar, I. and Teutsch-Dwyer, M., New York, Mouton de Gruyter.
- OSGOOD, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandanavian Journal of Psychology*, 7, 1-30.
- OTAKE, T., & CUTLER, A. (1999). Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics*, 25, 229-253.
- PAKOSZ, M. (1982). Intonation and Attitude. *Lingua*. 56. 153–178.
- PAVLENKO A. (2005). *Emotions and Multilingualism*. Cambridge, UK, Cambridge University Press.
- PIERREHUMBERT, J.B. & BECKMAN, M.E. (1988). *Japanese tone structure*, MIT Press.
- PIOT, O., & LYAGHAT, M. (2002). Expression et Reconnaissance de Onze Attitudes Assertives et Interrogatives en Persan Standard. *Proc. Speech Prosody 2002*, Aix-en-provence, France, 575-578.
- PLUTCHICK, R. (1984). Emotions : A general psychoevolutionary theory. *Approaches to Emotion*, ed. Scherer, K.R. & Ekman, P., Hillsdale, NJ : Lawrence Erlbaum Associates.
- PROGRESSIVE ENGLISH-JAPANESE DICTIONARY (2006). Syogakukan, Japan.
- RILLIARD, A., MARTIN, J.C., AUBERGE, V., & SHOCHI, T. (2008). Perception of French Audio-Visual Prosodic attitudes. *Speech prosody 2008*, Campinas, Brazil.

- RILLIARD, A., SHOCHI, T., MARTIN, J.C., ERICKSON, D., & AUBERGE, V. (en cours de révision). Multimodal indices to Japanese and French prosodically expressed social affects, *Language and Speech*.
- RINTELL, E. (1984). But did you feel about that? the learner's perception of emotion in speech. *Applied Linguistics*, 5, (3), 255-264.
- ROSEBERRY-MCKIBBIN, C., & BRICE, A. (1999). The perception of vocal cues of emotion by Spanish-speaking limited English proficient children. *Journal of Children's Communication Development*, 20 (2), 19-24.
- ROSENBERG, M.J., & HOVLAND, C.I. (1960). Cognitive, affective, and behavioral components of attitudes. *Attitude Organization and Change*, eds. C. I. Hovland & M. J. Rosenberg, New Haven and London, Yale University Press.
- ROSSATO, S., AUDIBERT, N., & AUBERGE, V. (2004). Emotional voice measurement : a comparison of articulatory –EGG and acoustic –amplitude parameters. *Proc. of Speech Prosody 2004*, Nara, Japan, 749-752.
- ROSSI, M. (1999). L'intonation, le système du français: description et modélisation, OPHRYS, Paris.
- ROSSI, M., di CRISTO, A., HIRST, D., MARTIN, P., & NISHINUMA, Y. (1981). *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.
- SADANOBU, T. (2003). Expressive Speech and Grammar — with Special Reference to Pressed Voice in Japanese. *CREST International Workshop on Expressive Speech Processing*, 55-59.
- SADANOBU, T. (2004a). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, Vol.8, No.1, 29-44.
- SADANOBU, T. (2004b). 「音声コミュニケーション教育の必要性と障害」 『日本語教育』 123号, 日本語教育学会, 1-16. [La nécessité de l'enseignement de la communication orale et les obstacles récents. *Teaching Japanese as Foreign Language*, n°123, ed. Society of Teaching Japanese as Foreign Language, 1-16.]
- SAGISAKA, Y., YAMASHITA, T., & KOKENAWA, Y. (2004). Speech synthesis with attitude. *Proceedings of Speech Prosody 2004*, Nara, Japan, 401-404.
- SAKATA, M., & HIROSE, K. (1995). Analysis and synthesis of prosodic features in spoken dialogue of Japanese. *Proc. 4th EUROSPEECH*, Madrid, 9, 1007-1010.
- SAWAMURA, K., DANG, J., AKAGI, M., ERICKSON, D., LI, A., SAKURABA, K., MINEMATSU, N., & HIROSE, K. (2007). Common factors in emotion perception among different cultures. *Proc. of Speech Prosody 2007*, Saarbrücken, Germany, 2113-2116.
- SCHERER, K.R. (1986). Vocal affect expression : A review and a model for future research. *Psychological Bulletin*, 99 (2), 143-165.

- SCHERER, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**(1-2), 227-256.
- SCHERER, K.R., BANSE, R., & WALLBOTT, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, **32**(1), 76-92.
- SCHERER, K.R., & ELLGRING, H. (2007). Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns? *Emotion*, **7**(1), 158-171.
- SHAHID, S., KRAHMER, E., & SWERTS, M. (2007). Real vs. acted emotional speech: comparing south-Asian and Caucasian speakers and observers. *Proc. of the International Conference on Spoken Language Processing (Interspeech 2007)*, Antwerp, Belgium.
- SHIBATA, T. (1961). 「日本語のアクセント」 『アクセント』 徳川宗賢編 有精堂, 7-14. [L'accent japonais. *L'accent*, ed. Tokugawa, M., Yusedo, 7-14.]
- SHIBATANI, M. (1990). *The languages of Japan*, Cambridge, Cambridge university Press.
- SHIGENO, S. (1998). Cultural similarities and differences in the recognition of audio-visual speech stimuli. *Proc. of International Conference on Spoken Language Processing 98*.
- SHIROTA, C. (1997). 「日本語アクセントの習得とイントネーション - フランス語母語話者による日本語発話の音調特徴とその要因 - 」 『世界の日本語教育』 7, 113-135. [L'acquisition de l'accent japonais et de l'intonation -caractéristiques tonales du japonais produites par les francophones. *Japonais dans le monde*, vol.7, 113-135.]
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2005). Because attitudes are social affects, they can be false friends. *Proc. of The first international Conference on Affective Computing & Intelligent Interaction*, LNCS 3784, 482-489.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2006). How prosodic attitudes can be false friends: Japanese vs. French social affects. *Proc. of Speech Prosody 2006*, Dresden, 692-696.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2007a). Cross-listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes. *Proceedings of 16th International Congress of Phonetic Science*, Saarbrücken, Germany, 2097-2100.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2007b). 「日本語韻律態度の知覚に与える日本語習得の影響」 『フランス日本語教育』 3, pp. 47-56.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2007c). 「発話態度の文化的特性と「偽の友達」」 『音声文法の対照』 定延利之、中川正之編, pp.55-78, くろしお出版

- SHOCHI, T., ERICKSON, D., RILLIARD, A, AUBERGÉ, V. & MARTIN, J.C. (2008). Recognition of Japanese attitudes in Audio-Visual speech. *Speech prosody 2008*, Campinas, Brazil.
- SHOCHI, T., RILLIARD, A, AUBERGÉ, V. & ERICKSON, D. (submitted). Intercultural Perception of Japanese, French and English Social Affective Prosody. *Affective Prosody*, (ed.) S. Hancil, Peter Lang, France.
- STOEZEL J. (1963). La conception actuelle de la notion d'attitude en psychologie sociale, *Bulletin de psychologie*, n°221, vol.XVI.
- SUGITO, M. (1983). 「日本語アクセントの研究の歴史」 『日本語音響学会誌』 39-4 : 266-273. [L'histoire des recherches sur l'accentuation japonaise. *Journal of acoustical science and technology*, 39-4, 266-273.]
- SUGITO, M. (1990). 「日本語と英語のアクセントとイントネーション」 『講座日本語と日本語教育 3 日本語の音声・音韻 (下)』 明治書院, 349-378. [L'accent et L'intonation en japonais et en anglais. *Japanese and Japanese teaching vol.3 : Japanese phonetics and phonology*. Meiji Syoin, 349-378.]
- SUGITO, M. (1997). 「話し言葉のアクセント, イントネーション, リズムとポーズ」 『アクセント・イントネーション・リズムとポーズ』 日本語音声[2] 杉藤美代子監修 国広哲弥・廣瀬肇・河野守夫編 三省堂, 169-202. [L'accent, l'intonation, le rythme et la pause dans la parole. *Japanese phonetics*, vol.2, ed. Sugito, M., Sanseido, 3-20.]
- SWERTS, M., & KRAHMER, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, **53**(1), 81-94.
- TABUCHI, S., & TAGAWA S. (2006). 「「答え」と「不満の答え」の弁別におけるイントネーションの影響」 『2006年度日本語教育学会秋大会予稿集』, 51-56. [L'influence de l'intonation pour la distinction entre la réponse affirmative et la réponse de mécontentement. Acte de symposium sur l'enseignement de la langue japonaise en 2006, 51-56.]
- HART, J., COLLIER, R., & COHEN, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.
- TOKI, S. (1989). 「音声の指導」 『日本語と日本語教育 第13巻 日本語教授法 (上)』 明治書院, 111-138. [L'enseignement de la prononciation. La langue japonaise et l'enseignement du japonais, n°13, Meijisyoin, 111-138.]
- THOMAS, W. I., & ZNANIECKI, F. (1918). *The Polish Peasant in Europe and America, Monograph of an immigrant group*, Chicago, III, University of Chicago Press.
- THORSEN, (1980). A study of the perception of sentence intonation –Evidence from Danish, *Journal of the Acoustical Society of America*, 67, 3, 1014-1030.

-
- TOMKINS, S. S. (1980). Affect as amplification: some modifications in theory. *Emotion, Theory, research and experience: theories of emotions*, ed. Plutchick, R. & Kellerman, H., NY : Academic Press (1), 141-165.
- UWANO, Z. (1989). 「日本語のアクセント」 『講座日本語と日本語教育 2 : 日本語の音声・音韻 (上) 』 明治書院, 177-205. [L'accent japonais. *Japanese and Japanese teaching, vol. 2 : Japanese phonetics and phonology*, Meiji Syoin, 177-205.]
- VAN BEZOOIJEN, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38, 3, 253-266.
- VAN BEZOOIJEN, R., OTTO, S., & HEENAN, T. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology*, 14, 387-406.
- VAN HEUVEN, V.J., HAAN, J., JANSE, E., & VAN DER TORRE, E.J. (1997). Perceptual identification of sentence type and the time distribution of prosodic interrogativity marker in Dutch. *ETRW Workshop on Prosody*, Athens, Greece, 317-320.
- WICHMANN, A. (2000). The attitudinal effects of prosody and how they related to emotion. *Proc. ISCA Workshop on Speech and Emotion*, Belfast, 143-147.
- WICHMANN, A. (2002). Attitudinal Intonation and the Inferential Process. *Proc. Speech Prosody 2002*, Aix-en-provence, France, 11-15.
- WILTING, J., KRAHMER, E., & SWERTS, M. (2006). Real vs. acted emotional speech. *Proc. of the 6th Interspeech*, Pittsburgh, PA, USA.

10. Annexes

ANNEXE 1 : INTERFACES DES TESTS PERCEPTIFS DANS LES DIFFERENTES LANGUES	288
EVALUATION DES ATTITUDES JAPONAISES	288
EVALUATION DES ATTITUDES FRANÇAISES	290
EVALUATION DES ATTITUDES ANGLAISES	291
ANNEXE 2 : CONTOURS DE F0 DU CORPUS JAPONAIS AUDIO- VISUEL.....	292
ANNEXE 3 : RESULTATS DES TESTS DE GATING	294
ANNEXE 4 : ANALYSES DES CORRESPONDANCES	300
ANNEXE 5 : FORMULAIRES D'ENQUETE PRELIMIANIRE	312
ETUDIANTS FRANCOPHONES DE JAPONAIS	312
ETUDIANTS JAPONAIS.....	314
ANNEXE 6 : QUESTIONNAIRES POST-EXPRIMENTAL	316

- Annexes 1 -

Interfaces du test perceptif (incluant la définition des attitudes) avec les attitudes japonaises en trois langues

TestAttitudes_Japonais_J (2) *

発話者の意図に最も適していると思われる発話態度を
12の中から一つだけ選択してください

☐ 感心	☐ 驚き	疑念 談話の際、発話相手の言ったこと、もしくは行動に対して否定的な考えを持ち、発話相手に対して懐疑的な態度を示すとき用いる発話態度
☐ ぞんざい	☐ 苛立ち	自明 発話者にとって自分が今から言おうとしていることは言うまでもない当然のことなのだという態度を示すときに用いる発話態度
☐ 高圧的	☒ 恐縮	驚き 談話の際、相手の言ったこと、もしくは行動を予測しておらず、その言動に対して意外だという驚きを示す際用いる態度
☐ 肯定	☐ 丁寧	高圧的 相手に対して高圧的であろうとする際用いる発話態度
☐ 疑念	☐ 疑問	苛立ち 相手の言動に対して不満および苛立ちを感じ、その苛立ちの態度を示す際用いる発話態度
☐ 自明	☐ 誠意	感心 見たり、聞いたり、味わったりしたことについて感心する際用いる発話態度（例：うまいな～）
		ぞんざい 相手に対して失礼な態度を示す際
		誠意 社会的立場が話者より上の人に対して、誠心誠意を表明する際用いる発話態度
		恐縮 社会的立場が話者より上の人に対して、「こんなことを言うのは申し訳ないですが」というときの恐縮した態度
		丁寧 相手の社会的地位、年齢に関係なく相手に対して丁寧に話そうとする際用いる発話態度
		肯定 特別な意図を込めずに、平叙文を言う際用いられる中立的な発話態度
		疑問 相手の判断を求める質問をするときに用いる発話態度

次の音

Test Attitudes F (2) *

Selon vous, quelle attitude le locuteur veut il exprimer ?

☐ Admiration	☐ Surprise	Admiration : expression utilisée pour manifester que l'on a beaucoup apprécié une chose vue, entendue ou encore goûtée. Arrogant / impoli : expression impolie d'un sentiment de supériorité vis-à-vis de son interlocuteur. Autorité : le locuteur veut imposer son autorité à son interlocuteur. Doute/incrédulité : le locuteur exprime son désaccord total ou partiel en reprenant dans son énoncé une affirmation que vient de donner l'interlocuteur. Evidence : le locuteur donne une information qui lui semble vraie, naturelle, et dont il est très sûr. Surprise : expression en réaction à quelque chose de soudain et d'inattendu. Irritation : lorsqu'un événement contrarie fortement le locuteur, au point qu'il éprouve le besoin de l'exprimer. Kyooshuku : cette attitude est typique de la culture japonaise. Elle apparaît quand le locuteur est dans une situation où son statut social est inférieur à celui de son interlocuteur, mais dans laquelle il a cependant un avis contraire à son interlocuteur, ou encore il désire lui demander un service. Dans ce contexte, le locuteur doit montrer sa souffrance honteuse et son embarras afin d'atténuer l'aspect arrogant et impoli qu'engendre sa désapprobation ou sa demande. Politesse : quelque soit le statut social du locuteur par rapport à son interlocuteur, cette attitude est utilisée pour exprimer la politesse et la courtoisie. Déclaration : le locuteur donne l'information simplement, sans exprimer aucun point de vue. Interrogation : le locuteur demande une information, sans exprimer de point de vue, et sans attendre autre chose qu'une réponse simple. Sérieux et sincérité : cette attitude japonaise est en relation avec le statut social. Le locuteur est socialement inférieur à son interlocuteur et exprime par cette attitude que son intention est sérieuse et sincère.
☐ Arrogance	☐ Irritation	
☒ Autorité	☐ Kyoshuku	
☐ Déclaration	☐ Politesse	
☐ Doute/Incrédulité	☐ Question simple	
☐ Evidence	☐ Sérieux, sincérité	

Suivant

Japanese attitudes (2) *

Which attitude do you think the speaker express?

☐ Declaration
☐ Arrogance

☐ Question
☐ Doubt-Incredulity

☐ Surprise
☒ Politeness

☐ Evidence
☐ Sincerity-Politeness

☐ Authority
☐ Kyoshuku

☐ Irritation
☐ Admiration

Next

Declaration: The speaker gives some simple information, without expressing any point of view.

Question: The speaker asks about information, without expressing any point of view, and without waiting any other thing than simple response.

Surprise: expression on the reaction when something unexpected has happened suddenly.

Evidence: The speaker says something he feels is self-evident.

Authority: The speaker expresses himself in order to influence other people or to impose his view or wish. He acts with authority and gives orders that must be obeyed.

Irritation: The speaker expresses his irritation towards what has been said; he protests because of this irritation.

Arrogant-impoliteness: Impolite expression that expresses a feeling of superiority towards the interlocutor.

Doubt-incredulity: The speaker expresses a feeling of being uncertain about something or of not believing something; he expresses a suspicion about what has previously been expressed.

Politeness: Regardless speaker's social states in relation to his interlocutor, the speaker says something in a polite way.

Sincerity-Politeness: This appears when the speaker is inferior to his interlocutor in Japanese society. The speaker expresses by this prosodic attitude that his intention is serious and sincere.

Kyoshuku: When the speaker wants to express a contradictory opinion or when he is willing to get a favor on a situation in which his social status is inferior to his interlocutor's, the speaker has to show an attitude which is a mixture of suffering ashamedness and embarrassment, in order to tone down the burden he imposes to his superior.

Admiration: The speaker expresses that he appreciates very much something he saw, heard or tasted.

Interfaces du test perceptif (incluant la définition des attitudes) avec les attitudes françaises en japonais et anglais

TestAttitude_Francais_J (2) *

発話者の意図に最も適していると思われる発話態度を12の中から1つだけ選択してください。

☐ 平叙	☐ 苛立ち
☐ 疑問	☐ 皮肉
☐ 驚き	☐ 軽蔑
☐ 自明	☒ 丁寧
☐ 疑念	☐ 誘惑
☐ 高圧的	☐ 称賛

次の音

平叙 特別な意図を含めずに、相手に対して情報を伝える際用いる発話態度
 疑問 特別な意図を含めずに、ある情報に対して相手のシンプルな返答を求める質問をするときに用いる発話態度
 驚き 談話の際、相手の言ったこと、もしくは行動を予測しておらず、その言動に対して、意外だという驚きを示す際用いる態度
 自明 発話者にとって自分が今から言おうとしていることは言うまでもない当然のことなのだという態度を示すときに用いる発話態度
 疑念 談話の際、発話相手の言ったこと、もしくは行動に対して否定的な考えを持ち、発話相手に対して懐疑的な態度を示すときに用いる発話態度
 高圧的 相手に対して高圧的であろうとする際用いる発話態度
 苛立ち 相手の言動に対して不満および苛立ちを感じ、その苛立ちの態度を示す際用いる発話態度
 皮肉 相手の言ったことに対して全く同意しておらず、又、聞いた内容について苛立っており、その相手に対して、イントネーションを用いて相手を非難する際に用いる。(例:「もちろんだよ。君の言ったことは正しいよ」と言いながら相手に自分の否定的な意見および相手に対する非難を浴びせる時)
 軽蔑 そんな、および軽蔑的な言い方で、相手に対して蔑んだ態度を示す際用いる韻律態度
 丁寧 相手の社会的地位、年齢に関係なく相手に対して丁寧に話そうとする際用いる発話態度
 誘惑 親密な間柄で、相手を誘惑する際に用いる発話態度
 称賛 相手の言動を是認し、それに対して喜びと驚きを表現する際に用いる発話態度

French attitudes (2) *

Which attitude do you think the speaker express?

☐ Declaration	☐ Irritation
☐ Question	☐ Sarcastic Irony
☐ Surprise	☒ Scorn-Aloofness
☐ Evidence	☐ Politeness
☐ Doubt-Incredulity	☐ Seduction
☐ Authority	☐ Admiration

Next

Declaration: The speaker gives some simple information, without expressing any point of view.
 Question: The speaker asks about information, without expressing any point of view, and without waiting any other thing than simple response.
 Surprise: expression on the reaction when something unexpected has happened suddenly.
 Evidence: The speaker says something he feels is self-evident.
 Doubt-incredulity: The speaker expresses a feeling of being uncertain about something or of not believing something; he expresses a suspicion about what has previously been expressed.
 Authority: The speaker expresses himself in order to influence other people or to impose his view or wish. He acts with authority and gives orders that must be obeyed.
 Irritation: The speaker expresses his irritation towards what has been said; he protests because of this irritation.
 Sarcastic Irony: The speaker calls into question (and even contradicts) a statement his interlocutor has said. He wants to show he doesn't agree and even that he's annoyed with what he's hearing, he condemns what is said or what happened, by using an intonation contour which could mean: 'Yeah, sure! I believe you... (meaning, I don't)'.
 Scorn-Aloofness: The speaker says something in an arrogant, aloof way, showing he gives no credit to what has been said.
 Politeness: The speaker says something in a polite way.
 Seduction: The speaker says something to seduce his interlocutor (seduction relationship).
 Admiration: The speaker expresses a feeling of pleasure, wonder, and approval.

Interfaces du test perceptif (incluant la définition des attitudes) avec les attitudes anglaises en anglais et en japonais

English attitudes (2) *

Which attitude do you think the speaker express?

☐ Declaration
 ☐ Question
 ☐ Surprise
 ☐ Evidence
 ☐ Doubt-Incredulity
 ☐ Command-Authority

☐ Irritation
 ☐ Sarcastic Irony
 ☒ Scorn-Aloofness
 ☐ Politeness
 ☐ Seduction

Next

Declaration: The speaker gives some simple information, without expressing any point of view.
 Question: The speaker asks about information, without expressing any point of view, and without waiting any other thing than simple response.
 Surprise: The speaker expresses his surprise at the moment he speaks.
 Evidence: The speaker says something he feels is self-evident.
 Doubt-incredulity: The speaker expresses a feeling of being uncertain about something or of not believing something; he expresses a suspicion about what has previously been expressed.
 Command-Authority: The speaker expresses himself in order to influence other people or to impose his view or wish. He acts with authority and gives orders that must be obeyed.
 Irritation: The speaker expresses his irritation towards what has been said; he protests because of this irritation.
 Sarcastic Irony: The speaker calls into question (and even contradicts) a statement his interlocutor has said. He wants to show he doesn't agree and even that he's annoyed with what he's hearing, he condemns what is said or what happened, by using an intonation contour which could mean: 'Yeah, sure! I believe you... (meaning, I don't)'.
 Scorn-Aloofness: The speaker says something in an arrogant, aloof way, showing he gives no credit to what has been said.
 Politeness: The speaker says something in a polite way.
 Seduction: The speaker says something to seduce his interlocutor (seduction relationship).

English_Attitudes_J (2) *

発話者の意図に最も適していると思われる発話態度を11の中から1つだけ選択してください。

☐ 平叙
 ☐ 苛立ち
 ☐ 皮肉
 ☒ 軽蔑
 ☐ 丁寧
 ☐ 誘惑

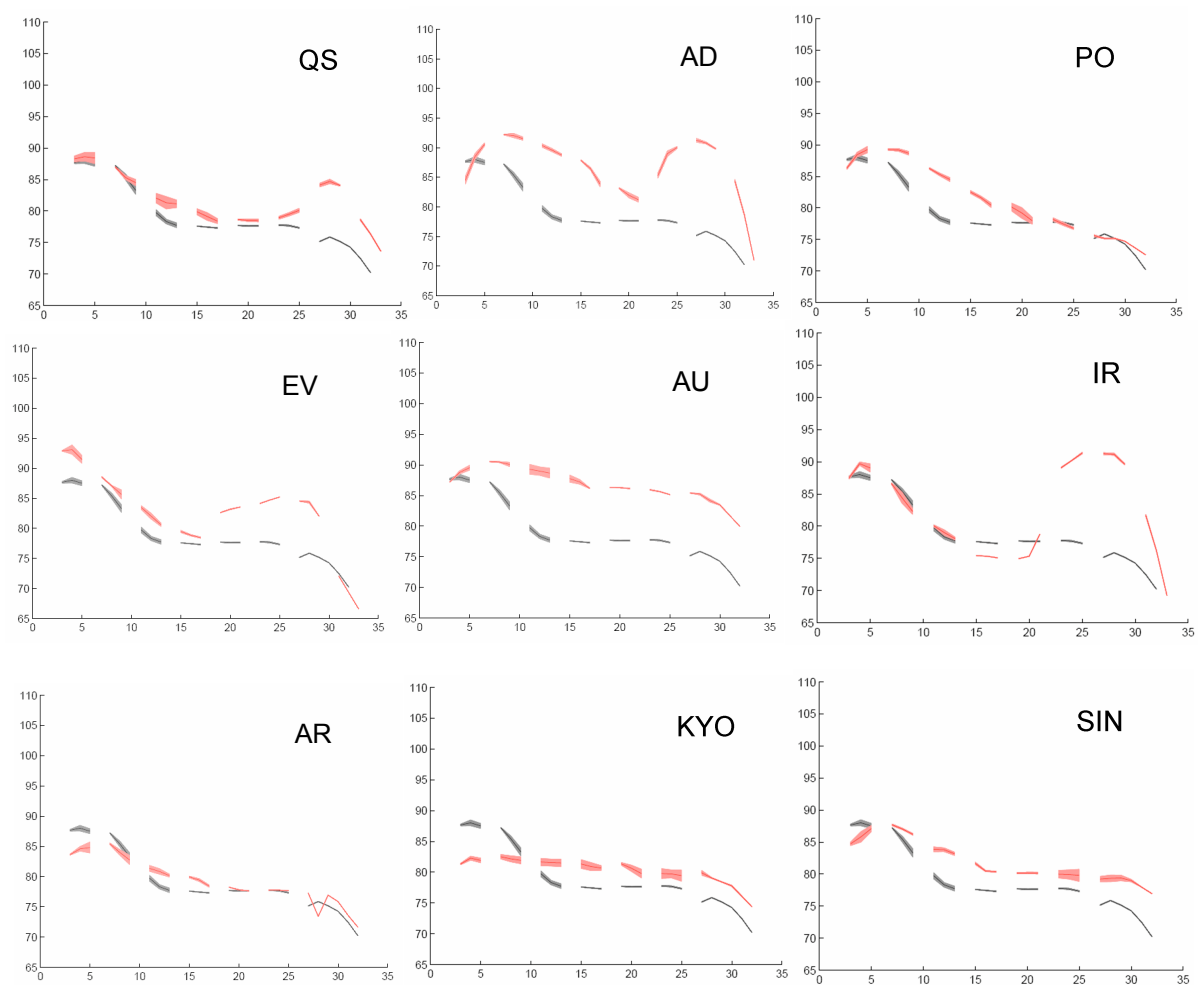
☐ 疑問
 ☐ 驚き
 ☐ 自明
 ☐ 疑念
 ☐ 高圧的

次の音

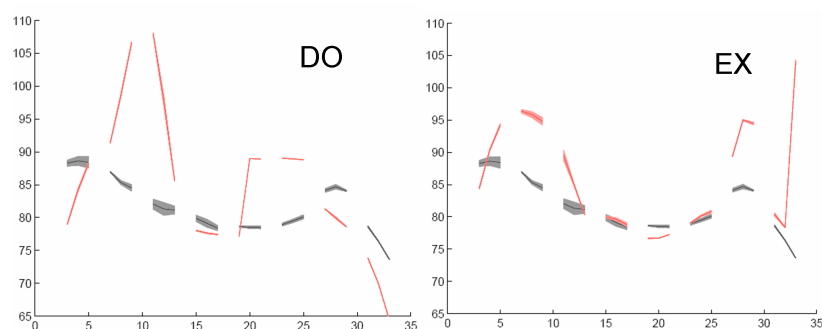
平叙 特別な意図を込めずに、相手に対して情報を伝える際用いる発話態度
疑問 特別な意図を込めずに、ある情報に対して相手のシンプルな返答を求める質問をするときにもちいる発話態度
驚き 談話の際、相手の言ったこと、もしくは行動を予測しておらず、その言動に対して意外だという驚きを示す際用いる態度
自明 発話者にとって自分が今から言おうとしていることは言うまでもない当然のことなのだという態度を示すときに用いる発話態度
疑念 談話の際、発話相手の言ったこと、もしくは行動に対して否定的な考えを持ち、発話相手に対して懐疑的な態度を示すとき用いる発話態度
高圧的 相手に対して高圧的であろうとする際用いる発話態度
苛立ち 相手の言動に対して不満および苛立ちを感じ、その苛立ちの態度を示す際用いる発話態度
皮肉 相手の言ったことに対して全く同意しておらず、又、聞いた内容について苛立っている。このような状況下で、その相手に対して、イントネーションを用いて相手を非難する際に用いる。(例:「もちろんだよ。君の言ったことは正しいよ」と言いながら相手に自分の否定的な意見および相手に対する非難を浴びせる時)
軽蔑 そんな、および軽蔑的な言い方で、相手に対して蔑んだ態度を示す際用いる発話態度
丁寧 相手の社会的地位、年齢に関係なく相手に対して丁寧に話そうとする際用いる発話態度
誘惑 親密な間柄で、相手を誘惑する際に用いる発話態度

- Annexe 2 -

Contour de F0 (en ½ ton) + intensité (présenté en épaisseur du trait) des attitudes japonaises produites par le locuteur A pour le corpus audio-visuel



Le contour de référence (déclaration) présenté en couleur noir et celui des attitudes étiquetées en haut à droite illustré en couleur rouge. L'axe des abscisses représente la durée en sec, l'axe des ordonnées représente F0 en demi-ton.



Le contour de référence (question-simple) présenté en couleur noir et celui des attitudes étiquetées en haut à droite illustré en couleur rouge. L'axe des abscisses représente la durée en sec, l'axe des ordonnées représente F0 en demi-ton.

- Annexe 3 -

Le taux de reconnaissance et le taux de confusions lors du test perceptif avec les attitudes japonaises dans le paradigme de gating.

<i>Kyoshuku</i>			<i>Kyoshuku</i>			Presented length of stimuli				
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5		
Admiration	0,0%	12,1%	Admiration	3,0%	9,1%	12,1%	3,0%	0,0%		
Arrogance	3,0%	0,0%	Arrogance	9,1%	9,1%	0,0%	6,1%	3,0%		
Authority	6,1%	0,0%	Authority	3,0%	0,0%	0,0%	0,0%	0,0%		
Declaration	6,1%	6,1%	Declaration	6,1%	6,1%	21,2%	3,0%	24,2%		
Doubt-Incredulity	21,2%	6,1%	Doubt-Incredulity	18,2%	12,1%	9,1%	9,1%	3,0%		
Evidence	6,1%	12,1%	Evidence	6,1%	9,1%	3,0%	3,0%	9,1%		
Surprise	3,0%	9,1%	Surprise	0,0%	0,0%	6,1%	3,0%	3,0%		
Irritation	18,2%	0,0%	Irritation	9,1%	3,0%	0,0%	6,1%	6,1%		
Kyoshuku	27,3%	27,3%	Kyoshuku	21,2%	30,3%	33,3%	42,4%	39,4%		
Politeness	0,0%	12,1%	Politeness	9,1%	12,1%	6,1%	6,1%	3,0%		
Question	0,0%	0,0%	Question	9,1%	0,0%	3,0%	3,0%	3,0%		
Sincerity-Politeness	9,1%	15,2%	Sincerity-Politeness	6,1%	9,1%	6,1%	15,2%	6,1%		

<i>Kyoshuku</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		0,0%	6,1%	12,1%	15,2%	3,0%	6,1%	3,0%	0,0%
Arrogance		3,0%	6,1%	0,0%	9,1%	18,2%	0,0%	3,0%	0,0%
Authority		0,0%	3,0%	0,0%	3,0%	0,0%	6,1%	0,0%	3,0%
Declaration		6,1%	12,1%	21,2%	9,1%	9,1%	18,2%	15,2%	3,0%
Doubt-Incredulity		9,1%	6,1%	9,1%	0,0%	9,1%	3,0%	6,1%	0,0%
Evidence		15,2%	6,1%	12,1%	9,1%	6,1%	9,1%	6,1%	0,0%
Surprise		0,0%	0,0%	3,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Irritation		15,2%	3,0%	3,0%	9,1%	3,0%	6,1%	3,0%	0,0%
Kyoshuku		12,1%	39,4%	24,2%	21,2%	33,3%	30,3%	27,3%	42,4%
Politeness		12,1%	9,1%	6,1%	18,2%	12,1%	9,1%	24,2%	21,2%
Question		18,2%	3,0%	3,0%	6,1%	3,0%	3,0%	0,0%	3,0%
Sincerity-Politeness		9,1%	6,1%	6,1%	0,0%	3,0%	9,1%	12,1%	27,3%

<i>simple-politeness</i>			<i>simple-politeness</i>			Presented length of stimuli				
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5		
Admiration	15,2%	0,0%	Admiration	0,0%	12,1%	6,1%	9,1%	3,0%		
Arrogance	3,0%	6,1%	Arrogance	18,2%	6,1%	6,1%	9,1%	0,0%		
Authority	6,1%	15,2%	Authority	3,0%	0,0%	0,0%	0,0%	0,0%		
Declaration	12,1%	18,2%	Declaration	18,2%	15,2%	39,4%	42,4%	48,5%		
Doubt-Incredulity	6,1%	9,1%	Doubt-Incredulity	3,0%	3,0%	0,0%	0,0%	0,0%		
Evidence	12,1%	18,2%	Evidence	9,1%	12,1%	3,0%	3,0%	9,1%		
Surprise	9,1%	9,1%	Surprise	12,1%	15,2%	6,1%	3,0%	3,0%		
Irritation	3,0%	3,0%	Irritation	3,0%	3,0%	0,0%	0,0%	0,0%		
Kyoshuku	0,0%	0,0%	Kyoshuku	12,1%	9,1%	3,0%	0,0%	6,1%		
Politeness	18,2%	12,1%	Politeness	9,1%	12,1%	21,2%	21,2%	12,1%		
Question	6,1%	3,0%	Question	12,1%	0,0%	9,1%	6,1%	3,0%		
Sincerity-Politeness	9,1%	6,1%	Sincerity-Politeness	0,0%	12,1%	6,1%	6,1%	15,2%		

<i>Simple-Politeness</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		3,0%	12,1%	6,1%	6,1%	6,1%	3,0%	6,1%	3,0%
Arrogance		3,0%	6,1%	3,0%	3,0%	3,0%	0,0%	3,0%	0,0%
Authority		3,0%	3,0%	0,0%	0,0%	0,0%	3,0%	0,0%	3,0%
Declaration		27,3%	27,3%	24,2%	39,4%	33,3%	54,5%	48,5%	54,5%
Doubt-Incredulity		6,1%	9,1%	6,1%	0,0%	0,0%	3,0%	3,0%	0,0%
Evidence		9,1%	6,1%	9,1%	0,0%	12,1%	3,0%	6,1%	9,1%
Surprise		3,0%	0,0%	3,0%	3,0%	0,0%	0,0%	3,0%	0,0%
Irritation		15,2%	3,0%	6,1%	6,1%	0,0%	0,0%	0,0%	0,0%
Kyoshuku		6,1%	3,0%	6,1%	6,1%	6,1%	0,0%	0,0%	0,0%
Politeness		6,1%	27,3%	15,2%	12,1%	18,2%	21,2%	30,3%	24,2%
Question		15,2%	0,0%	3,0%	3,0%	6,1%	3,0%	0,0%	0,0%
Sincerity-Politeness		3,0%	3,0%	18,2%	21,2%	15,2%	9,1%	0,0%	6,1%

<i>sincerity politeness</i>			<i>sincerity politeness</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	3,0%	9,1%	Admiration	12,1%	0,0%	3,0%	3,0%	3,0%
Arrogance	9,1%	0,0%	Arrogance	9,1%	3,0%	6,1%	3,0%	0,0%
Authority	0,0%	0,0%	Authority	0,0%	6,1%	0,0%	0,0%	0,0%
Declaration	12,1%	6,1%	Declaration	18,2%	21,2%	15,2%	15,2%	24,2%
Doubt-Incredulity	3,0%	9,1%	Doubt-Incredulity	0,0%	3,0%	3,0%	0,0%	0,0%
Evidence	15,2%	9,1%	Evidence	6,1%	3,0%	9,1%	0,0%	9,1%
Surprise	3,0%	3,0%	Surprise	3,0%	3,0%	0,0%	3,0%	0,0%
Irritation	3,0%	0,0%	Irritation	0,0%	3,0%	3,0%	0,0%	6,1%
Kyoshuku	24,2%	27,3%	Kyoshuku	15,2%	24,2%	24,2%	42,4%	30,3%
Politeness	12,1%	15,2%	Politeness	15,2%	18,2%	15,2%	21,2%	18,2%
Question	9,1%	12,1%	Question	12,1%	3,0%	15,2%	9,1%	6,1%
Sincerity-Politeness	6,1%	9,1%	Sincerity-Politeness	9,1%	12,1%	6,1%	3,0%	3,0%

<i>sincerity politeness</i>		Presented length of stimuli							
Phr. 8 mores	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	
Admiration	0,0%	3,0%	3,0%	6,1%	3,0%	0,0%	0,0%	0,0%	
Arrogance	9,1%	12,1%	9,1%	3,0%	3,0%	0,0%	0,0%	3,0%	
Authority	0,0%	0,0%	0,0%	0,0%	6,1%	3,0%	0,0%	0,0%	
Declaration	6,1%	9,1%	18,2%	15,2%	9,1%	21,2%	33,3%	6,1%	
Doubt-Incredulity	9,1%	6,1%	6,1%	0,0%	0,0%	0,0%	0,0%	3,0%	
Evidence	3,0%	6,1%	0,0%	3,0%	6,1%	9,1%	0,0%	0,0%	
Surprise	3,0%	0,0%	0,0%	0,0%	3,0%	3,0%	3,0%	3,0%	
Irritation	3,0%	6,1%	3,0%	3,0%	3,0%	3,0%	0,0%	0,0%	
Kyoshuku	30,3%	27,3%	39,4%	33,3%	30,3%	27,3%	36,4%	36,4%	
Politeness	15,2%	18,2%	9,1%	15,2%	15,2%	6,1%	18,2%	21,2%	
Question	6,1%	3,0%	3,0%	9,1%	6,1%	3,0%	0,0%	9,1%	
Sincerity-Politeness	15,2%	9,1%	9,1%	12,1%	15,2%	24,2%	9,1%	18,2%	

<i>Authority</i>			<i>Authority</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	0,0%	0,0%	Admiration	0,0%	0,0%	0,0%	0,0%	0,0%
Arrogance	3,0%	0,0%	Arrogance	3,0%	3,0%	6,1%	0,0%	6,1%
Authority	48,5%	63,6%	Authority	48,5%	60,6%	54,5%	57,6%	57,6%
Declaration	0,0%	0,0%	Declaration	0,0%	0,0%	0,0%	3,0%	0,0%
Doubt-Incredulity	3,0%	0,0%	Doubt-Incredulity	3,0%	0,0%	0,0%	3,0%	0,0%
Evidence	6,1%	6,1%	Evidence	3,0%	3,0%	0,0%	3,0%	6,1%
Surprise	3,0%	6,1%	Surprise	9,1%	3,0%	3,0%	3,0%	0,0%
Irritation	36,4%	24,2%	Irritation	33,3%	27,3%	27,3%	30,3%	24,2%
Kyoshuku	0,0%	0,0%	Kyoshuku	0,0%	0,0%	0,0%	0,0%	0,0%
Politeness	0,0%	0,0%	Politeness	0,0%	0,0%	3,0%	0,0%	0,0%
Question	0,0%	0,0%	Question	0,0%	0,0%	0,0%	0,0%	0,0%
Sincerity-Politeness	0,0%	0,0%	Sincerity-Politeness	0,0%	3,0%	6,1%	0,0%	6,1%

<i>Authority</i>		Presented length of stimuli							
Phr. 8 mores	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	
Admiration	0,0%	0,0%	3,0%	0,0%	0,0%	3,0%	0,0%	0,0%	
Arrogance	3,0%	3,0%	0,0%	3,0%	3,0%	3,0%	3,0%	3,0%	
Authority	57,6%	54,5%	51,5%	60,6%	63,6%	51,5%	48,5%	51,5%	
Declaration	0,0%	0,0%	9,1%	3,0%	6,1%	0,0%	0,0%	3,0%	
Doubt-Incredulity	0,0%	0,0%	0,0%	0,0%	3,0%	3,0%	3,0%	3,0%	
Evidence	0,0%	6,1%	3,0%	3,0%	3,0%	3,0%	0,0%	6,1%	
Surprise	9,1%	0,0%	3,0%	0,0%	0,0%	3,0%	3,0%	0,0%	
Irritation	21,2%	33,3%	21,2%	24,2%	12,1%	21,2%	36,4%	30,3%	
Kyoshuku	3,0%	0,0%	3,0%	6,1%	0,0%	3,0%	0,0%	0,0%	
Politeness	3,0%	3,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
Question	0,0%	0,0%	3,0%	0,0%	3,0%	0,0%	3,0%	0,0%	
Sincerity-Politeness	3,0%	0,0%	3,0%	0,0%	6,1%	9,1%	3,0%	3,0%	

<i>Admiration</i>			<i>Admiration</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	15,2%	54,5%	Admiration	18,2%	27,3%	42,4%	33,3%	60,6%
Arrogance	9,1%	3,0%	Arrogance	6,1%	0,0%	3,0%	0,0%	0,0%
Authority	6,1%	3,0%	Authority	0,0%	0,0%	0,0%	3,0%	0,0%
Declaration	3,0%	3,0%	Declaration	6,1%	6,1%	6,1%	6,1%	9,1%
Doubt-Incredulity	12,1%	9,1%	Doubt-Incredulity	6,1%	12,1%	6,1%	6,1%	0,0%
Evidence	6,1%	6,1%	Evidence	9,1%	0,0%	0,0%	9,1%	0,0%
Surprise	15,2%	12,1%	Surprise	6,1%	18,2%	15,2%	3,0%	3,0%
Irritation	15,2%	3,0%	Irritation	18,2%	9,1%	6,1%	9,1%	3,0%
Kyoshuku	9,1%	0,0%	Kyoshuku	3,0%	9,1%	9,1%	9,1%	3,0%
Politeness	3,0%	3,0%	Politeness	6,1%	9,1%	3,0%	18,2%	6,1%
Question	3,0%	3,0%	Question	15,2%	3,0%	3,0%	0,0%	3,0%
Sincerity-Politeness	3,0%	0,0%	Sincerity-Politeness	6,1%	6,1%	6,1%	3,0%	12,1%

<i>Admiration</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		18,2%	27,3%	39,4%	45,5%	42,4%	45,5%	45,5%	54,5%
Arrogance		3,0%	3,0%	0,0%	3,0%	0,0%	0,0%	0,0%	3,0%
Authority		3,0%	6,1%	0,0%	0,0%	6,1%	3,0%	0,0%	0,0%
Declaration		0,0%	0,0%	0,0%	0,0%	9,1%	3,0%	6,1%	9,1%
Doubt-Incredulity		9,1%	6,1%	9,1%	9,1%	0,0%	3,0%	0,0%	0,0%
Evidence		6,1%	6,1%	6,1%	9,1%	6,1%	3,0%	3,0%	12,1%
Surprise		24,2%	18,2%	15,2%	6,1%	3,0%	3,0%	6,1%	3,0%
Irritation		15,2%	0,0%	3,0%	3,0%	6,1%	3,0%	3,0%	3,0%
Kyoshuku		6,1%	0,0%	9,1%	6,1%	0,0%	3,0%	6,1%	0,0%
Politeness		6,1%	18,2%	9,1%	6,1%	27,3%	21,2%	18,2%	3,0%
Question		6,1%	6,1%	0,0%	6,1%	0,0%	3,0%	3,0%	0,0%
Sincerity-Politeness		3,0%	9,1%	9,1%	6,1%	0,0%	9,1%	9,1%	12,1%

<i>Arrogance</i>			<i>Arrogance</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	0,0%	3,0%	Admiration	0,0%	0,0%	0,0%	3,0%	3,0%
Arrogance	15,2%	48,5%	Arrogance	12,1%	18,2%	24,2%	36,4%	42,4%
Authority	9,1%	0,0%	Authority	6,1%	15,2%	12,1%	18,2%	9,1%
Declaration	21,2%	3,0%	Declaration	15,2%	15,2%	18,2%	12,1%	18,2%
Doubt-Incredulity	18,2%	12,1%	Doubt-Incredulity	12,1%	9,1%	3,0%	12,1%	0,0%
Evidence	6,1%	9,1%	Evidence	15,2%	12,1%	6,1%	3,0%	6,1%
Surprise	0,0%	3,0%	Surprise	0,0%	3,0%	0,0%	0,0%	0,0%
Irritation	18,2%	18,2%	Irritation	12,1%	12,1%	9,1%	6,1%	21,2%
Kyoshuku	3,0%	0,0%	Kyoshuku	3,0%	6,1%	3,0%	0,0%	0,0%
Politeness	0,0%	3,0%	Politeness	3,0%	3,0%	3,0%	3,0%	0,0%
Question	9,1%	0,0%	Question	18,2%	3,0%	15,2%	3,0%	0,0%
Sincerity-Politeness	0,0%	0,0%	Sincerity-Politeness	3,0%	3,0%	6,1%	3,0%	0,0%

<i>Arrogance</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		3,0%	0,0%	0,0%	0,0%	3,0%	0,0%	0,0%	3,0%
Arrogance		12,1%	24,2%	24,2%	15,2%	36,4%	42,4%	33,3%	51,5%
Authority		6,1%	12,1%	15,2%	18,2%	18,2%	9,1%	12,1%	3,0%
Declaration		15,2%	15,2%	27,3%	18,2%	21,2%	6,1%	9,1%	6,1%
Doubt-Incredulity		18,2%	12,1%	0,0%	6,1%	3,0%	3,0%	0,0%	0,0%
Evidence		3,0%	12,1%	6,1%	12,1%	9,1%	6,1%	18,2%	18,2%
Surprise		0,0%	3,0%	3,0%	0,0%	0,0%	3,0%	0,0%	0,0%
Irritation		9,1%	9,1%	9,1%	18,2%	3,0%	15,2%	15,2%	12,1%
Kyoshuku		9,1%	3,0%	6,1%	6,1%	0,0%	3,0%	0,0%	3,0%
Politeness		3,0%	0,0%	0,0%	3,0%	0,0%	3,0%	6,1%	0,0%
Question		18,2%	3,0%	6,1%	3,0%	3,0%	9,1%	0,0%	0,0%
Sincerity-Politeness		3,0%	6,1%	3,0%	0,0%	3,0%	0,0%	6,1%	3,0%

<i>Declaration</i>			<i>Declaration</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	3,0%	3,0%	Admiration	9,1%	3,0%	0,0%	0,0%	3,0%
Arrogance	12,1%	0,0%	Arrogance	21,2%	27,3%	18,2%	24,2%	3,0%
Authority	6,1%	9,1%	Authority	6,1%	6,1%	0,0%	3,0%	0,0%
Declaration	33,3%	48,5%	Declaration	36,4%	30,3%	39,4%	36,4%	57,6%
Doubt-Incredulity	3,0%	0,0%	Doubt-Incredulity	3,0%	3,0%	0,0%	6,1%	3,0%
Evidence	15,2%	21,2%	Evidence	0,0%	3,0%	15,2%	6,1%	15,2%
Surprise	3,0%	3,0%	Surprise	3,0%	0,0%	6,1%	0,0%	0,0%
Irritation	3,0%	6,1%	Irritation	0,0%	6,1%	0,0%	6,1%	3,0%
Kyoshuku	0,0%	0,0%	Kyoshuku	3,0%	0,0%	3,0%	12,1%	3,0%
Politeness	6,1%	3,0%	Politeness	12,1%	6,1%	9,1%	0,0%	0,0%
Question	3,0%	3,0%	Question	3,0%	12,1%	6,1%	3,0%	9,1%
Sincerity-Politeness	12,1%	3,0%	Sincerity-Politeness	3,0%	3,0%	3,0%	3,0%	3,0%

<i>Declaration</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		3,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Arrogance		21,2%	39,4%	21,2%	15,2%	15,2%	9,1%	9,1%	15,2%
Authority		0,0%	0,0%	0,0%	3,0%	3,0%	0,0%	0,0%	3,0%
Declaration		21,2%	36,4%	42,4%	36,4%	39,4%	45,5%	33,3%	48,5%
Doubt-Incredulity		9,1%	3,0%	3,0%	6,1%	3,0%	0,0%	3,0%	0,0%
Evidence		3,0%	9,1%	9,1%	9,1%	18,2%	9,1%	18,2%	9,1%
Surprise		0,0%	0,0%	3,0%	3,0%	0,0%	0,0%	0,0%	3,0%
Irritation		9,1%	9,1%	3,0%	3,0%	3,0%	6,1%	3,0%	6,1%
Kyoshuku		9,1%	0,0%	0,0%	9,1%	12,1%	12,1%	12,1%	3,0%
Politeness		12,1%	0,0%	3,0%	6,1%	0,0%	9,1%	12,1%	9,1%
Question		3,0%	0,0%	9,1%	6,1%	3,0%	0,0%	9,1%	0,0%
Sincerity-Politeness		9,1%	3,0%	6,1%	3,0%	3,0%	9,1%	0,0%	3,0%

<i>Doubt-Incredulity</i>			<i>Doubt-Incredulity</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	9,1%	0,0%	Admiration	6,1%	18,2%	0,0%	3,0%	0,0%
Arrogance	9,1%	18,2%	Arrogance	15,2%	6,1%	9,1%	15,2%	3,0%
Authority	3,0%	0,0%	Authority	3,0%	3,0%	0,0%	3,0%	0,0%
Declaration	18,2%	3,0%	Declaration	18,2%	12,1%	12,1%	0,0%	0,0%
Doubt-Incredulity	15,2%	27,3%	Doubt-Incredulity	12,1%	3,0%	30,3%	51,5%	57,6%
Evidence	3,0%	12,1%	Evidence	3,0%	30,3%	15,2%	9,1%	0,0%
Surprise	0,0%	3,0%	Surprise	0,0%	3,0%	0,0%	6,1%	6,1%
Irritation	18,2%	0,0%	Irritation	6,1%	6,1%	0,0%	6,1%	12,1%
Kyoshuku	3,0%	0,0%	Kyoshuku	6,1%	0,0%	3,0%	3,0%	0,0%
Politeness	3,0%	0,0%	Politeness	3,0%	3,0%	6,1%	3,0%	0,0%
Question	9,1%	33,3%	Question	21,2%	12,1%	24,2%	0,0%	21,2%
Sincerity-Politeness	9,1%	3,0%	Sincerity-Politeness	6,1%	3,0%	0,0%	0,0%	0,0%

<i>Doubt-Incredulity</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		0,0%	12,1%	0,0%	0,0%	0,0%	0,0%	0,0%	3,0%
Arrogance		15,2%	12,1%	12,1%	12,1%	18,2%	12,1%	9,1%	15,2%
Authority		0,0%	9,1%	3,0%	6,1%	0,0%	3,0%	0,0%	0,0%
Declaration		15,2%	15,2%	9,1%	6,1%	3,0%	3,0%	6,1%	0,0%
Doubt-Incredulity		18,2%	18,2%	33,3%	42,4%	36,4%	39,4%	39,4%	57,6%
Evidence		9,1%	9,1%	3,0%	0,0%	9,1%	12,1%	12,1%	0,0%
Surprise		3,0%	3,0%	0,0%	0,0%	0,0%	0,0%	3,0%	0,0%
Irritation		21,2%	12,1%	6,1%	3,0%	15,2%	3,0%	9,1%	0,0%
Kyoshuku		0,0%	0,0%	0,0%	3,0%	0,0%	0,0%	6,1%	0,0%
Politeness		6,1%	6,1%	3,0%	0,0%	6,1%	3,0%	3,0%	0,0%
Question		12,1%	3,0%	18,2%	21,2%	9,1%	18,2%	9,1%	24,2%
Sincerity-Politeness		0,0%	0,0%	12,1%	6,1%	3,0%	6,1%	3,0%	0,0%

<i>Evidence</i>			<i>Evidence</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	12,1%	12,1%	Admiration	21,2%	3,0%	9,1%	6,1%	3,0%
Arrogance	0,0%	9,1%	Arrogance	0,0%	6,1%	9,1%	3,0%	12,1%
Authority	21,2%	9,1%	Authority	6,1%	15,2%	15,2%	18,2%	15,2%
Declaration	0,0%	9,1%	Declaration	0,0%	6,1%	9,1%	15,2%	9,1%
Doubt-Incredulity	9,1%	6,1%	Doubt-Incredulity	12,1%	6,1%	3,0%	3,0%	6,1%
Evidence	6,1%	30,3%	Evidence	3,0%	18,2%	18,2%	42,4%	36,4%
Surprise	21,2%	6,1%	Surprise	36,4%	15,2%	21,2%	0,0%	0,0%
Irritation	9,1%	3,0%	Irritation	0,0%	0,0%	0,0%	3,0%	0,0%
Kyoshuku	6,1%	0,0%	Kyoshuku	3,0%	3,0%	0,0%	3,0%	0,0%
Politeness	3,0%	0,0%	Politeness	6,1%	3,0%	3,0%	3,0%	9,1%
Question	9,1%	12,1%	Question	6,1%	18,2%	6,1%	0,0%	6,1%
Sincerity-Politeness	3,0%	3,0%	Sincerity-Politeness	6,1%	6,1%	6,1%	3,0%	3,0%

<i>Evidence</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		12,1%	9,1%	12,1%	3,0%	6,1%	0,0%	0,0%	0,0%
Arrogance		6,1%	9,1%	9,1%	9,1%	3,0%	9,1%	24,2%	12,1%
Authority		3,0%	9,1%	6,1%	21,2%	21,2%	21,2%	12,1%	12,1%
Declaration		12,1%	21,2%	15,2%	6,1%	6,1%	12,1%	12,1%	6,1%
Doubt-Incredulity		3,0%	9,1%	6,1%	0,0%	9,1%	9,1%	3,0%	6,1%
Evidence		12,1%	12,1%	30,3%	21,2%	39,4%	24,2%	36,4%	42,4%
Surprise		15,2%	6,1%	3,0%	6,1%	0,0%	0,0%	0,0%	0,0%
Irritation		0,0%	6,1%	6,1%	3,0%	3,0%	6,1%	0,0%	6,1%
Kyoshuku		9,1%	0,0%	0,0%	6,1%	0,0%	3,0%	0,0%	0,0%
Politeness		9,1%	3,0%	3,0%	9,1%	6,1%	6,1%	3,0%	9,1%
Question		0,0%	15,2%	9,1%	15,2%	6,1%	3,0%	0,0%	3,0%
Sincerity-Politeness		18,2%	0,0%	0,0%	0,0%	0,0%	6,1%	9,1%	3,0%

<i>Exclamation of surprise</i>			<i>Exclamation of surprise</i>					
Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	12,1%	6,1%	Admiration	3,0%	15,2%	12,1%	6,1%	0,0%
Arrogance	3,0%	3,0%	Arrogance	0,0%	0,0%	0,0%	3,0%	0,0%
Authority	3,0%	0,0%	Authority	6,1%	3,0%	0,0%	0,0%	3,0%
Declaration	0,0%	0,0%	Declaration	6,1%	3,0%	0,0%	3,0%	0,0%
Doubt-Incredulity	18,2%	21,2%	Doubt-Incredulity	6,1%	0,0%	0,0%	3,0%	3,0%
Evidence	0,0%	6,1%	Evidence	9,1%	6,1%	3,0%	3,0%	0,0%
Surprise	48,5%	42,4%	Surprise	48,5%	57,6%	72,7%	69,7%	75,8%
Irritation	6,1%	0,0%	Irritation	9,1%	0,0%	3,0%	3,0%	0,0%
Kyoshuku	3,0%	0,0%	Kyoshuku	3,0%	0,0%	9,1%	0,0%	3,0%
Politeness	0,0%	0,0%	Politeness	0,0%	3,0%	0,0%	0,0%	0,0%
Question	3,0%	18,2%	Question	3,0%	3,0%	0,0%	6,1%	15,2%
Sincerity-Politeness	3,0%	3,0%	Sincerity-Politeness	6,1%	9,1%	0,0%	3,0%	0,0%

<i>Exclamation of surprise</i>		Presented length of stimuli							
Phr. 8 mores		1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration		9,1%	15,2%	3,0%	3,0%	9,1%	0,0%	0,0%	0,0%
Arrogance		0,0%	3,0%	3,0%	0,0%	3,0%	0,0%	0,0%	0,0%
Authority		0,0%	9,1%	0,0%	3,0%	3,0%	6,1%	0,0%	3,0%
Declaration		3,0%	0,0%	3,0%	0,0%	0,0%	6,1%	0,0%	0,0%
Doubt-Incredulity		6,1%	0,0%	15,2%	0,0%	3,0%	0,0%	9,1%	6,1%
Evidence		3,0%	3,0%	6,1%	0,0%	0,0%	3,0%	0,0%	3,0%
Surprise		57,6%	57,6%	63,6%	81,8%	78,8%	81,8%	81,8%	54,5%
Irritation		9,1%	3,0%	3,0%	6,1%	0,0%	0,0%	0,0%	0,0%
Kyoshuku		0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Politeness		0,0%	3,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Question		6,1%	6,1%	3,0%	3,0%	0,0%	3,0%	0,0%	30,3%
Sincerity-Politeness		6,1%	0,0%	0,0%	3,0%	3,0%	0,0%	9,1%	3,0%

Irritation

Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	6,1%	0,0%	Admiration	3,0%	6,1%	9,1%	6,1%	0,0%
Arrogance	6,1%	3,0%	Arrogance	3,0%	6,1%	3,0%	6,1%	15,2%
Authority	3,0%	18,2%	Authority	27,3%	30,3%	12,1%	15,2%	0,0%
Declaration	9,1%	3,0%	Declaration	0,0%	0,0%	9,1%	3,0%	0,0%
Doubt-Incredulity	9,1%	0,0%	Doubt-Incredulity	3,0%	3,0%	0,0%	3,0%	3,0%
Evidence	9,1%	0,0%	Evidence	12,1%	9,1%	9,1%	15,2%	6,1%
Surprise	6,1%	9,1%	Surprise	12,1%	6,1%	0,0%	0,0%	0,0%
Irritation	33,3%	66,7%	Irritation	33,3%	30,3%	24,2%	33,3%	69,7%
Kyoshuku	0,0%	0,0%	Kyoshuku	3,0%	0,0%	9,1%	3,0%	3,0%
Politeness	9,1%	0,0%	Politeness	0,0%	0,0%	15,2%	12,1%	0,0%
Question	6,1%	0,0%	Question	3,0%	3,0%	3,0%	0,0%	0,0%
Sincerity-Politeness	3,0%	0,0%	Sincerity-Politeness	0,0%	6,1%	6,1%	3,0%	3,0%

<i>Irritation</i>	Presented length of stimuli							
Phr. 8 mores	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration	12,1%	15,2%	9,1%	6,1%	6,1%	6,1%	3,0%	0,0%
Arrogance	9,1%	6,1%	6,1%	9,1%	21,2%	12,1%	12,1%	12,1%
Authority	6,1%	15,2%	33,3%	9,1%	15,2%	3,0%	3,0%	6,1%
Declaration	9,1%	6,1%	3,0%	3,0%	3,0%	0,0%	3,0%	0,0%
Doubt-Incredulity	6,1%	3,0%	6,1%	9,1%	0,0%	3,0%	3,0%	3,0%
Evidence	3,0%	0,0%	21,2%	12,1%	15,2%	9,1%	3,0%	12,1%
Surprise	9,1%	6,1%	3,0%	0,0%	0,0%	0,0%	3,0%	3,0%
Irritation	12,1%	21,2%	12,1%	24,2%	30,3%	63,6%	66,7%	57,6%
Kyoshuku	9,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,0%
Politeness	9,1%	12,1%	0,0%	15,2%	0,0%	0,0%	0,0%	0,0%
Question	9,1%	6,1%	3,0%	6,1%	3,0%	0,0%	0,0%	3,0%
Sincerity-Politeness	6,1%	9,1%	3,0%	6,1%	6,1%	3,0%	3,0%	0,0%

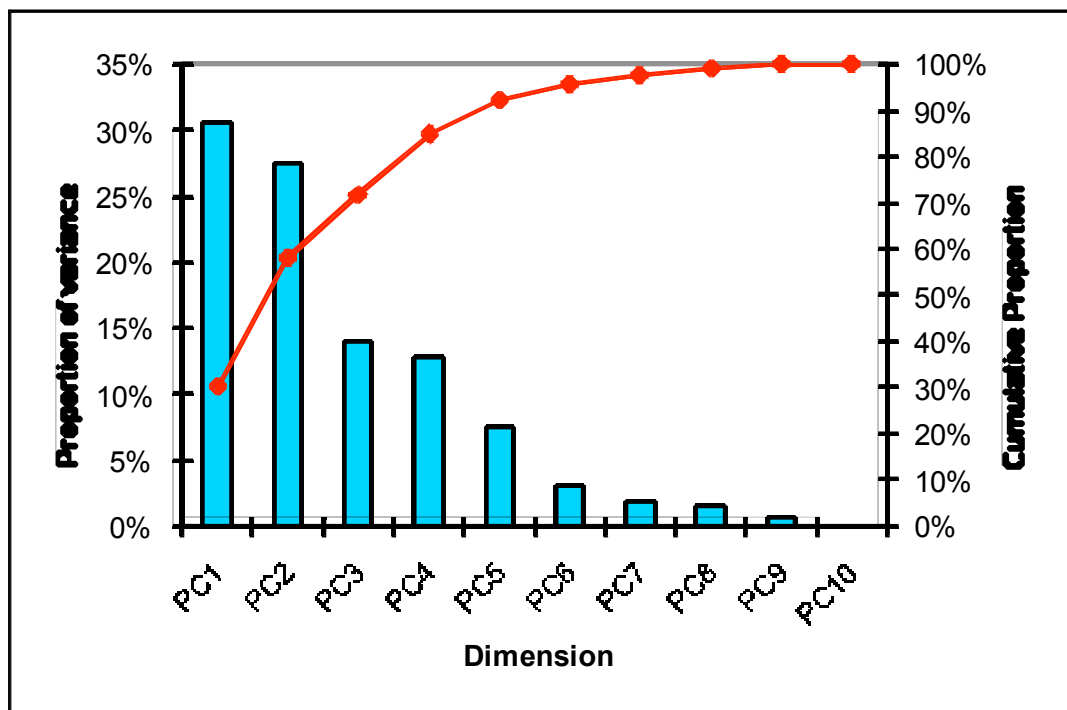
Question

Phr. 2 mores	1/2	2/2	Phr. 5 mores	1/5	2/5	3/5	4/5	5/5
Admiration	6,1%	3,0%	Admiration	3,0%	12,1%	3,0%	0,0%	0,0%
Arrogance	18,2%	15,2%	Arrogance	9,1%	18,2%	9,1%	21,2%	0,0%
Authority	0,0%	0,0%	Authority	0,0%	0,0%	0,0%	0,0%	6,1%
Declaration	18,2%	12,1%	Declaration	45,5%	36,4%	45,5%	30,3%	3,0%
Doubt-Incredulity	6,1%	12,1%	Doubt-Incredulity	3,0%	3,0%	6,1%	3,0%	15,2%
Evidence	12,1%	15,2%	Evidence	9,1%	3,0%	12,1%	6,1%	3,0%
Surprise	0,0%	3,0%	Surprise	3,0%	0,0%	0,0%	0,0%	0,0%
Irritation	6,1%	0,0%	Irritation	0,0%	0,0%	12,1%	6,1%	0,0%
Kyoshuku	15,2%	0,0%	Kyoshuku	9,1%	6,1%	6,1%	18,2%	0,0%
Politeness	12,1%	0,0%	Politeness	12,1%	21,2%	3,0%	9,1%	3,0%
Question	3,0%	39,4%	Question	3,0%	0,0%	0,0%	6,1%	69,7%
Sincerity-Politeness	3,0%	0,0%	Sincerity-Politeness	3,0%	0,0%	3,0%	0,0%	0,0%

<i>Question</i>	Presented length of stimuli							
Phr. 8 mores	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
Admiration	6,1%	9,1%	3,0%	0,0%	0,0%	0,0%	3,0%	0,0%
Arrogance	3,0%	9,1%	3,0%	18,2%	6,1%	12,1%	0,0%	3,0%
Authority	0,0%	3,0%	0,0%	0,0%	0,0%	3,0%	0,0%	0,0%
Declaration	15,2%	21,2%	27,3%	42,4%	33,3%	21,2%	30,3%	3,0%
Doubt-Incredulity	3,0%	6,1%	3,0%	3,0%	6,1%	3,0%	15,2%	21,2%
Evidence	3,0%	6,1%	0,0%	9,1%	9,1%	12,1%	9,1%	3,0%
Surprise	0,0%	0,0%	3,0%	0,0%	3,0%	0,0%	0,0%	0,0%
Irritation	0,0%	3,0%	3,0%	6,1%	0,0%	3,0%	3,0%	0,0%
Kyoshuku	15,2%	9,1%	24,2%	6,1%	12,1%	9,1%	12,1%	3,0%
Politeness	33,3%	18,2%	24,2%	12,1%	15,2%	21,2%	9,1%	9,1%
Question	12,1%	6,1%	0,0%	3,0%	6,1%	9,1%	12,1%	51,5%
Sincerity-Politeness	9,1%	9,1%	9,1%	0,0%	9,1%	6,1%	6,1%	6,1%

- Annexe 4 -

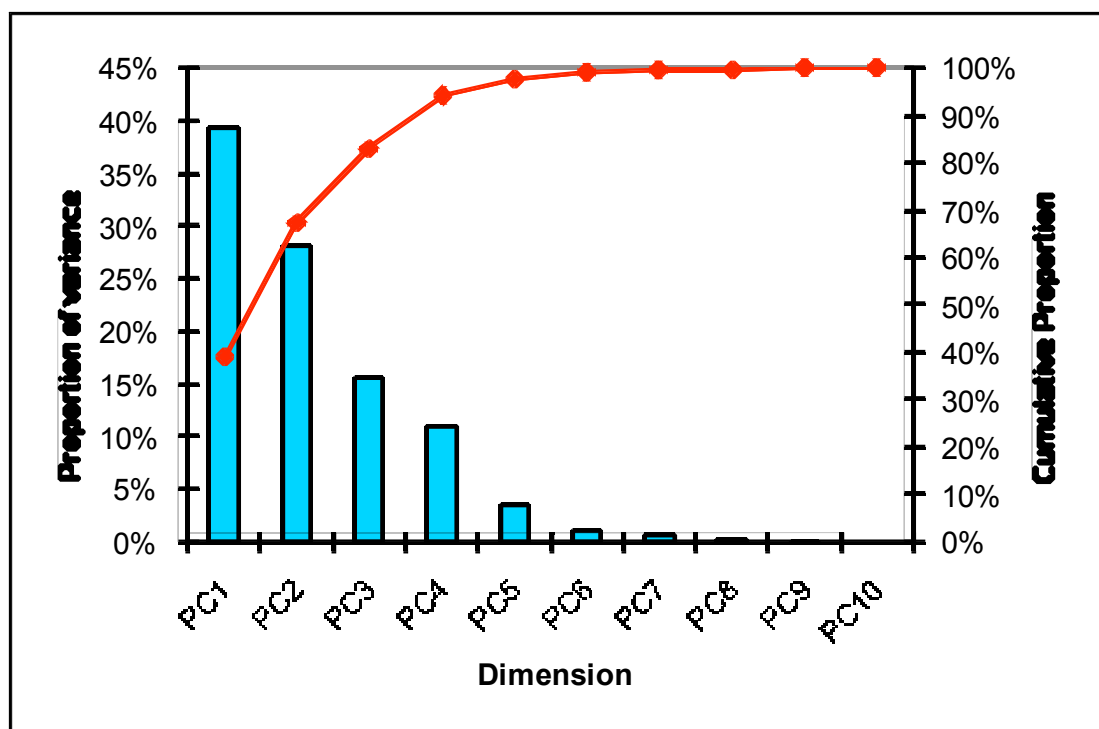
Proportion et proportion cumulative des variables expliquées sur chaque dimension par l'Analyse de Correspondance (AC)



Principal inertias (eigenvalues):

dim	value	%	cum%
[1,] 1	0.675173	30.6	30.6
[2,] 2	0.606455	27.5	58.2
[3,] 3	0.307551	14.0	72.1
[4,] 4	0.281120	12.8	84.9
[5,] 5	0.166889	7.6	92.5
[6,] 6	0.068281	3.1	95.6
[7,] 7	0.045037	2.0	97.6
[8,] 8	0.036091	1.6	99.3
[9,] 9	0.016464	0.7	100.0
[10,] 10	8e-06000	0.0	100.0
[11,]	-----	-----	-----
[12,] Total:	2.203069	100.0	

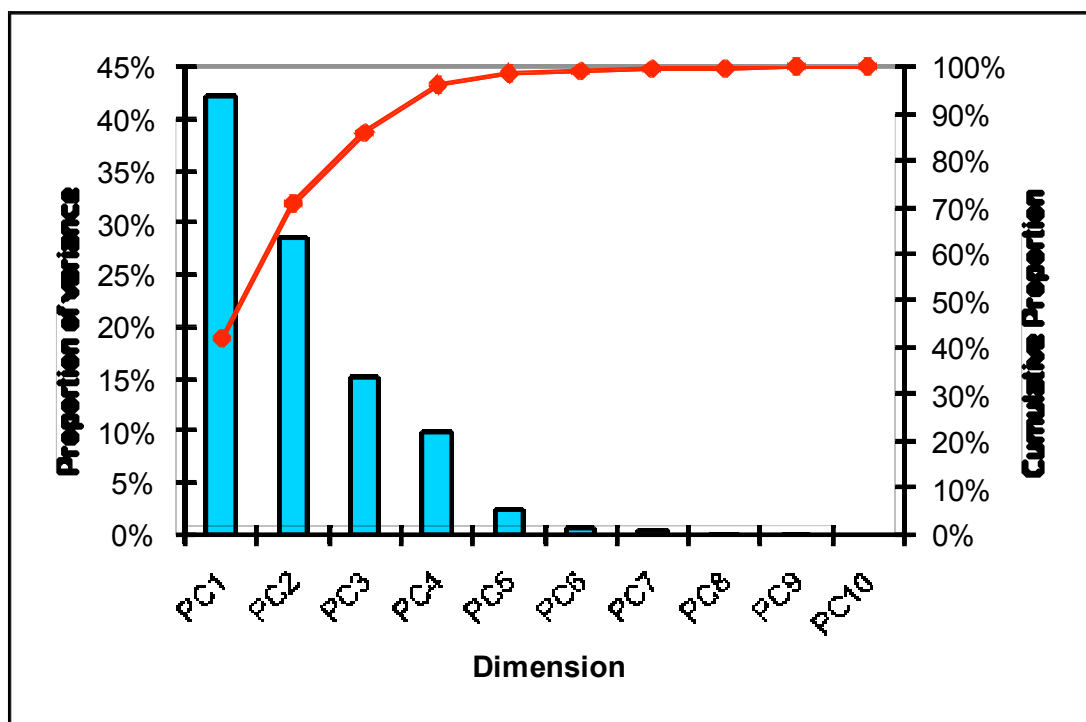
Résultats perceptifs des attitudes anglaises chez les sujets anglais britanniques.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.445831	39.4	39.4
[2,]	2	0.31929	28.2	67.6
[3,]	3	0.177923	15.7	83.3
[4,]	4	0.12441	11	94.3
[5,]	5	0.040273	3.6	97.9
[6,]	6	0.014097	1.2	99.1
[7,]	7	0.007007	0.6	99.7
[8,]	8	0.002483	0.2	99.9
[9,]	9	0.000936	0.1	100
[10,]	10	0.000105	0	100
[11,]		-----		
[12,]	Total: 1.132355 100			

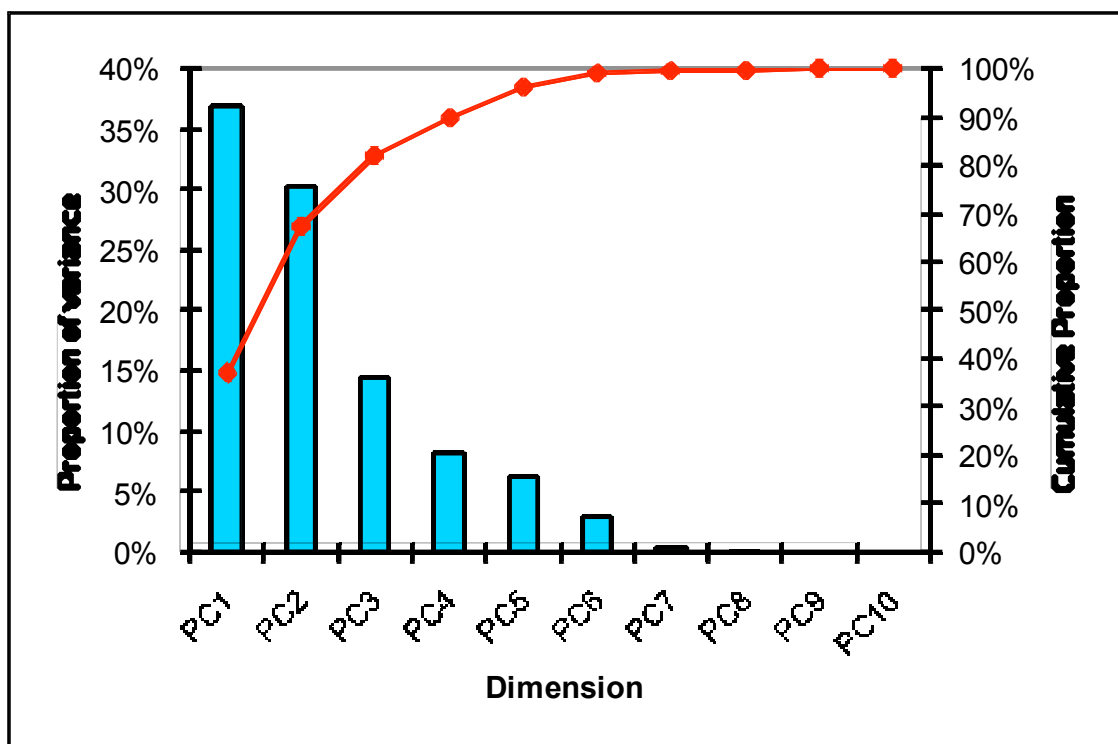
Résultats perceptifs des attitudes anglaises chez les sujets français.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.452080	42.3	42.3
[2,]	2	0.304979	28.6	70.9
[3,]	3	0.163103	15.3	86.2
[4,]	4	0.106556	10.0	96.1
[5,]	5	0.025603	2.4	98.5
[6,]	6	0.008052	0.8	99.3
[7,]	7	0.005170	0.5	99.8
[8,]	8	0.001324	0.1	99.9
[9,]	9	0.000911	0.1	100.0
[10,]	10	0.000169	0.0	100.0
[11,]		-----	-----	
[12,]	Total:	1.067947	100.0	

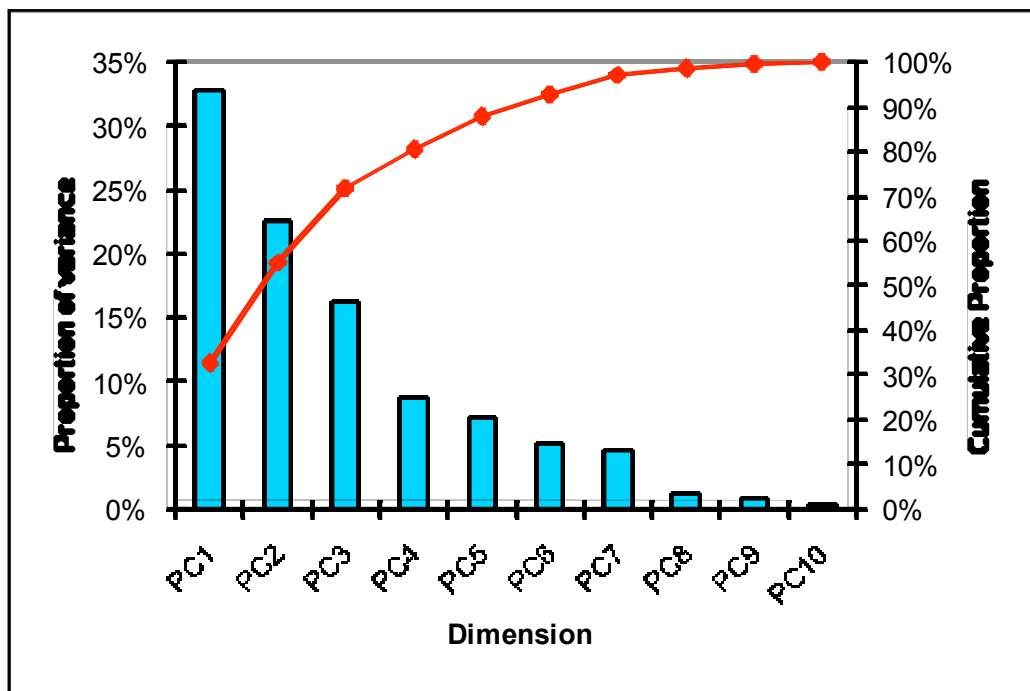
Résultats perceptifs des attitudes anglaises chez les sujets japonais.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.454971	37.0	37.0
[2,]	2	0.373467	30.4	67.4
[3,]	3	0.177331	14.4	81.9
[4,]	4	0.101178	8.2	90.1
[5,]	5	0.077683	6.3	96.4
[6,]	6	0.036484	3.0	99.4
[7,]	7	0.005425	0.4	99.8
[8,]	8	0.001238	0.1	99.9
[9,]	9	0.000611	0.0	100.0
[10,]	10	0.000329	0.0	100.0
[11,]		-----	-----	
[12,]	Total: 1.228717 100.0			

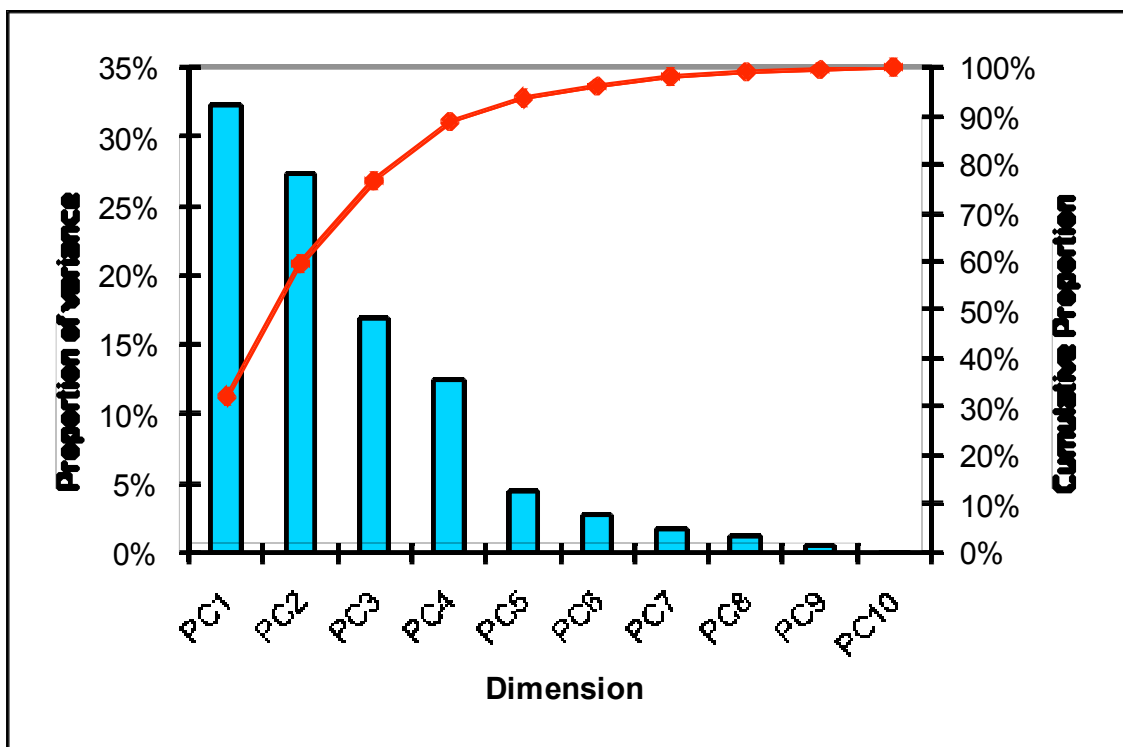
Résultats perceptifs des attitudes anglaises chez les sujets américains.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.691371	32.9	32.9
[2,]	2	0.475100	22.6	55.5
[3,]	3	0.341511	16.2	71.7
[4,]	4	0.184730	8.8	80.5
[5,]	5	0.154105	7.3	87.8
[6,]	6	0.106548	5.1	92.9
[7,]	7	0.096500	4.6	97.5
[8,]	8	0.027039	1.3	98.7
[9,]	9	0.018178	0.9	99.6
[10,]	10	0.008213	0.4	100.0
[11,]	11	1.4e-050	0.0	100.0
[12,]		-----	-----	
[13,]	Total: 2.103308 100.0			

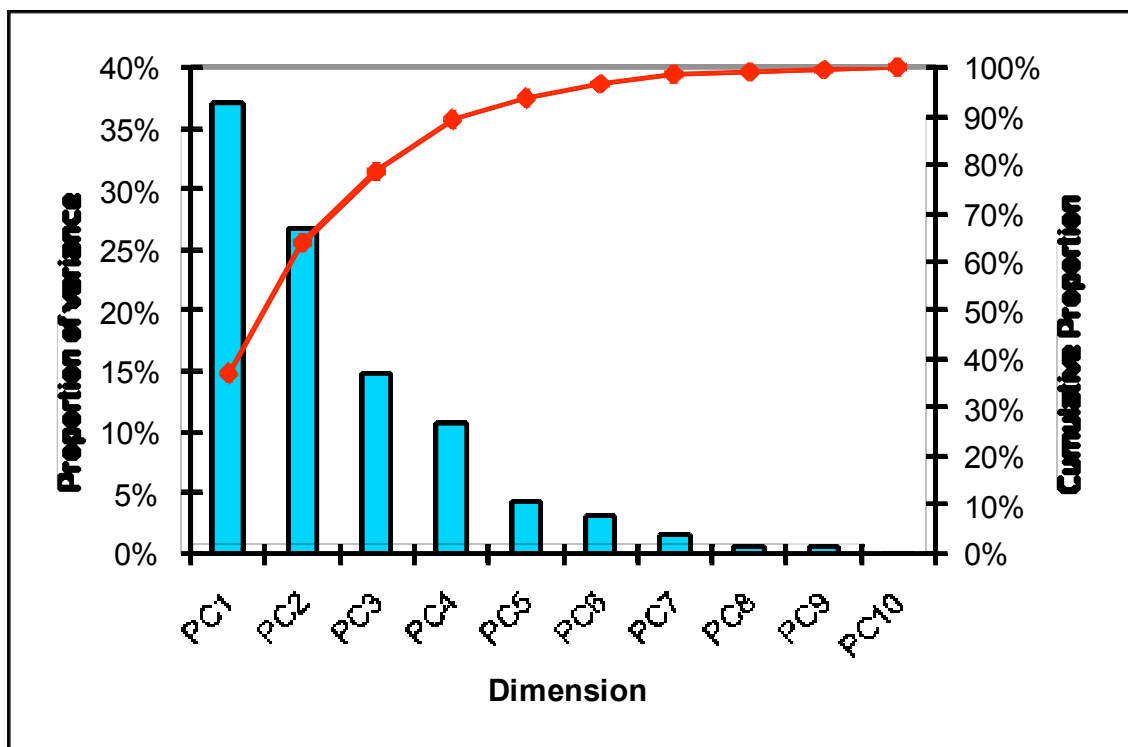
Résultats perceptifs des attitudes françaises chez les sujets français.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.536874	32.3	32.3
[2,]	2	0.455193	27.4	59.7
[3,]	3	0.282340	17.0	76.7
[4,]	4	0.208219	12.5	89.2
[5,]	5	0.074157	4.5	93.7
[6,]	6	0.044224	2.7	96.4
[7,]	7	0.028731	1.7	98.1
[8,]	8	0.020960	1.3	99.4
[9,]	9	0.008115	0.5	99.9
[10,]	10	0.001949	0.1	100.0
[11,]	11	0.000540	0.0	100.0
[12,]		-----	-----	
[13,]	Total: 1.661303 100.0			

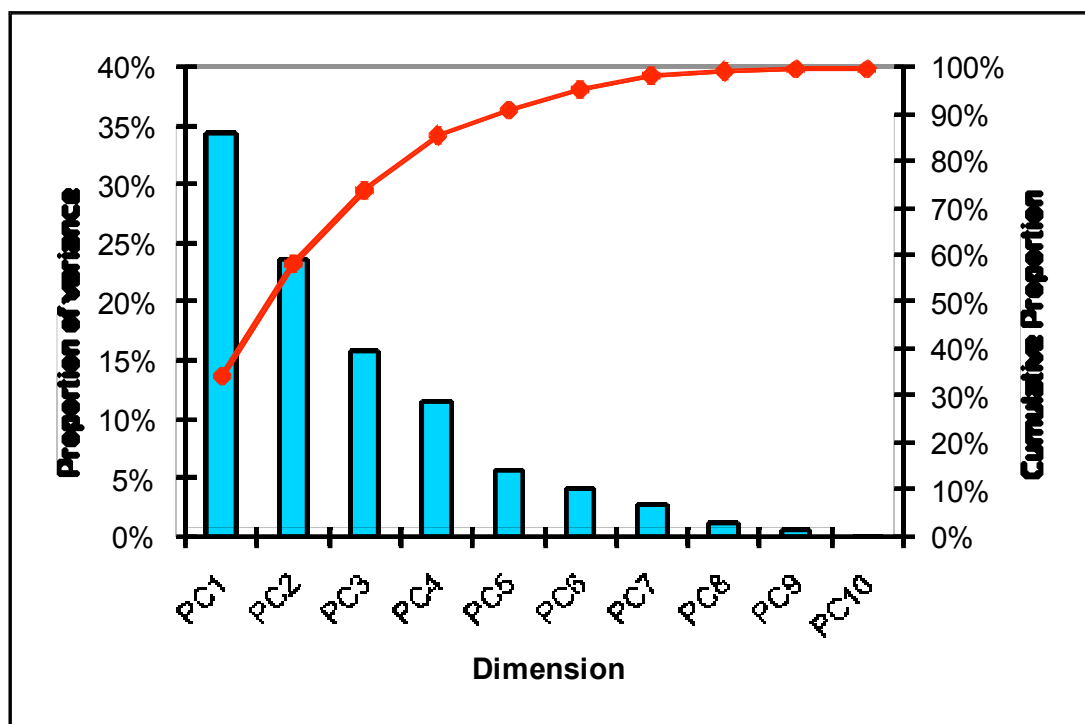
Résultats perceptifs des attitudes françaises chez les sujets japonais.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.474423	37.1	37.1
[2,]	2	0.341949	26.8	63.9
[3,]	3	0.188982	14.8	78.7
[4,]	4	0.137488	10.8	89.5
[5,]	5	0.055899	4.4	93.8
[6,]	6	0.040166	3.1	97.0
[7,]	7	0.021087	1.7	98.6
[8,]	8	0.009537	0.7	99.4
[9,]	9	0.007087	0.6	99.9
[10,]	10	0.000841	0.1	100.0
[11,]	11	2e-06000	0.0	100.0
[12,]		-----	-----	
[13,]	Total: 1.277461 100.0			

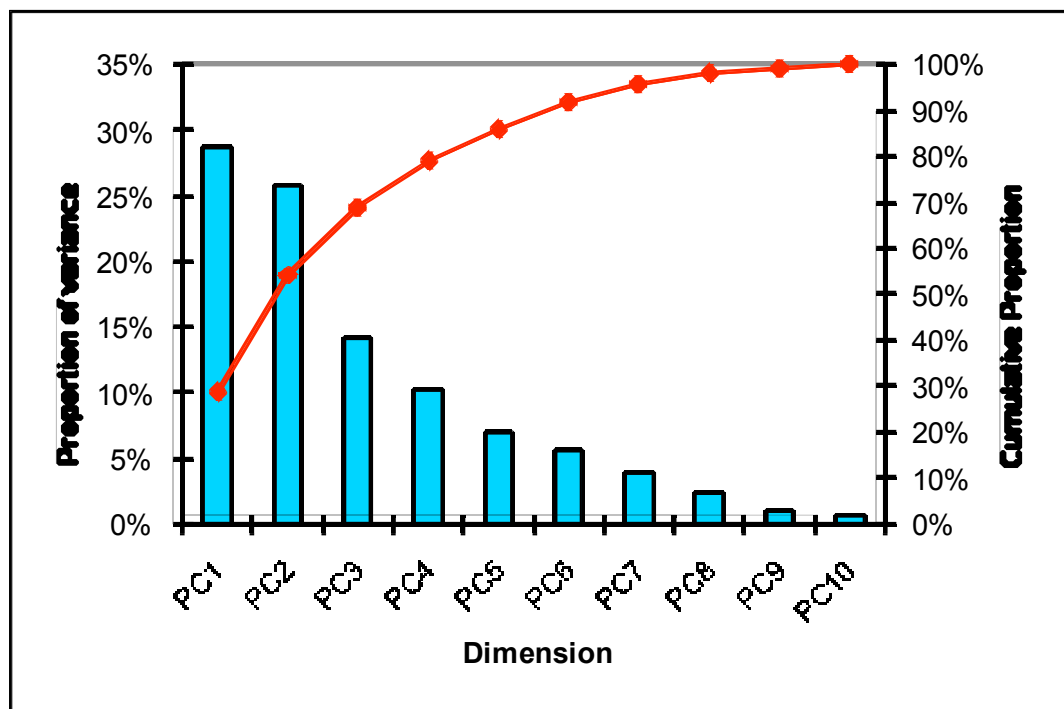
Résultats perceptifs des attitudes françaises chez les sujets américains.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.665175	34.4	34.4
[2,]	2	0.456324	23.6	58.0
[3,]	3	0.306552	15.8	73.8
[4,]	4	0.222601	11.5	85.3
[5,]	5	0.111264	5.8	91.1
[6,]	6	0.078840	4.1	95.2
[7,]	7	0.053817	2.8	98.0
[8,]	8	0.024057	1.2	99.2
[9,]	9	0.012314	0.6	99.8
[10,]	10	0.001796	0.1	99.9
[11,]	11	0.001369	0.1	100.0
[12,]		-----	-----	
[13,]	Total: 1.934109 100.0			

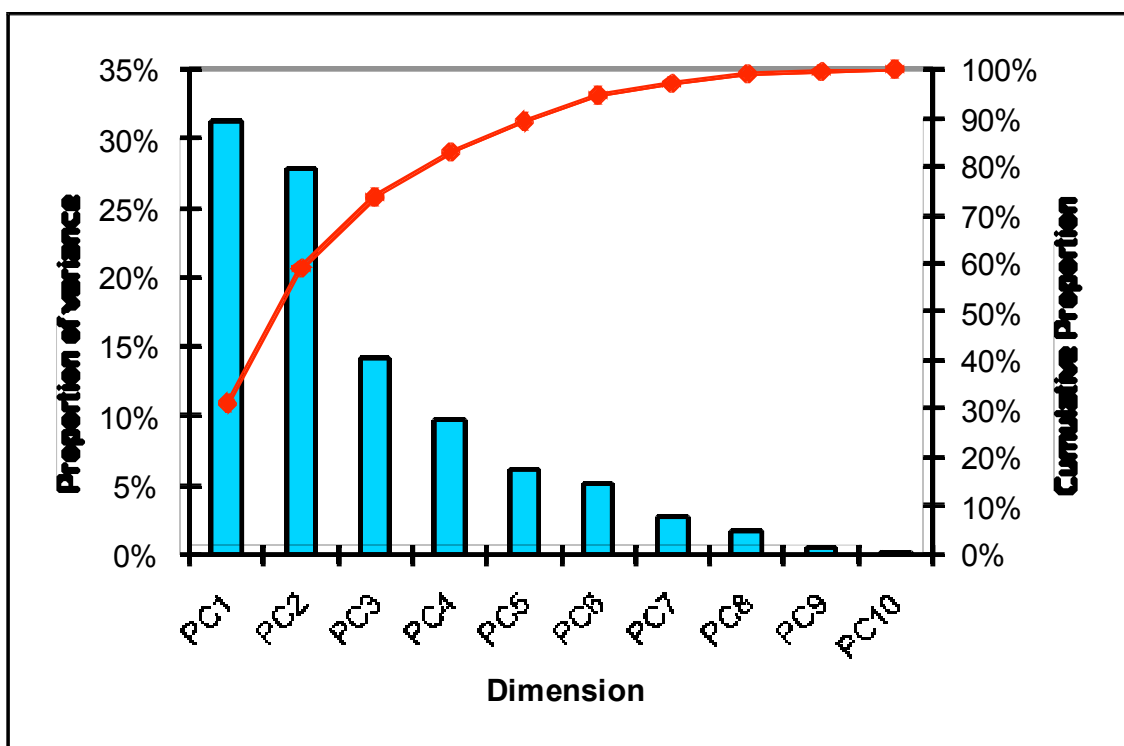
***Résultats perceptifs des attitudes japonaises chez les sujets français
niveau 0.***



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.525471	28.8	28.8
[2,]	2	0.470987	25.8	54.5
[3,]	3	0.260600	14.3	68.8
[4,]	4	0.188223	10.3	79.1
[5,]	5	0.128410	7.0	86.1
[6,]	6	0.102216	5.6	91.7
[7,]	7	0.072548	4.0	95.7
[8,]	8	0.043198	2.4	98.1
[9,]	9	0.020092	1.1	99.2
[10,]	10	0.014935	0.8	100.0
[11,]	11	0.000163	0.0	100.0
[12,]		-----	-----	
[13,]	Total: 1.826842 100.0			

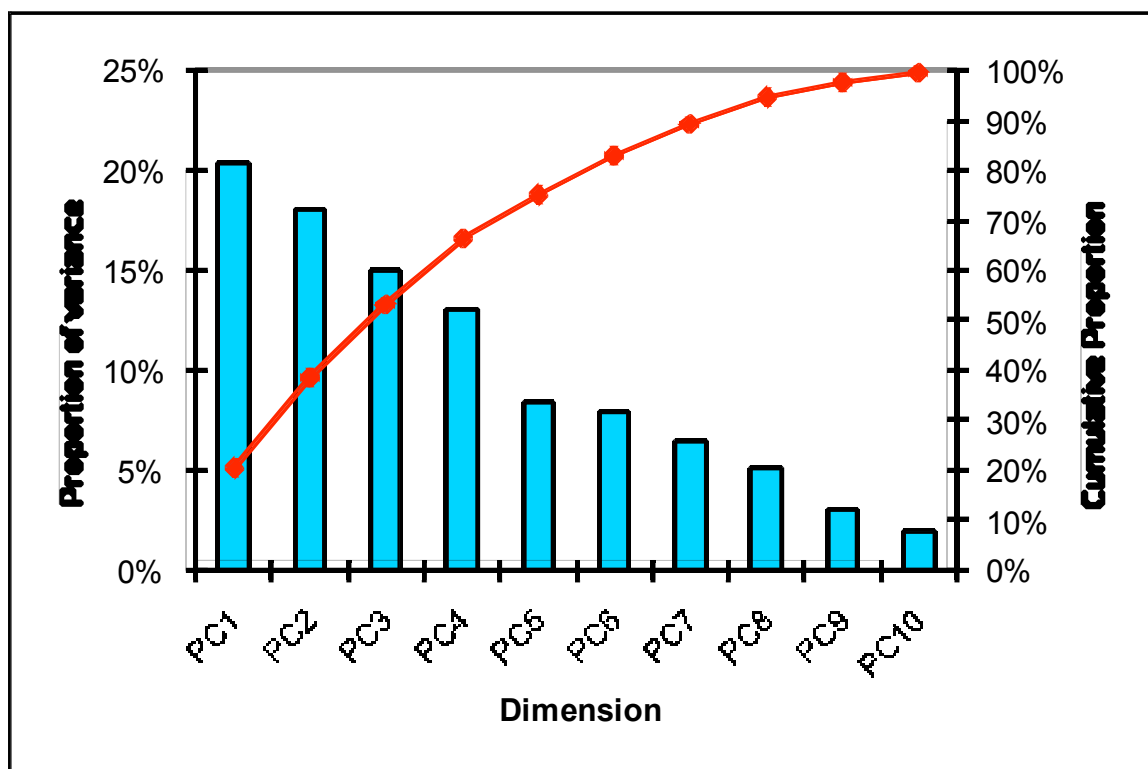
***Résultats perceptifs des attitudes japonaises chez les sujets français
niveau 1.***



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.612943	31.3	31.3
[2,]	2	0.546295	27.9	59.3
[3,]	3	0.278927	14.3	73.6
[4,]	4	0.188925	9.7	83.2
[5,]	5	0.121574	6.2	89.4
[6,]	6	0.100987	5.2	94.6
[7,]	7	0.055705	2.8	97.4
[8,]	8	0.035721	1.8	99.3
[9,]	9	0.010284	0.5	99.8
[10,]	10	0.003775	0.2	100.0
[11,]	11	0.000172	0.0	100.0
[12,]		-----	-----	
[13,]	Total: 1.955307 100.0			

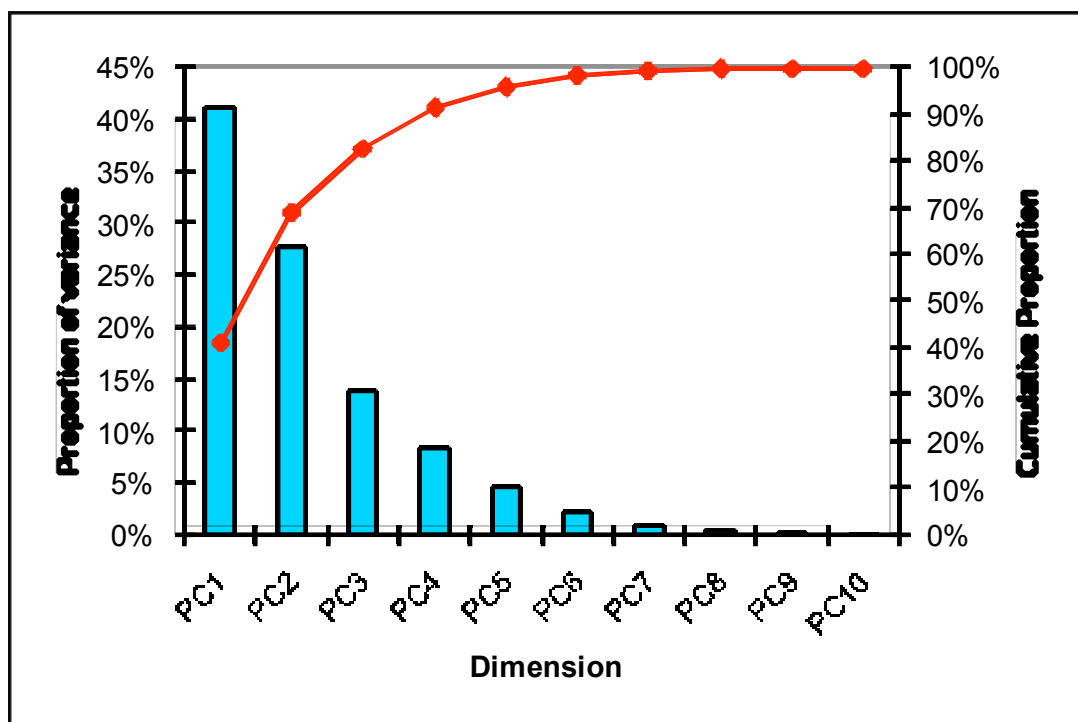
***Résultats perceptifs des attitudes japonaises chez les sujets français
niveau 2.***



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.705307	20.4	20.4
[2,]	2	0.626629	18.1	38.5
[3,]	3	0.517929	15	53.5
[4,]	4	0.451643	13.1	66.6
[5,]	5	0.293993	8.5	75.1
[6,]	6	0.277064	8	83.1
[7,]	7	0.226148	6.5	89.6
[8,]	8	0.176534	5.1	94.7
[9,]	9	0.107318	3.1	97.8
[10,]	10	0.069982	2	99.8
[11,]	11	0.000513	0	99.8
[12,]		-----		
[13,]	Total:	3.45306		99.8

Résultats perceptifs des attitudes japonaises chez les sujets japonais.



Principal inertias (eigenvalues):

	dim	value	%	cum%
[1,]	1	0.497547	41.2	41.2
[2,]	2	0.334406	27.7	68.9
[3,]	3	0.167076	13.8	82.7
[4,]	4	0.102335	8.5	91.2
[5,]	5	0.054939	4.6	95.8
[6,]	6	0.028078	2.3	98.1
[7,]	7	0.011324	0.9	99
[8,]	8	0.006105	0.5	99.5
[9,]	9	0.004197	0.3	99.8
[10,]	10	0.001196	0.1	99.9
[11,]	11	7.4e-05	0	99.9
[12,]		-----		
[13,]	Total:	1.207277	99.9	

Résultats perceptifs des attitudes japonaises chez les sujets américains.

- Annexe 5 -

Le papier d'enquête sur le test perceptif utilisé avant l'expérimentation auditive pour les apprenants français du japonais.

QUESTIONNAIRES (Ver. Française)

informations personnelles

1. Quel est votre prénom ?
2. Quelle est votre nationalité?
3. Quel âge avez-vous ?
4. Informations ayant trait à la langue japonaise
5. Parlez-vous japonais ?
6. Si oui, dans quel cadre utilisez-vous le japonais ?
7. Parlez-vous une autre langue étrangère ? Laquelle ?

→ Si vous êtes étudiant en japonais.

1. En quelle année êtes-vous ?
2. Combien d'heures de japonais avez-vous eu par semaine en 1^{ère} année, 2^{ème} année, ...
3. Avez-vous eu des cours dans lesquels vous pouviez, vous pouvez pratiquer à l'oral le japonais et l'autre langue étrangère?
4. Si oui, Combien d'heures en 1^{ère} année ? En 2^{ème} année ? ... Quel était l'intitulé de ces cours (civilisation, éloquence, ...)?

Informations concernant d'éventuels séjours linguistiques

1. Vous avez déjà été à l'étranger ?
2. Si oui, Où ?
3. Si oui sur Q1, En tant que touriste ou pour une autre raison (ex. séjour linguistique) ?
4. Pour combien de temps ?

→ Si vous êtes déjà allé au Japon :

1. Combien de temps ?
2. Dans quelle région?
3. En tant que touriste ou pour une autre raison ?
4. Vous habitiez chez une famille japonaise ?
5. Vous avez eu des cours de japonais sur place ?

Information portant sur le contexte et l'entourage japonais

1. Dans la vie quotidienne, vous avez des occasions de rencontrer la langue japonaise? (quel support, matériel, ...)
2. Si le support est oral (type dessin animé en V.O. ou chanson japonaise, etc...), à quelle fréquence entendez / écoutez-vous en japonais ces supports ?
3. Vous arrive-t-il de rencontrer des Japonais ? A quelle fréquence ?
4. Vous lui, leur parlez en quelle langue ?

QUESTIONNAIRES (Ver. japonaise)

Personal informations

氏名 ()

1. あなたが日常身の回りの人と会話するときに使用する言語をすべて記入してください。複数ある場合、だれとその言語を使用するかを詳しく教えてください。
2. あなたは何歳ですか。

Informations about foreign language

1. あなたは日本語以外の言葉を話しますか?
2. もし話す場合、その言語を使用するのはどのようなときですか。

→ もしあなたが外国語を勉強している場合

1. その言語を何年間勉強していますか?
2. 一週間に何時間ぐらい勉強していますか?
3. 語学学校や短大で外国語を勉強したことがありますか。
4. もしある場合、会話の授業はありましたか。また会話の練習の時間は一週間に何時間ぐらいでしたか?

Informations concerning occasional stay in the abroad

1. 今まで海外へ行ったことがありますか?
2. もしある場合、それはどこですか?
3. 観光でしたか、もしくは語学留学でしたか?
4. どのぐらいの期間いきましたか? (複数ある場合、個別に回答してください)

Information about linguistic environment

1. 日常、外国語に接する機会がありますか。(例：映画やビデオ、外国人など)
2. もし、外国語に接する機会がある場合、それはどのくらいの頻度ですか。
(例：毎日、ときどき、ほとんどない等)
3. 外国人と話す機会がありますか？ もしある場合それはどのくらいの頻度ですか？
4. 1.2.であると答えた方は、その外国人に対して何語で話しますか？

Le papier d'enquête sur le test perceptif utilisé après l'expérimentation auditive pour les apprenants français du japonais.

QUESTIONNAIRE APRES LE TEST PERCEPTIF
(Version française)

1. Est-ce qu'il y avait des attitudes pour lesquelles vous avez hésité ?
2. Y avait-il d'autres attitudes n'ayant pas été présentées mais que vous avez perçues ?
3. Avez-vous déjà entendu ces attitudes en japonais ?
4. Y a-t-il des attitudes que vous n'avez pas choisies ?
5. Est-ce que vous avez compris le sens des phrases que vous avez entendues ?
6. Si oui, est-ce que le sens de chaque phrase vous a gêné par rapport à l'attitude exprimée ?

QUESTIONNAIRE APRES LE TEST PERCEPTIF
(Version japonaise)

1. 選択した態度の間で迷ったものはありましたか？
2. 聞き取った態度がこの選択肢の中になかったといったことはありましたか？
3. 聞き取った態度は、いままで聞いたことがありましたか？
4. 選択しなかった態度はありましたか？
5. 聞こえてきた言葉の意味はわかりましたか？
6. もしわかった場合、その文の意味と聞き取った態度との間で違和感を感じることはありましたか？

以上です。ご協力ありがとうございました。

11. Résumé d'une sélection d'articles publiés

1. *Multimodal indices to Japanese and French prosodically expressed social affects (Language and Speech, en cours de révision)*

Bien que certains travaux ont exploré l'expression des émotions, peu d'étude ont été menées sur la manière dont sont combinées les modalités audio et visuelle lors de la production d'affects sociaux ou d'expressions « attitudinales ». Cet article présente une étude perceptive d'expressions audio-visuelles de 12 attitudes japonaises et de 6 attitudes françaises afin d'envisager la contribution des modalités audio et visuelle lors de la communication affective. L'importance relative de chaque modalité dans le décodage perceptif des expressions réalisées par quatre locuteurs est analysée dans cette première étape vers une meilleure compréhension de leur influence sur l'expression des affects sociaux. Ensuite, les productions audio-visuelles de deux locuteurs (le locuteur le plus performant de chaque langue) sont acoustiquement (F0, durée et intensité) et visuellement (en terme des Unités d'Action) analysées afin d'identifier la relation entre les paramètres objectifs et la perception des auditeurs pour ces affects sociaux.

Les caractéristiques objectives les plus pertinentes, qu'elles soient acoustique ou visuelle, sont discutées dans une perspective bilingue : ainsi est comparée l'influence relative de la fréquence fondamentale pour des expressions attitudinales en deux langues et l'apport de la dimension de qualité de voix en japonais est soulignée.

2. *Recognition of Japanese attitudes in Audio-Visual speech (Speech Prosody 2008, Campinas, Brésil.)*

L'objectif de cette recherche est d'étudier comment les auditeurs japonais reconnaissent 12 attitudes prosodiques japonaises en audio-visuel. Des influences significatives des deux locuteurs et de la modalité ont été observées. Généralement la condition audio-visuelle a montré le meilleur score de reconnaissance, et des spécificités de comportement pour les modalités audio et visuelle ont été observées. Les attitudes ont été regroupées au sein de catégories perceptives de haut-niveau pour le premier locuteur : les expressions de politesse, les attitudes « interrogatives » et les expressions d'imposition de l'avis du locuteur. Les attitudes de *kyoshuku* et de *surprise* sont particulièrement bien reconnues grâce aux informations visuelles.

3. *Perception of French Audio-Visual Prosodic Attitudes (Speech Prosody 2008, Campinas, Brésil.)*

Cette étude expérimentale permet de mieux comprendre la contribution des modalités audio et visuelle lors de la communication affective. L'article présente une étude perceptive de l'expression de six attitudes françaises audio-visuelles. L'importance relative de chaque modalité dans le décodage perceptif des expressions réalisées par deux locuteurs est analysée. Les expressions des deux locuteurs sont évaluées perceptivement afin de tester l'influence de la performance du locuteur et de chaque modalité sur la perception des auditeurs.

4. *Cross-listening of Japanese, English and French social affect : about universals, false friends and unknown attitudes (International Congress of Phonetic Science 2007, SaarBrücken, Germany)*

Sept phrases japonaises n'ayant pas de connotation émotionnelle particulière, qui ont été produites avec 12 attitudes différentes, sont analysées. Les auditeurs sont 15 Japonais natifs, 15 Français et 20 Américains. Tous les auditeurs non-natifs n'ont aucune connaissance sur la culture ou de la langue japonaise. Pour chaque phrase et chaque attitude, les auditeurs ont eu à choisir quelle expression attitudinale correspond à l'intention du locuteur parmi les 12 choix attitudinaux possibles. Les résultats montrent que les Japonais ont reconnu toutes les attitudes au-dessus du seuil de hasard, mais avec quelques confusions, surtout pour les expressions de politesse. A contrario, les Français et les Américains ne reconnaissent pas les deux expressions culturelles de politesse. Plus particulièrement, l'expression de *kyoshuku*, un type de la politesse qui n'apparaît pas comme expression conventionnelle dans la société occidentale, n'a pas été correctement décodé par les Français ni par les Américains. Ils l'ont reconnu comme une expression d'arrogance ou d'irritation.

5. *How prosodic attitudes can be false friends : Japanese vs. French social affects (Speech Prosody 2006, Dresden, Allemagne)*

Les affects sociaux produits par un locuteur durant l'interaction verbale sont liés aux intentions de ce locuteur et ce type d'expression est largement influencé par la langue et la culture du locuteur. Elles sont par ailleurs pour a plupart contrôlées volontairement. Sept phrases japonaises n'ayant aucune connotation particulière, qui ont été produites avec 12 attitudes différentes sont analysées. Les auditeurs sont 15 Japonais natifs, 15 Français n'ayant aucune connaissance de la langue et la culture japonaises, et 32 apprenants français du japonais (niveau débutant et intermédiaire). Ils leur ont été demandés de choisir l'attitude qui correspond à l'intention du locuteur parmi les 12 expressions attitudinales existantes. Les résultats montrent que les Japonais ont reconnu toutes les attitudes au-dessus du seuil de hasard. Cependant, quelques confusions sont relevées, principalement pour à l'intérieur des expressions de politesse (i.e. *sincérité-politesse* vs. *kyoshuku*). A contrario, l'expression de *kyoshuku*, un type de la politesse qui n'apparaît pas comme expression conventionnelle dans la société française, n'a pas été correctement décodé par les Français. Ils l'ont reconnu comme de

l'arrogance ou de l'irritation, même dans le cas des apprenants français ayant acquis suffisamment de compétence du japonais oral. Les apprenants débutants ont très bien reconnu l'*admiration*, l'*autorité* et l'*irritation*, mais ils n'arrivent pas encore à différencier la *question-simple* de la *déclaration*, ce que les apprenants au niveau intermédiaire ont très bien discriminé.

6. 発話態度の文化的特性と「偽の友達」－日仏語対照研究を通して－

Sept phrases japonaises n'ayant pas de connotation émotionnelle particulière, qui ont été produites avec 12 attitudes différentes, sont analysées. Les auditeurs sont 15 Japonais natifs, 15 Français n'ayant aucune connaissance de la langue et la culture japonaises, et 32 apprenants français du japonais qui ont été séparés en deux groupes selon leur compétence en japonais : niveaux débutant et intermédiaire, évalué grâce au test « SPOT ». Ils leur ont été demandés de choisir l'attitude qui correspond à l'intention du locuteur parmi les 12 expressions attitudinales existantes. Les résultats traités grâce à une « quantification IV de Hayashi », qui est une sorte de MultiDimensional Scaling (MDS) asymétrique, indiquent que le comportement perceptif des auditeurs japonais est différent de celui des auditeurs français. Cependant, le comportement perceptif de certaines attitudes chez les apprenants français de niveau intermédiaire ressemble à celui des auditeurs japonais en comparaison du comportement des auditeurs français naïfs et débutants. Notamment, les auditeurs japonais montrent des confusions entre les expressions de politesse, alors qu'au contraire, les sujets français reconnaissent le *kyoshuku*, une des expressions de politesse, comme de l'*arrogance* ou de l'*irritation*, et ceci même pour les apprenants français ayant acquis suffisamment de compétence en japonais oral. Par la suite, l'analyse acoustique des expressions attitudinales clarifie la cause des confusions entre attitudes.

12. Sélection d'articles publiés

1. MULTIMODAL INDICES TO JAPANESE AND FRENCH PROSODICALLY EXPRESSED SOCIAL AFFECTS (LANGUAGE AND SPEECH, EN COURS DE RÉVISION)	321
2. RECOGNITION OF JAPANESE ATTITUDES IN AUDIO-VISUAL SPEECH (SPEECH PROSODY 2008, CAMPINAS, BRÉSIL.)	343
3. PERCEPTION OF FRENCH AUDIO-VISUAL PROSODIC ATTITUDES (SPEECH PROSODY 2008, CAMPINAS, BRÉSIL.)....	347
4. CROSS-LISTENING OF JAPANESE, ENGLISH AND FRENCH SOCIAL AFFECT : ABOUT UNIVERSALS, FALSE FRIENDS AND UNKNOWN ATTITUDES (INTERNATIONAL CONGRESS OF PHONETIC SCIENCE 2007, SAARBRUCKEN, GERMANY)	351
5. HOW PROSODIC ATTITUDES CAN BE FALSE FRIENDS : JAPANESE VS. FRENCH SOCIAL AFFECTS (SPEECH PROSODY 2006, DRESDEN, ALLEMAGNE)	355
6. 発話態度の文化的特性と「偽の友達」 - 日仏語対照研究を通して -	359

MULTIMODAL INDICES TO JAPANESE AND FRENCH

PROSODICALLY EXPRESSED SOCIAL AFFECTS

Albert Rilliard (1), Takaaki Shochi (2), Jean-Claude Martin (1), Donna Erickson (3) & Véronique Aubergé (2)
(1) LIMSI-CNRS, Orsay, France; (2) GIPSA-Lab, Grenoble, France; (3) Showa Music University, Kawasaki City, Japan
Corresponding author:
Albert RILLIARD
LIMSI-CNRS
Bat. 508 – BP 133
91403 Orsay Cedex
France
Tel.: +33 169858102
Email: albert.rilliard@limsi.fr

ABSTRACT

Whereas several studies have explored the expression of emotions, little is known on how the visual and audio channels are combined during production of what we call the more controlled social affects, e.g., “attitudinal” expressions. This paper presents a perception study of the audio-visual expression of twelve Japanese and six French attitudes in order to understand the contribution of audio and visual modalities for affective communication. The relative importance of each modality in the perceptual decoding of the expressions of four speakers is analysed as a first step toward a deeper comprehension of their influence on the expression of social affects. Then, the audio-visual productions of two speakers (one for each language) are acoustically (F0, duration and intensity) and visually (in terms of Action Units) analysed, in order to match the relation between objective parameters and listeners’ perception of these social affects.

The most pertinent objective features, either acoustic or visual are then discussed, in a bilingual perspective: e.g. the relative influence of fundamental frequency for attitudinal expression in both languages is discussed, and the importance of a certain aspect of the voice quality dimension in Japanese is underlined.

KEY WORDS

Audio-visual Prosody; Attitudes; Perception; Japanese; French;

INTRODUCTION

Some theoretical models of affect claim that affective expression may be controlled at different levels of cognitive processing (e.g. Scherer & Ellring, 2007), from involuntarily to intentionally, deliberately controlled expression of the speaker's attitudes (Danes, 1994; Ohala, 1996). Our position (Aubergé, 2002) is to distinguish *attitudes* vs. *emotions* as a difference in the nature of the control by the speaker (voluntary vs. involuntary), not the affective value carried by the expressions. Moreover, attitudinal expressions are considered here as affects conventionally encoded in a culture and a language, while emotional expressions may not have been socially constructed to this same extent. According to Damasio (1994), spontaneous emotions involve physiological changes, and are spontaneously expressed inside a "body loop". But emotions can also be simulated, outside this body loop, using the memory of the somatic states associated to a past emotion. These simulation capabilities allow one to express emotion during conversation, and allow actors to act them. It may be that some people are able to detect the imitation, while others are not (Audibert et al., 2008). We make the hypothesis that this competence of playing emotions could be used for simulated emotions and that different types of processes exist for authentic vs. simulated expressions. In this view, reproduction of an emotion, sincerely or not, when it occurs outside this "body loop", involves the same type of control as that which produces attitudinal affects that have been specifically built by the language and the culture. Therefore, as all socially encoded tools, such attitudinal expressions must be learned during the child's developmental phase, and they also must be learned by foreign language learners when these attitudes are not shared by L1 and L2 languages (Danes, 1994).

Recent finding by Paulmann et al. (2008), based on an ERP study, have shown that listeners are able to distinguish neutral prosody from an emotional one (whatever the valence and the label of the emotion) as soon as 200 ms (approximately one syllable) after the stimulus onset. This finding critically differs from the gating experiments held on attitudinal speech (Shochi, 2008) where

listeners may decode the attitudinal meaning at various points in time within the utterance, e.g., from the beginning of the utterance (in the same way reported by Pullman et al., *ibid* for emotions), or gradually throughout the utterance, or even at a specific point in the utterance, either at the middle or end. We therefore propose that attitudes and emotions can be encoded by the producer, using the same dimensions, but – to some extent – may be perceived differently by the receiver (cf. Aubergé & Cathiard, 2003). In other words, the same acoustic and visual dimensions may convey amongst other things both attitudinal and emotional information, but these two kinds of information may be differentiated by the variation of the parameters through time, which are controlled differently. For example the affect labelled as "*surprise*" may be either an attitude or an emotion, according to the control exerted by the speaker on his expression. He may want to voluntarily express some surprise in order to perform an illocutory act (Danes, 1994), and therefore the prosodic parameters carrying the surprise information will vary according to the linguistic structure of the utterance. Or he may be really surprised and express this surprise involuntarily: the same prosodic dimensions may be used to carry the emotional information, but they may have a different morphology, according to the time course of this emotional surprise. These different expressions of surprise may be recognized as an attitude or an emotion due to the existing correlation among the variations of the prosodic dimensions, the linguistic structure of the utterance, and/or the occurrence of a "surprising" event.

Studies on such attitudes, or *social affects*, are important because an attitude is part of the global meaning of speech acts (Danes, 1994): even if the speaker does not express any attitude by performing a simple declaration, he can still be expressing an attitude, e.g., "*the speaker decides not to give information on his attitude*".

But, as claimed by Pavlenko (2005) while advocating the necessity of cross-cultural studies on language and emotions, the role of languages and cultures on the production and perception of affects has to be taken into account: some affects tend towards universality, whereas others seem

more specific to one culture (see also Danes, 1994) for their meaning as well as for their encoding. Therefore, similar to emotions, the study of attitudinal expressions may benefit from a cross-cultural approach (e.g. Shigeno, 1998): attitudinal values can exist or not in one language, and their audiovisual expression in a specific language may not be recognized by foreign speakers – or may be ambiguous in the learner's language (Scherer et al., 2001; Shochi et al., 2006 for a comparison of Japanese, English and French).

Such a cross-cultural approach recalls the works of Ekman (1999) on facial expression of emotions amongst cultures, which led us to examine what modalities carry the affects to the interlocutor. Attitudinal expressions are linked to language and as such are classically described as one of the basic functions of prosody (Rossi et al., 1981; Fónagy, 2003). This expressive function has been studied for a long time with regard to its acoustical aspect (cf. Allerton & Crutenden, 1978; Fónagy et al., 1984; Danes, 1994). But more recent works emphasised the multimodal nature of prosody's functions: e.g. the one of demarcation (Barkhuysen et al., 2006), or the feeling of knowing (Swerts & Kramer, 2005). It seems obvious that in face-to-face interactions, attitudes are perceived within the multimodality of speech (Barkhuysen et al., 2007).

Numerous studies of attitudes or social affects have been done (e.g. van Heuven et al., 1997; Morlec et al., 2001; Bänziger & Scherer, 2005; Campbell, 2005; de Moraes, 2008) as well as in cross-linguistic contexts (Barkhuysen et al., 2007; Shochi et al., 2007), but mainly in their acoustic modality only. As the study of the multimodal expression of affects (in the broad sense) is still a recent field of research (Scherer, 2003; Scherer & Ellring, 2007), only a few studies question directly the specific topic of social affects, as differentiated from that of emotions (Granström & House, 2005).

As attitudes belong to language, are not the expression of an emotional state, are learned by the speaker, and are able to be produced voluntarily during an interaction, we assume that attitudinal expressions recorded in a lab might not suffer from the same bias as much as do an acted

emotional corpus, as has been observed to be different for spontaneous emotional speech (Wiling et al., 2006; Barkhuysen et al., 2007).

Given the particular importance of multimodality in affective communication (Scherer & Ellring, 2007), this research examines the differences between the production and perception of multimodal expressions of social affect in both Japanese and French. This study aims at investigating (1) the relative contribution of audio and visual modalities to prosodic attitudes, (2) whether the facial indices may have a significant impact on the perception of prosodic attitudes; and (3) the influence of speaker's performance on the recognition of attitudes. The recognition tests used in this study are described by Shochi et al. (2008) for Japanese and Rilliard et al. (2008) for French. We reported the observed results, including acoustical and video analyses of the corpora, in order to compare them to the subjective results. We also discuss the differences and the similarities in audio and visual modalities for both Japanese and French data.

JAPANESE & FRENCH CORPUS

A limited set of audio-visual material was recorded in both Japanese and French for this study. The attitudinal expressions were recorded on utterances with an affectively neutral meaning in order to avoid interference between lexicon and prosody (Banse & Scherer, 1996; Mozziconacci, 1998). Lexically expressive sentences may introduce a bias both in the location of prosodic information on the lexically loaded word and in the perception test, as listeners may be influenced by lexical cues, as well as prosodic ones (see Grichkovitsova et al., 2007, for examples of such interaction between prosodic and lexical levels). In order to ease the objective comparison of prosodic parameters, speakers produced each utterance with all the attitudes chosen for one language.

The choice of attitudes for each language is based on didactic foreign language literature (see details for each language in subsequent sections), which provides an empirical field analysis of

the culturally controlled expressions for a language. Our selection of attitudes for the two languages is therefore based on this knowledge, which explains the different sets of attitudes observed for the two languages. Since the literature is intended to help teachers express these attitudes as naturally as possible for the learners, hints for the communication situation as well as for the way to produce the expressions in speech, gesture and face, are also provided. In addition, the recordings used in this study are based on previous works on prosodic attitudes (Shochi et al., 2006, for Japanese and Morlec et al., 2001, for French). The main differences between the present work and the two preceding ones are (1) audio-visual recording of the attitudes, (2) recording of two different speakers in order to measure the influence of individual performance over recognition scores, and (3) recording paradigm designed to set the speaker in a somewhat natural condition of production for each attitude. Whereas both Morlec et al. (2001) and Shochi et al. (2006) recorded the audio-only attitude by eliciting them sitting alone in a sound-proof room producing the required sentence with the required attitude, in the present study, the speakers were instructed to produce each sentence in order to express one attitude, as an answer to a statement produced by a partner – i.e. we tried to reproduce a basic communication context suitable for eliciting the production of the expression. A training session, in which speakers were asked to behave as naturally as possible, without any constraints on their expressive audiovisual strategy, preceded the actual recordings. Since this work aimed at recording those attitudes specifically prescribed by didactic language teaching materials, we attempted to use trained language teachers, not actors.

Selection of 12 Japanese Attitudes

A set of 12 Japanese attitudes which were validated in Shochi et al. (2006) was used. These attitudes were selected according to the literature (Mackawa, 1998; Erickson et al., 2003; Sadanobu, 2004) and Japanese language teaching methods (Mizutani & Mizutani, 1979): *doubt-incredulity* (DO), *obviousness* (EV), *exclamation of surprise* (SU), *authority* (AU), *irritation* (IR), *arrogance* (AR), *sincerity-politeness* (SIN), *admiration* (AD), *kyoshuku* (KYO), *simple-politeness* (PO),

declaration (DC) and *interrogation* (IN) (see Shochi et al. 2006, for definitions). Some of these attitudes are specific or specifically important for the Japanese culture, especially those linked to the politeness strategy: *simple-politeness*, *sincerity-politeness* and *kyoshuku* vs. *arrogance*. The sincerity-politeness attitude appears when a socially inferior speaker is talking to someone superior to him in Japanese society: the speaker expresses a serious and sincere intention by using this prosodic attitude. The *kyoshuku* attitude (there is no lexical entry to translate this in English) is a typically Japanese cultural attitude. Even if such situations occur in all cultures, the Japanese language has chosen to encode this situation as a prosodic attitude (“attitudinene”). A speaker uses *kyoshuku* when he wants to express a conflicting opinion to an interlocutor considered socially superior aiming to not disturb him but to help him, or when the speaker desires to get a favor from his superior. It is described by Sadanobu (2004, p. 34) as “a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker’s consciousness of the fact that his/her utterance of request imposes a burden to the hearer”.

Selection of 6 French Attitudes

Following the work done by Morlec et al. (2001) on French prosodic attitudes, and based on studies by Fónagy et al. (1984), Calbris & Montredon (1981), Calamand (1973) and Calbris & Porcher (1989), 6 attitudinal expressions were selected for recording a French audio-visual corpus: *declaration* (DC), *simple question* (IN), *obviousness* (EV), *surprise exclamation* (SU), *doubt-incredulity* (DO), *suspicious irony* (SC). These attitudes are defined as follow. *Declaration* is used by the speaker to give some simple information, without expressing any point of view. With *question*, the speaker asks for information without expressing any point of view, and waits for a simple answer. *Surprise* is an expression of amazement when something unexpected happens suddenly. The speaker expresses *obviousness* when he says something he feels is self-evident. *Doubt-incredulity* is used to express a feeling of being uncertain about something or of not believing something that has previously been expressed. With *suspicious irony*, the speaker calls

into question (and even contradicts) a statement his interlocutor has said. He wants to show he doesn't agree, that he's annoyed with what he's hearing, and condemns what is said or what happened by using an intonation contour which could mean: "Yeah, sure! I believe you... (meaning, I don't)".

Corpus recording

Two male Japanese native language speakers (SJ1 & SJ2) produced each sentence (detailed later) with the twelve attitudes. SJ1 is a Japanese native language teacher who teaches various attitudes in his class using pragmatic explanations; SJ2 is a native native speaker with no teaching experience. Two male French native speakers, SF1 & SF2, the former an experienced teacher, the latter not experienced, recorded the French corpus. All four speakers were recorded in a soundproof room at LIMSI. Three sessions were recorded for each speaker (either Japanese or French speaker), separated by pauses and with the possibility for speakers to watch their performance. During the recording, speakers were standing in front of a video camera, with an omnidirectional AKG C414B microphone placed 40 cm to their mouth. The microphone was connected to an USBPre-sound device connected to a computer outside the room, recording the speech signal at 44,1 kHz, 16bits. A digital DV camera (Canon XM1 3CCD) recorded the speakers' performances. Hand claps between each sentence, recorded both by the camera and the microphone, allowed a post-processing, a replacement of the camera sound by the high quality sound recorded by the microphone, synchronized with the claps. Video clips were encoded with a cinepak codec with a 784 x 576 pixels resolution.

For the Japanese attitudes, one 8-mora sentence was selected from a corpus developed and validated in a previous study (Shochi et al., 2006). This sentence ("Nagoyade nomimas.", [nagojade nomimas], meaning "He drinks in Nagoya.") was constructed on a verb-object syntactic structure. The lexical stress position was located on the first mora. In order to express some attitudes like *doubt* or *surprise*, the vowel [u] may be inserted at phrase final position, and in this case, the lexical

stress would be realized at the seventh mora, also. The sentence was constructed in order to avoid any particular affective connotation in any region of Japan.

The French attitudinal corpus is based on three sentences of 4, 5 and 7-syllable length, without any specific affective meanings that can bootstrap or forbid one of the 6 attitudes. The sentences were borrowed from a previous corpus intended to evaluate prosodic expressivity controlled by gesture (d'Alessandro et al., 2007), and constructed with the same principle used for the Japanese corpus. After the recording and the post-processing, the speakers' performances were judged by each of the speakers, and only the 5-syllable length sentence was kept for the perception test, because of its naturalness and rich audio-visual cues: "*Nicolas revenait.*" [nikola kəvɛnɛ] ("Nicolas was coming back"), played with the 6 attitudes. 6 short videos were thus produced for each speaker.

PERCEPTION TEST

Experimental Design

An evaluation test was designed in order to evaluate the relative efficiency of the two modalities to carry the attitudinal information, for each language separately. Listeners were all native speakers of the language they listened to, i.e. French listeners listened to French attitudes whereas Japanese listeners judged Japanese attitudes. The factors controlled during the experiment are:

- the attitudes (6 for French or 12 for Japanese);
- the speakers (2 for each language);
- the modality (Audio, Visual or Audio-Visual);
- the modalities' presentation order (audio first, then visual and audio-visual; or visual first, then audio and audio-visual);

Subjects listened to each stimulus only once for each modality, presented in a random order. For each stimulus, they had to select the perceived attitude as well as its intensity on an open scale ranging from “*hardly perceptible*” to “*very marked*” (encoded on a 1-100 scale, with the 0 score used for the 5 not-selected attitudes). Subjects had to answer to the test on a PC without any time constraint.

For each language, two groups of subjects took the experiment. The first group listened to the audio-only stimulus first, and then watched the video-only stimulus, and finally the audio-video stimulus. The second group started with the video only stimulus, continued with the audio only stimulus and finally ended with the audio-video stimulus. This enabled us to counter-balance a possible effect of the presentation order of the modalities. During the presentation of one modality, the stimuli corresponding to all attitudes and to the two speakers were randomized in a different order for each listener.

Subjects

For the Japanese part, 46 Tokyo dialect speakers (mean age = 18,7) participated in this experiment. They were separated into 2 groups (Audio-only first and Visual-only first) with 28 subjects (4 males and 24 females) for the audio-only condition, and 18 subjects (7 males and 11 females) for the visual-only condition.

For the French part, 32 French listeners (17 males and 15 females, mean age = 32) took the experiment, 16 in each group (7 females and 9 males in the Audio-only-first group and 8 females and males in the Visual-only-first group).

Statistical Processing

Results given by listeners are expressed by two measures: a simple categorical choice (the perceived attitude), and a relative intensity score for the selected attitude. Two kinds of results are analyzed. (1) The recognition rate of each attitude, expressed either as the sum of the categorical

choice of the attitude by listeners (percentage of good recognitions), or as the mean intensity rating of good answers. And (2) the confusion matrices, grouping the categorical answers given by listeners for each of the presented attitudes – expressed either as categorical recognition rate received by each possible attitude, or as the relative intensity received by each possible attitude compared to the total intensity rating received by the stimuli. Only the categorical answer (and not intensity scores) will be detailed, as there is no major difference between the two types of measures.

For each language, recognition rates (either categorical or intensity) are analysed using a repeated-measure ANOVA, which takes as a dependant variable the recognition rate of each attitude (expressed in percentages), *subjects* as a random effect, one between-subject factor (the *group*, i.e. the order of presentation of the Audio and Video modality for each group of subjects), and three within-subject factors (the 6 or 12 *Attitudes*, the 2 *Speakers* and the 3 *Modalities*). ANOVA analysis mainly aims at measuring the relative importance of the different factors of the listener's behaviour.

Confusion matrices are analysed using a correspondence analysis and a cluster analysis (Benzecri, 1973). Both of these methods are based on data-reduction techniques that allow a more simple and comprehensive representation of the data. A correspondence analysis allows a graphical representation of the perception results in order to analyse the self-recognition of a particular attitude and its relative dispersion. A cluster analysis hierarchically groups the different stimuli in clusters. The distances between the clusters indicate the perceptive distances between the corresponding attitudes (the Ward distance metric is used for clustering), and thus allow distance judgements.

Anova Results

Results of the ANOVAs for Japanese and French are summarized in Table 1. For both languages the effect of the controlled factors are almost identical: all factors show a strong effect on the results, except the *group* factor (corresponding to the different order of presentation of modality

for each group), which is never significant at $p<0.1$. Interactions between factors (except the group) are all significant, except the interaction between *speaker* and *modality*, for French only.

Insert Table 1 about here

The *attitude* and *modality* factors and their interaction are significant, for both languages. Results of the interaction are presented in figure 1. In their multimodal presentation, all attitudes receive recognition scores higher than chance level, indicating an overall good recognition of the performances even if for some speakers, in some modalities, and for some attitudes, scores may be low.

Amongst the five attitudinal labels represented in both languages (i.e. *declaration*, *doubt*, *obviousness*, *interrogation* and *surprise*), all except *declaration* show a similar influence of the modality on perception results for Japanese and French. Best performances for *doubt* are achieved with visual information; *obviousness* and *surprise*, mainly with audio-visual; and *interrogation* mainly with acoustic information. This is a first indication of the coherence of the multimodal data recorded in this corpus. The relative contribution of modalities to the different attitudes will be studied later, together with the analysis of these modalities.

Insert Figure 1 about here

Another factor having a main influence on the perception results is the speaker's performance. For Japanese, the naïve speaker (SJ2) received lower overall recognition scores than the trained one. The strategy of SJ2 regarding the use of each modality is significantly different from the one of SJ1. For French, there is also one speaker (SF1) who systematically received better recognitions scores than the other speaker (SF2), even though their strategy towards the two modalities seemed coherent (non significant interaction between the *speaker* and the *modality* for French). Therefore, for the objective analysis, only the data from SJ1 and SF1 will be discussed, in comparison with their perception scores. The reason for this is because we focus on evaluating the possible influences of the visual parameter on prosodic attitude, and not speaker differences. Even if our experiment may show that speaker performance is critical, our data is not sufficient for extracting relevant information concerning this point.

Analysis of Confusion Matrices

Analysis of the confusion matrices leads to interesting parallels between attitudinal expressions: data reduction techniques allow grouping the attitudinal expression inside a more general cluster, thus enhancing the main perceptive proximities. The results of the clustering analysis are presented here, for both Japanese and French, and for each modality (cf. fig. 2).

For the Japanese speaker, SJ1, the recognition of each attitudinal expression is over the chance level in each modality, with few exceptions (especially the audio-only *admiration*). *Admiration* is perceived as *arrogance* in the audio-only condition, but visual information allows a clear distinction, even if they stay at a global level in the same cluster. Three general clusters emerge from the audio-only and the audio-video analyses. The first cluster groups the politeness expressions plus declaration; whereas the remarkably well-recognized video-only *kyoshuku* is outside this cluster, and the video-only *obviousness* is mixed up with *sincerity-politeness*. The second cluster groups the expression of “query” (*interrogation*, *doubt*, *surprise*), with video-only *surprise*, also recognized without confusion. The third cluster groups *admiration* with attitudes

described by Shochi et al. (2008) as “*express[ing] the imposition of the speaker’s opinion*” (*arrogance, authority, obviousness, irritation*).

For the French speaker SF1, each attitude is recognized over chance, whatever the modality. The main confusion arises for the audio- and video-only expressions of *doubt*, mixed with *surprise* for the audio-only modality, and with *interrogation* for the video-only modality. At a global level, two main clusters appear, plus the *suspicious irony* expression, always well distinguished from the other attitudes. The first cluster groups the declarative expression of *obviousness* and *declaration*, whereas the second cluster groups the interrogative attitudes of *surprise*, *doubt* and *interrogation*. This configuration remains stable for each modality, except for the visual-only modality, where *declaration* is perfectly recognized but *obviousness* and *suspicious irony* show some confusions.

=====

Insert Figure 2 about here

=====

OBJECTIVE PARAMETERS AND ANALYSIS

Objective Analysis

Extraction of acoustic parameters. Acoustic parameters of prosody were extracted automatically from the hand-labelled signals, with Matlab scripts using the yin algorithm of fundamental frequency (F0) extraction (cf. de Cheveigné & Kawahara, 2002). Three parameters are extracted: F0 (in semitones), duration (in seconds – either syllabic for French or moraic for Japanese), and intensity (in dB). For both F0 and intensity parameters, three values for each vowel were calculated.

For each parameter, we calculated the maximum, the minimum, the mean, and the range, and also the slope of the curve (i.e., the mean value of the first vowel minus the mean value of the

last one). These five values for each of the three acoustic parameters are recorded for each attitudinal expression, and used as entries for the principal component analysis.

Due to the complexity and as yet not well-known approach to analyzing voice quality (e.g., (d’Alessandro, 2006), the study of voice quality cues is reserved for future work. Moreover, although it is clear that voice quality has a significant impact on perception of Japanese attitudes (Shochi, 2008), it is probably not that useful for French attitudes (Morlec et al., 2001).

Extraction of visual Action Units (AU). A researcher with some knowledge about the Facial Action Coding System (FACS) (not a certified FACS coder) from the research team viewed each video and marked appearance of Action Units related to the upper face, lower face and head positions based on appearance changes (intensity was not scored) according to the FACS Manual (Ekman et al., 2002). The list of AU used for this analysis, and their labels, are presented in table 2.

=====

Insert Table 2 about here

=====

Principal Component Analysis (PCA) on audio and video parameters

Using the objective parameters described above as features, separate Principal Components Analyses were run. Results (cf. figure 3 & 4) present the relative ability of these features to separate each set of attitudes. We compare the results of the PCAs and the results of the clustering analysis (based on perception results).

Japanese attitudes. PCAs based on F0 parameters show the following distinctive features. Both a high range of F0 and an important maximum F0 value characterize *doubt* (with the highest range) and *surprise* expressions. Expressions with a lower F0 values are *authority*, *sincerity* and *kyoshuku*.

The general slope of the curve relates to *obviousness* and to a lesser degree to *declaration*. The F0 mean does not have any particular distinctive power.

=====

Insert Figure 3 about here

=====

Moraic duration patterns mainly separate *admiration* from the others, with the most important duration range, due to an impressive lengthening of the last syllable, also resulting in the highest maximum of duration (and therefore the most important negative slope). A slightly positive duration slope corresponds systematically to sentences pronounced with the last mora based on a simple [s], without added [u], i.e. to *arrogance, authority, declaration, sincerity*, and *kyoshuku*.

The intensity parameter mainly separates the most intense expressions, with important intensity mean and maximum (mainly *authority*, and also *obviousness* and *surprise*) from the others. It also separates expressions with a positive slope (i.e. lowest voice at the end): *arrogance, declaration* and *obviousness*; from the expressions with a rising voice strength: *kyoshuku, interrogation, surprise*, and *doubt*.

Finally, the visual AUs clearly distinguish *irritation* and *kyoshuku* as specific expressions. Then, it groups the questioning attitudes (*interrogation, surprise* and *doubt*); makes a clear cluster composed of *politeness, admiration, sincerity* and *arrogance* around AU1. *Declaration* and *authority* do not feature any particular facial AU. The somewhat restricted amount of information obtained from the AUs' labelling does not allow a sufficiently clear distinction of Japanese attitudes, even though listeners showed good results. Future work along these lines may require a more detailed analysis of visual information. .

French attitudes. Analysis of the F0 parameters underlines the main influence of F0 in the attitudinal expression of French attitudes, already shown by Morlec et al. (2001). *Suspicious irony* is apart from the other attitudes, with the highest F0 range, an important F0 maximum and a decreasing F0 (positive slope). Declaration is also separated from the other attitudes, with a positive slope and a generally low F0 curve. Interrogative expressions (*surprise, doubt* and *question*) are grouped together, with a negative slope – i.e. a rising final F0.

Syllabic duration parameters also separate *suspicious irony* from other, with the highest duration range and the longest syllable (at the end). Other expressions generally show small changes in their durational patterns, except for *surprise* with a positive slope.

Intensity parameters distinguish two expressions: *suspicious irony* and *surprise*. Both attitudes have a high intensity (mean and maximum), but *suspicious irony* has a large range of intensity, whereas intensity for *surprise* is more constant.

Four kinds of facial expressions are distinguished: 1) the *doubt* and *question* expressions, 2) the *surprise* and *obviousness* expressions, that are both expressed with AU1+2 and AU57, 3) *suspicious irony*, and 4) finally *declaration*, that does not display any facial AU.

=====

Insert Figure 4 about here

=====

Analysis & Comparison with Subjective Results

By comparing the distribution of attitudinal expressions made on the basis of objective and subjective results, it is obvious that results are more coherent for French than for Japanese. That can be explained by the two main drawbacks of our analysis: the absence of voice quality measurement for Japanese, and the fact that our AU labeller may have been less acute on Japanese stimuli, as he

is not a speaker of Japanese, and may have been “blind” to some pertinent cues. An automatic detection of the visual parameters which we consider for future research may be more suited to such a study.

For French, F0, duration and AU, as well as perception results, clearly distinguish *suspicious irony* from others, and opposes interrogative expression and declarative ones. The main difference between objective and subjective results concerns the facial expressions of *obviousness* and *surprise*, performed with two identical AU, amongst others. One of the main contributions of facial information is the very clear distinction allowed by these cues of the expressions of *doubt*, *obviousness* and *suspicious irony*. In preceding works (Morlec et al., 2001), both *doubt* and *suspicious irony* show some confusion (in audio), while for our speaker, *doubt* and *obviousness* are confused in the audio-only condition. It seems then obvious that visual information brings a clear distinction between the more expressive attitudes, while they are less relevant for the two modalities (non-expressive attitudes) of *declaration* and *interrogation*.

The analysis of the Japanese data gives some hints about the main characteristics of attitudinal prosody, but the description of prosodic parameters seems too gross to catch the complexity of all these expressions; specifically it suffers from the size of the corpus, which is too limited to extract statistically relevant data. Two essential measurements may need to be done: a measure of the voice quality dimension, of primary importance for the perception of some Japanese attitudes like *kyoshuku* (cf. Shochi, 2008); and a description of intonation in terms of contours, not only using F0 values averaged on the utterances. However, the results obtained in the audio-only and with the audio-visual condition are close to those obtained by Shochi et al (2006) in an audio-only condition. Furthermore, a set of AU has been identified for each attitudinal expression. These AUs allow a straightforward identification of the *kyoshuku* expression, and this finding is important in the perspective of a cross-cultural perception of this culturally built expression – absolutely not recognized in an audio-only condition by occidental listeners (Shochi, 2008). Other information

carried by the visual modality allows also a clear distinction between assertive vs. dubitative expressions.

There are differences between Japanese and French, but also similarities. In both languages, *declaration* doesn't use a specific AU but is recognized in the video-only modality. This is a bridge to Danes (1994) assertion, that “*We must rather assume that any utterance or higher discourse unit has an emotional value in its communicative situation, both on the producer's and the receiver's side (though not necessarily for both of them in a given case, or in the same measure). Thus, even the alleged 'absence of emotional involvement' represents, in fact, an instance of the category of emotional state, and an utterance of this nature may carry - in certain contexts - a high emotional value.*” (p. 258). Evidence in both languages is difficult to characterize, and seems to heavily rely on the speaker's performance, whereas *surprise* expressions are performed with a rising F0 at the end of the sentence, and use the AU1+2, and *surprise* is one of the few attitudes cross-culturally recognized by Japanese, American and French in Shochi (2008). The dubitative expressions, *interrogation*, *surprise* and *doubt* seem to behave similarly in Japanese and French: in the audio-only condition, *surprise* and *doubt* are close together, with relative proximity to *interrogation*, while with the visual information (visual-only and audio-visual conditions), *interrogation* and *doubt* show a higher proximity and *surprise* receives better discrimination scores, even if it is generally still in the same cluster. The results of the audio-only condition recalls very interestingly those obtained by Shochi (2008), who obtained similar confusions between *doubt* and *surprise*, with a proximity to *interrogation* on Japanese, French and British English, with native listeners as well as in cross-cultural conditions with Japanese, French and American English listeners (with some minor differences for the Japanese listeners). This strong perceptual effect obtained on audio-only cues seems to be counterbalanced by the visual cues, which discriminate mainly the *surprise* expression. Finally, in both languages, there is a systematic opposition between the assertive and the dubitative expression, that can be enlightened by the classification of speech acts expressivity proposed by

Brandt (2008), from a linguistic point of view. The Japanese language, moreover, adds a dimension of dominance, or of imposition of the speaker, orthogonal to the assertive-dubitative dimension. The French attitude of *suspicious irony* may be seen as one expression of this dimension.

CONCLUSIONS

Production and perception cues to Japanese and French attitudinal expressions display interesting similarities. First, the two modalities are widely used by all speakers, and are decoded adequately by listeners: almost all attitudes are recognized over the chance level in each single modality, and audio-visual performances generally outperform those obtained with individual modalities. The different expressivity and strategy of the speakers also have a strong impact on the subjects' performances. Along these lines, future work would benefit from preparation of more naturalistic situations of interaction, and from more speakers in order to assess a wider range of possible strategies. This would allow us to determine some main coherences, which were not easily seen from the limited set of speakers we had in the current study.

Approaches of attitudinal expressions based on a collection of discrete categories may also raise some issues. It seems promising to restrict such a study to the analysis of one particular dimension that comes out of the results described in this paper (such as the politeness – impoliteness dimension for Japanese in comparison with politeness in other languages – Loveday, 1981, such as French). Such a dimension may be more easily tested amongst languages (cf. Sagisaka et al., 2004), and its acoustic and visual correlates could also be extracted. We also intend to investigate the impact of speakers and listeners' personality traits and emotional intelligence on the audiovisual expression and perception of these attitudes as a means of understanding the individual differences that we observe, since some research has shown that personality traits of listeners might affect perception of attitudes (e.g., Cooper, 2002; Matthews et al. 2002).

Objective analysis based on static cues – either acoustic or visual ones – seems not to be sufficient to understand the performances of listeners: some of the main divergences can be explained, but not all of them. It seems of primary importance to add dynamic and global information to such an objective analysis (Aubergé et al., 2004) using an automatic labelling tool. For example, rapid eye movement might have a dramatic influence on the subjects' ratings, but are not taken into account in the current analysis. In a similar way, the F0 contour of the sentence was of importance for the expression of French attitudes (Morlec et al., 2001), but is unsatisfactorily described by the mean F0 values, used as descriptors of intonation in this study. The main problem for using such information is the difficulty in our current data reduction technique to deal with contour shapes, not just scales. Future work is required in order to improve the objective analysis, e.g. by adding an intensity rating to the AUs, and analysing the dynamics of the audiovisual signs.

Once coherent sets of acoustic and visual cues will be at our disposal, we plan to create controlled animated audio-visual replays of the attitudinal expressions, using facial animation (Martin et al., 2006; Buisine et al., 2006) and copy-synthesis approaches, in order to have a better control of the perceptual relevance of the different cues. Virtual characters, the expressivity of which would be controlled with hand-driven interfaces (d'Alessandro et al., 2007; Martin et al., 2007) may allow experimenters to design Wizard-of-Oz scenarios (Dahlbäck et al. 1993) of man-machine interaction. In this situation, the time-course of the perceptually validated cues, relative to other information such as the syntactic structure of the utterance or the occurrence of an emotionally loaded event, may be manipulated in order to evaluate if the dynamics of both audio and visual cues can induce the perception of the difference between spontaneous vs. controlled expressive states. Such studies on social affects provide additional knowledge about the audiovisual expression and perception of affects.

ACKNOWLEDGEMENTS

We are deeply grateful to T. Sadanobu from Kobe University for his helpful advice and contribution on Japanese attitudes, and to J. Beskow and an anonymous reviewer for their helpful comments and advice on an earlier version of this paper. We also thank K. Yoshimoto for his very precious performance, and students at Kanagawa University and people at LIMSI and GIPSA-lab for their participation in the perception tests. This work was supported in part by the CNRS-SHS ATIP grant "La prosodie des affects : des attitudes aux émotions" to the first author, in part by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research, (2007-2010):19520371 to the fourth author, and also by SCOPE (071705001) of Ministry of Internal Affairs & Communications (MIC), Japan.

REFERENCES

- d'ALESSANDRO, C. (2006). Voice Source Parameters and Prosodic Analysis. In Sudhoff, S., Letemova, D., Meyer, R., Pappert, S., Augurzy, P., Mleinek, I., Richter, N., Schlieser, J. (Eds). *Methods in Empirical Prosody Research*, Berlin-New York: Walter de Gruyter, 63-87
- d'ALESSANDRO, C., RILLIARD, A., & Le BEUX, S. (2007). Computerized chironomy: evaluation of hand-controlled Intonation reiteration. *Proceedings of Interspeech 2007*, Antwerpen, Belgium, 1270-1273.
- ALLERTON, D.J., & CRUTTENDEN, A. (1978). Syntactic, illocutionary, thematic and attitudinal factors in the intonation of adverbials. *Journal of Pragmatics*, **2**, 155-188.
- AUBERGÉ, V. (2002). A Gestalt Morphology of Prosody Directed by Functions: the Example of a Step by Step Model Developed at ICP. *Proceedings of Speech Prosody 2002*, Aix-en Provence, France, 151-154.
- AUBERGÉ, V., & CATHIARD, M. (2003). Can we hear the prosody of smile? *Speech Communication*, **40** (1), 87-97.

- AUBERGE, V., AUDIBERT, N., & RILLIARD, A. (2004). Acoustic morphology of expressive speech: what about contours? *Proceedings of Speech Prosody 2004*, Nara, Japan, 201-204.
- AUDIBERT, N., AUBERGE, V. & RILLIARD, A. (2008). How we are not equally competent for discriminating acted from spontaneous expressive speech. *Proceedings of Speech Prosody 2008*, Campinas, Brasil, 693-696.
- BANSE, R. & SCHERER, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, **70**, 614-636.
- BÄNZIGER, T., & SCHERER, K.R. (2005). The role of intonation in emotional expressions. *Speech Communication*, **46**, 252-267.
- BARKHUYSEN, P., KRAHMER, E., & SWERTS, M. (2006). How Auditory and Visual Prosody is Used in End-of-Utterance Detection. *Interspeech*, Pittsburgh, USA.
- BARKHUYSEN, P., KRAHMER, E., & SWERTS, M. (2007). Incremental perception of acted and real emotional speech. *Interspeech*, Antwerp, Belgium, 1262-1265.
- BARKHUYSEN, P., KRAHMER, E., & SWERTS, M. (2007). Cross-modal perception of emotional speech. *Proceedings of 16th International Congress of Phonetic Sciences*, Saarbruecken, Germany, 2133-2136.
- BENZECRI, JP. (1973). *L'analyse des données*. Paris: Bordas.
- BRANDT, P.A. (2008). Thinking and language. A view from cognitive semio-linguistics. *Proceedings of Speech Prosody 2008*, Campinas, Brasil, 649-654.
- BUISINE, S., ABRILIAN, S., NIEWIADOMSKI, R., MARTIN, J.-C., DEVILLERS, L., & PELACHAUD, C. (2006). Perception of Blended Emotions: from Video Corpus to Expressive Agent. *6th International Conference on Intelligent Virtual Agents*. Marina del Rey, USA, Springer: 93-106.
- CALLAMAND, M. (1973). *L'intonation expressive*. Collection le français dans le monde, B.E.L.C. Librairies Hachette et Larousse.

- CALBRIS, G., & MONTREDON, J. (1981). *Oh là là. Expression intonative et mimique*. Paris: CLE international.
- CALBRIS, G., & MONTREDON, J. (1986). *Des gestes et des mots pour le dire*. Paris: CLE international.
- CALBRIS, G., & PORCHER, L. (1989). *Geste et communication*. Paris : Hatier-Crédif.
- 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), 111-120.
- de CHEVEIGNE, A., & KAWAHARA, H. (2002). “YIN, a fundamental frequency estimator for speech and music”. *Journal of the Acoustical Society of America*, **111**, 1917-1930.
- COOPER, C. (2002). *Individual Differences*. (2nd edition) London: Arnold.
- DAHLBÄCK, N., JÖNSSON, A., & AHRENBERG, L. (1993). Wizard of Oz studies - Why and how. *Workshop on Intelligent User Interfaces*, Orlando, FL.
- DAMASIO, A.R. (1994). *Descartes' error. Emotion, reason, and the human brain*. New-York: G.P. Putnam.
- DANES, F. (1994). Involvement with language and in language. *Journal of Pragmatics*, **22**, 251-164.
- EKMAN, P. (1999). Facial Expressions. In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion* (pp. 301-320). Sussex, U.K.: John Wiley & Sons, Ltd.
- EKMAN, P., FRIESEN, W. C., & HAGER, J. C. (2002). *Facial Action Coding System*. The Manual on CD ROM, Research Nexus division of Network Information Research Corporation.
- ERICKSON, D., OHASHI, S., MAKITA, S., KAJIMOTO, N., & MOKHTARI, P. (2003). Perception of naturally-spoken expressive speech by American English and Japanese listeners. *CREST International Workshop on Expressive Speech Processing*, 31-36.
- FÓNAGY, I., BÉRARD, E., & FÓNAGY, J. (1984). Clichés mélodiques. *Folia Linguistica*, **17**, 153-185.
- FÓNAGY, I. (2003). Des fonctions de l'intonation : essai de synthèse. *Flambeau*, **29**, 1-20.
- GRANSTRÖM, B., & HOUSE, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, **46**, 473-484.
- GRICHKOVTSOVA, I, LACHERET, A., MOREL, M., BEAUCOUSIN, V., TZOURIO-MAZOYER, N. (2007). Affective speech gating. *Proceedings of the International Congress of Phonetic Sciences*, Saarbrücken, Germany, 805-808.
- LAVER, J. (1980). *The phonetic description of voice quality*. Cambridge University Press, New York.
- LOVEDAY, L. (1981). Pitch, politeness and sexual role : An exploratory investigation into the pitch correlates to English and Japanese politeness formulae. *Language and Speech*, **24** (1), 71-89.
- MAEKAWA, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, 635-638.
- MARTIN, J.-C., NIEWIADOMSKI, R., DEVILLERS, L., BUISINE, S., & PELACHAUD, C. (2006). Multimodal Complex Emotions: Gesture Expressivity And Blended Facial Expressions. *Special issue of the Journal of Humanoid Robotics on "Achieving Human-like Qualities in Interactive Virtual and Physical Humanoids"*. Eds: C. Pelachaud, L. Canamero. **3**(3): 269-291.
- MARTIN, J.-C., d'ALESSANDRO, C., JACQUEMIN, C., KATZ, B., MAX, A., POINTAL, L. & RILLIARD, A. (2007). 3D Audiovisual Rendering and Real-Time Interactive Control of Expressivity in a Talking Head. *7th International Conference on Intelligent Virtual Agents (IVA'2007)*. Paris, France September 17-19.

- MATTHEWS, G., ZEIDNER, M. & ROBERTS, R. D. (2002). *Emotional Intelligence - Science and Myth*. Cambridge, MA: MIT Press.
- MIZUTANI, O., & MIZUTANI N. (1979). *Aural Comprehension Practice in Japanese*. The Japan Times.
- de MORAES, J.A. (2008). The Pitch Accents in Brazilian Portuguese: analysis by synthesis. *Proceedings of Speech Prosody 2008*, Campinas, Brasil, 389-397.
- MORLEC, Y., BAILLY, G., & AUBERGÉ, V. (2001). Generating prosodic attitudes in French: Data, model and evaluation. *Speech Communication*, **33** (4), 357-371.
- MOZZICONACCI, S. J. L. (1998). *Speech variability and emotion: Production and perception*. Ph.D. thesis, Eindhoven, the Netherlands.
- OHALA, J.J. (1996). Ethological theory and the expression of emotion in the voice. *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, 1812-1815.
- PAULMANN, S., SCHMIDT, P., PELL, M. & KOTZ, S. (2008). Rapid processing of emotional and voice information as evidenced by ERPs. *Proceedings of Speech Prosody 2008*, Campinas, Brasil, 205-209.
- PAVLENKO A. (2005). *Emotions and Multilingualism*. Cambridge, UK: Cambridge University Press.
- RILLIARD, A., MARTIN, J.C., AUBERGE, V., SHOCHI, T. (2008). Perception of French Audio-Visual Prosodic attitudes. *Speech prosody 2008*, Campinas, Brasil, 685-688.
- ROSSI, M., DI CRISTO, A., HIRST, D., MARTIN, P., NISHINUMA, Y. (1981). *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.
- SADANOBU, T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, **8** (1), 29-44.
- SAGISAKA, Y., YAMASHITA, T., & KOKENAWA, Y. (2004). Speech synthesis with attitude. *Proceedings of Speech Prosody 2004*, Nara, Japan, 401-404.
- SCHERER, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**(1-2), 227-256.
- SCHERER, K.R., BANSE, R., & WALLBOTT, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, **32**(1), 76-92.
- SCHERER, K.R., & ELLGRING, H. (2007). Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns? *Emotion*, **7**(1), 158-171.
- SHIGENO, S. (1998). Cultural similarities and differences in the recognition of audiovisual speech stimuli. *Proceedings of 5th International Conference on Spoken Language Processing*, Sydney, 149-152.
- SHOCHI, T. (2008). *Prosodie des affects socioculturels en japonais, français et anglais : à la recherche des vrais et faux-amis pour le parcours de l'apprenant*. PhD thesis, University Grenoble 3, France.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2006). How prosodic attitudes can be false friends: Japanese vs. French social affects. *Proceedings of Speech Prosody 2006*, Dresden, 692-696.
- SHOCHI, T., AUBERGÉ, V., & RILLIARD, A. (2007). Cross-listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes. *Proceedings of 16th International Congress of Phonetic Sciences*, Saarbruecken, Germany, 2097-2100.
- SHOCHI, T., ERICKSON, D., RILLIARD, A., AUBERGÉ, V. & MARTIN, J.C. (2008). Recognition of Japanese attitudes in Audio-Visual speech. *Speech prosody 2008*, Campinas, Brasil, 689-692.
- SWERTS, M., & KRAHMER, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, **53**(1), 81-94.

van HEUVEN, V.J., HAAN, J., JANSE, E., & van der TORRE, E.J. (1997). Perceptual identification of sentence type and the time distribution of prosodic interrogativity marker in Dutch. *ETRW Workshop on Prosody*, Athens, Greece, 317-320.

WILTING, J., KRAHMER, E., & SWERTS, M. (2006). Real vs. acted emotional speech. *Proceedings of the 6th Interspeech*, Pittsburgh, PA, USA.

Table 1: ANOVAs results for Japanese and French expressions, and for categorical and intensity results. Factors used are Grp for the listener group (either audio-only or video-only first); Spk for the 2 speakers; Mod for the 3 modalities; Att for 6 or 12 attitudinal expressions. Factors having a significant influence on results ($p<.01$) are marked with a star.

	df	French				Japanese			
		% Reco		Intensity		% Reco		Intensity	
		F	p	F	p	F	p	F	p
Grp	1	0.5	0.508	0.3	0.605	1	0.08	0.779	6.8 0.012
Spk	1	45.3	0.000 *	105.5	0.000 *	1	163.6	0.000 *	211.5 0.000 *
Grp:Spk	1	1.9	0.178	7.1	0.012	1	0.02	0.899	0.5 0.481
Mod	2	14.9	0.000 *	25.6	0.000 *	2	44.8	0.000 *	44.0 0.000 *
Grp:Mod	2	3.8	0.028	3.7	0.030	2	1.2	0.299	1.3 0.297
Att	5	6.4	0.000 *	11.8	0.000 *	11	23.5	0.000 *	22.5 0.000 *
Grp:Att	5	2.2	0.055	0.9	0.466	11	1.5	0.134	0.8 0.597
Spk:Mod	2	1.7	0.199	3.0	0.059	2	6.8	0.000 *	12.9 0.000 *
Grp:Spk:Mod	2	1.0	0.364	0.0	0.994	2	0.15	0.856	0.09 0.911
Spk:Att	5	3.8	0.003 *	7.5	0.000 *	11	11.0	0.000 *	17.9 0.000 *
Grp:Spk:Att	5	0.7	0.589	0.5	0.764	11	1.4	0.188	2.1 0.018
Mod:Att	10	7.9	0.000 *	8.6	0.000 *	22	3.7	0.000 *	5.0 0.000 *
Grp:Mod:Att	10	1.3	0.246	1.4	0.195	22	1.2	0.227	1.0 0.442
Spk:Mod:Att	10	5.4	0.000 *	3.6	0.000 *	22	7.3	0.000 *	8.6 0.000 *
Grp:Spk:Mod:Att	10	1.5	0.146	1.7	0.085	22	0.95	0.522	1.0 0.441

Figure 1: Percentage of recognition obtained by each attitude, for each modality, in Japanese (top) and French (bottom). The plain horizontal bar on each graph represents the chance level.

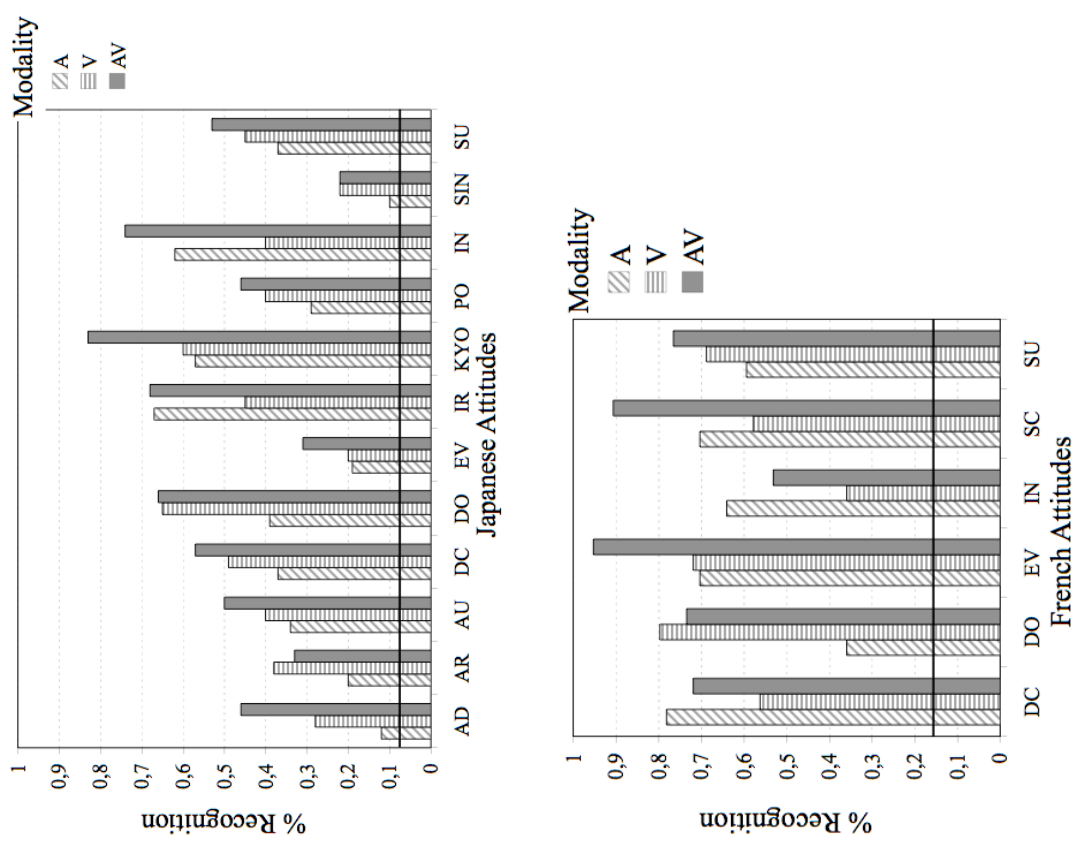
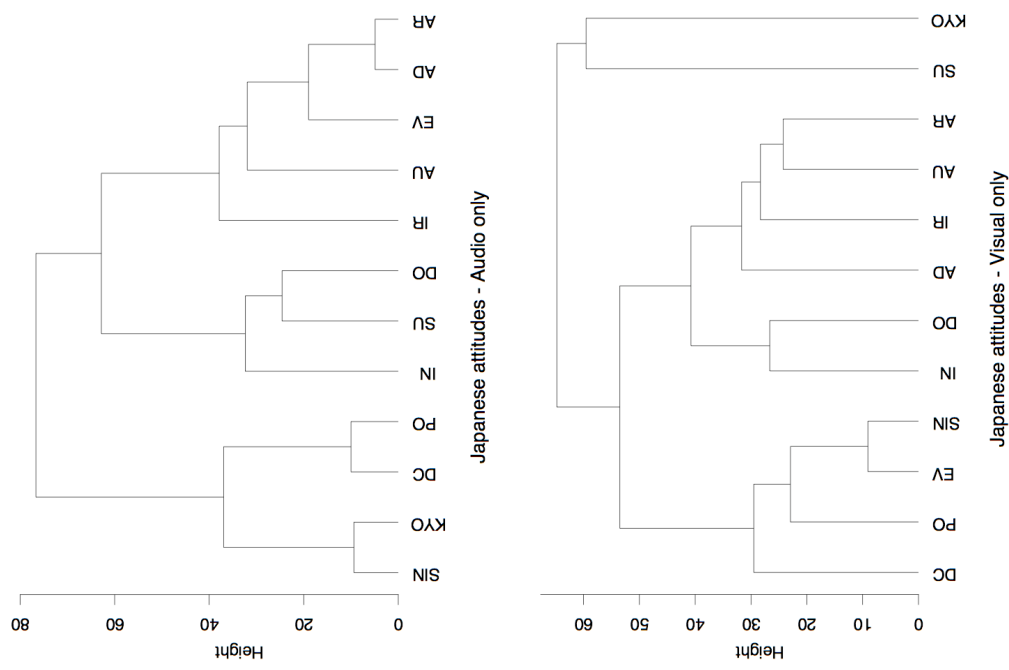


Figure 2: hierarchical clustering of attitudinal expressions obtained from recognition scores, for speakers SJ1 and SF1, and for each modality (audio, visual and audio-visual). The height indicates the relative distances between each attitudes or clusters of attitudes.



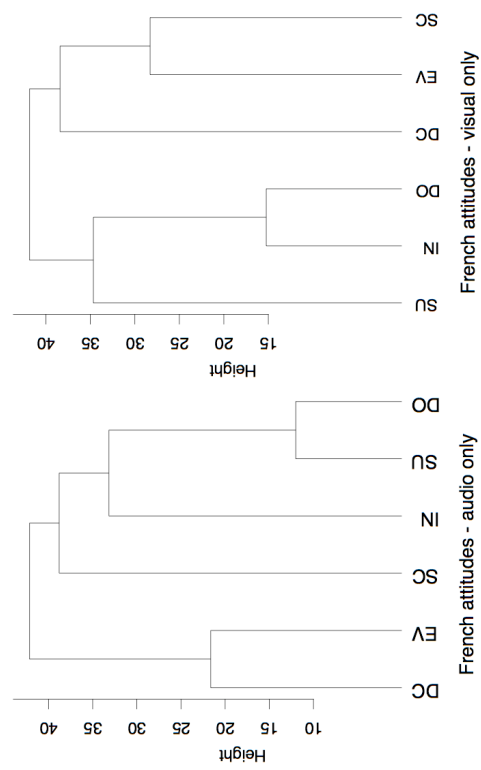
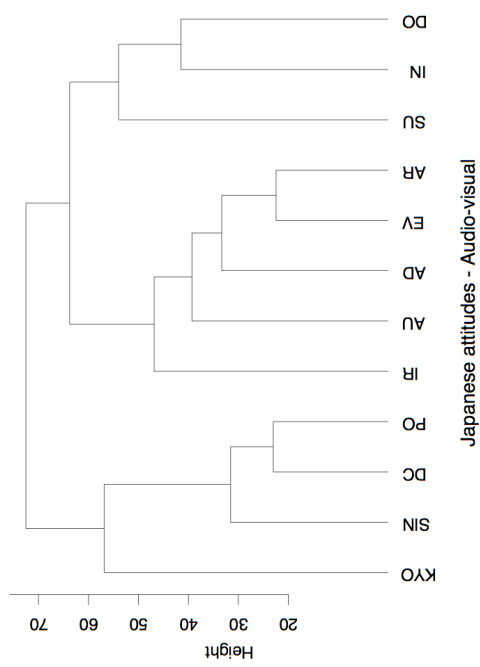
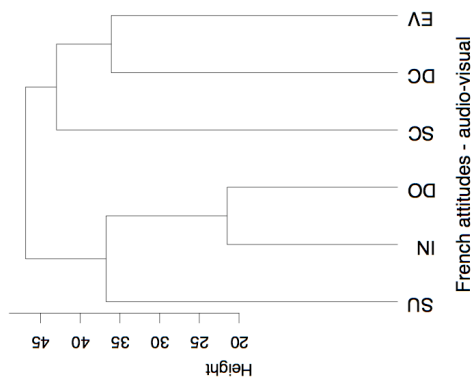
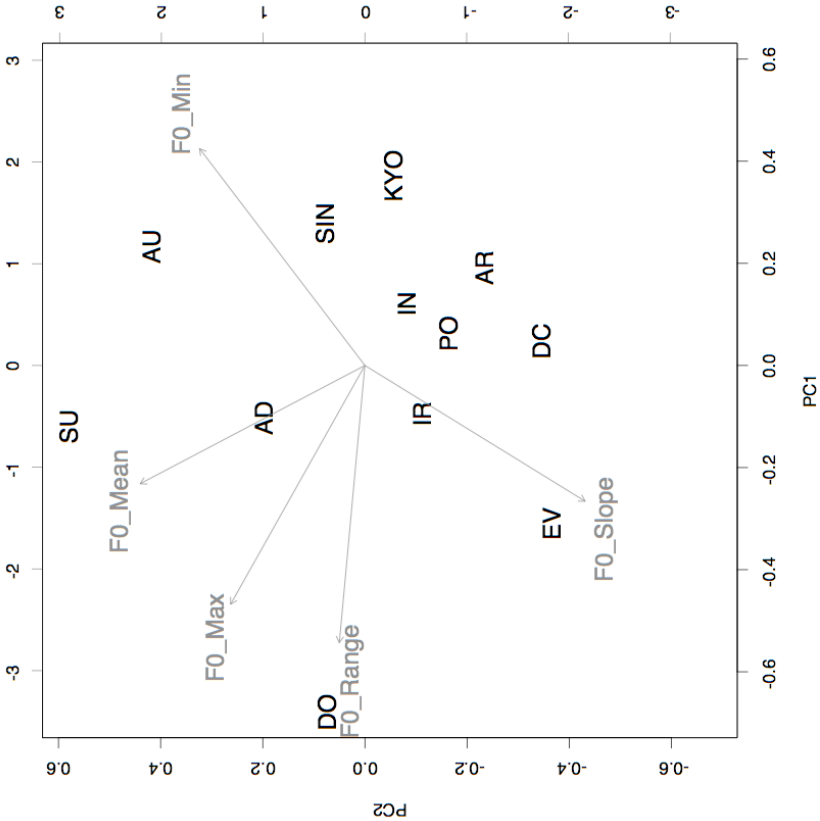
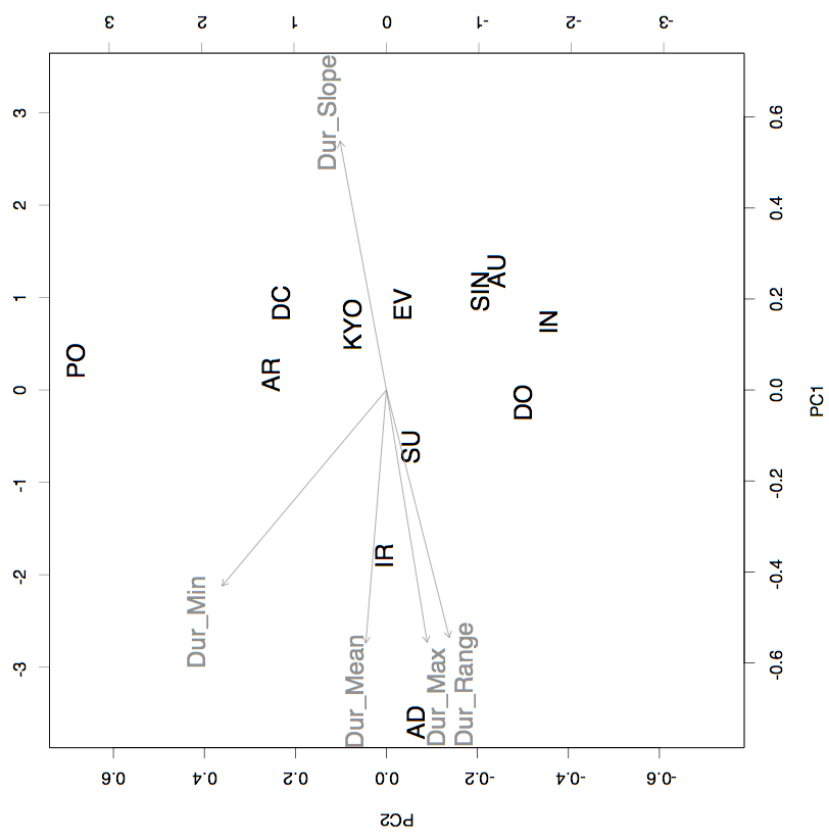
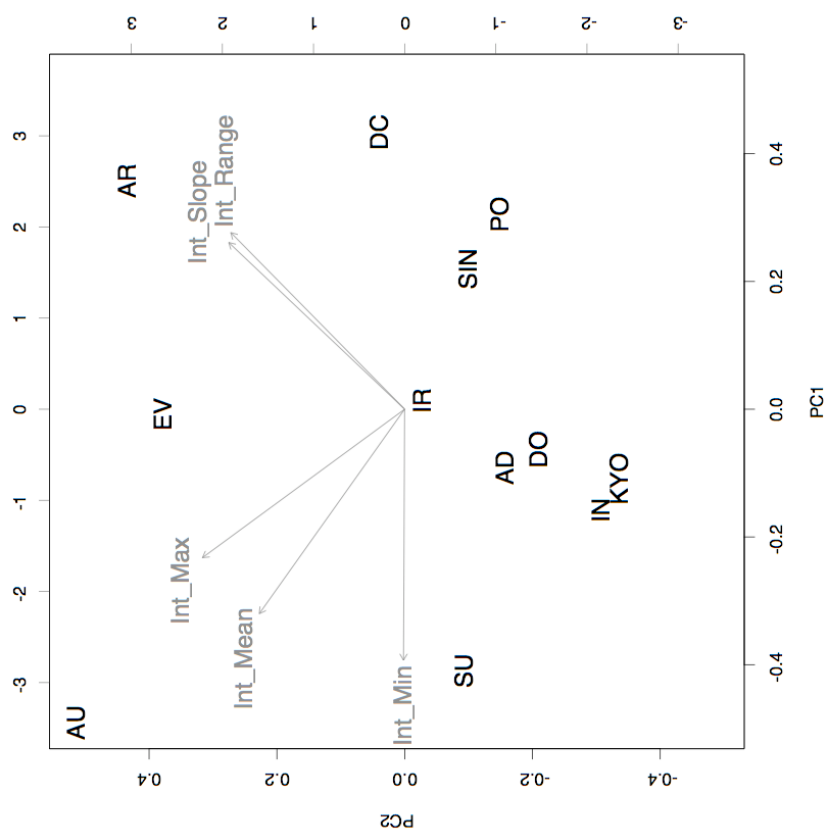


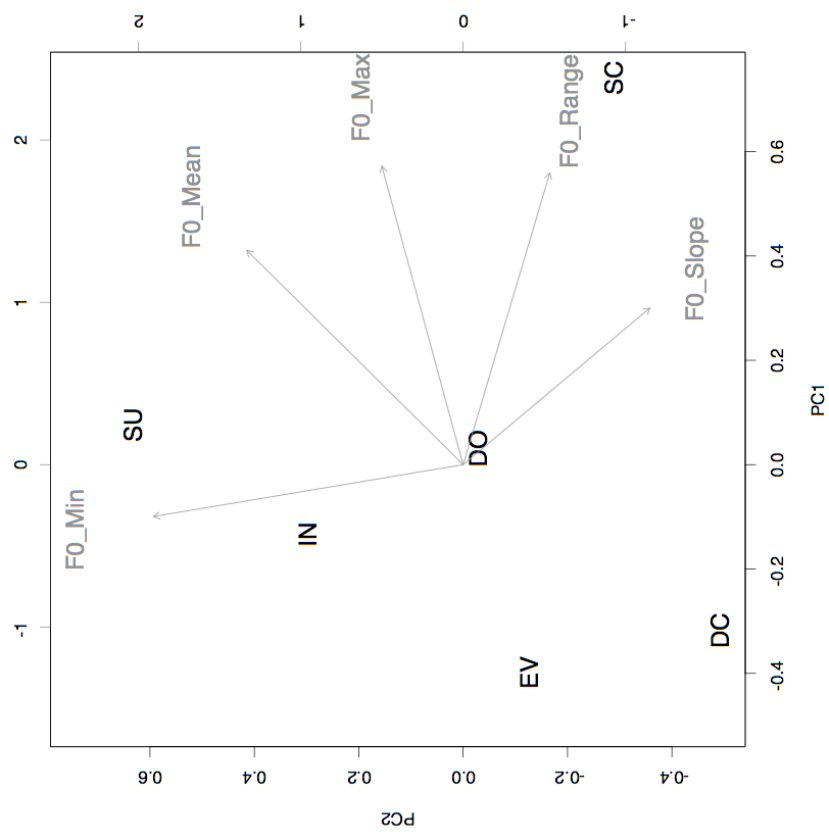
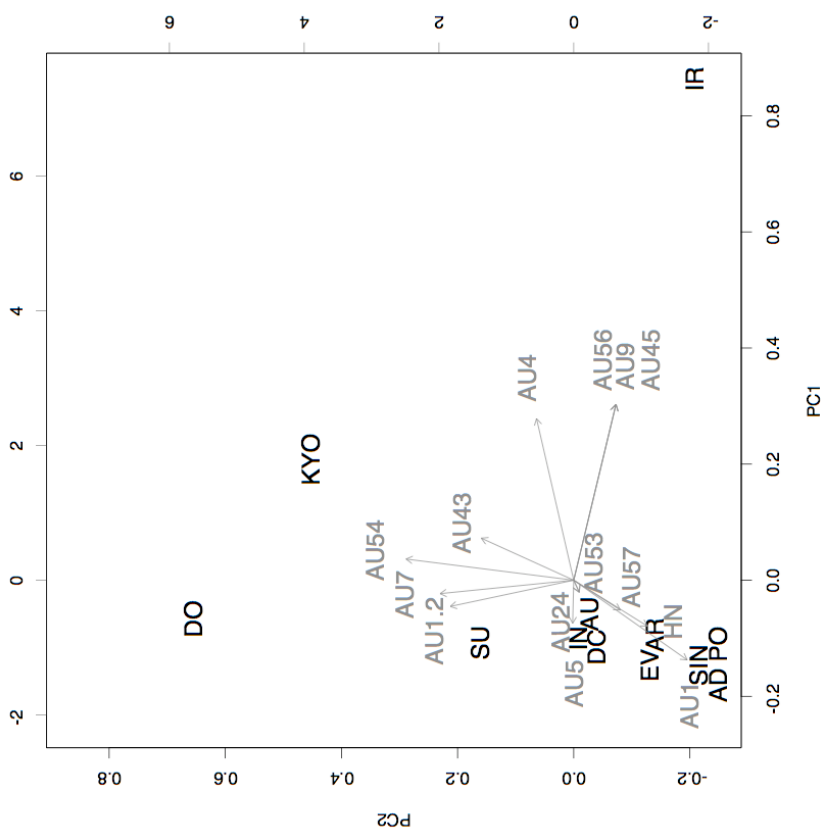
Table 2: list of visual Action Units used during the labelling – labels and description.

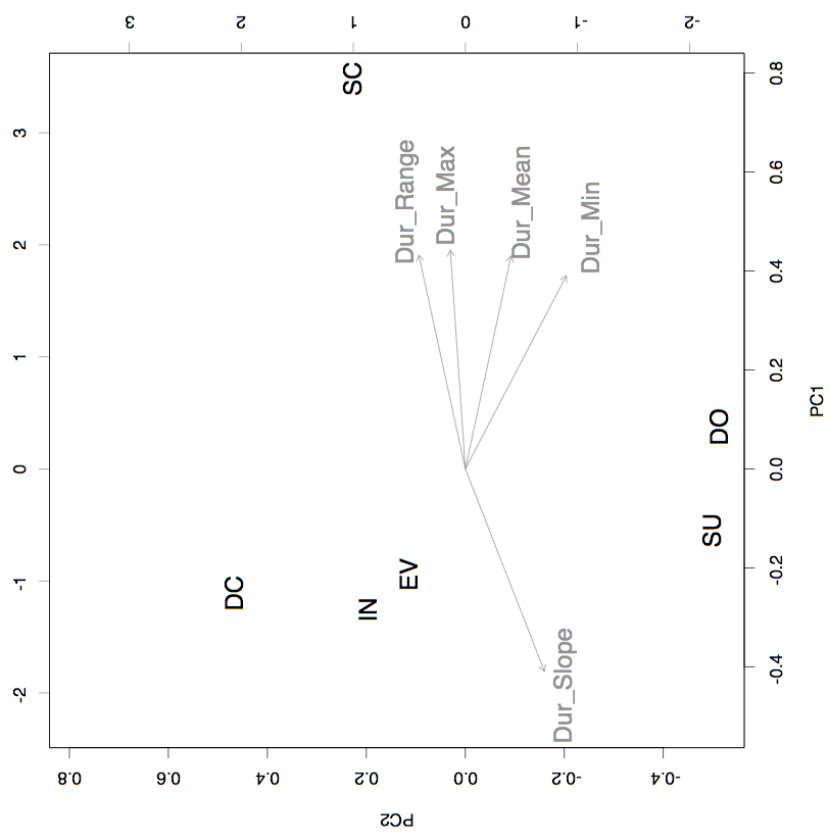
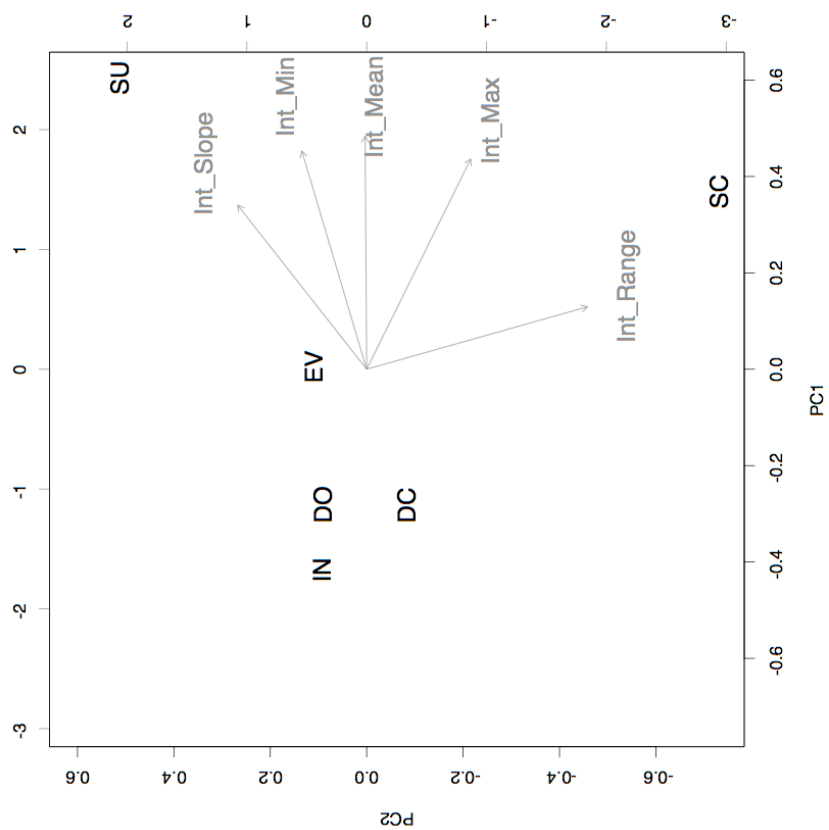
AU's label	Description of the AU
AU1+2	Inner + outer Brow Raiser
AU4	Brow Lowerer
AU5	Upper Lid Raiser
AU6	Cheek Raiser
AU11	Nasolabial Deepener
AU12	Lip Corner Puller
AU15	Lip Corner Depressor
AU26	Jaw Drop
AU43	Eyes Closed
AU51	Head turn left
AU57	Head forward
SH	Shoulder shrug
HN	Head nod

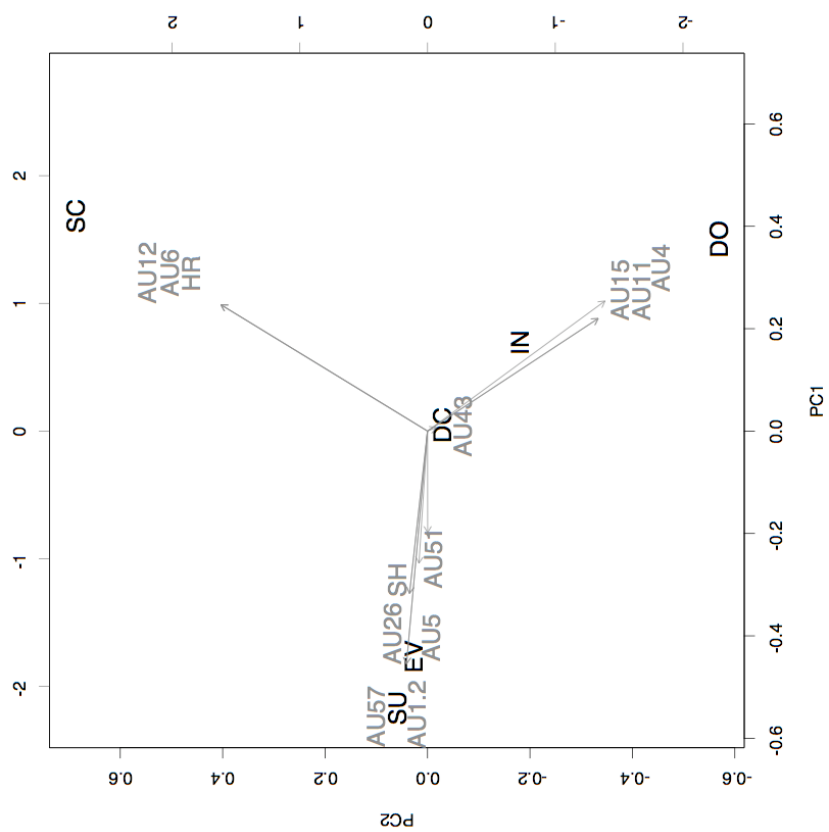
Figure 3: Two main dimensions of PCAs for speaker SJ1, showing the relative contribution of the following parameters to each attitude: F0, moraic duration, intensity (mean value, maximum value, minimum value and slope for each parameter), AUs (as listed in table 2).











Recognition of Japanese attitudes in Audio-Visual speech

Takaaki Shochi¹, Donna Erickson², Albert Rilliard³, Véronique Aubergé¹ & Jean-Claude Martin³

¹ GIPSA-Lab, Grenoble France, ² Showa Music Univ., Kawasaki City Japan, ³ LIMSI, Orsay France
¹{shochi;auberge}@gipsa-lab.inpg.fr, ²EricksonDonna2000@gmail.com, ³{rilliard;martin}@limsi.fr

Abstract

The aim of the present work is to investigate how Japanese listeners recognize 12 audio-visual prosodic attitudes of Japanese. Significant influences of the two speakers and three modalities were observed. Generally the audio-visual condition showed the best recognition score, and interesting behavior for audio and visual modality was observed. Attitudes were regrouped into 3 higher-level perceptual categories for the first speaker: polite expressions, the attitudes of “query”, and the expressions of imposition of one’s own opinion. The attitude of *kyoshuku* and *surprise* are particularly well recognized by visual information.

1. Introduction

Multimodal expression of affects seems to be controlled at different cognitive processing levels [11], from involuntary controlled expressions to the intentional, deliberate control of the speaker’s attitudes [4; 1]. Our position [1] is to distinguish attitudes vs. emotions as a different nature of control by the speaker: voluntary vs. involuntary, and not following the affective value carried by the expressions. In this view, to reproduce, sincerely or not, an emotion outside the body loop described by Damasio, is in the same control as to produce attitudinal affects that have been specifically built by the language and the culture, and then must be learned in development or by foreign speakers when these attitudes are not shared by the native and learned languages. Our proposal is that attitudes and emotions evolve in all the acoustic (and perhaps audio-visual) space but are identified by their anchoring in the speech timing, as a difference of the nature (speech vs. under speech control) – e.g. the affect value “surprise” may be either an attitudinal expression or an emotion, according to the nature of the control exerted by the speaker on his expression, and its acoustic encoding will be anchored or not in the linguistic organization of the speech. In this view, study on attitudes are important, because this is a part of the global meaning of speech acts [4]: even if speaker does not express any attitude by performing a simple declaration, it correspond to the attitude described by [1] as “the speaker decides not to give information on his attitudes”.

In the face-to-face interaction, attitudes are expressed within the multimodality of speech [3]. As attitudes are highly linked to language and culture, the study of attitudinal expression may benefit from a cross-cultural approach [12]. Like all language specifications, and because attitudes are constructed socially for and by the language, some attitudes can be expected to have universal values [4; 1] (e.g. authority or surprise), and possibly a universal prosodic morphology. Attitude values can also exist or not from one language to the other, and their realization in a specific language may not be recognized (or may be ambiguous in the learner’s language [10; 13]. This paper presents a study made in the continuation of a wider work on the expression of Japanese attitudes, in a

cross-cultural context [13]. It is a first study of the audio-visual expression of Japanese attitudes, in parallel to a study on audio-visual attitudes of French presented in a companion paper [8].

The multimodal nature of speech prosody has been shown for several functions [14, 2], and has a particular importance in affective communication [11]. Thus, this study tries to investigate the relative contribution of both visual and acoustic cues in the expression of prosodic attitudes, for both Japanese and French languages (this paper is specifically devoted to Japanese). As a first step, the two papers try to measure the nature of quantity of information retrieved by native listeners, thanks to each modality on a set of attitudes. A cross-cultural comparison may be envisaged.

After presenting the Japanese corpus and the set of attitudes studied here, the perception experiment and its experimental setting is described. Analyses of the recorded data thanks to various statistical tests are made, and leads to a set of conclusions for follow-up in future works.

2. Method

2.1. Selection of 12 Japanese attitudes

A set of 12 Japanese attitudes which were validated in [13] were used for this experiment. These attitudes were selected according to the literature [5], [6], [9] and Japanese language teaching methods [7]: doubt-incredulity (DO), evidence (EV), exclamation of surprise (SU), authority (AU), irritation (IR), arrogance (AR), sincerity-politeness (SIN), admiration (AD), *kyoshuku* (KYO), simple-politeness (PO), declaration (DC) and interrogation (IN) (see [13] for definitions). Some of these attitudes are specific or specifically important for the Japanese culture, especially those linked to the politeness strategy: simple-politeness, sincerity-politeness and *kyoshuku* vs. arrogance. The sincerity-politeness attitude appears when a socially inferior speaker is talking to someone superior to him in the Japanese society: the speaker expresses a serious and sincere intention by using this prosodic attitude. The *kyoshuku* attitude (there is no lexical entry to translate this in English) is a typically Japanese cultural attitude. Even if such situations occur in all cultures, the Japanese language has chosen to encode this situation as a prosodic attitude (“attitudineme”). A speaker uses *kyoshuku* when he wants to express a conflicting opinion to an interlocutor considered socially superior aiming to not disturb him but to help him, or when the speaker desires to get a favor from his superior. It is described by [9] as “a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker’s consciousness of the fact that his/her utterance of request imposes a burden to the hearer” (p.34).

2.2. Corpus

We selected one sentence of 8 moras from the corpus developed and validated in previous research [13]. This

sentence (“*Nagoyade nomimas.*”, [nagojade nomimas], meaning “*He drinks in Nagoya.*”) is constructed on a verb-object syntactic structure. The lexical stress position is located on the first mora. In order to express some attitudes like *doubt* or *surprise*, the vowel [u] may be inserted at phrase final position, and in this case, the lexical stress will be realized at the seventh mora, too. The sentence was constructed in order to have no particular affective connotations in any region of Japan. Two male Japanese native language speakers produce each sentence with all the fourteen attitudes. The first speaker is a Japanese native language teacher who teaches various attitudes in his class by pragmatic explanations; the second, a naïve native speaker. A total of 24 utterances (1 sentence x 12 attitudes x 2 speakers) were digitally recorded in a soundproof room at LIMSI. Both speakers were standing in front of the video camera, with an omni-directional AKG C414B microphone placed 40 cm to their mouth. The microphone was connected to an USBPre sound device connected to a computer outside the room, recording the speech signal at 44,1 kHz, 16bits. A digital DV camera (Canon XM1 3CCD) recorded the speakers’ performances. Hand claps between each sentence, recorded both by the camera and the microphone, allows a post-processing a replacement of the camera sound by the high quality sound recorded by the microphone, synchronized thanks to the claps. The corpus of 3 modalities (i.e. audio-alone, visual-alone and audio-visual) of 24 utterances were created, thus a total of 72 stimuli (1 sentence x 12 attitudes x 2 speakers x 3 modalities) were used for the following perceptual experiment.

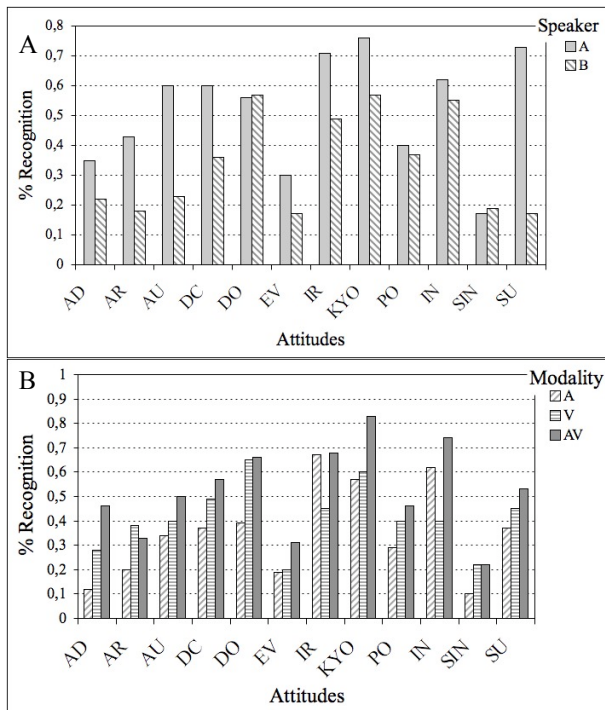


Figure 1: Effect of speaker (A) or of modality (B) on the recognition of attitudes. X-axis shows attitudes. Y-axis shows recognition rate of 46 Japanese listeners.

2.3. Experiment protocol

There are three presentation conditions – auditory alone, visual alone, and audio-visual. In the auditory-alone

condition, subjects were instructed to listen to what the speaker said and to judge the attitude he expressed. In the visual-alone condition, subjects were required to carefully watch his facial and body movements without sound and to judge the attitude he expressed. In audio-visual condition, subjects were required to watch the speaker’s face and body movements while listening to his voice and to judge the attitude he expressed.

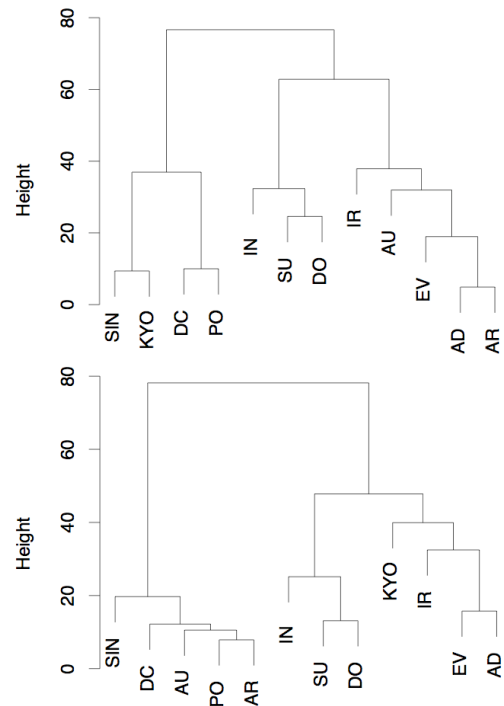


Figure 2: clustering analysis for 12 attitudes (A: speaker A; B: speaker B) for the audio-only modality.

The 46 Tokyo dialect speakers who participated in this experiment are separated into 2 groups. Group A consisted of 28 subjects (4 males and 24 females mean age = 18.8) who tested first with audio-only, then visual-only, and finally audio-visual modalities. Group B consisted of 18 subjects (7 males and 11 females mean age = 18.6) who tested first with visual-only, audio-only, and finally audio-visual modalities.

The perception tests were administered in a quiet room, using headphones, and on a Windows-based computer for running the interface program. The test interface gave an explanation of each label, and instructions were read in class. No subject expressed any trouble to understand the concepts referred to by the labels. All subjects listened to (and/or looked at) each stimulus one time only. For each stimulus, they were asked to indicate the perceived attitude among the twelve. Then they were asked to evaluate the intensity of expressiveness of the perceived attitude by moving a cursor located on the bottom of each label on a free scale ranging from “hardly perceptible” to “very marked” (encoded on a 1-100 scale, with the 0 score for the 11 not selected attitudes). The presentation order was randomized in a different order for each subject. No listeners participating in this experiment reported any listening disorder or any visual problem.

3. Results

Results, either the percentage of recognition obtained by each attitude or their mean intensity rating, were analyzed with repeated-measure ANOVAs for testing the relative influence of each factor: the modality's presentation order (i.e. the listeners' group), the 3 modalities, the 12 attitudes and the 2 speakers. Therefore, confusions between attitudes were analysed with data-reduction techniques (a correspondence analysis, and a cluster analysis). For more details on the statistical analyses, refer to [8].

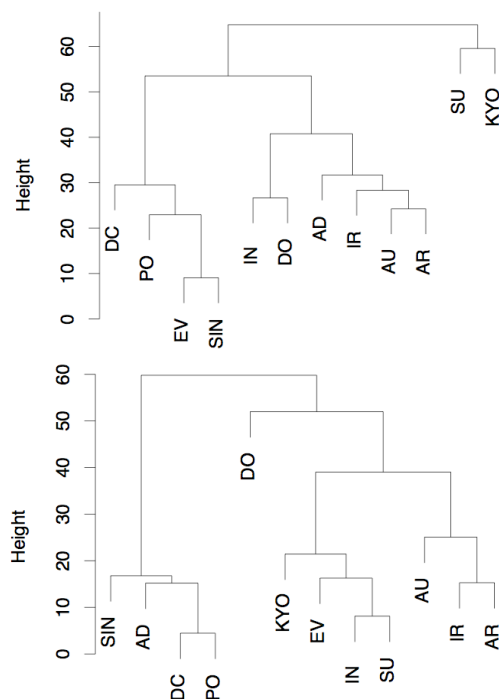


Figure 3: clustering analysis for 12 attitudes (A: speaker A; B: speaker B) for the video-only modality.

3.1. Results of the ANOVAs

For both recognition percentages and intensity scores, the Mauchly's test of sphericity is not significant ($p > .01$). Therefore, the repeated-measures ANOVAs were computed, assuming compound symmetry ($p > .01$). First we investigated whether the modality's presentation order influenced the perception of attitude, and evaluated the results of the two groups by ANOVA. The results show no significant influence of the order of presentation on listeners' answer ($F = 0.0795$, $p = .78$).

Results show a significant influence of the speaker on the recognition scores, expressed either as categorical choice ($F = 163.6$, $p < .0001$) or as intensity ratings ($F = 211.5$, $p < .01$). Figure 1A presents listeners' recognition rate for the two speakers, and for each attitude. This graph shows that the attitudes expressed by speaker A, who is the Japanese native language teacher, were globally better recognized than speaker B's expressions. Especially the perception of AU and SU was quite different.

The modality also influenced the listeners' perception: a significant effect was observed with both categorical choices ($F = 44.8$, $p < .01$) and intensity ratings ($F = 44$, $p < .01$).

Figure 1B presents listeners' recognition rates of the three modalities and for each attitude. The audio-visual condition shows the best recognition score for all the attitudes except AR. The recognition score of AR, DO and SIN in visual alone condition are mostly the same as in the audio-visual condition, indicating a predominance of the visual information over the auditory one for these attitudes. However, the attitude of IR seems mostly recognized by auditory information rather than visual information. For most attitudes (AD, AU, DC, EV, KYO, PO, IN, SU), audio and visual information cooperate, as audio-visual score are better than audio or visual only information. Auditory and visual information complemented each other.

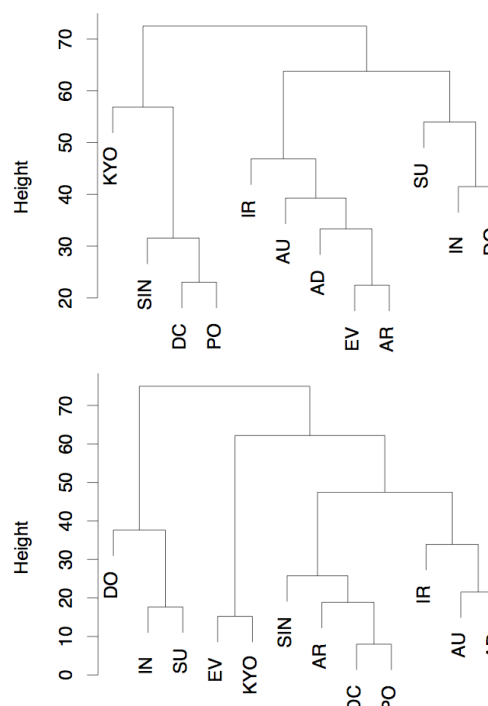


Figure 4: clustering analysis for 12 attitudes (A: speaker A; B: speaker B) for the audio-video modality.

3.2. Confusion analysis by data reduction techniques

Similarities between listeners' recognition between the 12 attitudes will be detailed (for each modality and each speaker) hereafter using the clustering analysis.

Audio alone modality

For speaker A, the results regroup in 3 main categories (cf. fig. 2A). The first group consists of SIN, KYO, PO and DC. Listeners tend to perceive in a similar way all the politeness expressions, plus DC. The second group contains IN, DO and SU, and therefore constitutes a more global group of query expressions. The third group consists of IR, AU, EV, AD and AR. All these attitudes except admiration express the imposition of the speaker's opinion.

Similarly, there are 3 categories observed for speaker B (cf. fig. 2B), but the components of each category are not the same. The first group consists of SIN, PO, DC, AU and AR. Contrary to the first group for speaker A, AU and AR are included instead of KYO. The second group contains IN, DO and SU, which are the same components seen with speaker A: all these attitudes express a query. The third group consists of

KYO, IR, EV and AR. As it has already been suggested above, the performances of untrained speaker B are lower than those of speaker A, and that can lead subjects to more confusions, even between expressions that are supposed to be very different.

Visual alone modality

For speaker A, the perceptual categories of attitudes in the visual modality (cf. fig. 3A) are the same as those in the audio modality, except KYO and SU in that each behave very specifically in the visual modality.

Results for speaker B (fig. 3B) show a radical change of perceptual behavior in the visual modality. The first group consists of 2 polite expressions plus DC and AD. However KYO is added to the second group, with EV, IN and SU, and mainly seems to group the most confused expressions for this modality and this speaker. A third group consists of AU, IR, and AR, all expression of dominance. DO does not show any confusion.

Audio-visual modality

According to 3 perceptual categories shown in fig. 4, the attitudes belonging to each category are almost the same as in the audio-alone modality (fig. 2), for speaker A. However, the results of speaker B in fig. 4 are quite different from both audio-alone and visual-alone except for the category of “query” observed in audio-alone. In addition, KYO and EV form one independent category.

Such predominance of the audio modality over the visual one for the confusion of attitudinal expressions is coherent with the results observed on French (see [8]).

4. Conclusion

The aim of the present work is to investigate how Japanese listeners recognize audio-visual attitudinal expression, and the completion or redundancy of the two modalities to each attitude represented in this work. The results show no significant difference due to presentation order, whereas the speaker performance, the modality and the expressed attitudes did have a significant effect on listeners’ recognition scores. Generally the audio-visual modality showed the best recognition score for all attitudes. For some attitudes, the one modality alone seems to carry all the needed information, whereas the other doesn’t add more information, but for most attitudes, auditory and visual information complemented each other.

The listeners generally group the set of 12 attitudes into 3 more general perceptual categories for the first speaker. The first group consists of the polite expressions, the second, the attitudes of “query”, and the third, the expressions of imposition of one’s own opinion. The attitude of KYO and SU are particularly well recognized by visual information for speaker A. Worst recognition scores and more important confusions shown on speaker B’s results may be due to several factors. As a naïve speaker, he may have experienced more difficulties performing the 12 attitudes, and may have chosen different strategies than speaker A. He may also have expressed these attitudes in a more subtle and less intense way, leading to stimuli harder to recognize out of context – and this is reflected by the difference between the mean intensity ratings obtained by both speaker A and B.

In order to better understand some of the anomalies in the results presented here, future work will involve acoustic and visual analysis of the productions of each of the speakers. Also analysis-resynthesis of these audio-visual attitudes are

planned in order to extract the most pertinent parameters for perception.

5. Acknowledgements

We are deeply grateful to T. Sadanobu from Kobe University for his helpful advice and contribution on Japanese attitudes. We also thank students at Showa Music University for their participation in the perception tests. This work was performed as collaboration between the Graduate School of Cultural Studies and Humane Science of Kobe University and the GIPSA-Lab of Grenoble under the auspices of the “College Doctoral Franco-Japonais”. It was supported in part by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research, (2007-2010):19520371 to the second author, and also by SCOPE (071705001) of Ministry of Internal Affairs & Communications (MIC), Japan.

6. References

- [1] Aubergé V. 2002. A Gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP. *Speech Prosody*, Aix-en-Provence, France, 151-155.
- [2] Barkhuysen, P., Krahmer, E. & Swerts, M. 2006. How Auditory and Visual Prosody is Used in End-of-Utterance Detection. *Interspeech*, Pittsburgh, USA.
- [3] Barkhuysen, P., Krahmer, E. & Swerts, M. 2007. Cross-modal perception of emotional speech. *ICPhS*, Saarbruecken, Germany, 2133-2136.
- [4] Danes, F. 1994. Involvement with language and in language. *Journal of Pragmatics*, 22, 251-164.
- [5] Erickson, D., Ohashi, S., Makita, S., Kajimoto, N., Mokhtari, P. 2003. Perception of naturally-spoken expressive speech by American English and Japanese listeners. *CREST International Workshop on Expressive Speech Processing*, 31-36.
- [6] Maekawa, K., 1998. Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *ICSLP*, 635-638.
- [7] Mizutani O., Mizutani N., 1979. *Aural Comprehension Practice in Japanese*. The Japan Times.
- [8] Rilliard, A., Martin, J.C., Aubergé, V. & Shochi, T. (submitted). Perception of French Audio-Visual Prosodic Attitudes. *Speech Prosody*, Campinas, Brasil.
- [9] Sadanobu, T. 2004. A natural history of Japanese pressed voice. *J. of the Phonetic Society of Japan* 8 (1), 29-44.
- [10] Scherer, K. R., Banse, R., Wallbott, H. G., 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76-92.
- [11] Scherer, KR & Ellgring, H. 2007. Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns? *Emotion*, 7(1), 158-171.
- [12] Shigeno, S. 1998. Cultural similarities and differences in the recognition of audio-visual speech stimuli. *ICSLP98*,
- [13] Shochi, T., Aubergé, V., Rilliard, A., 2006. How prosodic attitudes can be false friends: Japanese vs. French social affects. *Speech Prosody*, Dresden, 692-696.
- [14] Swerts, M. & Krahmer, E. 2005. Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1), 81-94.

Perception of French Audio-Visual Prosodic Attitudes

A. Rilliard¹, J.C. Martin¹, V. Aubergé² & T. Shochi²

¹ LIMSI-CNRS, Orsay, France; ² GIPSA-Lab, Grenoble, France
{rilliard; martin}@limsi.fr ; {auberge; shochi}@gipsa-lab.inpg.fr

Abstract

Experimental studies are required to understand the contribution of audio and visual modalities during affective communication. This paper presents a perception study of the audio-visual expression of six French attitudes. The relative importance of each modality in the decoding of these expressions is analysed, as a first step toward a deeper comprehension of their influence on social affects expression. Two speakers are evaluated, in order to test the influence of speaker's performances on listeners' perception.

1. Introduction

Classical works on prosody describe it as conveying different levels of information, ranging from linguistics one to expressive ones [17, 12]. The variety of expressive functions has been studied for a long time (cf. [1, 11, 10]). In a similar way, audio-visual speech prosody has more recently been recognized as carrying functions such as the feeling of knowing [22] and the signalling of end of utterances [5].

About the nature of affects, the work of [23] have raised question on possible differentiation between social affects and emotions. A similar differentiation is used by [16]. Aubergé [3] start from this distinction between social affects (or attitudes) and emotions to question the existence of different controls underlying the expression of these different affects, i.e. the parameters carrying the expressivity may be organized differently by the subject (in timing, shape or range), and this allows his interlocutor to decode his emotional state.

Such attitudes or social affects have already been studied by several scientists [24, 14, 4, 9] and also in cross-linguistic contexts [20, 7], but mainly in their acoustic modality only. As the study of the multimodal expression of emotions (in their broad sense) is still a recent field of research [18, 19], only a few works question directly the specific question of social affects, specifically differentiated from emotions [13]. This paper, and the companion paper describing a similar work on Japanese attitudes [21] intend to analyse the specificities and the complementarities of visual and audio modalities in the expression and perception of a specific kind of social affect: attitudes. It also tries to measure the influence of speaker's performance on perception judgements, as many works only rely on the performance of only one speaker.

The first part of the paper describes the construction of the corpus, and explains the choices of expressions; then the experimental setting is tackled. Second part approaches the results analysis and discussion, before some conclusions and perspectives for future works.

2. Corpus & Experimental design

2.1. French Audio-Visual attitudes

Following the work done by Morlec [14] on French prosodic attitudes, based on [11], 6 attitudinal expressions were selected

for recording a French audio-visual corpus: *declaration* (DC), *simple question* (QS), *obviousness* (EV), *surprise exclamation* (EX), *doubt-incredulity* (DO), *suspicious irony* (SC). The main differences between the work done by [Mor01] and this one reside in (1) the audio-visual recording of the 6 attitudes, (3) a work on two different speakers, in order to measure the influence of individual performance over recognition scores, and (2) the recording paradigm, designed to set the speaker in a somewhat more natural condition of production of these 6 attitudes: speakers were instructed to produce each of these sentences in order to express one attitude, as an answer to a statement produced by a partner. They had already been trained to produce these attitudes in a preceding session, and had to behave as naturalistic as possible, without any constraints on their expressive strategy.

Two speakers, S1 & S2, both male native French speakers, were recorded in a soundproof room at LIMSI. They were standing in front of a video camera, with an omnidirectional AKG C414B microphone placed 40 cm to their mouth. The microphone was connected to an USBPre sound device connected to a computer outside the room, recording the speech signal at 44,1 kHz, 16bits. A digital DV camera (Canon XM1 3CCD) recorded the speakers' performances. Hands claps between each sentence, recorded both by the camera and the microphone, allows as a post-processing a replacement of the camera sound by the high quality sound recorded by the microphone, synchronized thanks to the claps. Video clips were encoded thanks to a cinepack codec with a 784 x 576 pixels resolution, either using both MOV and AVI video file formats (respectively for display on Apple or Windows platforms).

The corpus is based on three sentences respectively of 4, 5 and 7-syllable length, without any specific meanings that can bootstrap or forbid one of the 6 attitudes. After the recording and the post-processing, the speakers' performances were judged by both of them, and only the 5-syllable length sentence was kept for the perception test: "*Nicolas revenait.*" [nikola ʁəvənɛ] ("Nicolas was coming back"), played with the 6 attitudes. 6 short videos were thus produced.

2.2. Perception test

An evaluation test was designed in order to evaluate the relative efficiency of the two modalities to carry the attitudinal information. The factors that have been controlled during this experiment are:

- the 6 attitudes;
- the speaker (S1 or S2);
- the modality (Audio, Visual or Audio-Visual);
- the modalities' presentation order (Audio or Visual first);

Subject listen to each stimuli only once for each modality, presented in a random order. For each stimulus, they had to select the attitude they perceived in the stimulus as well as its intensity on an open scale ranging from "*hardly perceptible*" to "*very marked*" (encoded on a 1-100 scale, with the 0 score

for the 5 not selected attitudes). Subjects had to fill the questionnaire on the PC without any time constraint.

Two groups of subjects passed the experiment. The first group first listened to the audio only stimulus, and then watched the video only stimulus, and finally the audio video stimulus. The second group started with the video only stimulus, continued with the audio only stimulus and finally ended with audio-video stimulus. This enabled to counter-balance a possible effect of the presentation order of the stimuli's modality. During the presentation of one modality, the stimuli corresponding to all attitudes and to the two speakers are randomized – in a different order for each listener.

2.3. Subjects

32 French listeners (17 male and 15 female, mean age = 32) pass the experiment, 16 in each group (Audio only first and Visual only first).

Table 1: results of the 2 ANOVAs, on the percentage of recognition of each attitude and on the intensity scores. Significant effect ($p < .01$) are in bold. Grp stands for the Group factor, Spk for the Speaker, Mod for the Modality and Att for the Attitudes.

	df	% Reco.		Intensity	
		F	p	F	p
Grp	1	0.5	0.508	0.3	0.605
Spk	1	45.3	0.000	105.5	0.000
Grp:Spk	1	1.9	0.178	7.1	0.012
Mod	2	14.9	0.000	25.6	0.000
Grp:Mod	2	3.8	0.028	3.7	0.030
Att	5	6.4	0.000	11.8	0.000
Grp:Att	5	2.2	0.055	0.9	0.466
Spk:Mod	2	1.7	0.199	3.0	0.059
Grp:Spk:Mod	2	1.0	0.364	0.0	0.994
Spk:Att	5	3.8	0.003	7.5	0.000
Grp:Spk:Att	5	0.7	0.589	0.5	0.764
Mod:Att	10	7.9	0.000	8.6	0.000
Grp:Mod:Att	10	1.3	0.246	1.4	0.195
Spk:Mod:Att	10	5.4	0.000	3.6	0.000
Grp:Spk:Mod:Att	10	1.5	0.146	1.7	0.085

3. Results analysis

3.1. Results processing

Results given by listeners are expressed by two measures: as categorical answers (the perceived attitude), and as a relative intensity score given to one category of attitude. Two kinds of results are analyzed: (1) the recognition rate of each attitude, expressed either as the sum of the categorical choice of the attitude by listeners (percentage of good recognitions), or as the mean intensity rating of good answers; and (2) the confusions matrix grouping the categorical answer given by listeners for each presented attitudes – expressed either as categorical recognition rate received by each possible attitude, or as the relative intensity received by each possible attitude compared to the total intensity rating received by the stimuli.

Recognition rates (either categorical or intensity) are analysed thanks to two repeated-measure ANOVA analyses (one for each kind of measure). Each ANOVA takes as a dependant variable the recognition rate of each attitude (expressed with percentages or mean intensity), subjects as a

random effect, one between-subject factor (the group, or the order of presentation of the Audio and Video modality), and three within-subject factors (the 6 Attitudes, the 2 Speakers and the 3 Modalities). ANOVA analysis mainly aims at measuring the relative importance of the different factors on the listener's behaviour.

Confusion matrices are analysed thanks to a correspondence analysis and a cluster analysis [8]. Both of these methods are based on data-reduction techniques that allow a more simple and comprehensive representation of our data. The first one allows to graphically representing the perception results, in order to analyse the self-recognition of one attitude and their relative dispersion. The second one hierarchically regroups the different stimuli in clusters, the distances of which indicate the perceptive distance between the attitudes (the Ward distance metric is used for clustering), and thus allows to make distance judgements.

3.2. ANOVA results

For both recognition percentages and intensity scores, the Mauchly's test of sphericity is not significant ($p > .01$). Therefore, the repeated-measures ANOVAs were computed, assuming compound symmetry. Results are displayed in table 1 for both ANOVAs.

The first consideration of this analysis is the complete coherence obtained for both kind of measures: all affects receive coherent significance level, whatever the measure. Therefore, in the remaining of the analysis, no difference between the two measures will be made. The order of presentation of the modality (the group factor) is not significant. It only has a small interaction marginally significant with the Modality factor: scores for the Audio only condition are slightly higher when the audio modality is not presented first. The main effects are received by the three main factors: the Speaker, the Modality and the attitudes. Scores for these three factors are presented on the figure 2.

Speaker S1 receives higher recognitions scores than S2, whatever the modality or the expressed attitude. The difference between speakers is even amplified by the intensity ratings – S2's expressions being perceived with a lower activation.

As expected, Audio-Visual stimuli receive the highest ratings. Audio- and Visual-only modalities received comparable mean scores.

Mean recognition rate for each individual attitude is far above the chance level (16,6%), despite important differences between them (cf. figure 2).

Interesting interactions between these three factors can also be noticed. Figure 2A presents the percentage of recognition for each attitude, according to the presentation modality. Audio information appears particularly efficient alone for the expressions of *declaration* (DC) and *simple question* (QS), and video information for *doubt-incredulity* (DO). The Audio-Visual modality receives almost always the best score (or at least nearly the best), for all attitudes, and shows a particularly important synergy between the two modalities for the expressions of *obviousness* (EV), *suspicious irony* (SC) and *surprise exclamation* (EX).

Figure 2B presents the relative performances of the two speakers for each attitude. If S1 scores are almost always higher, he is particularly more efficient for the SC attitude, and surprisingly for DC. For this attitude, speaker S2 moves his eyes up at the beginning of the recording (trying to

remember the sentence to produce), and it gives a strong notion of doubt or question, especially on the video-only stimuli: DC is absolutely not recognized for the video-only stimuli, whereas it receives good scores for the audio-only one. The difference between the two speakers for the SC expression is mainly due to the relative efficiency of their production.

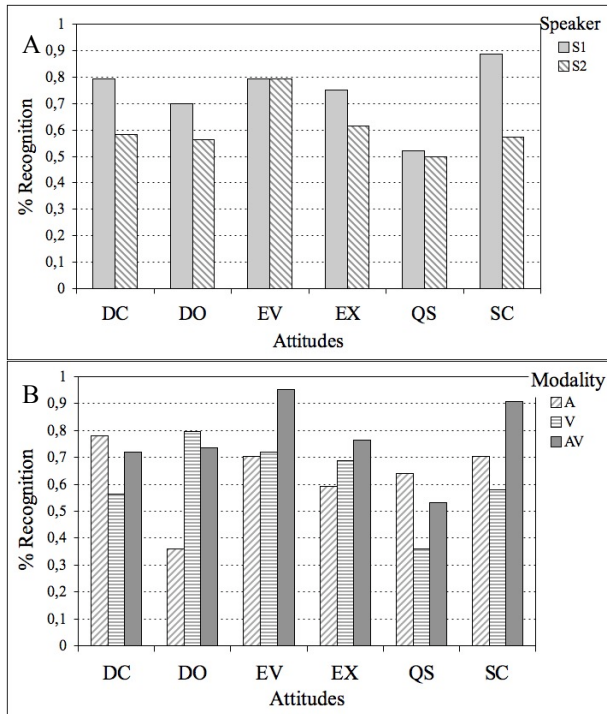


Figure 2: (A) presents the mean percentage of recognition obtained by each attitude in each of the 3 modalities (Audio-Visual, Audio and Visual). (B) presents the percentages obtained by each attitude for each speaker (S1 and S2).

3.3. Analysis of confusion matrices

The analysis of the confusion matrices leads to interesting parallel between attitudinal expressions. In order to obtain the main differences and similarities, two data reduction techniques are applied.

3.3.1. Correspondence analysis (CA)

Such analysis extracts the main abstract dimensions that explained most of the variance in the original data. Applied to the confusion matrix for each speaker and for each modality (and for both the percentage of recognition and the intensity scores), it raised the most important divergences between stimuli and also the main proximities. Moreover, as it plots both the rows of the matrices (corresponding to the proposed stimuli) and its columns (i.e. the answered attitudes, corresponding to the concept subject have of each attitude), the proximity between the rows and column dots may indicate a good recognition of the attitude, whereas a great distance indicate lower recognition scores. As the results obtained by recognition percentages and by intensity scores gives completely coherent results, they will not be analysed separately.

The first two dimensions of these analyses explain between 51 to 69% of the variance, for each speaker and each modality; and adding the third one generally explains about 80%. Interestingly, dimensions extracted by the analyses are very comparable (as they grouped or opposed the same attitudes) from one speaker to the other; even if their order sometimes changes¹. If we compare the first 4 dimensions of each CA, it then gives a very precise sketch of the contribution to French attitudes of each modality.

The first result is the good recognition scores obtained by each attitude, for each speaker and in each modality – unless the DC by S2 in the video-only condition, as already mentioned.

Audio-only information. First dimension opposed affirmative attitudes (DC and EV) to questioning ones (QS, DO, EX). Second one opposed SC to the others. The third and the fourth dimensions opposed the attitudes regrouped by the main opposition of the first dimension: respectively QS vs. EX and DO; and DC vs. EV. For this modality, EX and DO show some perceptive similarities.

Video-only information. The first dimension for the facial information opposes EX to statements expressions. The second dimension opposes SC and EV to DC. The third opposes QS and DO to EV and EX. And finally the fourth opposes SC to EV. Such results show that the four non-linguistically encoded attitudes (i.e. DO, EV, EX and SC) have strong visual cues, whereas expressions of DC and QS are respectively opposed to the first four², or have strong confusions with one other (QS is mixed with DO in this modality).

Audio-video information. The first dimension for the audio-video stimuli opposes affirmative to questioning expressions, and the second one opposes SC to the others, as for the audio-only condition.

The third dimension opposes obviousness to declaration, and the fourth opposes EX to DO and QS. The last two dimensions show the main differences existing inside the two main groups drawn by the first dimension, as for the audio condition. But for the AV condition, confusions exist for QS, recognized as DO, as for the visual-only modality.

3.3.2. Cluster analysis

The cluster analysis leads to different considerations. Its main purpose is to analyse the distances between the stimuli. It has already been said that most attitudes are well recognized, so this analysis will focused on the confusions between attitudes, and on the differences and similarities between speakers by modalities.

Audio-only stimuli. Speakers receive coherent result in modality. DO and EX are closely related, and show a looser connection with QS. This result differs from preceding ones obtained by [2], where DO was mixed up with SC, a confusion that is totally absent of our data. Other attitudes are clearly dissociated.

Video-only stimuli. Visual information clearly differs from one speaker to the other. For S1, QS is mixed up with DO, while all the others expressions are clearly recognized. For S2, the already mentioned problem with the DC make it close

¹ For the sake of simplicity, the order of dimension will follow the one obtain on the analysis of speaker S1. When dimensions differ between speakers, it will be explicitly mentioned.

² We don't refer here to the results obtained by speaker S2 for the visual-only declaration, as they may be biased.

to three other attitudes, each of which also showing confusion with another one. QS is mixed up with EX, and SC with DC.

Audio-video stimuli. There is a clear interference between both modalities in these results. For S1, the visually clear expression of EX helps listeners to distinguish it from its acoustically similar DO; whereas the acoustic of QS separates it from the more similar visual DO. For S2, acoustic DC prevails over the problematic eyes movements observed in visual modality and the visual difference between DO and EX allows listeners to differentiate them where acoustics fails to.

4. Conclusions

This work on the audio-visual expression of French social affects gives interesting results on several dimensions, and open interesting research questions.

By comparing these results to the ones obtained by [2], the importance of the speaker's strategy appears clearly. Even if most of the data is coherent for the three speakers (the 2 of this study and the one cited above), it is clear that both strategic choices in the use of available acoustic and visual parameters, as well as individual performance to achieve a particular attitude have an influence on the perception results. More works, like [15] may be devoted to data collection on important amounts of social affect expressions, with a specific attention to naturalness.

The comparison of each modality's account to expressivity is coherent with [7] works, which points out the relative importance of multimodality in interaction. Moreover, the relative contribution of audio and visual cues to each attitude is coherent: audio information seems primarily important for *declaration* and *simple question*, both linguistically encoded, whereas the others (and especially *surprise exclamation* and *doubt-incredulity*) seems primarily influenced by visual cues, with strong speaker-dependant changes. Moreover, the synergy between modalities is important for almost each attitude, and especially *obviousness* and *suspicious irony*.

We are currently working on the analysis of perception results filtered by the extraversion rating of listeners, and also on the analysis of the acoustic and visual cues to each attitude. Correlations between these cues (and their evolution with time) and perception results will allow a deeper understanding of the modalities' contribution.

Finally, audio-visual replay on a talking head of these affects may allow investigating the indices that lead to naturalness and spontaneity in expressivity. Spontaneity of multimodal expression is still a great challenge, and some improvement of the recording paradigm may be introduced (or spontaneous data collected) in order to acquire more information on this point. Moreover, it may be interesting to specifically design a perception test in order to be able to monitor the recognition speed of listeners, as it has been related to the valence of multimodal stimuli [6].

5. References

- [1] Allerton, D.J. & Cruttenden, A. 1978. Syntactic, illocutionary, thematic and attitudinal factors in the intonation of adverbials. *Journal of Pragmatics*, 2, 155-188.
- [2] Aubergé, V., Grépillat, T. & Rilliard, A. 1997. Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. *EuroSpeech*, Rhodes, 871-877.
- [3] Aubergé, V. & Cathiard, M.A. 2003. Can we hear the prosody of smile? *Speech Communication*, 40 (1), 87-97.
- [4] Bänziger, T. & Scherer, K.R. 2005. The role of intonation in emotional expressions. *Speech Communication*, 46, 252-267.
- [5] Barkhuysen, P., Krahmer, E. & Swerts, M. 2006. How Auditory and Visual Prosody is Used in End-of-Utterance Detection. *Interspeech*, Pittsburgh, USA.
- [6] Barkhuysen, P., Krahmer, E. & Swerts, M. 2007. Incremental perception of acted and real emotional speech. *Interspeech*, Antwerp, Belgium, 1262-1265.
- [7] Barkhuysen, P., Krahmer, E. & Swerts, M. 2007. Cross-modal perception of emotional speech. *ICPhS*, Saarbrücken, Germany, 2133-2136.
- [8] Benzecri, J.P. 1973. *L'analyse des données*. Paris: Bordas.
- [9] 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), 111-120.
- [10] Danes, F. 1994. Involvement with language and in language. *Journal of Pragmatics*, 22, 251-164.
- [11] Fónagy, I., Bérard, E. & Fónagy, J. 1984. Clichés mélodiques. *Folia Linguistica*, 17, 153-185.
- [12] Fónagy, I. 2003. Des fonctions de l'intonation : essai de synthèse. *Flambeau*, 29, 1-20.
- [13] Granström, B. & House, D. 2005. Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, 46, 473-484.
- [14] Morlec, Y., Bailly, G. & Aubergé, V., 2001. Generating prosodic attitudes in French: Data, model and evaluation. *Speech Communication*, 33 (4), 357-371.
- [15] Moroni, V. 1997. *Enquête sur les attitudes du français: définition et interprétation*. Master thesis, Univ. Grenoble 3, France.
- [16] Ohala, J.J. 1996. Ethological theory and the expression of emotion in the voice. *ICSLP*, Philadelphia, 1812-1815.
- [17] Rossi, M., Di Cristo, A., Hirst, D., Martin, P., Nishinuma, Y. 1981. *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.
- [18] Scherer, K.R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2), 227-256.
- [19] Scherer, K.R. & Ellgring, H. 2007. Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns? *Emotion*, 7(1), 158-171.
- [20] Shochi, T., Aubergé, V. & Rilliard, A. 2007. Cross-listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes. *ICPhS*, Saarbrücken, Germany.
- [21] Shochi, T., Erickson, D., Rilliard, A., Aubergé, V. & Martin, J.C. (submitted). Recognition of Japanese attitudes in Audio-Visual speech. *Speech prosody*, Campinas, Brasil.
- [22] Swerts, M. & Krahmer, E. 2005. Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1), 81-94.
- [23] Tomkins, S. S. 1984. Affect theory. In *Approaches to emotion*, K. R. Scherer & P. Ekman (eds.). Hillsdale, N.J.: Erlbaum, 163-196.
- [24] van Heuven, V.J., Haan, J., Janse, E., van der Torre, E.J. 1997. Perceptual identification of sentence type and the time distribution of prosodic interrogativity marker in Dutch. *ETRW Workshop on Prosody*, Athens, Greece, 317-320.

CROSS-LISTENING OF JAPANESE, ENGLISH AND FRENCH SOCIAL AFFECT: ABOUT UNIVERSALS, FALSE FRIENDS AND UNKNOWN ATTITUDES

Takaaki Shochi[†], Véronique Auberge[†], Albert Rilliard[‡]

[†]GIPSA-lab, Grenoble, France - [‡]LIMSI, Orsay, France

Takaaki.Shochi@icp.inpg.fr, Veronique.Auberge@icp.inpg.fr, Albert.Rilliard@limsi.fr

ABSTRACT

Seven affectively-neutral Japanese sentences uttered with 12 different attitudes are investigated. The listeners were 15 Japanese, 15 French and 20 American. Both groups of non-native listeners had no Japanese language skill. They were asked to choose the speaker's intended attitudes among the 12. Results showed that Japanese recognized all attitudes above chance, but with some confusion, especially for the politeness expressions. However, French and American do not recognize the two cultural politeness expressions. Especially *kyoshuku*, a type of politeness that does not occur as a conventional expression in occidental society, was incorrectly decoded by French and American as *arrogance* or *irritation*.

1. INTRODUCTION

Affects in speech are expressed according to different cognitive processing levels, from involuntarily controlled expressions of emotion to the intentionally, voluntarily controlled expressions of attitudes of the speaker. Attitudes (as “labels” or “notions”) are sometimes confused with Emotions (as “labels” or “notions”) since both use specialized prosodic encoding in the acoustic signal. Since the prosodic emotional expressions are carried in parallel with the prosodic linguistic structures, then we can view expressions of attitudes as an integral part of “language interaction building” (i.e. communication). Some affects such as *surprise*, are classified either as attitudes or emotions. Our position is to assume that *surprise* can be an emotion if it is involuntarily processed, and can be an attitude when it is voluntarily processed and needs to be learned. When a speaker does not produce any attitude on an utterance, it must be considered a special attitude (hereafter referred to as *declaration*): the speaker doesn't intend to express any information about his intentions. We consider that studying affects in spoken interaction is crucial for

expressive speech modelling, and we need to describe precisely what the attitudes of a speaker in a language and a culture can be and how they are linguistically encoded. A cross-cultural study is a helpful method to examine this problem. Some attitudes can be expected to be universal in part, both in terms of what is expressed and how it is expressed. But because attitudes are socially constructed for and by the language, they can exist in one language but not another; moreover the prosodic realization of one specific attitude in a specific language may not be recognized (or may be ambiguous) in another language. This often constitutes problems for foreign language learners. After presenting our corpus of Japanese attitudes, we will describe how listeners either Japanese, French or American perceived them.

2. CORPUS OF JAPANESE ATTITUDES

A set of 12 Japanese attitudes were selected for their representativeness according to the literature [3], [7], [8] and Japanese language teaching methods [6]: *doubt-incredulity* (DO), *evidence* (EV), *exclamation of surprise* (SU), *authority* (AU), *irritation* (IR), *arrogance* (AR), *sincerity-politeness* (SIN), *admiration* (AD), *kyoshuku* (KYO), *simple-politeness* (PO), *declaration* (DC) and *interrogation* (IN) (see [10] for definitions). Some of these attitudes are specific or specifically important for the Japanese culture, especially those linked to the politeness strategy: *simple-politeness*, *sincerity-politeness* and *kyoshuku* vs. *arrogance*. The *sincerity-politeness* attitude appears when a speaker considered inferior is talking to someone superior in the Japanese society: the speaker expresses a serious and sincere intention by using this prosodic attitude. The *kyoshuku* attitude (there is no lexical entry to translate this in English) is a typically Japanese cultural attitude. Even if such situations occur in all cultures, the Japanese language has chosen to encode this situation as an attitude (“attitudineme”). A speaker uses *kyoshuku*

when he wants to express a conflicting opinion to an interlocutor considered socially as superior – aiming at not disturbing him but to help him, or when the speaker is desiring to get a favour from his superior. It is described by [8] as “a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker’s consciousness of the fact that his/her utterance of request imposes a burden to the hearer” (p.34).

The corpus consists of seven sentences of 1, 2, 5 or 8 moras, with either a single word or a verb-object structure as the syntactic structure. For the eight-mora utterances, the lexical stress position may be on the first, second, and third mora, or absent [4]. In order to express some attitudes like *doubt* or *surprise*, the vowel [u] may be inserted at phrase final position, and the lexical stress will be realized at the seventh mora, too. The sentences were constructed in order to have no particular affective connotations in any region of Japan. Each sentence is produced with all the attitudinal functions by a male Japanese native language teacher. A total of 84 stimuli were digitally recorded at 44.1 kHz sampling rate stereo with Sony DATTCD-D7 recorder and SHURE WBH53 headset microphone.

Table 1: Corpus of Japanese attitudes – the lexical stress is marked with a star.

mora	Utterance	Translation
1	Me	The eye
2	Na *ra	Nara
5	Na *rade neru	He sleeps in Nara
8	Na *goyade nomimas	He drinks in Nagoya
8	Nara *shide nomimas	He drinks in Nara Town
8	Matsuri *de nomimas	He drinks at the party
8	Naniwade nomimas	He drinks at Naniwa

3. EXPERIMENTAL PROTOCOL

In order to cross-linguistically evaluate the recorded attitudes, a computer displayed interface was developed and used amongst several groups of listeners. The test interface proposes a translation of each label in the listener’s native language, and was introduced by a long oral explanation about each attitude, giving many examples of relevant situations. No subject expressed any trouble to understand the concepts referred to by the labels. All the subjects of these experiments listened to each stimulus one time only. For each stimulus, they were asked to indicate the perceived attitude amongst the twelve. The presentation order was randomized in a different order for each subject.

Listeners from three languages were tested during the experiment: Japanese, English and French native speakers. Japanese listeners were used in effect to validate the quality of the recorded attitudes; French and Americans languages were used to represent cultures and languages distant from Japanese. In all, the subjects were 15 Japanese (11 females and 4 males, Tokyo dialect, mean age = 29.5), 15 French (10 females and 5 males, level 0 in Japanese, mean age = 25.4), and 20 American listeners (12 females and 8 males, level 0 in Japanese, mean age = 24.9). No listeners participating in this experiment reported any listening disorder.

4. RESULTS

4.1. Validation with Japanese listeners

According to a chi-square test, the distribution of answers for each attitude was above chance ($p < .05$ df:11). A possible effect of the stimuli length for the distributions of selected attitudes was tested with a chi-square test. The results show a significant effect of length between two and five-mora sentences (23.1 $p < .05$ df:11) and also between five and eight-mora ones (34.5 $p < .05$ df:11). There were no effects of the lexical stress (16.5 $p > .05$ df:33).

In order to determine which attitude listeners recognized over chance, the following criterion was used: the mean identification rate must be over twice the theoretical chance level. According to this criterion, seven attitudes (i.e. *arrogance*, *declaration*, *doubt-incredulity*, *simple-politeness*, *exclamation of surprise*, *irritation* and *interrogation*) were recognized without any particular confusion.

Authority was confused with *evidence*, (e.g., feeling confident in oneself). One possible explanation is that these two attitudes may be similar, since when imposing authority, the speaker is sure of himself.

Evidence was also confused with *arrogance*. Since *evidence* shows that the speaker is confident of himself, this expression of certainty can sometimes be perceived as disrespect to the interlocutor.

The two typical Japanese attitudes of *sincerity-politeness* and *kyoshuku* were also confused with each other. These two attitudes express essentially the humility of a speaker facing a superior person in the social hierarchy. *Sincerity-politeness* was also confused with *simple-politeness*, whereas this

confusion with *simple-politeness* was not seen for *kyoshuku*.

Concerning the attitude of *admiration*, we observed confusion with *simple-politeness*. These two attitudes are interconnected in Japanese society as can be seen by the lexical polysemy of items like *sonkee* [admiration / politeness], and *keifuku* [admiration / politeness]

Table 2: Percentage of recognition of attitudes by Japanese listeners. Good answers are in bold and significant confusions in *italic bold*.

Perceived attitudes	Presented attitudes											
	AD	PO	KYO	SIN	AR	AU	IR	DO	EX	QS	EV	DC
Admiration	21,9	4,8	1,9	3,8	0,0	0,0	0,0	1,9	1,0	0,0	0,0	0,0
Politeness	26,7	64,8	2,9	17,1	1,0	4,8	0,0	0,0	0,0	8,6	1,0	11,4
Kyoshuku	11,4	9,5	24,8	27,6	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0
Sincerity	14,3	6,7	26,7	32,4	0,0	5,7	0,0	0,0	0,0	1,9	0,0	1,0
Arrogance	1,9	0,0	1,9	0,0	72,4	5,7	2,9	5,7	0,0	1,0	21,0	10,5
Authority	0,0	0,0	15,2	1,0	9,5	51,4	11,4	1,0	1,0	1,9	9,5	2,9
Irritation	1,0	0,0	12,4	1,0	5,7	4,8	85,7	13,3	4,8	1,0	2,9	1,0
Doubt	0,0	1,0	0,0	0,0	0,0	0,0	0,0	56,2	14,3	3,8	0,0	0,0
Surprise	9,5	1,9	1,0	1,9	0,0	0,0	0,0	14,3	59,0	1,0	4,8	0,0
Interrogation	0,0	1,0	0,0	1,9	0,0	0,0	0,0	7,6	14,3	77,1	2,9	0,0
Evidence	8,6	1,9	10,5	3,8	5,7	17,1	0,0	0,0	5,7	1,0	45,7	7,6
Declaration	4,8	8,6	2,9	9,5	5,7	10,5	0,0	0,0	0,0	1,9	12,4	65,7

4.2. Behavior of French listeners – level 0

According to a chi-square test, the distribution of answers for all attitudes was above chance ($p < .05$ df:11). By the same test, a significant effect of length was identified between the one and two-mora (32.2 $p < .05$ df:11), and between the five and eight-mora sentences (34.5 $p < .05$ df:11). It was not possible to identify any significant effect of lexical stress for French subjects (39.9 $p > .05$ df:33).

According to the same criterion used for Japanese listeners, the results presented in figure 1 were extracted: *authority*, *irritation* and *admiration* were perceived with no significant confusion. But, the attitude of *arrogance* showed a weak identification score by French listeners. This attitude was confused with *declaration* and *authority*. French listeners did not recognize the two attitudes of politeness particular to Japanese society, namely, *sincerity-politeness* and *kyoshuku*. Moreover, *sincerity-politeness* was confused with *simple-politeness* and *kyoshuku* which represent degrees of politeness. On the contrary, the attitude of *kyoshuku* was recognized as *irritation*, *arrogance* and *authority*. The result that these attitudes of politeness were not well-recognized could be expected, since this prosodic attitude does not exist in French society, nor does the voice quality, especially of *kyoshuku* not match any politeness expression in French. French listeners also confused *interrogation* with *declaration*. This

result shows a possibility for French people to perceive Japanese interrogation as simple declaration. They show significant reciprocal confusions between *declaration* and *evidence*, between *doubt-incredulity* and *exclamation of surprise*, and between *simple-politeness* and *sincerity-politeness*.

Figure 1: Confusion graph for French listeners (level 0): percentages outside the circles indicate the confusion rate. Percentages under the labels of attitude represent the identification rates.

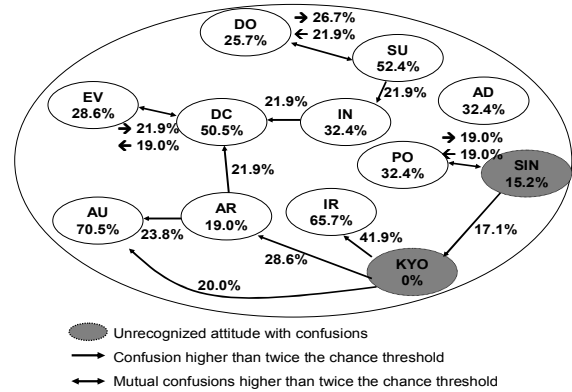
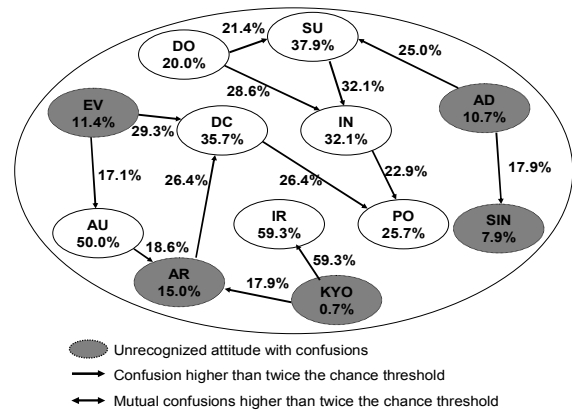


Figure 2: Confusion graph for American listeners (level 0): percentages outside the circles indicate the confusion rate. Percentages under the labels of attitude represent the identification rates.



4.3. Behavior of American listeners – level 0

The distribution of all attitudes was above chance ($p < .05$ df:11) by a chi-square test. A significant effect of length was identified only between the two and five-mora sentences (35.88 $p < .05$ df:11). No significant effect of lexical stress was identified for American subjects (40.75 $p < .05$ df:33).

The results showed that American listeners have a globally lower recognition score compared to French listeners. According to the confusion graph

(figure 2.), American listeners, like French listeners, did not recognize the two expressions of Japanese politeness (*kyoshuku* and *sincerity-politeness*): *kyoshuku* was recognized as *arrogance* or *irritation*. However, they did not recognize *sincerity-politeness* as an expression of politeness, whereas French listeners did. Americans recognized *arrogance* as *declaration*, and *evidence* as *authority* or *declaration*.

Unlike French listeners who recognized *admiration* well without any confusion, American listeners did not recognize this attitude but confused it with *surprise* and *sincerity-politeness*.

5. CONCLUSION

According to a chi-square test, the distribution of attitudes is above chance for main attitudes, even for French and American listeners who have no skill in Japanese. By the same statistical test, a length effect can be observed for all three groups, but in a different way (between two and five-moras, and also five and eight-moras sentences for Japanese listeners, then between one and two-mora, and also between five and eight-mora sentences for French listeners, and between two and five-mora sentences for American listeners). No effect of lexical stress was observed for any of the three groups of listeners.

Perceptual behaviour of the three groups revealed some curious perceptual differences between Japanese listeners vs. occidental listeners. The two Japanese cultural expressions of politeness (i.e. *sincerity-politeness* and *kyoshuku*) were confused within the politeness class for Japanese listeners; however, one of them (*kyoshuku*) was absolutely not recognized as a politeness expression by the occidental listeners. *Kyoshuku* is difficult to identify, even for fluent Japanese speakers as reported in [9]. This weak recognition rate for occidental subjects might come from the absence of this concept in occidental society, but may more surely come from the particular voice quality used by *kyoshuku*, a vocal fry/creak, that conveys a completely different signification (or intention) for French and American speakers. This is typically what we call a “false friend,” that is, an expression whose similar acoustic characteristics lead non-native listeners to understand it according to their own language, but actually is different from what was intended by the native speaker.

In addition, only French listeners showed confusion between *interrogation* and *declaration*.

[9] compared F0 contour and intensity of these two attitudes in Japanese and French in order to find a cause of such a phenomenon but no similarity either between Japanese interrogation and declaration, nor between Japanese and French interrogation/declaration prototypes could be observed. This point must be studied with more stimuli in order to identify the causes (global contours vs. local stress) of ambiguities.

A perception test for Japanese listeners using a gating paradigm is under way, to ascertain how early in the utterance listeners can predict attitudinal values and to confirm that attitudinal processing involves global integration [1].

6. ACKNOWLEDGEMENT

We are deeply grateful to T. Sadanobu and M. Nakagawa (Kobe University) for their financial support and their helpful advices on Japanese attitudes. We also thank D. Erickson and students at Black Hills State University, Spearfish, S.D. for the American perception tests. This work was performed as collaboration between Graduate School of Cultural Studies and Humane Science, Kobe university and GIPSA-lab in the frame of the “College Doctoral Franco-Japonais”, directed by French and Japanese foreign offices, and also as a part of Crest-JST “Expressive Speech Processing” Project, directed by Nick Campbell, ATR, Japan.

7. REFERENCES

- [1] Aubergé, V., Grépillat, T., Rilliard, A. 1997. Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. *5th Eurospeech*, 871-874.
- [2] Campbell, N. 2004. Perception of Affect in Speech – towards an Automatic Processing of Paralinguistic Information in Pro. *8th ICSLP*, Jeju, Korea, 881-884.
- [3] Erickson, D., Ohashi, S., Makita, S., Kajimoto, N., Mokhtari, P. 2003. Perception of naturally-spoken expressive speech by American English and Japanese listeners. *CREST International Workshop on Expressive Speech Processing*, 31-36.
- [4] Japanese accent dictionary 1960. Hirata, T. (eds), Tokyo-do shuppan (in Japanese).
- [5] Maekawa, K., 1998. Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *ICSLP*, 635-638.
- [6] Mizutani O., Mizutani N., 1979. *Aural Comprehension Practice in Japanese*. The Japan Times.
- [7] Ofuka, E., McKeown, J.D., Waterman, M.G., Roach, P.J. 2000. Prosodic cue for rated politeness in Japanese speech. *Speech Communication*, 32, 199-217.
- [8] Sadanobu, T. 2004. A natural history of Japanese pressed voice. *J. of the Phonetic Society of Japan* 8 (1), 29-44.
- [9] Shochi, T., Aubergé, V., Rilliard, A., 2006. How prosodic attitudes can be false friends: Japanese vs. French social affects. *Speech Prosody*, Dresden, 692-696.

How prosodic attitudes can be false friends: Japanese vs. French social affects

Takaaki Shochi, Véronique Aubergé, Albert Rilliard

Institut de la Communication Parlée, UMR CNRS 5009, Grenoble, France

{Takaaki.Shochi, Veronique.Auberge, Albert.Rilliard}@icp.inpg.fr

Abstract

The attitudes expressed by a speaker during a verbal interaction are linked to what the speaker intends to say, and are much influenced by the speaker's language and the culture. To a large part, they are voluntarily controlled. We investigated 7 affectively-neutral Japanese sentences as uttered with 12 different attitudes. The listeners were 15 Japanese listeners, 15 French listeners who had no Japanese language skill, and 32 French listeners learning Japanese (lower level and higher level). They were asked to choose the speaker's intended attitudes among the 12 attitudes. Results showed that Japanese listeners recognized all attitudes above chance, but there were some confusions, especially for the expressions of politeness (i.e. "sincerity-politeness" vs. "kyoshuku"). Moreover, "kyoshuku", a type of politeness that does not occur in French society, was incorrectly decoded by French listeners, and was often identified as impoliteness, even by listeners who had a fairly good command of Japanese. French listeners with some Japanese language skill could well recognize admiration, authority and irritation, but not interrogation vs. declaration, except by those listeners in the higher level group (i.e. level 2).

1. Introduction

The affects in speech are expressed following different cognitive processing levels, from involuntary controlled expressions to the intentionally, voluntary, control of the attitudes of the speaker. "Attitudes" (as "labels" or "notions") are sometimes confused with "Emotions" (as "labels" or "notions") since both use specialized prosodic encoding in the acoustic signal. But if the prosodic emotional expressions are carried in parallel with the prosodic linguistic structures, then we can view expressions of attitudes as an integral part of "language interaction building" (i.e. "communication"). Some affects such as "surprise", are classified either as attitudes or emotions. Our position is to assume that "surprise" can be an emotion if it is involuntary processed, and can be an attitude when it is voluntary processed, in which case, it needs to be learned. When a speaker does not produce any attitude on an utterance, it must be considered as a special attitude: that this speaker has the intention to not express any information about his intentions, that is the "no attitude" attitude. It is why we consider that studying affects in speech interaction are crucial for the expressive speech modelling, and we need to describe precisely what can be the attitudes of a speaker in a language and a culture and how they are linguistically encoded. A cross-cultural study is a helpful method to examine this problem. As for all language specifications, attitudes can be expected to be universal in part, both in terms of what is expressed and also in how it is expressed (for instance, *authority*, *surprise*). Because attitudes are constructed socially for and by the language, they can exist or not in one language but not another, and

moreover, prosodic realization of one specific attitude in a specific language may not be recognized or may be ambiguous in the language the speaker is learning.

After presenting the Japanese corpus on which is based this paper, we summarize different perceptual experiments. The first is the validation of the Japanese attitudes by native Japanese listeners. Then we present (1) how French learners of Japanese (level 0) perceive the Japanese attitudes (2) what French learners of Japanese, (level 1) learned about Japanese attitudes from their language learning process (3) what kinds of attitudes French learners of Japanese in fluent level (level 2) still cannot identify despite their language skill.

2. Selection of 12 Japanese Attitudes

We have selected a set of 12 attitudes for Japanese, supposed to be representative from the literacy, especially from the Japanese teaching methods: "doubt-incredulity", "evidence", "exclamation of surprise", "authority", "irritation", "arrogance-impoliteness", "sincerity-politeness", "admiration", "kyoshuku", "simple-politeness", "declaration" and "interrogation". Several of these attitudes are specific or specifically important to the Japanese culture, especially those linked to the Japanese politeness strategy; "simple politeness", "sincerity-politeness" and "kyoshuku" vs. "arrogance-impoliteness". The attitude of "sincerity-politeness" appears when a speaker who is considered to be inferior is talking to his interlocutor who is considered to be superior in Japanese society. The speaker expresses that his intention is serious and sincere by using this prosodic attitude. The attitude of "kyoshuku" (there is no lexical entry to translate this into English) is a typical Japanese cultural attitude: even if such situations occur in all cultures, the Japanese language has chosen to specially encode this situation as an attitude ("attitudineme"). A speaker uses this when he wants to express his contradictory opinion on a situation in which his social status is inferior to that of his interlocutor, not with the aim of disturbing his superior but to help him, or when the speaker is willing to get a favour from his superior. It is described by T. Sadanobu [11] as "*a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker's consciousness of the fact that his/her utterance of request imposes a burden to the hearer*" (Sadanobu [11] p.34).

3. The corpus

Since our aim is to measure the perceptive behavior of French listeners, we need reference data about Japanese attitudes. These utterances must be free of lexico-syntactic information about attitudes (only prosodic attitudinal information can be used). The other prosodic variations must be balanced, in order to control and measure the possible bias of the Japanese

intonation or lexical stress, which could be interpreted by French listeners as cues for some attitudes. The first step is thus the recording of a very controlled corpus, constructed according to a few structure principles. The construction of opposed minimal pairs allows us to observe the effect of the targeted factor only. On the basis of such controlled corpora, an acoustic analysis leads to a statistical model of prosodic variations, which can be used to synthesize the captured prosodic variations and then in turn be further explored by using analysis-resynthesis techniques. In order to validate this work, a compared analysis of the natural prosody contained in the corpus *vs.* that of the synthetic prosody extracted by analysis will be done in order to define a description of the observed prosody.

The corpus is based on seven sentences ranging from one to eight moras. The syntactic structure of the sentences used in the corpus is either a single word, or a simple verb-object structure. For the eight-mora utterances, the position of lexical accent may be on the first, second, and third mora, or absent. In order to express some attitudes like “doubt” or “surprise”, the vowel [u] may be inserted at phrase final position, and lexical accent will be realized at the seventh mora, too. The sentences were constructed in order to have no particular connotations in any region of Japan. Each sentence is produced with all the attitudinal functions.

The sentences were constructed in order to have no particular connotations in any region of Japan. Each sentence is produced with all the attitudinal functions. One male Japanese native language teacher produced recordings of all the attitudes for this corpus. We already mentioned that the attitudes are socially constructed for each language. Therefore foreign language learners have to learn prosodic attitudes corresponding to the target language and culture, and languages teachers are able to elicit different types of attitudes for didactic/pragmatic reasons. The complete corpus contains 84 stimuli, i.e. seven utterances produced with twelve different attitudes. All stimuli are used for the perception test.

Table 1. Corpus of Japanese attitudes: 7 utterances of different length with different position of the lexical accent, which is marked with a star, each produced with the 12 different attitudes.

Nb mora	Utterance	Translation
1	Me	The eye
2	Na*ra	Nara
5 (3+2)	Na*rade neru	He sleeps in Nara
8 (4+4)	Na*goyade nomimas	He drinks in Nagoya
8 (4+4)	Nara*shide nomimas	He drinks in Nara Town
8 (4+4)	Matsuri*de nomimas	He drinks at the party
8 (4+4)	Naniwade nomimas	He drinks at Naniwa

4. Experimental protocol

The first step is the validation of the corpus attitudes by Japanese native listeners. 15 Japanese listeners (11 females and 4 males) who speak the Tokyo dialect, whose mean age is 29.5, had to choose attitudes in a forced choice of the 12 attitudes. For the Japanese test interface, each label is described and illustrated by examples of situations in which such an attitude can happen. French was chosen since it is culturally distant from the Japanese culture and language. - A first experiment was held with 15 French listeners (10 females and 5 males) naïve in Japanese, that is who have never heard the Japanese language. They are the “level 0” group. The

mean age is 25.4. Listeners do not report any listening disorder. The French test interface proposes a French translation of each label, introduced by a long explanation about each attitude, giving many examples of relevant situations. No subject expresses any trouble to “understand” the attitude label. - A second experiment was held with 16 French listeners, learning the Japanese language, and evaluated to be in the same coherent level (labelled level 1 in our Japanese courses at Grenoble University), that is a beginner level: they can speak and understand Japanese, but are still not fluent. They use the same test interface. - A third experiment was held with 16 French listeners, evaluated to be in the same level 2, that is that they are quite fluent in Japanese. They use the same test interface. All the subjects of these experiments listen to each stimulus one time only. For each stimulus, they were asked to answer the perceived attitude among twelve possible attitudes. The presentation order of the stimuli was randomized in a different order for each subject.

5. Results

5.1. Validation with Japanese Listeners

According to a chi-square test, all attitudes were above the chance level. Then, we tested a possible effect of the stimuli length for the distributions of selected attitudes. The results show a significant difference of the length between two and five-moras sentences and also between the five and eight-moras one. But, there is no effect of the lexical accent.

In order to determine which attitude listeners recognized over chance, a criterion was used: the mean identification rate must be over twice the theoretical chance level.

According to this criterion, seven attitudes (i.e. “*arrogance-impoliteness*”, “*declaration*”, “*doubt-incredulity*”, “*simple-politeness*”, “*exclamation of surprise*”, “*irritation*” and “*interrogation*”) have been recognized without any particular confusion.

“*Authority*” was confused with “*evidence*”. One possible explanation is that, concerning the self-confidence of speaker, these two attitudes are, in fact, similar: when imposing authority, the speaker is certainly sure of himself.

“*Evidence*” was confused with “*arrogance-impoliteness*”. “*Evidence*” shows that speaker is confident of himself, and this expression of certainty can sometimes be perceived like disrespect to interlocutor.

Curiously, two typical Japanese attitudes like “*sincerity-politeness*” and “*kyoshuku*” were confused with each other. These two attitudes express essentially the humility of a speaker facing a superior person in the social hierarchy. It is important to note that “*sincerity-politeness*” was also confused with “*simple-politeness*”, whereas this confusion with “*simple-politeness*” was absent for “*kyoshuku*”.

Concerning the attitude of “*admiration*”, we observed confusion with “*simple-politeness*”. These two attitudes are interconnected each other in the Japanese society. This evidence can be explained by the lexical polysemy of items like “*sonkee*” [admiration / politeness], and “*keifuku*” [admiration / politeness]

Table 2. Attitudes recognition and confusion percentages for 15 Japanese listeners. The values in the diagonal cells of the confusion matrix, which are recognized above chance level (over 16.6%), are in bold. Attitudes which showed significant confusion are in italic bold.

Perceived attitudes	Presented attitudes											
	AD	PO	KYO	SIN	AR	AU	IR	DO	EX	QS	EV	DC
Admiration	21,9	4,8	1,9	3,8	0,0	0,0	0,0	1,9	1,0	0,0	0,0	0,0
Politeness	26,7	64,8	2,9	17,1	1,0	4,8	0,0	0,0	0,0	8,6	1,0	11,4
Kyoshuku	11,4	9,5	24,8	27,6	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0
Sincerity	14,3	6,7	26,7	32,4	0,0	5,7	0,0	0,0	0,0	1,9	0,0	1,0
Arrogance	1,9	0,0	1,9	0,0	72,4	5,7	2,9	5,7	0,0	1,0	21,0	10,5
Authority	0,0	0,0	15,2	1,0	9,5	51,4	11,4	1,0	1,0	1,9	9,5	2,9
Irritation	1,0	0,0	12,4	1,0	5,7	4,8	85,7	13,3	4,8	1,0	2,9	1,0
Doubt	0,0	1,0	0,0	0,0	0,0	0,0	0,0	56,2	14,3	3,8	0,0	0,0
Surprise	9,5	1,9	1,0	1,9	0,0	0,0	0,0	14,3	59,0	1,0	4,8	0,0
Interrogation	0,0	1,0	0,0	1,9	0,0	0,0	0,0	7,6	14,3	77,1	2,9	0,0
Evidence	8,6	1,9	10,5	3,8	5,7	17,1	0,0	0,0	5,7	1,0	45,7	7,6
Declaration	4,8	8,6	2,9	9,5	5,7	10,5	0,0	0,0	0,0	1,9	12,4	65,7

5.2. Behavior of the French Listeners - level 0

The distribution of all attitudes was above of chance. A significant effect of the length was identified between the one and two-moras sentences. It was not possible to identify any significant effect of the lexical stress for French subjects. According to the same criterion for the Japanese listeners, the following results were extracted. Figure 1. shows that “authority”, “irritation” and “admiration” have been perceived with no significant confusion according to our criteria. But, attitude of “arrogance-impoliteness” showed weak identification score by French listeners. This attitude was confused with “declaration” and “authority”. French listeners did not recognize two particular attitudes of politeness connected in Japanese society like “sincerity-politeness” and “kyoshuku”. Moreover, “Sincerity-politeness” was confused with “simple-politeness” and “kyoshuku” which are other degrees of politeness. On the contrary, attitude of “kyoshuku” was recognized like “irritation”, “arrogance-impoliteness” and “authority”. This result was expected, since these attitudes do not exist at all in the French society, nor such voice quality does not match any politeness expression in French.

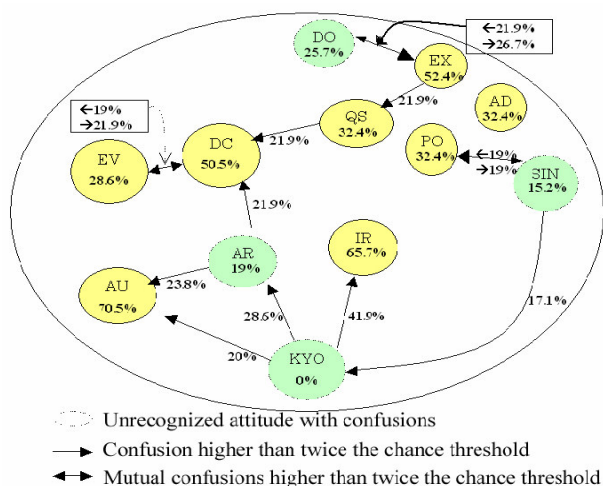


Figure 1. Confusion graph for 15 French listeners (level 0): percentages outside the circles indicate the confusion rate in percentages. Other percentages written under the labels of each attitude represent the identification rates of attitude in percentages. NOTE: AD(admiration), AR(arrogance-impoliteness), AU(authority), DC(declaration), DO(doubt-incredulity), EV(evidence), EX(exclamation of surprise), IR(irritation), KYO(kyoshuku), PO(politeness), QS(Interrogation) and SIN(sincerity-politeness)

French listeners confused also “interrogation” with “declaration”. This result shows a possibility for French people to perceive Japanese interrogation as simple declaration. They show significant reciprocal confusions between “declaration” and “evidence”, between “doubt-incredulity” and “exclamation of surprise”, and between “simple-politeness” and “sincerity-politeness”.

5.3. Behavior of the French Listeners - level 1

At this level of Japanese (beginners), the subjects have learned to identify “sincerity-politeness” and “doubt-incredulity”. They have changed some confusion and misinterpret. “Sincerity-politeness” was confused with “declaration”, and “arrogance-impoliteness” was also confused with “evidence”. There is still mutual confusion between “doubt-incredulity” and “exclamation of surprise”. They have learned to discriminate “arrogance-impoliteness” vs. “authority”, “simple-politeness” vs. “sincerity-politeness”, “declaration” vs. “evidence”. However they still confused “kyoshuku” with “irritation”, “arrogance-impoliteness” and “authority”, they confused also “arrogance-impoliteness” with “declaration”, and mainly “interrogation” with “declaration”, that can be a strong communicative handicap for such subjects starting to interact in Japanese.

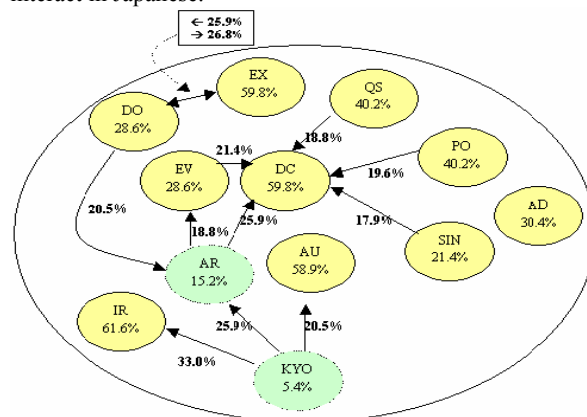


Figure 2. Confusion graph for 16 French listeners (level 1): percentages outside the circles indicate the confusion rate in percentages. Other percentages written under the labels of each attitude represent the identification rates of attitude in percentages. NOTE: AD(admiration), AR(arrogance-impoliteness), AU(authority), DC(declaration), DO(doubt-incredulity), EV(evidence), EX(exclamation of surprise), IR(irritation), KYO(kyoshuku), PO(politeness), QS(Interrogation) and SIN(sincerity-politeness)

5.4. Behavior of the French Listeners - level 2

At this level, where the French subjects start to be fluent in Japanese, they did not learn to identify any other attitudes. “arrogance-impoliteness” and “kyoshuku” are still not recognized. They however learned at least to discriminate “interrogation” and “declaration”, but “kyoshuku” is still confused with “irritation”, “arrogance-impoliteness” and “authority”, and “arrogance-impoliteness” with “declaration” and “evidence”. They have changed confusion of “declaration” with “sincerity-politeness”. “Doubt-incredulity” and “exclamation of surprise” are still mutually confused (which is not the case for Japanese listeners, and not the case for French in French “doubt” and “surprise” [1])

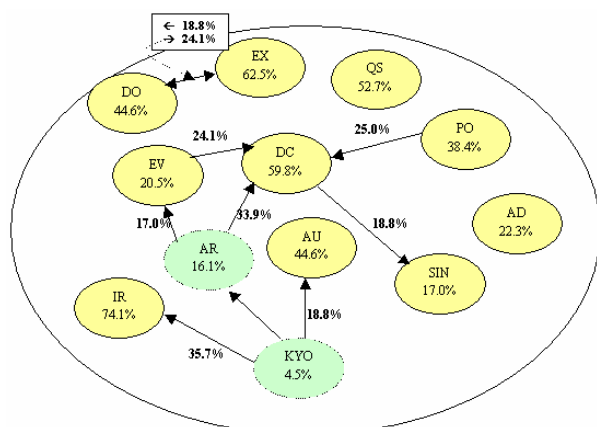


Figure 3. Confusion graph for 16 French listeners (level 2): percentages outside the circles indicate the confusion rate in percentages. Other percentages written under the labels of each attitude represent the identification rates of attitude in percentages. NOTE: AD(admiration), AR(arrogance-impoliteness), AU(authority), DC(declaration), DO(doubt-incredulity), EV(evidence), EX(exclamation of surprise), IR(irritation), KYO(kyoshuku), PO(politeness), QS(Interrogation) and SIN(sincerity-politeness)

6. Conclusion

The distribution of attitudes is above chance for main attitudes, even for French listeners who have any skill in Japanese. A length effect can be observed between one and two-mora sentences for French listeners (level 0), and between two and five-mora, and also five and eight-mora sentences for Japanese listeners. The effect of lexical accent was not observed for both groups of listeners. We analyzed the F0, intensity and duration contour of declaration and question in order to understand if such a stress effect could explain the confusion between question and declaration by French listeners, which disappears at level 2. But no peak similarity neither between some stressed Japanese question and declaration, nor between Japanese and French question/declaration prototypes, could be observed. This point must be studied with more stimuli in order to identify the causes (global contours vs. local stress) of ambiguities. The Japanese subjects showed confusions of three politeness levels inside the politeness class. “Sincerity-politeness” and “kyoshuku” are confused each other. “Sincerity-politeness” is also confused over chance with “simple politeness” which is well identified. The “evidence” attitude is confused over chance with “arrogance-impoliteness” (and “authority” with “evidence”) for Japanese listeners. On the contrary, the concept of degree of politeness is not processed by French subjects, even for fluent Japanese speakers: they recognize the “simple-politeness” (as well as “admiration”, “authority” and “irritation”), but they do not identify the typical Japanese politeness degrees “kyoshuku” and “sincerity-politeness”. “Sincerity-politeness” is confused with “simple politeness” and “kyoshuku”, then “simple politeness” is confused with “sincerity-politeness”. This weak recognition might come from the absence of this concept in French society, on one hand, but also, on the other hand, this particular voice quality which expresses this Japanese politeness may manifest a completely different signification (or intention) for French native speakers. It is to be noted that “kyoshuku” is confused by French listeners with “arrogance-impoliteness”, and also with “authority” and “irritation”. This is typically what we

call a “false friend”. After this result, an acoustic analysis of the corpus should reveal the prosodic characteristics of each attitude. It is important to test confusions due to cross-cultural differences with the stimuli composed of French sentences and superposed Japanese prosodic attitudes.

Some complementary experiments are under way: on one hand we are implementing a perception test for these two groups of listeners using a gating paradigm to envisage if listeners can predict attitudinal values early in the utterance, that is if this processing is confirmed to be a global integration [1], and on the other, we are observing the behaviors of American native subjects from level 0 to 3 in Japanese, in parallel to the study of Japanese subjects from level 0 to 3 in French and English.

7. Acknowledgment

We especially thank T. Sadanobu, from Kobe University, for his help in studying Japanese attitudes. This work was held as a part of the Crest-JST “Expressive Speech Processing” Project, directed by Nick Campbell, ATR, Japan.

8. References

- [1] Aubergé, V., Grépillat, T., Riiliard, A.: Can we perceive attitudes before the end of sentences ? The gating paradigm for prosodic contours. In 5th European Conference On Speech Communication And Technology, Vol.2, Greece (1997) 871-874.
- [2] Ayuzawa, T.: Nihongo no gimonbun no inritsuteki tokucyou. In Nihongo no inritsu ni mirareru bogo no kansyou(2), Grand-in-aid for Scientific research on Priority areas (D1), research report 1992 (1992) 1-20.
- [3] Campbell, N.: Modelling Affect in Speech Communication, Beijing (2003).
- [4] Diaferia, M.L.: Les Attitudes de l'Anglais : Premiers Indices Prosodiques. Mémoire de DEA en Science Cognitives, Institut National Polytechnique de Grenoble France (2002).
- [5] Erickson, D., Ohashi, S., Makita, S., Kajimoto, N., Mokhtari, P.: Perception of naturally-spoken expressive speech by American English and Japanese listeners. In CREST International Workshop on Expressive Speech Processing (2003) 31-36.
- [6] Ito, M.: The Contribution of Voice Quality to Politeness in Japanese. In VOQUAL'03, Geneva (2003) 157-162
- [7] Ko, M.: Teinei hyougen ni mirareru nihongo onsei no inritsuteki tokucyou. In The Phonetic Society of Japan 1993 Annual Convention (1993) 35-40.
- [8] Matsumoto, E., Sadanobu, T.: Nihongo no inritsu niokeru rikimi to, nihongo gakushuusha no rikaido. In: Departement of Japanese Studies, The Chinese University of Hong Kong and Society of Japanese Language Education (eds.): Quality Japanese Studies and Japanese Language Education in Kanji-Using Areas in the New Century, Hong Kong Himawari Publishing Co. (2001) 455-461.
- [9] Morlec, Y., G. Bailly, and V. Aubergé *Generating prosodic attitudes in French: data, model and evaluation*. Speech Communication, (2001) 33(4): p. 357--371.
- [10] Ofuka, E., McKeown, J.D., Waterman, M.G., Roach, P.J.: Prosodic cue for rated politeness in Japanese speech. In Speech Communication, 32 (2000) 199-217.
- [11] Sadanobu, T.: A natural history of Japanese pressed voice. In Journal of the Phonetic Society of Japan, Vol.8. No.1 (2004) 29-44.
- [12] Van Bezooijen, R.: Sociocultural aspects of pitch differences between Japanese and Dutch women. In Language and Speech 38, 3 (1995) 253-266

発話態度の文化的特性と「偽の友達」

——日仏語の対照研究を通して——

昇地 崇明（フランス GIPSA-lab）・Aubergé Véronique（フランス国

立科学研究所）・Rilliard Albert（フランス国立科学研究所）

1. はじめに

人間は属する社会の中で、様々な人々と接しながら喜怒哀楽が体からこみ上げてきたり、または意図的に喜怒哀楽を表現したりしながらコミュニケーションをとっている。このように人間の「気持ち」は様々な場面で現れるが、前者のように、人間の意志に関わりなく、外的要因（環境や経験）によって引き起こされる人間の「気持ち」は話者が制御できないものである。その典型的な例は、緊張したときに声や手が震えたり、汗を掻いたりする、などであろう。一方、後者のように具体的な（または抽象的な）意図を（一人の、または複数の）相手に伝える際の話者の「気持ち」は、上に述べたような抑え切れない「気持ち」とは異なったものである。具体的に言えば、母親がわざと怒ったフリをして、子供をしかる時や、社会的立場が上の人に対して敬意の気持ちを伝えたりするとき、人間は属する社会の慣習、もしくは話者が得た体験から獲得したストラテジーを用いて「気持ち」を伝える。

このような2つの「気持ち」は表出の仕方が生理的にも心理的にも異なっていることが近年の研究から明らかになっている（Ohala 1984, Damasio 1994, Sherer 1986, Auberge 2002 他）。本稿では、抑え切

本研究を行うにあたり、貴重なご助言を下さった神戸大学の定延利之教授、ATRの Dr. Nick Campbell、昭和音楽大学の Dr. Donna Erickson、国立国語研究所の Dr. Caroline Menezes に感謝の意を表する。

また、SPOT を実施するにあたり、資料の提供及び、ご助言をくださった筑波大学小林典子教授にも深い感謝の意を表したい。

れない「気持ち」を「感情」(emotion)，伝える「気持ち」を「態度」(attitude)とし，言語コミュニケーションにおいて大変重要な役割を担う「態度」に対する考察に絞って議論を進める¹。

2. 発話態度

2.1. 声によって表される話者の態度

話者の態度である「苛立ち」や「疑念」は，ジェスチャーや体の向き，相手との距離，顔の表情，視線，声などさまざまな媒体を通して表現される。この中でも，声によって態度を伝える場合，語彙によって態度を明確に表現することもあるが，実際のコミュニケーションでは「韻律 (Prosody)」と呼ばれるものが深く関わってくる。この「韻律」とは，レキシカルなもの（例えば高低アクセントによって区別される橋と箸の違い等）より上位のレベルで音の長さ・高さ・強さ・声質の時間的变化により表れる現象を指し，コミュニケーション上大変重要な役割を担っている。例えば「ありがとう」という感謝の意味を表す文であっても「皮肉」の態度を韻律によって表現すれば，聞こえてくる言葉の意味とは裏腹に，聞き手は話者の「皮肉」を聞き取る。

このように韻律に話者の態度が表れることは日本語に限ったことではない。スコットランド英語ではクリーキーな声は冷静さ，又は安心を与えたいといった態度を示す際に表れ，タミル語では普段よりピッチの高い声で話すことで相手に対する敬意の念を示す態度を表すことが報告されている (Brown&Levinson 1986)。

このような韻律によって表される態度（以下，発話態度とする）は，言語獲得の比較的遅い段階²で獲得される (Clément 1999) ことが報告されており，表現される様々な発話態度は，精神及び身体の発達に伴

¹この「態度」はその話者のキャラクタと密接に結びついている。例えば，キザなキャラクタの人は緊張した状況であっても，頭を搔いておどけたりする態度は示さない。つまり，コンテキストに応じて話者が選択する態度は話者のキャラクタによって限定されると思われる。

²フランス語のいくつかの発話態度は 7～11 歳で獲得されることが報告されている。

って、属する社会文化から要求され、比較的長い時間をかけて学習していくものであると推測できる。

近年、ヨーロッパの諸言語、英語、日本語、中国語、韓国語の発話態度に関する研究が行われ (Sherer 1986, Maekawa 1998, Erickson et al. 2003 他), 様々な発話態度の韻律的特徴が明らかになってきている。

2.2. Cross cultural research の重要性

先行研究から、各言語文化の様々な発話態度は、その言語文化を共有する人々の間で声からの情報だけでも正しく認識されることがわかった。一方、このように母語話者の間で共通に認識される発話態度も、同じ言語文化を共有しない非母語話者には正しく認識されない可能性がある。このような異文化・非母語環境ではコミュニケーション上の障害や話者と同一文化を共有しない聞き手の怒りや混乱を招く恐れがある。その原因は、各言語文化固有の態度が非母語話者に正しく認識されていないことに起因している (Shochi et al. 2005, 松本・定延 2001) と思われる。また、言語間で共有される発話態度であっても言語によって生成方法が異なっており、結果として異なった発話態度として認識される可能性も考えられる (Sherer et al. 2001, エリクソン・昇地 2006 他)。

2.3. 偽の友達 (False friends)

非母語話者に異なった意味として認識される発話態度は「偽の友達 (False friends)」である可能性が考えられる。この「偽の友達」とはヨーロッパ言語間の表記と意味のズレを指す場合によく使用される用語である。例えば、英語の「Deception」という語の意味は「うそ」という意味であるが、同じ表記であるフランス語の「Déception」の意味は「失望」という意味である。このような事象と全く同じことが発話態度にも存在すると思われる。仮に、日本語とフランス語共に「驚き」の発話態度が存在したとする。しかしながら、日本語の「驚き」を多くのフランス語母語話者が「驚き」と認識せずに別の発話態度と判断したとき、両言語の「驚き」の発話態度は「偽の友達」であると

言える。このような「偽の友達」が見られるケースは言語コミュニケーション上最も注意しなければならないと思われる。なぜなら、発話者の伝えたい意図とは裏腹に、発話の聞き手は聞き取った韻律に母語の発話意図を知覚してしまい、コミュニケーション上の大きな障害となる恐れがあるからである。

2.4. 日仏語を対象とした発話態度の対照研究

本稿では、上記したような「偽の友達」が文化間で存在するのかを調査するため、フランス語と日本語を対象言語とし、特に日本語の発話態度に見られるフランス語母語話者と日本語母語話者の認識のズレについて調査を行った。具体的には、10の発話態度（感心、苛立ち、高圧的、丁寧、自明（evidence）、恐縮、誠意、驚き、疑念、ぞんざい）に、特別な意図がないことを示す態度である Yes-No 疑問と平叙を加えたものを両言語母語話者に聞かせた。この結果から、母語話者と非母語話者との間で、どのような認識のズレが見られるのかについて考察し、日本語の発話態度でフランス語母語話者にとって「偽の友達」になり得るものがあるのかどうかを調べた。さらに、同様の聴覚実験をフランス在住のフランス人日本語学習者を対象に行い、日本語学習経験のないフランス語母語話者の結果とどの程度異なっているのかを調べることで、日本語習得が発話態度の知覚に影響を与えているのかについても考察した。

3. 実験資料

3.1. 発話態度のラベルと定義

聴覚実験に先立って、韻律によって表現される態度（「感心(AD)」 「ぞんざい(AR)」 「高圧的(AU)」 「疑念(DO)」 「自明(EV)」 「驚き(EX)」 「苛立ち(IR)」 「恐縮(KYO)」 「丁寧(PO)」 「誠意(SIN)」）に加え、明確な意図を持たないことを示す態度である「平叙(DC)」と「YES-NO 疑問(QS)」を併せた 12 の発話態度³を選定した。各発話態度の定

³ 発話態度の後ろの括弧内の表記は便宜上用いた発話態度のラベルを表す。

義については先行研究を参考に、以下のように設定した。

感心 (admiration) : 見たり聞いたり味わったりしたことについて感心する際用いる発話態度

ぞんざい (arrogance) : 相手に対して失礼な態度を示す際用いる発話態度

高圧的 (authority) : 相手に対して高圧的であろうとする際用いる発話態度

疑念(doubt) : コミュニケーションの際、発話相手の言ったこと、もしくは行動に対して否定的な考えを持ち、発話相手に対して懐疑的な態度を示すとき用いる発話態度

自明(evidence) : 発話者にとって自分が今から言おうとしていることは言うまでもない当然のことなのだという態度を示すとき用いる発話態度

驚き(surprise) : コミュニケーションの際、相手の言ったこと、もしくは行動を予測しておらず、その言動に対して意外だという驚きを示す際用いる発話態度

苛立ち (irritation) : 相手の言動に対して不満および苛立ちを感じ、その苛立ちの態度を示す際用いる発話態度

丁寧(politeness) : 相手の社会的地位、年齢に関係なく相手に対して丁寧に話そうとする際用いる発話態度

誠意(sincerity) : 社会的立場が話者より上の人に対して、誠意を表明する際用いる発話態度

恐縮(kyoshuku) : 社会的立場が話者より上の人に対してお願いをするときなどに、「こんなことを言うのは申し訳ないのですが…」という恐縮した態度を示すとき用いる発話態度

3.2. 発話者と音声資料

上記の発話態度を発話するにあたって、まず発話文を作成した。この実験資料は統語構造などの言語学的情報や音声的条件が完全にコントロールされており、観察対象以外のものは完全に制限されたものと

なっている。具体的には，可能な限り有声音で構成された，長さとアクセント核の位置の異なる 7 つの文を作成した（表 1 参照）

表 1 実験資料 1.2.5.8 モーラでそれぞれ構成されている。表中のアスタリスクはアクセント核を表す。アクセント核の位置については平山（1960）を参照した。

モーラ数	文	音韻環境
1	目	/me/
2	奈良	/na*ra/
5 (3+2)	名古屋で寝る	/na*goyadeneru/
8 (4+4)	名古屋で飲みます	/na*goyadenomima*su/
8 (4+4)	奈良市で飲みます	/nara*shidenomima*su/
8 (4+4)	祭りで飲みます	/matsuri*denomima*su/
8 (4+4)	「ナニワ」で飲みます	/naniwadenomima*su/

発話者は，声優ではなく日本語母語話者の日本語教師である。Campbell（2000）で指摘されているように声優によって生成される感情音声は日常の生活の中で使用される感情音声とは異なったものである可能性が高い。一方，語学教師は様々なコンテクストに応じたスピーチアクトの指導を行っており，教師の提示するモデルは声優の発話のような誇張されたものではなく，それが目標言語の社会において理解されなければならない点に重点をおいているので，現実的日常会話に見られる発話により近いものであると思われる。

3.3. 録音

録音は無響録音室で行った。発話者には前もって各態度の定義を提示し，理解したことを確認してから，用意した 7 文を 12 通りの態度で発話してもらった。当然のことながら，一つの態度を表現する方法は多様であると思われるが，ここでは発話者にとってもっとも自然であると思われる方法で発話してもらった。なお，本実験では Sony 製 DATTCO-D7 リコーダーに Sennheiser 製 HSP-2 ヘッドセットマイクを接続して行い，サンプリングレート 44.1kHz ステレオでデジタル録音した。

4. 聴覚実験

4.1. 被験者

本稿で対象とする被験者は日本語母語話者（東京方言話者）15名、日本語に関する知識を全く持たないフランス語母語話者（以下レベル0）15名、フランス人日本語学習者32名である。

4.2. 刺激音の選定

収録した発話から84の刺激音（7種類の文×前述した12の発話態度）を選定し、フランス滞在歴1年未満の日本人留学生5名にパイロット実験を行った結果、発話者の発話態度の認識率に有意差がみられたので、これらをそれぞれの発話態度のステレオタイプ⁴とし、次節以降述べる聴覚実験の音声資料とした。

4.3. 実験方法

84の刺激音を被験者に聞いてもらい、常に提示されている各発話態度の定義を参考に、12の発話態度から1つを選択してもらう方法を取った。なお、前述の発話態度のラベルと定義は日本人被験者へ提示する為のものであり、フランス人被験者にはそのフランス語訳を提示した。ただし、「恐縮」のラベルについては、フランス語で対応する語が存在しないため、ローマ字で「Kyoshuku」とそのまま提示した。各発話態度の定義についても、「恐縮」と「誠意」の2つの発話態度については以下のような補足説明を加えた。まず、「誠意」については「社会的立場が話者より上の人に対して、話者の伝えたい意図に誠意・真剣さが込められている際現れるものである」と定義した。「恐縮」については、Sadanobu (2004) に倣い「社会的立場が話者より上

⁴ ある態度を声で表現する際、その表現の「仕方」は様々である。したがって、態度とその表現方法の関係は必ずしも1対1ではない。この観点から、本実験で選択した発話態度は「プロトタイプ」ではなく、むしろ「ステレオタイプ」と認定した。

の人に対して，異なった意見を伝えようとする際，又はお願いをする際，失礼になることがないよう羞恥の苦しみと困惑を伝えるとき現れるものである⁵」と定義した。

聴覚実験では刺激音の提示順序はランダムに設定しており，被験者ごとに異なったものを用いた。さらに，被験者が刺激音に慣れ，学習するのを避けるため被験者が一つの刺激音を聞ける回数は1回だけと設定した。

5. 実験結果

5.1. 日本語母語話者

被験者が12の発話態度をどのように認識したかを示す傾向を考察する前に，日本人被験者が選択した発話態度の分布に有意な傾向が見られるかどうかをカイ二乗検定によってテストした。その結果，選択した発話態度の分布には有意な傾向がみられた。次に，文の長さが発話態度の選択に影響を与えているのかどうかを同じくカイ二乗検定によってテストしたところ，2モーラ文と5モーラ文，そして5モーラ文と8モーラ文の間で有意な文の長さの影響が確認できた。さらに，語

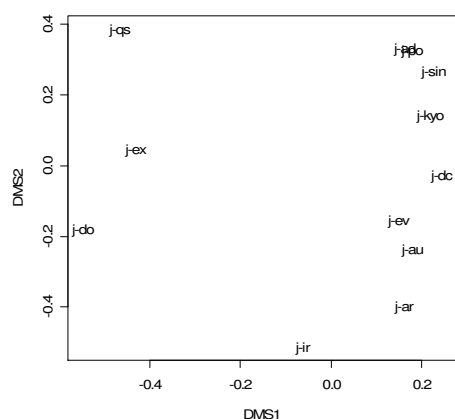


図1 12の発話態度に対する日本語母語話者の知覚行動（j-は日本語母語話者を，後続のラベルは発話態度を指している）

アクセントが発話態度の選択に与える影響もテストしたが，ここでは有意差はみられなかった。このことから本実験において語アクセントは発話態度の判定に影響を与えなかったと考えられる。

次に，日本語母語話者が各発話態度をどのように知覚したのかを調べるために，数量化I V類を用いて2次元の知覚空間を描いた（図1参照）。図1の第1次元（DMS1）は「Yes-No疑問(j-qs)，驚き(j-ex)，疑念(j-do)」とそれ以外の発話態度を分離している。この3つの発話態度

⁵ Sadanobu (2004) p.34 からの引用。原文は英語。

は文末でピッチの上昇が見られ、聞き手の反応を要求している点で他の発話態度とは異なった反応が見られたと推測できる。第2次元(DMS2)は「ポライトネス」の軸と解釈できる。「恐縮(j-kyo), 誠意(j-sin), 丁寧(j-po), 感心(j-ad)」が上部に位置するのに対して, 「高圧的(j-au), ぞんざい(j-ar), 苛立ち(j-ir)」は下部に集中していることが確認できる。

さらに, 日本語母語話者が12の発話態度をどのように認識したかを示す結果(表2参照)を考察したところ, 「ぞんざい」, 「平叙」, 「疑念」, 「驚き」, 「苛立ち」, 「丁寧」, 「Yes-No疑問」の7つの発話態度は他の発話態度と混同することなく認識できていることがわかった⁶。

表2 日本語母語話者15名の発話態度認識率と混同を示した表 横軸は提示された発話態度, 縦軸は被験者が知覚した発話態度をそれぞれ示している。認識率(16.6%)以上の知覚を示した発話態度は太字で, 有意な混同を示したものはイタリック表示している。

Perceived attitudes	Presented attitudes											
	AD	PO	KYO	SIN	AR	AU	IR	DO	EX	QS	EV	DC
Admiration	21,9	4,8	1,9	3,8	0,0	0,0	0,0	1,9	1,0	0,0	0,0	0,0
Politeness	26,7	64,8	2,9	17,1	1,0	4,8	0,0	0,0	0,0	8,6	1,0	11,4
Kyoshuku	11,4	9,5	24,8	27,6	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0
Sincerity	14,3	6,7	26,7	32,4	0,0	5,7	0,0	0,0	0,0	1,9	0,0	1,0
Arrogance	1,9	0,0	1,9	0,0	72,4	5,7	2,9	5,7	0,0	1,0	21,0	10,5
Authority	0,0	0,0	15,2	1,0	9,5	51,4	11,4	1,0	1,0	1,9	9,5	2,9
Irritation	1,0	0,0	12,4	1,0	5,7	4,8	85,7	13,3	4,8	1,0	2,9	1,0
Doubt	0,0	1,0	0,0	0,0	0,0	0,0	0,0	56,2	14,3	3,8	0,0	0,0
Surprise	9,5	1,9	1,0	1,9	0,0	0,0	0,0	14,3	59,0	1,0	4,8	0,0
Interrogation	0,0	1,0	0,0	1,9	0,0	0,0	0,0	7,6	14,3	77,1	2,9	0,0
Evidence	8,6	1,9	10,5	3,8	5,7	17,1	0,0	0,0	5,7	1,0	45,7	7,6
Declaration	4,8	8,6	2,9	9,5	5,7	10,5	0,0	0,0	0,0	1,9	12,4	65,7

「感心」, 「高圧的」, 「自明」, 「恐縮」, 「誠意」の5つの発話態度は他の発話態度との混同を示したものの, 正しく認識できた発話態

⁶本稿では発話態度を認識したとする基準を期待値の2倍である16.6%に設定してある。

度である。「高圧的」な発話態度は「自明」と混同する傾向が見られ、「自明」の発話態度は「ぞんざい」な発話態度と混同を示した。さらに、フランス語に存在しない2つの日本語特有の発話態度である「誠意」と「恐縮」は、この2つの発話態度間で非常に高い率の混同を示した。この2つの発話態度は本質的に話者より社会的立場の高い人に対して敬意を示す際に用いられる発話態度である点で共通していることから、被験者間で混同を起こしたと考えられる。「誠意」は「恐縮」の他に「丁寧」とも混同を示した。しかしながら、「恐縮」と「丁寧」とを混同する有意な結果は見られなかった。

一方、「感心」は「丁寧」との混同を示した。この混同の原因に関しては音響分析の項で詳しく述べるが、この2つの発話態度は共にピッチが高く、テンポがゆっくりしているという点があり、それに起因しているのではないかと思われる。

以上、日本語母語話者の結果では、いくつかの発話態度の混同は見られたが、日本人被験者は韻律的特徴を頼りに12の発話態度をすべて認識することができた。このことは、本実験に用いた音声資料の各発話態度が日本人被験者に共通して認識されたことを示している。

5.2. 日本語学習歴のないフランス語母語話者（レベル0）の反応

フランス人被験者の回答の分布も、同様にカイ二乗検定したところ、選択した発話態度の分布には有意な傾向が見られた。

次に、文の長さが発話態度の選択に与える影響について検定したところ1モーラ文と2モーラ文の間でのみ有意な影響が見られた。又、日本語の語アクセントが発話態度の選択に与える影響についても検定したところ、フランス人被験者の発話態度の選択には、語アクセントの有意な影響は見られなかった。

また、数量化I V類で構築したフランス語母語話者の知覚行動（図2参照）を観察すると、日本語母語話者の知覚行動に見られた第1次元の「問合せ」でも、第2次元の「ポライトネス」でも同様の傾向は見られなかった。この図からは大きく分けて①「苛立ち(f0-ir)」，②「恐縮(f0-kyo)，ぞんざい(f0-ar)，高圧的(f0-au)」，③「疑念(f0-do)，感心

(f0-ad), 驚き(f0-ex)」, ④「誠意(f0-sin), 丁寧(f0-po), 自明(f0-ev), 平叙(f0-dc), Yes-No疑問(f0-qs)」の4つのグループが確認できる。

このうち, グループ②では「恐縮」の態度が, 敬意の程度が低い「ぞんざい」, 「高圧的」と同様の傾向を示したことは興味深い。

次に, フランス語母語話者が示した発話態度間の混同を考察すると, 「高圧的」, 「苛立ち」, 「感心」は他の発話態度と混同を示すことなく認識された。一方, 「ぞんざい」に関しては, 日本人被験者とは対照的に,

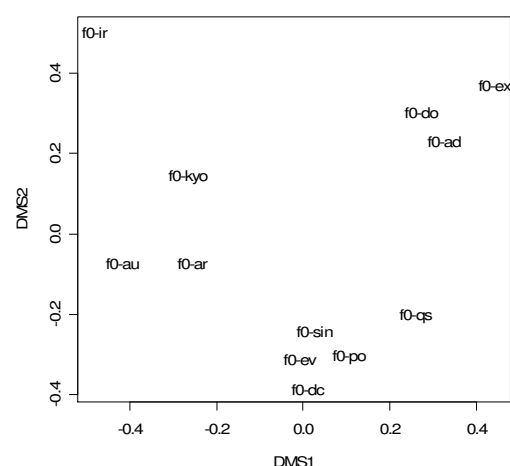


図 2 フランス語母話者(レベル 0) 15 名の発話態度認識率と混同を示した図。円の内側に表示された値は認識率, 円の外側に表示された値は混同の値をそれぞれ示している。

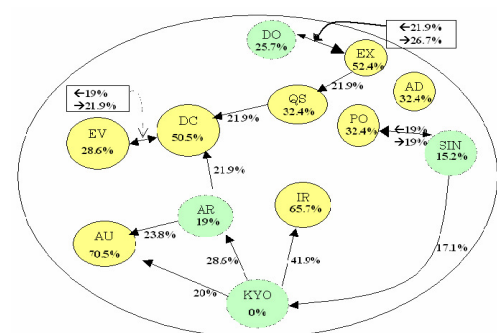


図 3 フランス語母語話者(日本語学習初期段階で, 日本語の「Yes-No疑問」を, 「平叙」と誤って認識する可能性があることを示している。その他にも,

フランス語母語話者には正しく認識されなかった。フランス語母語話者は「ぞんざい」を「高圧的」, 又は「平叙」と混同を示した。さらに, フランス語母語話者は, 日本社会特有の発話態度である「誠意」と「恐縮」を認識することができなかった。「誠意」は, 「丁寧」, 又は「恐縮」との間で混同を示した。「誠意」とは対照的に, 「恐縮」の発話態度は, 「苛立ち」, 「ぞんざい」, または「高圧的」としてフランス語母語話者に認識された。この「恐縮」という発話態度はフランス社会には存在しておらず, このりきみを含んだ音色は, フランス語では敬意を払った発話には使用されることがないため, このように「苛立ち」や「ぞんざい」, または「高圧的」といった意味に誤解されたと思われる。さらに, 驚いたことに, フランス語母語話者は「Yes-No 疑問」を「平叙」と認識する傾向が見られた。この結果は日本語学習初期段階で, 日本語の「Yes-No疑問」を, 「平叙」と誤って認識する可能性があることを示している。その他にも,

「平叙」と「自明」の間、「疑念」と「驚き」の間、さらに「丁寧」と「誠意」の間でも有意な混同が見られた。

一方、レベル0のフランス語母語話者は日本文化と関係の深い発話態度である「恐縮」と「誠意」を認識することができなかった。「ぞんざい」は認識されたが19%と大変低い認識率であった。

6. フランス人日本語学習者の反応

6.1. 被験者の日本語運用能力測定

ここで対象とする被験者は、大学で日本語を主専攻、もしくはオプションで選択している日本語学習者32名である。聴覚実験に先立ち、日本語能力簡易テスト(以下、SPOTとする)を対象の被験者に対して行い、正答率60%未満の学習者16名をレベル1(低)に区分し、正答率60%以上の学習者16名をレベル2(高)に区分した。

6.2. レベル1のフランス人日本語学習者の反応

レベル1に属するフランス人学習者が12の発話態度をどのように認識したかを考察する前に、レベル1の被験者が選択した発話態度の分布に有意な傾向が見られるかどうかを統計的に検定した。

レベル1の被験者が選択した発話態度の分布をカイ二乗検定したところ、選択した発話態度の分布には有意な傾向がみられた。また、文の長さが発話態度の選択に影響を与えているのかどうかを同じくカイ二乗検定によってテストしたところ、文の長さは発話態度の選択に影響を与えていることが確認できた。さらに、語アクセントが発話態度の選択に与える影響もテストしたが、ここでは有意差はみられなかった。このことから本実験において語アクセントは発話態度の判定に影響を与えなかったと考えられる。

次に、数量化I V類によってレベル1のフランス人学習者の2次元知覚空間を構築したところ、レベル0のフランス人被験者同様に、日本語母語話者のような知覚行動は見られなかった。レベル0の被験者同様に、3つの発話態度「疑念(f1-do)、感心(f1-ad)、驚き(f1-ex)」は同様の領域に分布している。また「誠意(f1-sin)、丁寧(f1-po)、自明(f1-ev)」も

比較的同じ領域に分布している。しかしながら，その他の発話態度「苛立ち(f1-ir)，恐縮(f1-kyo)，ぞんざい(f1-ar)，高圧的(f1-au)」の分布は学習歴のないフランス語母語話者のそれとは異なり，拡散的に分布している。

次に，レベル 1 のフランス人学習者が 12 の発話態度をどのように認識したかを示す結果（図 6 参照）を考察したところ，2 つの文化的側面の強い発話態度である「恐縮」と「ぞんざい」を認識することができなかった。「ぞんざい」は，「自明」，又は「平叙」との間で混同を示した。「恐縮」の発話態度は，レベル 0 の被験者同様に「苛立ち」，「ぞんざい」，または「高圧的」として認識された。先にも述べたように，この「恐縮」という発話態度はフランス社会には存在しておらず，このりきみを含んだ音色は，フランス語では敬意を払った発話には使用されることがないため，このように「苛立ち」や「ぞんざい」，または「高圧的」といった意味に誤解されたと思われる。さらに，レベル 1 の被験者は相手に対して敬意を示す表現である「誠意」と「丁寧」を平叙文と混同する傾向が見られた。

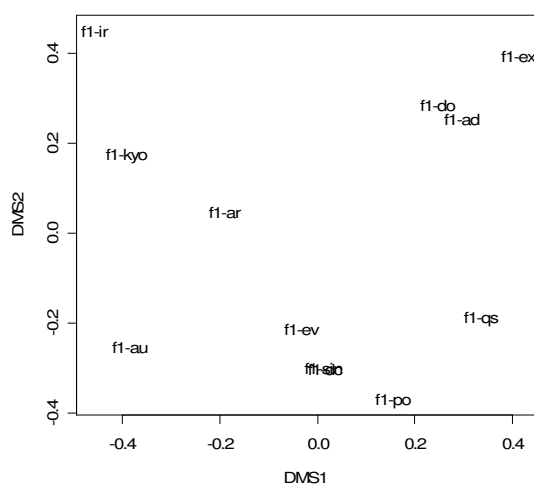


図 4 フランス人日本語学習者（レベル 1）の日本語発話態度に対する知覚行動。ラベルにある f1 はフランス語母語話者日本語レベル 1 を指す。

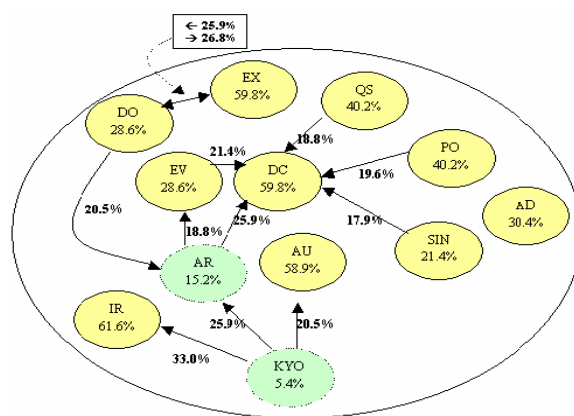


図 5 フランス語母話者（レベル 1）16 名の発話態度認識率と混同を示した図。円の内側に表示された値は認識率，円の外側に表示された値は混同の値をそれぞれ示している。

6.3. レベル 2 のフランス人日本語学習者

レベル 2 のフランス人日本語学習者の回答の分布も、同様にカイ二乗検定したところ、選択した発話態度の分布には有意な傾向が見られた。さらに、発話態度に与える文の長さの影響は 1 モーラ文と 2 モーラ文の間、5 モーラ文と 8 モーラ文の間で確認できた。また、日本語

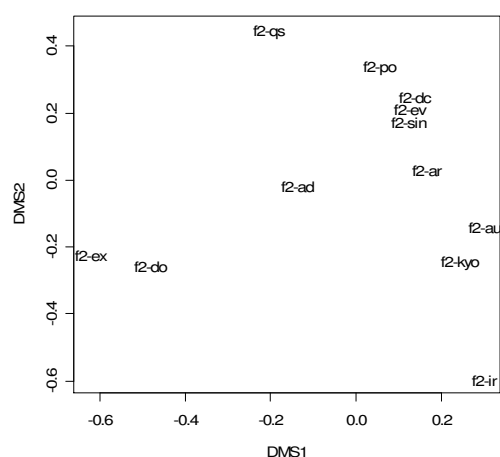


図 6 フランス人日本語学習者（レベル 2）の日本語発話態度に対する知覚行動。ラベルにある f2 はフランス語母語話者日本語レベル 2 を指す。

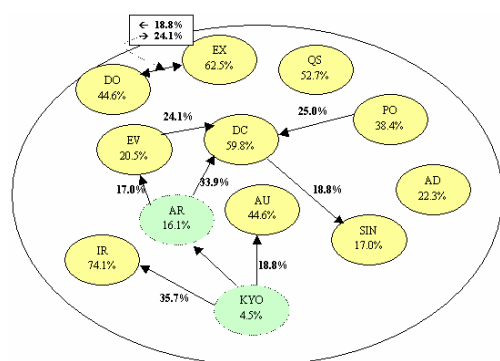


図 7 フランス語母話者（レベル 1）16 名の発話態度認識率と混同を示した図。円の内側に表示された値は認識率，円の外側に表示された値は混同の値をそれぞれ示している。

の語アクセントが発話態度の選択に与える影響についても検定したところ、レベル 2 のフランス人学習者の発話態度の選択にも、語アクセントの有意な影響は見られなかった。次に、数量化 I V 類によってレベル 2 のフランス人学習者の 2 次元知覚空間を構築したところ、第 1 次元（DMS1）で「驚き(f2-ex)、疑念(f2-do)」が他の発話態度と離れた地点に分布していることが確認できる。ただこの領域に「Yes-No 疑問(f2-qs)」がないのは、フランス語母語話者が「Yes-No 疑問」を「平叙」と混同してしまう傾向が関係していると考えられる。第 2 次元（DMS2）の「ポライトネス」の軸では、「誠意(f2-sin)、丁寧(f2-po)、自明(f2-ev)」が上部に分布している。また、ぞんざいは、あまり丁寧度の低い態度とは認識されず、中間地点に分布している。そして、「恐縮(f2-kyo)」は、「高圧的(f2-au)、苛立ち(f2-ir)」と同様に下部に分布しており、敬意のない、ぞんざいな発話態度として認識されていると解釈できる。

次に、フランス語母語話者が示した

発話態度間の混同（図 7 参照）を考察すると、「誠意」は他の発話態度との有意な混同はなく知覚された。この発話態度は、レベル 0 及びレベル 1 の段階では認識することが難しかった発話態度である。この結果から、日本語運用能力が向上するに従って、「誠意」を正しく認識できるようになると推測できる。

この「誠意」とは対照的に、レベル 2 の被験者もレベル 1 の被験者同様、文化的側面の強い 2 つの発話態度「恐縮」、「ぞんざい」は認識することができなかった。この 2 つの発話態度と混同した他の発話態度のパターンもレベル 1 と酷似している。「ぞんざい」は、「自明」、又は「平叙」との間で混同を示し、「恐縮」の発話態度は、「苛立ち」、「ぞんざい」、または「高圧的」として認識された。文化的側面の強い発話態度は、日本語運用能力が向上しても、母語の知覚パターンの干渉が強く残ることを裏付けている。

7. 音響分析

次に、12 の態度で発話された、文頭の語にアクセント核を持たない 8 モーラ文を音響分析した。具体的には、音響分析ソフト Praat を用いて、発話を構成する母音の開始地点、中間地点、終了地点の基本周波数（F0）値を測定した後、聴覚的音の高さの指標である Quarter-tone（以下、1/4tone とする）で各発話態度のピッチ曲線、ピッチレンジ、ピッチの平均を示した。さらに、発話全体の持続時間、声の大きさ（dB）も抽出した。声の大きさは、発話を構成する各分節音の開始地点、中間地点、終了地点の大きさ（dB）を測定し、その平均を求めた。なお、分析した音声のサンプリングレートは 44.1kHz、量子化 16bits、ステレオで wav ファイルとして保存したものを用いた。

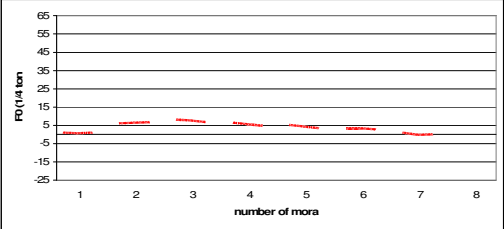
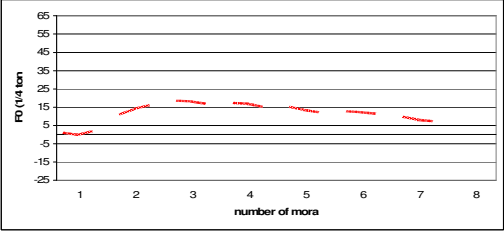
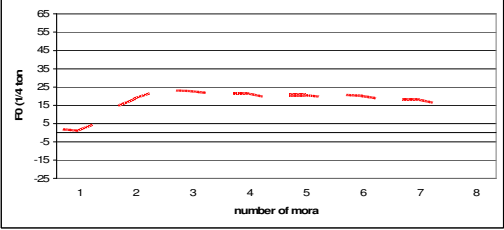
7.1. 「丁寧」と「感心」にみられる韻律

日本語母語話者は「丁寧」と「感心」との間で混同を示したがこの 2 つの発話態度は共に平叙文に比べ、発話時間が長く、ピッチの平均も高い（表 3 参照）。ピッチ曲線を観察すると文頭から 3 モーラの頭にかけてのピッチ上昇の度合いで、平叙文とこの 2 つの発話態度の間

に顕著な違いがみられる。ただし，この2つの発話態度にも若干の相違点が観察された。文末に注目すると，「丁寧」は平叙文同様ピークから文末まで 10 1/4tone 程度下降するのに対して，「感心」は文末まで半分の 5 程度しか下降しない。

さらに，この3つの発話態度のピッチレンジ（f0 最大値から f0 最小値を引いた値）を比べて見ると，「丁寧」な発話では平叙文に比べ約 2 倍，「感心」では平叙文に比べ約 2.5 倍の開きが観察できる。

表 3 「丁寧」と「感心」，平叙文の3つの発話態度の音響的特徴

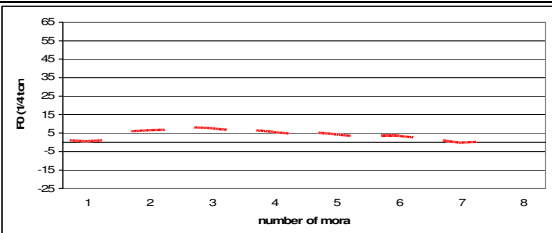
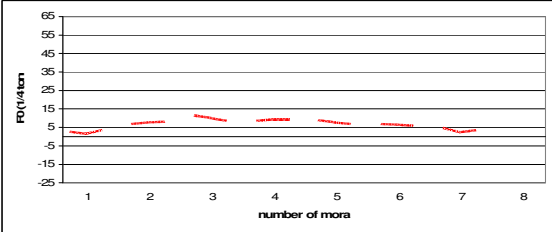
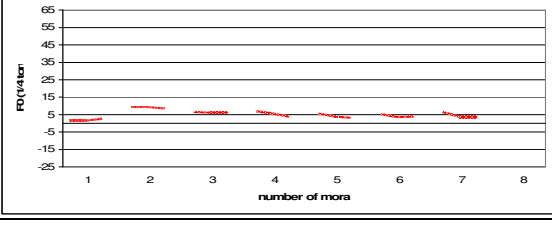
態度	ピッチ曲線 (1/4 tone)	発話持続時間 (ms)	F0 _{1/4tone} max	F0 _{1/4tone} min	F0 _{1/4tone} 平均値	大きさ (dB) 平均値
平叙		950	8.42	0.02	4.24	61.35
丁寧		1133	18.90	-0.01	12.07	63.57
感心		1364	23.43	1.07	17.68	67.50

7.2. 「誠意」と「恐縮」の韻律に見られる特徴

聴覚実験の結果で述べたように日本語母語話者は「誠意」と「恐縮」との間に混同を示したが，この2つの発話態度の音響的特徴（発話持続時間，ピッチ曲線，ピッチレンジ，ピッチの平均，発話を構成する各文節音の大きさの平均）からは際立った相違点は見られず，類似した特徴を共有しているといえる（表 4 参照）。「恐縮」と「誠意」は類似した音響的特徴を持っているにもかかわらず，フランス語

母語話者にはまったく異なった発話態度として誤認識される原因を調べるため、H1-H2 により 5 つの発話態度（「恐縮」、「誠意」、「高圧的」、「苛立ち」）の声質を測定した。

表 4 敬意を示す発話態度「誠意」、「恐縮」と「平叙」の音響的特徴

態度	ピッチ曲線 (1/4 tone)	発話持続時間 (ms)	F0 _{1/4tone} Max	F0 _{1/4tone} min	F0 _{1/4tone} 平均値	大きさ (dB) 平均値
平叙		950	8.42	0.02	4.24	61.35
誠意		1005	11.69	1.74	6.92	62.70
恐縮		1032	9.79	1.61	5.32	64.04

各発話態度の H1-H2 の値を考察する前に、H1-H2 は何を測定するものなのかを簡単に述べておく。H1-H2 は第 1 高調波(first harmonic, 以下 H1 とする)と第 2 高調波(second harmonic, 以下 H2 とする)との間にみられるスペクトル上のパワー(dB)を測定することで得られ、声を生成する際の声門がどのような開き方をするのかを知る手がかりになることが知られている。H1 は声が生成される際の声門の開口度(Open glottis)の指標であり、H2 は声門の閉口度の指標となる⁷。H1-H2 の値が増大する (H1 のパワー(dB)が H2 のそれに比べ大きい) と

⁷ H1-H2 による声質の解釈についての詳細な説明は Menezes&Maekawa (2006) を参照。

いうことは声門の開きが広いことを示し、より **breathy**（息の漏れた）な声になる。反対に、**H2** のパワーが **H1** のそれに比べ増大すると、より **tense**（緊張した）な声になる。さらに、**H2** のパワーの値が **H1** のそれを上回る（**H1-H2** の値がマイナスになる）時、声はより **creaky**（きしみ音）なものになる。

ここでは、「恐縮」、「誠意」、「苛立ち」、「高圧的」に「平叙」を加えた5つの発話態度で発話された8モーラ文の各母音の中間地点での **H1-H2** を測定した（表5）。図8は5つの発話態度（平叙文、「恐縮」、「誠意」、「苛立ち」、「高圧的」）の **H1-H2** 値を比較したものである。図8左の折れ線グラフは第1母音 /a/ 中間地点の **H1-H2** 値であるが、「平叙」、「恐縮」、「誠意」の3つの発話態度は文頭で **creaky** な声質であることがわかる。

表5 8モーラ文「ナニワで飲みます」の各母音中間地点の **H1-H2** の値（dB）

発話態度	1st	2nd	3rd	4th	5th	6th	7th	8th
平叙文 (DC)	-5.2	0	-0.9	-1	0	-1.8	-0.4	
恐縮 (KYO)	-0.6	1.6	1	-7.2	-1.8	0	1.2	
誠意 (SIN)	-2	7.2	1.2	0.6	-3.5	-1	2.8	
苛立ち (IR)	2.7	5.4	6.2	1.7	5.8	6.1	2.3	-4.4
高圧的 (AU)	1.1	1.7	-6	-1.2	-4.5	0.4	1.6	

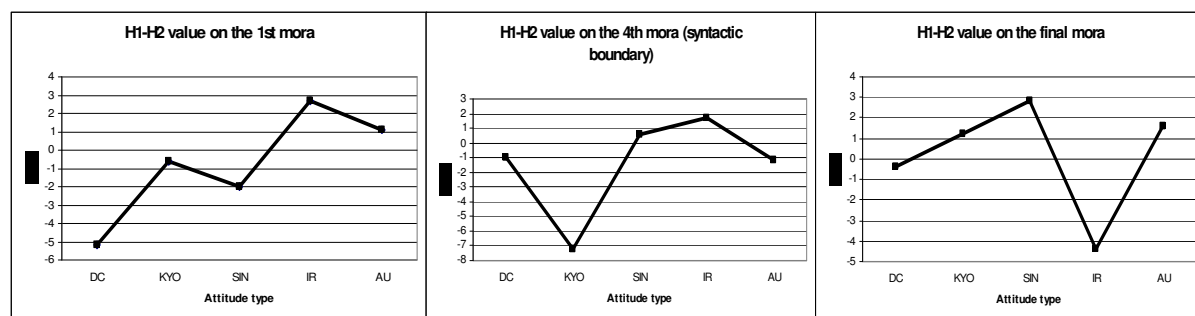


図8 5つの発話態度（右から「平叙」、「恐縮」、「誠意」、「苛立ち」、「高圧的」）の **H1-H2** 値（左から第1母音 /a/ 中間地、第4母音 /e/ 中間地点、文末母音中間地点 /a/ [「苛立ち」だけ /u/]の **H1-H2** 値をそれぞれ示している）

このうち「平叙」の文頭第1母音が最もクリーキーの程度が強く、次に「誠意」，「恐縮」が続く。「苛立ち」や「高圧的」な発話態度にはこの傾向は見られなかった。次に，句末に位置する第4母音 /e/ 中間地点の H1-H2 値をみると，他の発話態度に比べ，「恐縮」だけ極端に H1-H2 値が低い。これは，ここでの「りきみ」を帯びた「恐縮」の声質が **harsh**（荒々しさ）を帯びた **creaky** な声質になっているからであると推測できる。最後に，図の右側の文末母音中間地点 /a/

（「苛立ち」だけ /u/ の脱落がないので，この母音の中間点）の H1-H2 値を観察した。その結果，他の発話態度に比べ，「誠意」は文末でより **breathy** に発話されていることがわかる。また，「苛立ち」は文末の母音でのみ H1-H2 値が大幅に下落している。この文末での声質は **harsh** の伴った **creaky** なものである。

この結果から，「恐縮」の発話には非常に **harsh** で **creaky** な声質が確認できるのに対して，「誠意」は **creaky** な声質は句頭でのみ確認でき，文中，文末では「恐縮」に比べ，より **breathy** な声質が確認できる。このことから，「恐縮」と「誠意」間で見られた混同は声質以外の音響的特徴が優先的に影響を与えている可能性が考えられる。また，「苛立ち」の発話態度は日仏双方の被験者に変よく認識されたものであるが，この文末の **harsh** で **creaky** な音質も影響を与えたのではないかと推測される。

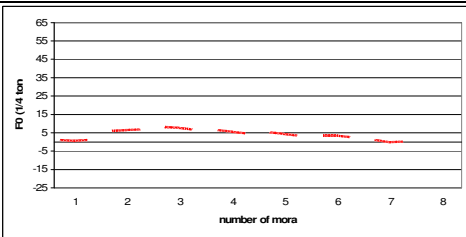
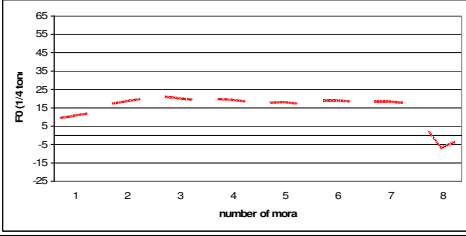
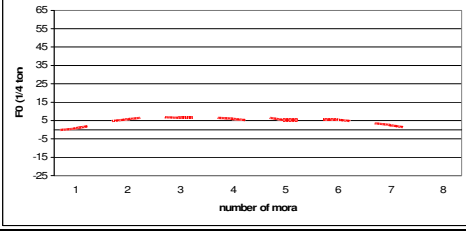
7.3. 「ぞんざい」な態度にみられる韻律

「ぞんざい」は日本語母語話者にはよく認識された発話態度であるが，フランス語母語話者には認識が難しかった発話態度の一つである。多くのフランス人被験者は「ぞんざい」と明確な意図がないことをしめす「平叙」との間で混同を示したが，表6を参照すると，「平叙」と「ぞんざい」は共にピッチが平坦であり，ピッチレンジ，ピッチ平均も際立った違いはみられない。ただ発話持続時間と声の大きさの値をみると「平叙」に比べ，「ぞんざい」の発話持続時間は長く，声の大きさも「平叙」に比べ大きい。この2点の相違に日本語母語話者は

敏感であったが，フランス語母語話者はそれよりもピッチから得られる情報に知覚の判断基準を置いたのではないかと推測される。

一方，日本語母語話者は「自明」を「ぞんざい」と混同する傾向もみられたが，音響的特徴を観察すると「自明」は「ぞんざい」とは音響的にかなり異なっていることがわかった。「自明」は「平叙」や「ぞんざい」に比べピッチ平均が高く，文末モーラでのピッチの急下降が見られる点に特徴があった。このことから，「ぞんざい」と「自明」の混同は音声的特徴から影響を受けたものというより，コンセプトレベルでの影響が考えられると推測される。

表 6 平叙文と比較した「ぞんざい」，「自明」の発話態度の音響的特徴

態度	ピッチ曲線 (1/4 tone)	発話持続時間 (ms)	F0 _{1/4tone} Max	F0 _{1/4tone} min	F0 _{1/4tone} 平均値	大きさ (dB) 平均値
平叙		950	8.42	0.02	4.24	61.35
自明		995	21.38	-6.78	15.23	64.77
ぞんざい		1165	7.21	0.07	4.81	65.55

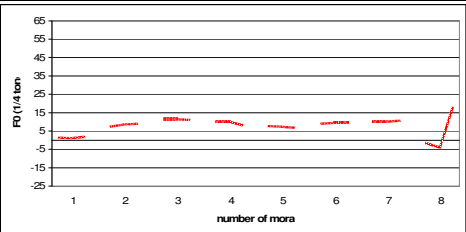
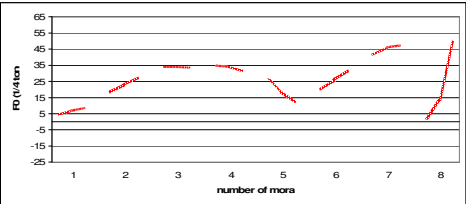
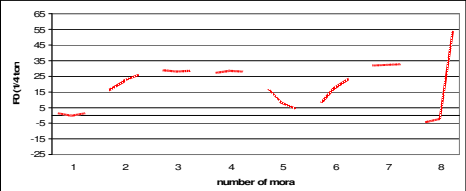
7.4. 問いかかけのモダリティの発話態度に見られる韻律のパターン

日本語母語話者は3つの発話態度（「Yes-No 疑問」，「驚き」，「疑念」）を他の発話態度と混同することなく正しく認識できたが，知覚行動の分布を描いた2次元空間では，この3つの発話態度は比較的近い地点に分布していることが観察できた。表7をみると，この3

つの発話態度は共に文末でのピッチ上昇がみられる点で共通している。ただし、そのピッチ上昇の度合いは「Yes-No 疑問」が 8.4 1/4tone でもっとも小さく、「驚き」は 47 1/4tone, 「疑念」は 48.9 1/4tone と「Yes-No 疑問」に比べピッチの急上昇が確認できる。文中のピッチ曲線も、「Yes-No 疑問」が文末以外の部分ではあまり起伏が激しくないのに対して、「驚き」と「疑念」ではピッチの起伏が激しい。声の大きさも「Yes-No 疑問」に比べ、「驚き」と「疑念」は約 10dB ほど大きい。このような音響的特徴がフランス語母語話者の「驚き」と「疑念」の間にみられた混同に影響を与えた可能性が考えられる。

一方、3つの発話態度の発話持続時間をみると「Yes-No 疑問」の発話持続時間は最も短く、「驚き」, 「疑念」の順に持続時間は長かった。聴覚実験の結果からは、日本語母語話者は「驚き」を「疑念」として認識する傾向はみられなかったが、これは「驚き」の発話態度は「疑念」に比べ発話速度が速いことが影響していると思われる。

表 7 「Yes-No 疑問」と比較した「驚き」, 「疑念」の発話態度の音響的特徴

態度	ピッチ曲線 (1/4 tone)	発話持続時間 (ms)	F0 _{1/4tone} e Max	F0 _{1/4tone} e min	F0 _{1/4tone} e 平均値	大きさ (dB) e 平均値
Yes-No 疑問		1013	8.42	0.02	7.87	59.80
驚き		1233	49.09	2.04	26.07	69.69
疑念		2088	53.10	-4.20	19.19	67.91

8. 結論と今後の課題

本稿では、まず日本語母語話者に対して選定した 12 の日本語の発話態度を聞かせ正しく認識できるかを調べた。その結果、日本語母語話者は「恐縮」と「誠意」を示す態度間で混同を示したものの、12 の発話態度をすべて正しく認識することができた。次に、日本語学習経験のないフランス語母語話者（レベル 0）に対して同様の聴覚実験を実施した結果、被験者は敬意を示す発話態度である「恐縮」と「誠意」を認識することが出来なかった。さらに、日本語学習経験のないフランス語母語話者及び、日本語母語話者の結果と比較し、フランス人日本語学習者の日本語運用能力が発話態度の知覚にどのような影響を与えているのかについても考察した。その際、同じ学習歴をもつ被験者でも日本語のレベルに格差があることが考えられるので、聴覚実験に先立ち、日本語能力簡易テスト(SPOT) を対象の学生に対して行い、2 つのグループ（レベル 1（低）とレベル 2（高））に分け、聴覚実験を行った。その結果、フランス語母語話者に共通して見られる知覚の傾向と、日本語学習の影響で発話態度の知覚が異なって現れるものが確認できた。

フランス人日本語学習者（レベル 1 とレベル 2）は 12 の発話態度のうち「ぞんざい」と「恐縮」を認識することができなかった。特に、「恐縮」は、日本語学習歴を問わず「高圧的」、「ぞんざい」「苛立ち」と混同する傾向がフランス語母語話者に見られた。

一方、レベル 1 とレベル 2 ではいくつかの異なる傾向も見られた。レベル 1 の学習者は、レベル 0 が認識することができなかった「誠意」の発話態度を認識することができたものの、「平叙」との混同も示した。レベル 2 の学習者はこの発話態度を他の発話態度と混同することなく認識することができた。

また、数量化 I V 類を用いて各グループの知覚行動を 2 次元で示し比較した結果、各発話態度に対する日本語母語話者の反応はフランス語母語話者のそれとは異なっていることが確認できた。ただし、フランス語母語話者の反応でもレベル 0、レベル 1 の被験者に比べ、レベル 2 の被験者の反応は日本語母語話者の知覚パターンに似たものが確認できた。

日本語母語話者がある発話態度間で混同を示した理由を調査するため音響分析をした結果、「感心」と「丁寧」間でみられた混同の原因は、「平叙」に比べ発話時間が長く、ピッチの平均、文頭のピッチ上昇の度合いも「平叙」に比べ高いという特徴を共通して持っている点が影響を与えていると推測できる。また、高い混同を示した「恐縮」と「誠意」は類似した音響的特徴を持っているが、声質が異なっていることがわかった。次に、「自明」と「ぞんざい」を音響分析した結果、「自明」は「ぞんざい」に比べピッチ平均が高く、文末モーラでのピッチの急下降が見られ、音響的にかなり異なっていることがわかった。このことから、「ぞんざい」と「自明」の混同は音声的特徴から影響を受けたものというより、コンセプトレベルでの影響が考えられる。また、日本語母語話者の知覚行動の分布を描いた 2 次元空間で比較的近い地点に分布していた「Yes-No 疑問」、「驚き」、「疑念」の 3 つの発話態度は、共に文末でのピッチ上昇がみられる点で共通している。ただし、そのピッチ上昇の度合いは「Yes-No 疑問」に比べ、「驚き」、「疑念」で非常に大きい。文中のピッチ曲線も、「Yes-No 疑問」が文末以外の部分ではあまり起伏が激しくないのに対して、「驚き」と「疑念」ではピッチの起伏が激しい。声の大きさも「Yes-No 疑問」に比べ、「驚き」と「疑念」は約 10dB ほど大きい。このような音響的特徴がフランス語母語話者の「驚き」と「疑念」の間にみられた混同に影響を与えた可能性が考えられる。

今後、フランス語母語話者だけでなく 英語母語話者や日本語と文化的にも近くに位置する中国語や韓国語の母語話者に対して同様の聴覚実験を行うことで、多言語間に潜む「偽の友達」の実態が解明でき、多言語文化間でのコミュニケーション理解に貢献できるのではないかとと思われる。

参考文献

エリクソンドナ・昇地崇明（2006）「性差、および母語が感情音声に与える影響—日本語、韓国語、英語母語話者を対象として—」『音声文法 V』pp.31-46、くろしお出版。

小林典子, フォード丹羽順子, 山元啓史 (1996) 「日本語能力の新しい測定法 [SPOT]」 『世界の日本語教育』 6, pp.201-218.

平山輝男 (1960) 『全国アクセント辞典』 東京堂出版.

松本恵美子・定延利之 (2001) 「日本語の韻律におけるりきみと日本語学習者の理解度」 In. Departement of Japanese Studies, The Chinese University of Hong Kong and Society of Japanese Language Education (eds.) *Quality Japanese Studies and Japanese Language Education in Kanji-Using Areas in the New Century*, pp.455-461, Hong Kong Himawari Publishing Co..

Aubergé, Véronique (2002) A gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP. *Proc. Speech Prosody 2002*, pp.151-154. Aix-en Provence, France.

Aubergé, Véronique, Grépillat, Tuuliki, Rilliard, Albert (1997) Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. *5th European Conference on Speech Communication and Technology Vol.2*, pp. 871-874, Greece.

Brown, Penelope, Levinson, Stephen (1986) *Politeness: Some Universals in Language Usage*, Cambridge, Cambridge University Press.

Campbell, Nick (2000) Databases of emotional speech. *Proc. International Speech Communication Association 2000*, pp.34-38, Northern Ireland, UK.

Clément, Juliette (1999) Structure des Représentations Prosodiques. Développement Normal et Pathologique du Traitement de la Prosodie, Doctoral Thesis, Université de Paris V.

Damasio, Antonio R. (1994) Descartes' Error. Emotion, Reason, and the Human Brain, Putnam books.

Erickson, Donna, Ohashi, Seiko, Makita. Yuko, Kajimoto. Natsuki, Mokhtari, Parham (2003) Perception of naturally-spoken expressive speech by American English and Japanese listeners. *Proceeding of CREST International Workshop on Expressive Speech*, pp.31-36, Kobe, Japan.

- Ito, Mika (2003) The contribution of voice quality to politeness in Japanese. *VOQUAL'03*, pp.157-162, Genève.
- Maekawa, Kikuko (1998) Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *International Conference on Spoken Language Proceeding*, pp.635-638.
- Menezes, Caroline, and Maekawa, Kikuo (2006) Paralinguistic effects on voice quality: A study in Japanese. *Proceeding of Speech Prosody 2006*, pp.656-659. Desden, Gernamy.
- Ofuka, Etsuko, Mckeown, J. Denis, Waterman, Mitch G., and Roach, Peter J. (2000) Prosodic cue for rated politeness in Japanese speech. *Speech Communication* 32, pp.199-217.
- Ohala, John J. (1984) An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, pp.1-16.
- Sadanobu, Toshiyuki (2004) A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan* 8, No.1, pp.29-44.
- Scherer, Klaus R. (1986) Vocal affect expression : A review and a model for future research. *Psychological Bulletin* 99 (2), pp.143-165.
- Scherer, Klaus R., Banse, Rainer, and Wallbott, Harald G. (2001) Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32 (1), pp.76-92.
- Shochi, Takaaki, Aubergé, Vérinique, and Rilliard, Albert (2005) Because attitudes are social affects. they can be false friends. *The First International Conference on Affective Computing & Intelligent Interaction*, LNCS 3784, pp.482-489, Beijin, China.

Résumé : Les affects sociaux (ou attitudes) exprimés lors des interactions orales sont intimement liées à ce que le locuteur veut exprimer et sont influencées par sa langue et sa culture. Aubergé (2002) défend un contrôle volontaire de ces expressions. À la suite des travaux sur les attitudes du français et de l'anglais britannique menées au GIPSA-lab, cette thèse analyse 12 attitudes prosodique japonaises. Elle décrit leur réalisation et perception par des japonais. L'influence de l'information visuelle est mesurée au moyen d'un corpus multimodal. La localisation des informations acoustiques critiques est évaluée grâce à paradigme de gating. Pour certaines attitudes, les informations sont réparties entre le début et la fin de la phrase, tandis que d'autres sont reconnues immédiatement, et que certaines sont reconnues à un point clés au milieu ou à la fin des énoncés.

La perception interculturelle des attitudes est analysée pour trois langues (japonais, français, anglais britannique) afin de mieux comprendre l'importance de la culture sur chacune des attitudes. Les résultats montrent que deux expressions (*exclamation de surprise* et *déclaration*) sont reconnues interculturellement, tandis que d'autres se comportent comme des faux amis – en particulier l'expression de politesse japonaise de « *kyoshuku* ». Finalement, l'effet d'apprentissage des langues est mesuré sur des apprenants français de japonais. Ils augmentent progressivement performances, malgré le problème récurrent de l'expression de « *kyoshuku* ».

Title: Prosody of Social Affects in Japanese, French and English:
in search of true- and false-friends for the learner's path

Resume: Social affects (or attitudes) expressed by a speaker during a verbal interaction are linked to what the speaker intends to say, and are influenced by the speaker's language culture. Aubergé (2002) postulates a voluntary control of such expressions. Following works on attitudinal expression in French and British English at GIPSA-lab, the current thesis investigates 12 Japanese prosodic attitudes. It describes their realization and their perception by Japanese speakers. The influence of visual information is measured through a multimodal corpus. Critical acoustic information for the attitudes' recognition is gathered thanks to a gating paradigm. For some attitudes, information is spread from the start until the end of the utterances, but some attitudes are recognized immediately, while others are recognized at a critical point at the middle or the end of utterances.

Cross-cultural perception of social affects in three languages (Japanese, French and British English) is investigated in order to study cultural cues in social affective expressions. Results show that two attitudes (exclamation of surprise and declaration) are recognized cross culturally, while some others behave like false friends – especially the “kyoshuku” Japanese expressions of politeness. Finally, the language learning effect is measured on French learners of Japanese. Learners gradually increase their performances, but still have problems with the most culturally encoded expressions (i.e. “kyoshuku”).

Spécialité : Sciences du Langage

Mot-clés :

prosodie, affects sociaux, perception interculturelle, apprentissage des langues, multimodalité

Laboratoire :

Laboratoire Grenoble Images Parole Signal Automatique

Université Stendhal – 1180 Av. Centrale, B.P. 25 – 38040 Grenoble Cedex 9