

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences du Langage / Français Langue Étrangère**

Arrêté ministériel : 7 août 2006

Présentée par

Thi Thuy Hien TRAN

Thèse dirigée par **Nathalie VALLÉE** et
co-dirigée par **Quang Thuan NGUYEN**

préparée au sein du **Laboratoire GIPSA**
dans l'**École Doctorale Langues, Littératures et Sciences Humaines**

**Processus d'acquisition des clusters et
autres séquences de consonnes en langue
seconde : de l'analyse acoustico-perceptive
des séquences consonantiques du
vietnamien à l'analyse de la perception et
production des clusters du français par des
apprenants vietnamiens du FLE**

Thèse soutenue publiquement le **12 décembre 2011**,
devant le jury composé de :

M. Didier DEMOLIN

Professeur de l'Université Stendhal, Grenoble, Président

M. Bernard HARMEGNIES

Professeur de l'Université de Mons, Belgique, Rapporteur

M. Philippe HILIGSMANN

Professeur de l'Université Catholique de Louvain, Belgique, Rapporteur

M. Marc BRUNELLE

Assistant Professor de l'Université d'Ottawa, Canada, Examineur

Mme Nathalie VALLÉE

Chargée de Recherche, CNRS, GIPSA-lab, Grenoble, Directrice de thèse

M. Quang Thuan NGUYEN

Professeur de l'Université Nationale de Hanoï, Vietnam, Co-directeur de thèse



À ma famille

À mes indispensables Nếp et Đức

REMERCIEMENTS

« Ăn quả nhớ kẻ trồng cây »

Pense à celui qui a planté l'arbre dont tu manges le fruit (t.p.n)

(Proverbe vietnamien)

Je tiens à remercier ici les personnes qui, de près ou de loin, ont contribué à la concrétisation de ce travail de thèse.

Mes premiers remerciements s'adressent à Nathalie Vallée pour l'encadrement de cette thèse, pour son investissement dans ce travail même après la belle gamelle, ses rigueurs intellectuelles, ses relectures minutieuses et les corrections soigneuses. Les heures qu'elle m'a consacrées tout au long de ces années, et plus particulièrement les derniers temps, m'ont été des plus précieuses. Nathalie, je te remercie de m'avoir formée, de m'avoir fait confiance jusqu'au bout et de m'avoir donné l'envie de continuer dans cette voie. Merci tout simplement de m'avoir transmis le goût de la recherche.

J'adresse mes plus sincères remerciements aux membres du jury pour avoir accepté d'évaluer ce travail. À M. Didier Demolin, qui m'a fait l'honneur de présider le jury de thèse, pour l'intérêt et le soutien chaleureux dont il a toujours fait preuve. Je suis très reconnaissant à M. Bernard Harmegnies et M. Philippe Hiligsmann d'avoir accepté de se déplacer de la Belgique pour assurer le rôle de rapporteur. Un grand merci également à M. Marc Brunelle de l'Université d'Ottawa pour son appui à notre recherche et pour avoir accepté de faire partie de ce jury. Je remercie vivement M. Nguyễn Quang Thuân d'avoir accepté de diriger officiellement cette thèse, merci d'avoir toujours soutenu notre travail, notamment lors de mes deux séjours à l'Université Nationale de Hanoi en 2008 et 2009.

Je remercie l'Agence Universitaire de la Francophonie de m'avoir attribué une bourse de formation à la recherche (n° PC – 411/2460) qui m'a permis de réaliser mes recherches sur le terrain pendant deux ans. Je remercie également le Département Parole et Cognition du GIPSA-lab d'avoir financé les trois derniers mois de ma thèse.

Je tiens à exprimer ma reconnaissance à Dominique Abry qui par son cours en Master, m'a donné le goût de la phonétique corrective et m'a orienté vers ce thème de recherche lors de mon mémoire de fin d'études en Master 1.

Je remercie Lionel Granjon, René Carré, Christophe Savariaux, Lê Việt Bắc, Trần Đỗ Đạt pour l'aide matérielle qu'ils ont chacun apportée à ce projet de recherche et pour les échanges constructifs et discussions très fructueuses. Merci spécialement à toi, Lionel, de m'avoir supportée jusqu'à la fin, de ta disponibilité dont tu as fait preuve pour résoudre des difficultés que j'ai rencontrées lors du traitement des données. Merci pour tes conseils et ton aide inestimable en fin de thèse. Tes beaux « dessins » avec lesquels tu m'expliques les notions, que ce soit en statistique ou en acoustique, ont très bien marché ! Je les garderai pour longtemps !

Je tiens à exprimer ma reconnaissance à Mme Phạm Ngọc Yến, M. Eric Castelli pour leur accueil chaleureux au centre MICA (Multimédia, Information, Communication et Applications) de l'Institut Polytechnique de Hanoi lors de mes deux séjours sur le terrain pendant trois mois chaque année en 2008 et 2009. Les séjours ont été très riches de renseignements, aussi bien sur le plan théorique que pratique. Merci à les amis de MICA notamment Đạt, Lan, Son qui ont fait de mon séjour à Hanoi des souvenirs inoubliables.

Merci à thầy Bùi Ngọc Oánh et son équipe du Centre Multimédia de l'Ecole Supérieure des Langues Etrangères de l'Université Nationale de Hanoi pour leur accueil chaleureux et pour leur aide lors de mes différents enregistrements des données en 2008 et 2009.

Un grand merci à tous mes « sujets » : aux locuteurs Christophe Savariaux, Nicolas Audibert, Émilie Troille qui ont prêté leur voix ; à une centaine d'étudiants du Département de Français de l'Ecole Supérieure des Langues Etrangères de l'Université Nationale de Hanoi qui ont subi les contraintes de l'enregistrement et sans qui je n'aurais pu obtenir ces résultats. Merci à Hà Tiên Cường, pour son aide précieuse pour trouver des sujets lors de mes différents tests à Hanoi. Merci à mes amis de Grenoble qui ont passé les tests de perception et production du vietnamien et qui ont toujours manifesté un grand intérêt à l'égard de mes travaux : Việt Bắc, Hoàng Nam, Việt Anh, Hồng Thái, Ngọc Diệp, Đặng Khoa, Đặng Toàn, Quỳnh Chi, Thái Hà, Trung Hiếu, Hà Thái, Vũ Ngà, Đỗ Hiếu, Thanh Hương, Hải Sơn, Thủy Vân, Sỹ Lâm, Hoàng Ngân, Vũ Thủy, Kim Quy.

Ma gratitude s'adresse à Solange Rossato, qui a guidé mes premiers pas dans l'enseignement de la phonétique phonologie à l'UFR Sciences du Langage de l'Université Stendhal. Solange, je te dois la grande majorité des connaissances en acoustique que j'ai acquises. Merci pour les jeux de lectures de spectrogrammes « à la Jacqueline » qu'on a passé ensemble après l'École de Nasalité de Porquerolles. Tes conseils, tes discussions notamment autour de la transition formantique, du traitement statistique des données sous SPSS qui m'ont bien aidé à avancer cette thèse.

Un grand merci à tous les personnels que j'ai rencontrés du laboratoire et qui ont participé de près ou de loin à l'aboutissement de cette thèse : Carole, Coriandre, Christophe, Giovanni, Nathalie H., Jean-Pierre, Hélène, Jean-Luc, Marc, Tom, Sonia, Frédéric E., Pierre, Gérard. Elisabetta et Anne Vi., merci à vous de m'avoir transmis la passion d'enseignement, merci pour de nombreuses discussions à propos des cours que nous avons eu ensemble pendant toutes ces années. Merci à toi, Elisabetta pour ton message de soutien au beau matin du jour de soutenance qui m'a directement touché. Merci à toute l'équipe administrative et technique depuis mon entrée au DPC en 2006 : Marie-Thé, Alain, Dominique, Marie-Ange, Liva, Houria. Je tiens à remercier en particulier Marie-Thé, qui a toujours su faciliter mes démarches administratives, qui a été une oreille attentive et une parole très douce et réconfortante. Marie-Thé, ta présence au labo même brève avant ton départ à la retraite a été très importante pour moi, merci pour ton soutien et tes encouragements répétitifs pendant toutes ces années. Merci aussi à des agents de Sécurité et aux gardes du campus que j'ai croisés, merci pour leur enthousiasme et leurs encouragements qui m'ont bien monté le moral pendant les soirées et les weekends de rédaction.

Pendant mes années de thèse, j'ai eu le plaisir de participer à l'organisation des Rencontres des Jeunes Chercheurs en Parole 2011 à Grenoble. Merci à Sandra, Mathilde, Ben, Amélie, Rosario et Atef d'avoir partagé avec moi cette aventure unique et très riche dans son genre !

Merci aux thésards, stagiaires et aussi aux anciens thésards de deux sites Stendhal et Ampère pour les moments qu'on a passés ensemble, surtout pour vos soutiens sans faille à la fin de la rédac : Anne, Amélie, Atef, Audrey, Benjamin, Chloé, Ibrah, Julie, Krystyna, Marine, Maily, Manon, Nico, Omran, Paolo, Rosario, Sylvia, Yan. Et évidemment merci à ma Maria irremplaçable avec son enthousiasme, son beau sourire de tous les jours qui m'ont beaucoup manqué depuis son départ en post-doc à Avignon ! Et puis il y a bien sûr Claire et Sandra (et les deux Louis), mes co-bureaux incontournables, quelle chance de vous avoir à mes côtés pendant mes années de thèse ! Merci pour tous les moments magiques qu'on a passés ensemble que ce soit au labo ou ailleurs, merci à vous d'avoir toujours su trouver les bons mots quand j'en avais besoin. « Les thèses sont parfois lourdes et embêtantes mais sans les nôtres, nous ne nous serions pas rencontrées. Alors merci à nos thèses !!! Restons amies même au-delà des frontières ».

Mes remerciements vont également à mes amis en France et au Vietnam qui m'ont gentiment aidée et m'ont soutenue à accomplir ce travail. Un grand merci à bác Linh, bác Thúy Dom, bác Hà Thanh, avec l'aide de plusieurs amis de l'AEVG, pour avoir organisé un magnifique pot de thèse. La qualité du pot m'a valu de nombreux compliments qui auraient dû leur revenir. Merci à anh Dom d'avoir sacrifié un weekend pour la relecture d'une partie de la thèse. Merci à tonton Bí, Thủy, Nhu et spécialement à bác Linh pour leurs aides précieuses dans la vie quotidienne surtout après l'arrivée de Nếp et pour leur soutien sans faille les derniers jours de la rédaction.

Merci à ma famille et belle-famille pour leur soutien même à distance, leurs coups de téléphone et leurs lettres émotionnelles. Merci à mes parents, ma sœur Hải Anh et mon frère Tiễn qui ont toujours cru en moi, pour leur confiance, leur fierté et leur amour.

Merci à tous ceux que j'ai probablement oubliés et qui m'ont soutenue et encouragée dans cette aventure !

Enfin, c'est à vous, Nếp et Đức, que j'adresse mes derniers remerciements et que je dédie cette thèse. C'est bien à vous que je la dois pour être présents dans ma vie, pour m'avoir toujours encouragée et soutenue. Đức, merci pour ta présence et ton soutien de tous les instants, ton aide aussi bien technique que morale, tes idées en Matlab, tes paroles réconfortantes et ton bonheur de me voir réussir. Merci d'avoir supporté les contraintes d'une fin de thèse laborieuse, d'avoir assuré non seulement la mise en page du manuscrit mais aussi la petite, la cuisine, les courses, le ménage pendant ces dernières semaines. Ton amour, ta patience et ta confiance m'aident à avancer. Merci !

« Có công mài sắt có ngày nên kim »

A force d'affuter le fer, un jour on auras une aiguille

(Proverbe vietnamien)

Table des matières

Liste des tables	vi
Liste des figures	x
Introduction.....	1
Chapitre 1.....	6
Acquisition de la phonologie en langue seconde : le cas des groupes de consonnes.....	6
1.1. Influence de la langue native.....	7
1.1.1. La notion de « Crible Phonologique ».....	7
1.1.2. Hypothèse de l'analyse contrastive.....	9
1.1.3. Perception en langues première et seconde : théories et modèles	12
1.2. Influence de principes universels	22
1.2.1. La notion de marquage	22
1.2.2. L'Hypothèse de la Différence de Marquage.....	23
1.2.3. L'Hypothèse de la Conformité Structurale.....	25
1.3. Acquisition des groupes de consonnes en langue seconde	27
Chapitre 2.....	35
Phonologie du vietnamien : description et comparaison avec le français	35
2.1. Éléments de linguistique historique.....	36
2.2. Structures des unités lexicales du vietnamien	42
2.2.1. Mot et Syllabe	42
2.2.2. Structure syllabique : différents points de vue théoriques	46
2.3. Inventaire et distribution des phonèmes.....	50
2.3.1. L'inventaire des consonnes en attaque	50
2.3.2. La prétonale /w/	51
2.3.3. Le système du noyau	52
2.3.4. Les consonnes finales.....	54
2.4. Le système tonal	56
2.4.1. Description des tons du vietnamien (en syllabe isolée).....	56
2.4.2. Variation des tons dans la chaîne parlée.....	61
2.4.3. Ton et intonation	62
2.4.4. Accentuation.....	64
2.5. Comparaisons avec le système phonémique du français.....	65

2.5.1. Structures syllabiques	66
2.5.2. Systèmes consonantiques.....	66
2.5.3. Systèmes vocaliques.....	69
Chapitre 3.....	71
De la constitution d'un lexique du vietnamien à l'étude comparative des séquences de consonnes et de leur distribution en vietnamien - comparaison avec le français.....	71
3.1. Construction d'un lexique du vietnamien	72
3.1.1. Recueil des données lexicales	72
3.1.2. Intégration du lexique dans la base G-ULSID	74
3.2. Analyses quantitatives.....	75
3.2.1. Nombre de syllabes par item lexical	75
3.2.2. Rendement syllabique.....	77
3.2.3. Gabarits lexicaux du vietnamien.....	79
3.2.4. Structures syllabiques du vietnamien.....	80
3.3. Analyses quantitatives des séquences consonantiques du vietnamien	83
3.3.1. Séquences C + /w/ en vietnamien	86
3.3.2. Séquences consonantiques à l'inter-syllabique des mots composés en vietnamien	87
3.3.3. Comparaison avec les groupements de consonnes intra- et inter-syllabe du français	92
Chapitre 4.....	94
Corrélat acoustico-perceptifs des consonnes simples et en séquences du vietnamien	94
4.1. Expérience 1 : Étude acoustique des consonnes en séquences du vietnamien	95
4.1.1. Méthodologie.....	95
4.1.1.1. Constitution du corpus	95
4.1.1.2. Traitement des données	100
4.1.1.3. Analyses statistiques	102
4.1.2. Résultats.....	102
4.1.2.1. Effet de la durée	102
4.1.2.2. Caractéristiques acoustiques des bruits d'explosion (burst)	105
4.1.2.3. Caractéristiques du VOT	107
4.1.2.4. Transitions voyelle-plosive.....	107
4.1.2.4.1. Transitions de F1, F2 et F3	109
4.1.2.4.2. Transitions de l'intensité (I).....	114
4.1.2.4.3. Transitions de la fréquence fondamentale (F0)	116
4.1.3. Discussion.....	117
4.2. Expérience 2 : Identification des consonnes finales du vietnamien	121
4.2.1. Méthodologie.....	121
4.2.1.1. Stimuli.....	121

4.2.1.2. Déroulement du test.....	123
4.2.2. Résultats.....	125
4.2.3. Discussion.....	127
4.2.3.1. Fréquence lexicale.....	127
4.2.3.2. Indice acoustique : le burst.....	130
4.2.3.3. Indice acoustique : durée de la voyelle.....	131
4.2.3.4. Indice acoustique : la transition VC.....	132
4.2.3.5. Indice acoustique : les coefficients MFCC.....	135
4.2.4. Conclusion.....	136
4.3. Expérience 3 : Identification du type de frontière syllabique en vietnamien.....	137
4.3.1. Méthodologie.....	137
4.3.2. Résultats.....	138
4.3.3. Discussion.....	139
4.3.3.1. Durée de la syllabe.....	140
4.3.3.2. Durée de la consonne finale.....	141
4.3.3.3. Durée de la voyelle.....	143
4.3.4. Conclusion.....	144
4.4. Expérience 4 : Perception des constituants de la syllabe.....	144
4.4.1. Méthodologie.....	144
4.4.2. Résultats.....	146
4.4.2.1. Erreur d'identification de la voyelle.....	146
4.4.2.2. Erreur d'identification de la consonne finale.....	147
4.4.3. Conclusion.....	149
Chapitre 5.....	151
Perception et production des séquences de consonnes en langue seconde : le cas du français par des apprenants vietnamiens.....	151
5.1. La perception des consonnes simples et en séquences du français par des apprenants vietnamiens.....	153
5.1.1. Expérience 1 : Perception des consonnes simples du français.....	153
5.1.1.1. Méthodologie.....	154
5.1.1.1.1. Constitution du corpus.....	154
5.1.1.1.2. Participants.....	155
5.1.1.1.3. Déroulement du test.....	155
5.1.1.2. Résultats.....	156
5.1.1.2.1. Effet de la distribution.....	156
5.1.1.2.2. Effet du mode d'articulation.....	157
5.1.1.2.3. Effet du lieu d'articulation.....	158
5.1.1.3. Effet de similarités entre les deux langues : le cas des consonnes /p t k m n ŋ/.....	159
5.1.1.3.1. Effet de la position.....	160
5.1.1.3.2. Effet du lieu d'articulation.....	160

5.1.1.3.3. Analyse des erreurs	161
5.1.1.4. Discussion	162
5.1.2. Expériences 2 : Perception des groupements biconsonantiques en intra- et inter-syllabe du français	165
Expérience 2A : Perception des groupes de consonnes dans les pseudo-mots	166
5.1.2.1. Méthodologie.....	166
5.1.2.1.1. Constitution du corpus	166
5.1.2.1.2. Participants	167
5.1.2.1.3. Déroulement du test	167
5.1.2.2. Résultats.....	168
5.1.2.2.1. Effet de la position	169
5.1.2.2.2. Temps de réaction.....	170
5.1.2.2.3. Types de consonnes dans les combinaisons	171
5.1.2.3. Discussion	173
Expérience 2B : Perception des groupements biconsonantiques dans des mots du français.....	176
5.1.2.4. Méthodologie.....	177
5.1.2.4.1. Corpus	177
5.1.2.4.2. Participants	177
5.1.2.4.3. Déroulement du test	177
5.1.2.5. Résultats.....	178
5.1.2.5.1. Effet de la position	178
5.1.2.5.2. Temps de réaction.....	179
5.1.2.5.3. Types de consonnes dans les combinaisons	180
5.1.2.6. Discussion	182
5.2. Expérience 3 : Production de clusters et de séquences biconsonantiques du français par des apprenants vietnamiens	185
5.2.1. Méthodologie.....	187
5.2.1.1. Corpus.....	187
5.2.1.2. Participants.....	187
5.2.1.3. Déroulement du test.....	188
5.2.1.4. Traitement des données	188
5.2.2. Résultats.....	188
5.2.2.1. Effet de la position des clusters	189
5.2.2.2. Effet du type de consonne dans les combinaisons	190
5.2.2.2.1. Les combinaisons en attaque.....	190
5.2.2.2.2. Les combinaisons en coda	191
5.2.2.2.3. Les combinaisons inter-syllabe	192
5.2.2.3. Les stratégies de simplification	193
5.2.2.3.1. Modification par non relâchement.....	193
5.2.2.3.2. Suppression.....	196
5.2.2.3.3. Epenthèse	199
5.2.2.3.4. Substitution	202
5.2.3. Discussion	205

Discussion générale et perspectives	213
Bibliographie.....	222
Annexes.....	247

Liste des tables

Table 2.1. L'origine des tons en vietnamien de Haudricourt (1954)	40
Table 2.2. Représentation schématique de l'organisation interne de la syllabe	47
Table 2.3. Représentation schématique de l'organisation interne de la syllabe en vietnamien d'après Đoàn Thiện Thuật (1999).	48
Table 2.4. Structure syllabique du vietnamien avec entre parenthèses le nombre de phonèmes pouvant apparaître dans chaque position (axe paradigmatique)	50
Table 2.5. Les types de structure syllabique du vietnamien.....	50
Table 2.6. Les consonnes du vietnamien en position d'attaque.	51
Table 2.7. Combinaisons possibles entre la prétonale et les initiales	52
Table 2.8. Combinaisons possibles entre la prétonale et les voyelles	54
Table 2.10. Combinaisons possibles entre voyelles et semi-consonnes finales.	55
Table 2.11. Système des tons du vietnamien (* : B1 et B2 à sonantes finales, D1 et D2 à plosives finales selon Michaud, 2004).	56
Table 2.12. Compatibilité entre les tons et le type de consonne finale en vietnamien.....	57
Table 3.1. Structures des gabarits lexicaux et leur fréquence d'apparition dans le lexique du vietnamien (5 000 entrées).....	79
Table 3.2. Structures représentant les gabarits lexicaux et leur fréquence d'apparition dans le lexique du français	79
Table 3.3. Structures représentant plus de 2 % des gabarits de l'ensemble des 16 lexiques de G-ULSID.	80
Table 3.4. Structures syllabiques classées par type et leur proportion dans le lexique du vietnamien .	81
Table 3.5. Structures syllabiques classées par type et leur proportion dans le lexique du français	81
Table 3.6. Types de séquences consonantiques dans le lexique du vietnamien	84
Table 3.7. Structures des gabarits lexicaux qui contiennent des groupes consonantiques en français...	84
Table 3.8. Regroupement des consonnes du vietnamien selon leurs lieu (à gauche) et mode d'articulation (à droite).....	85
Table 3.9. Nombre d'occurrences et pourcentage des séquences C+/w/ classées selon le lieu d'articulation.....	86
Table 3.10. Nombre d'occurrences et pourcentage des séquences C+/w/ classées selon le mode d'articulation.	87
Table 3.11 : Cooccurrences C ₂ et C ₃ selon le lieu d'articulation.....	89

Table 3.12. Matrice de cooccurrences entre consonnes dans les séquences C ₂ .C ₃ selon leur mode d'articulation.....	91
Table 3.13. Cooccurrences des consonnes dans les séquences C ₂ .C ₃ selon leur lieu d'articulation en vietnamien.....	92
Table 4.1. Liste des 62 items sélectionnés à partir du lexique vietnamien pour l'analyse acoustique des consonnes plosives et nasales. Les consonnes ciblées par l'étude sont mises en évidence en caractère gras.....	97
Table 4.2a. Phrase porteuse pour les mots cibles terminés par une sonante. Ici le mot cible est « ma ».	99
Table 4.2b. Phrase porteuse pour les mots cibles terminés par une plosive sourde. Ici le mot cible est « mât ».....	99
Table 4.3. Valeur des seuils du test de significativité pour les valeurs moyennes de F1, F2, F3 selon le lieu des plosives en C ₂ vs. C ₂ .C ₃ mesurée de 40 % à 90 % avec un pas de 10 de la durée de la voyelle (**** si p ≤ 0,001, *** si p ≤ 0,005, ** si p ≤ 0,01, * si p ≤ 0,05).....	109
Table 4.4. Valeur des seuils du test de significativité pour les valeurs moyennes de F1, F2, F3 selon le ton de la syllabe mesurée de 40 % à 90 % avec un pas de 10 de la durée de la voyelle (**** si p ≤ 0,001, *** si p ≤ 0,005, ** si p ≤ 0,01, * si p ≤ 0,05).....	110
Table 4.5. Valeur des seuils de significativité des différences de ΔF1, ΔF2, ΔF3 selon les facteurs lieu, et ton de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si p ≤ 0,001, *** si p ≤ 0,005, ** si p ≤ 0,01, * si p ≤ 0,05).....	113
Table 4.6. Comparaison des deltas F1, F2, F3 entre consonne orale et nasale (* différence significative (p < 0,05)).	113
Table 4.7. Valeur des seuils de significativité des différences de ΔI selon les facteurs lieu, ton de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si p ≤ 0,001, *** si p ≤ 0,005, ** si p ≤ 0,01, * si p ≤ 0,05).....	115
Table 4.8. Valeur des seuils de significativité des différences de ΔF0 selon les facteurs « lieux », « tons » de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si p ≤ 0,001, *** si p ≤ 0,005, ** si p ≤ 0,01, * si p ≤ 0,05).....	117
Table 4.9. Liste des 55 stimuli du vietnamien pour le test de perception des consonnes finales tirés du corpus acoustique	122
Table 4.10. Scores d'identification des plosives.	125
Table 4.11. Scores d'identification des nasales.	125
Table 4.12. Score d'identification en fonction de la présence/absence d'un burst et de la fréquence d'occurrences des plosives dans les stimuli du test.....	131
Table 4.13. Valeurs moyennes et écarts-types des pentes de F1, F2, F3, F0 (Hz) et de l'intensité (dB) en fonction de la consonne en coda (/p t k/) calculés pour les productions du locuteur retenu pour établir les stimuli du test perceptif.	132

Table 4.14. Seuils de significativité des différences des deltas F1, F2, F3, F0, I dans la partie de transition VC (noyau et coda) (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).....	133
Table 5.1. Liste des mots et pseudo-mots utilisés pour le test de perception et production des consonnes simples du français.....	154
Table 5.2. Taux de bonnes productions des consonnes (du plus fort au plus faible) selon leur lieu d'articulation, tout niveau d'apprentissage confondu.	159
Table 5.3. Traits d'erreurs dans la production des consonnes /p t k m n ŋ/, niveaux d'apprentissage confondus.	161
Table 5.4. Traits d'erreurs dans les productions des consonnes /p t k m n ŋ/ en fonction de leur position dans les pseudo-mots (totalisés pour les deux niveaux confondus).	162
Table 5.5. Séquences de consonnes communes aux deux langues sélectionnées pour le test, classées selon leur position (en initiale (#CC) et en finale (CC#) uniquement pour le français, en inter-syllabique (C.C) pour les deux langues, le point indique la frontière syllabique attendue en français.	166
Table 5.6. Résultats statistiques des contrastes intra-sujet en fonction des positions des séquences /p/+C et /k/+C dans les pseudo-mots.	170
Table 5.7. Unités lexicales du français sélectionnées pour l'expérience et classées selon la position des groupes de consonnes dans le mot et dans la syllabe.....	176
Table 5.8. Unités lexicales du français sélectionnées pour l'expérience et classées selon leur position dans le cluster ou la séquence consonantique ainsi que dans le mot et dans la syllabe.....	187
Table 5.9. Taux de non relâchement des C ₁ initiant des clusters en coda en fonction du type de consonnes impliquées en C ₂	196
Table 5.11. Taux de suppression dans les clusters en coda en fonction du mode des consonnes impliquées.	197
Table 5.12. Taux de suppression dans les clusters en coda en fonction du lieu des consonnes impliquées.	197
Table 5.13. Taux de suppression dans les séquences inter-syllabe en fonction du mode des consonnes impliquées.	197
Table 5.14. Taux de suppression dans les séquences inter-syllabe en fonction du lieu des consonnes impliquées.	198
Table 5.15. Taux de suppression des clusters en attaque en fonction du mode des consonnes impliquées.	198
Table 5.16. Fréquence de suppression des consonnes des groupes biconsonantiques.....	198
Table 5.17. Taux d'épenthèse dans les clusters en coda en fonction du mode d'articulation des consonnes successives.	200

Table 5.18. Taux de substitution selon la position des groupes consonantiques et le niveau des étudiants	202
Table 5.19. Taux de substitution des clusters en coda de mot en fonction de la combinaison des consonnes.	202
Table 5.20. Taux de suppression des séquences inter-syllabe en fonction du type de consonnes impliquées.	203
Table 5.21. Taux de substitution dans les clusters en attaque de mot en fonction de la combinaison des consonnes (C ₁ = plosives).....	204

Liste des figures

Figure 1.1. Les types d'assimilation perceptive des contrastes non natifs (Best, 1995).	17
Figure 1.2. Echelle de sonorité (selon Broselow et Finer, 1991).....	31
Figure 2.1. Représentation arborescente de l'organisation interne de la syllabe vietnamienne selon Hoàng Tuệ et Hoàng Minh (1975) (d'après Cao Xuân Hạo, 1985, p. 271).....	47
Figure 2.2. Description en traits de lieu et d'aperture du système vocalique du vietnamien.	52
Figure 2.3. Evolution temporelle des tons du vietnamien (d'après Andreev & Gordina, 1957).....	57
Figure 2.4. Représentation temporelle de l'évolution des tons du vietnamien (Michaud, 2004).....	59
Figure 2.5. Illustration laryngoscopique de la constriction épiglotale (Brunelle et al., 2010).....	60
Figure 2.6. Représentation temporelle de l'évolution des tons du dialecte du Nord et du Sud (Brunelle, 2009).....	60
Figure 2.7. Représentation schématique de l'organisation interne de la syllabe en français (Rousset, 2004).....	66
Figure 2.8. Représentation schématique de l'organisation interne de la syllabe en vietnamien (selon Đoàn Thiện Thuật, 1999 et Trần Đỗ Đạt & al., 2005)	66
Figure 2.9. Sonagrammes des syllabes [ga] en français et [ɣa ¹] en vietnamien.	67
Figure 2.10. Spectrogrammes de « loupe » [lup] français et vietnamien.....	68
Figure 2.11. Description en traits d'aperture et de lieu des systèmes vocaliques du vietnamien et du français.....	69
Figure 3.1. Représentation syllabique de l'unité lexicale [? a n . n i ŋ].	74
Figure 3.2. Typologie des 16 langues de G-ULSID, élaborée sur la répartition du lexique en fonction du nombre de syllabes par item lexical (Rousset, 2004, p. 90).....	76
Figure 3.3. Distribution du lexique vietnamien en fonction du nombre de syllabes par unité lexicale.	77
Figure 3.4 Rendements syllabiques des langues de G-ULSID (Rousset, 2004, p. 121)	78
Figure 3.5. Proportions des différents types syllabiques du vietnamien et du français.....	82
Figure 3.6. Proportions des structures syllabiques avec et sans séquences C + /w/ du vietnamien.	84
Figure 3.7. Pourcentage d'occurrences des séquences C +/w/ sur l'ensemble des suites de consonnes en vietnamien, classés selon le mode d'articulation.	86
Figure 3.8. Pourcentage des séquences consonantiques à l'inter-syllabique des mots composés sur l'ensemble des suites de consonnes du vietnamien.....	87

Figure 3.9. Pourcentage d'occurrences des différentes modes consonantiques rencontrées en position de coda C_2 d'un mot composé $C_1VC_2.C_3VC_4$	88
Figure 3.10. Pourcentage d'occurrences des consonnes en coda (C_2) classée selon leur lieu d'articulation sur l'ensemble des suites de consonnes en vietnamien.	88
Figure 3.11. Pourcentage d'occurrences des consonnes en position C_3 classées selon leur lieu d'articulation sur l'ensemble des suites de consonnes en vietnamien.	89
Figure 3.12. Pourcentage d'occurrences des consonnes C_3 classées selon leur mode d'articulation sur l'ensemble des suites de consonnes en vietnamien.	90
Figure 3.13. Pourcentage d'occurrences des clusters et des séquences de consonnes classés selon leur mode d'articulation.....	93
Figure 3.14. Pourcentage d'occurrences des clusters et des séquences de consonnes classés selon leur lieu d'articulation.....	93
Figure 4.1. Exemple de segmentation des mots et consonnes cibles sous Praat.	100
Figure 4.2. Durée moyenne des plosives et des nasales en fonction de leurs positions dans la syllabe.	103
Figure 4.3. Durée moyenne de la voyelle /a/ suivie par des plosives et des nasales en fonction de la position finale de mot vs. finale de syllabe 1 de mot composé.....	104
Figure 4.4. Durée moyenne de l'occlusion des plosives selon leur position dans la syllabe.....	105
Figure 4.5. Pourcentage de plosives réalisées avec burst en fonction de leur position dans la syllabe et dans le mot.	106
Figure 4.6. Valeurs moyennes de la durée et de l'intensité des bursts	106
Figure 4.7. Durée moyenne du VOT des différentes plosives selon le type de frontière syllabique (inter-mot vs. intra-mot).	107
Figure 4.8. Spectrogramme du mot « gác bút » [ɣak ⁵ -but ⁵].	108
Figure 4.9. Evolution de $\Delta F1$ estimée de 40% jusqu'à 90% de la durée de la voyelle en fonction de la plosive en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible.....	111
Figure 4.10. Evolution de $\Delta F2$ estimée de 40% jusqu'à 90% de la durée de la voyelle en fonction de la consonne en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible	111
Figure 4.11. Evolution de $\Delta F3$ estimée de 40% jusqu'à 90% de la durée de la voyelle en fonction de la consonne en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible	112
Figure 4.12. Analyse en composantes principales des coefficients MFCC de la transition à 90 % de la durée totale de /a/ en contextes des plosives /t k p/ et des nasales /ŋ n m/ où N = /ŋ/.....	114
Figure 4.13. Evolution de l'intensité estimée de 40 % jusqu'à 90 % de la durée totale de la voyelle en fonction de la plosive en coda (p, t, k) et du ton de la syllabe qui suit.....	115

Figure 4.14. Evolution de ΔF_0 estimée de 40% jusqu'à 90% de la durée de la voyelle en fonction de la plosive en coda (/p/, /t/, /k/) et du ton de la syllabe qui suit	116
Figure 4.15. Spectrogramme du mot « phát hành » [fa ⁵ .həŋ ²] (<i>diffuser</i>).	119
Figure 4.16. Spectrogramme du mot « phá thành » [fa ⁵ .t ^h əŋ ²] (<i>démolir le rempart</i>).	119
Figure 4.17. Interface du test perceptif utilisé pour les plosives.	124
Figure 4.18. Interface du test perceptif utilisé pour les nasales.	124
Figure 4.19. Score d'identification des consonnes finales par lieu et mode d'articulation.	126
Figure 4.20. Score d'identification des consonnes finales selon leur position dans la syllabe.	126
Figure 4.21. Score d'identification des consonnes finales selon le sexe des sujets.	127
Figure 4.22. Calcul des occurrences des consonnes en coda dans le lexique du vietnamien de 5 000 entrées.	128
Figure 4.23. Corrélation entre scores de bonnes réponses et nombre d'occurrences des plosives et des nasales en coda.	128
Figure 4.24. Nombre d'occurrences des consonnes en coda en comparaison avec la position d'attaque dans les unités lexicales du vietnamien.	129
Figure 4.25. Fréquence lexicale des consonnes en coda en contexte de la voyelle /a/ et le ton D1.	129
Figure 4.26. Corrélation entre scores de bonnes réponses et fréquence d'occurrences des plosives et des nasales en coda, en contexte de la voyelle /a/ et du ton montant D1.	130
Figure 4.27. Score de bonnes réponses en fonction de la présence vs. absence du burst et du type de frontière : intra-mot (C ₂ .C ₃), inter-mot (C ₂), et des lieux d'articulation.	131
Figure 4.28. Pentes (et écart-types) des transitions formantiques de F1, F2, F3 (Hz) mesurées entre 50 % et 90 % de la durée de la voyelle en fonction de la consonne en coda (/p t k/).	133
Figure 4.29. Analyse en composantes principales des coefficients MFCC-PCA de la transition à 90% de /a/ dans les contextes consonantiques /p t k m n ŋ/ pour les productions du locuteurs natif qui font l'objet du stimuli du test perceptif.	135
Figure 4.30. Interface du test perceptif utilisé pour les plosives.	138
Figure 4.31. Scores d'identification du type de frontière syllabique pour des syllabes à plosives finales (à gauche) et des syllabes à nasales finales.	138
Figure 4.32. Scores d'identification du type de frontière syllabique pour des syllabes à consonnes finales /p t k m n ŋ/.	139
Figure 4.33. Durée des syllabes à plosive finale et à nasale finale	140
Figure 4.34. Durée des syllabes à plosive finale et à nasale finale.	141
Figure 4.35. Durée moyenne des plosives et des nasales en finale	142
Figure 4.36. Durée moyenne des consonnes en finale en fonction de la structure syllabique	142

Figure 4.37. Durée moyenne de la voyelle suivie d'une plosive finale et d'une nasale finale	143
Figure 4.38. Durée moyenne de la voyelle en fonction de la consonne finale de syllabe	143
Figure 4.39. Interface du test perceptif de la syllabe.....	145
Figure 4.40. Résultats pour l'identification de la voyelle.....	146
Figure 4.41. Identification de la voyelle /a/ comme brève en fonction du mode et du type de structure syllabique.....	147
Figure 4.42. Perception de la voyelle /a/ comme brève en fonction de la structure syllabique pour chaque consonne.....	147
Figure 4.43. Erreur dans l'identification de la consonne finale selon le mode et la structure syllabique et selon le lieu.....	148
Figure 4.44. Erreur dans l'identification de la coda en fonction du type de consonne.....	148
Figure 5.1. Taux de production correcte des consonnes selon leur position dans les pseudo-mots et selon le niveau des étudiants en FLE.....	157
Figure 5.2. Taux de bonnes productions des consonnes simples du français selon leur position dans les pseudo-mots par les étudiants du niveau intermédiaire.....	157
Figure 5.3. Taux de bonnes productions des consonnes simples du français selon leur position dans les pseudo-mots par les étudiants du niveau avancé.....	158
Figure 5.4. Taux de bonnes productions des consonnes selon leur mode d'articulation et le niveau des apprenants.....	158
Figure 5.6. Taux de production correcte des consonnes /p t k m n ŋ/ pour les trois positions dans les pseudo-mots par les étudiants des niveaux intermédiaire et avancé.....	160
Figure 5.7. Les trois échantillons [aŋ] du locuteur français intégrés dans la liste des stimuli.....	164
Figure 5.8. Interface du test utilisé pour l'expérience perceptive des séquences de consonnes dans les pseudo-mots.....	168
Figure 5.9. Score d'identification correcte des séquences biconsonantiques en fonction de C ₁ , de la position dans le pseudo-mot entendu et le niveau des apprenants.....	169
Figure 5.10. Taux d'identification correcte des initiales complexes de pseudo-mot selon le type de combinaison	171
Figure 5.11. Taux d'identification correcte des codas complexes selon le type de combinaison.....	172
Figure 5.12. Taux d'identification correcte des séquences inter-syllabe selon le type de combinaison.....	173
Figure 5.13. L'échantillon [pna] du locuteur français intégré dans la liste des stimuli.	174
Figure 5.14. Scores d'identification correcte des clusters et des séquences de consonnes du français en fonction de la position de leur C ₁ et du niveau des apprenants	178
Figure 5.15. Taux d'identification correcte des clusters en initiale selon le type de combinaison.....	180

Figure 5.16. Taux d'identification correcte des clusters en finale selon le type de combinaison.	181
Figure 5.17. Taux d'identification correcte des séquences consonantiques inter-syllabe selon leur mode articulatoire	181
Figure 5.18. Proposition de clusters produits modifiés par rapport à la cible selon leur position.	189
Figure 5.19. Propositions de modifications des clusters /p/+C et /k/+C par rapport à la cible selon leur position dans le mot et selon le niveau des apprenants.	190
Figure 5.20. Taux d'erreurs de production des attaques complexes selon le type de combinaison des consonnes dans les séquences et le niveau des étudiants.	191
Figure 5.21. Taux d'erreurs dans la production des codas complexes du français selon le type de combinaison consonantique et le niveau des étudiants.	192
Figure 5.22. Taux d'erreur de production de séquences de consonnes inter-syllabe selon le type de combinaison des consonnes impliquées et le niveau des étudiants.	192
Figure 5.23. Nombre de modifications observées dans les groupes de consonnes (clusters et séquences inter-syllabe) en fonction des stratégies utilisées par les apprenants des deux niveaux.....	193
Figure 5.24. Production du mot « hot-dog » par un apprenant de niveau intermédiaire. Cas du non relâchement de la première consonne du groupe [td].....	194
Figure 5.25. Production du mot « adepte » par un apprenant de niveau intermédiaire. Cas du non relâchement de la première consonne du groupe [pt].....	195
Figure 5.26. Taux de non relâchement des plosives en C ₁ des groupes inter-syllabe et en coda, selon le niveau intermédiaire et avancé des apprenants.....	195
Figure 5.27. Taux de suppression d'un ou des deux éléments du groupe en fonction de la position des groupes consonantiques dans le mot, la syllabe et le niveau des apprenants.	195
Figure 5.28. Exemple de prononciation du mot « phénix » par un apprenant de niveau avancé où l'élément moins sonore [k] du groupe est supprimé.....	199
Figure 5.29. Exemple de production du mot « jeans » par un apprenant de niveau intermédiaire.	199
Figure 5.30. Taux d'épenthèse en fonction de la position des clusters dans le mot et le niveau des étudiants.....	200
Figure 5.31. Exemple de production du mot « adepte » par un apprenant de niveau intermédiaire. .	201
Figure 5.32. Exemple de production du mot « cycle » par un apprenant de niveau avancé.....	201
Figure 5.33. Exemple de production du mot « ptose » par un apprenant de niveau intermédiaire.....	201
Figure 5.34. Exemple de production du mot « rythme » par un apprenant de niveau avancé.....	203
Figure 5.35. Exemple de production du mot « rumba » par un apprenant de niveau avancé.....	204
Figure 5.36. Illustration d'une des productions du mot « psychique » par un apprenant de niveau intermédiaire.....	204

Introduction

« Pour apprendre à prononcer, il faut des années et des années. Grâce à la science, nous pouvons y parvenir en quelques minutes »

« La leçon », Eugène Ionesco

Dans le contexte de l'ouverture des frontières et de la mondialisation des échanges, du développement des transports et des voyages longue-distance, la pratique fluente de langues étrangères est devenue un enjeu essentiel, économique, professionnel et culturel. L'enseignement des langues étrangères connaît, dans cette situation, un essor important.

Dans le domaine de la didactique, les apprentissages sont souvent valorisés de façon traditionnelle par la réflexion logique et la mémorisation (qui concernent surtout la grammaire et le vocabulaire). Pourtant, l'apprentissage d'une langue vivante ne se conçoit pas en dehors de l'apprentissage de la prononciation en ce que la mise en œuvre de la langue se fait le plus souvent à l'oral. C'est alors que la prononciation prend toute sa valeur. Or la parole, notamment lors des échanges avec des natifs, a reçu, en théorie didactique ces dernières décennies, peu d'attention par rapport aux autres domaines de la langue (Lauret, 2007). Des explications variées peuvent être apportées à ce constat, parmi lesquelles le fait que la prononciation d'une langue étrangère est le plus souvent une matière peu valorisée par l'institution enseignante. Cet état de chose résulte du manque de formation/d'information des enseignants en phonétique corrective et d'une faible présence des aspects phonético-phonologiques dans les manuels généraux actuels d'apprentissage des langues étrangères, comme ils sont peu présents aujourd'hui dans les programmes scolaires et l'évaluation (Lauret, 2007, p. 13).

Il est évident qu'une prononciation correcte est essentielle dans l'apprentissage d'une langue étrangère, car la maîtrise de la prononciation facilite le processus de compréhension et donc de communication. Depuis les années 30, les études sur l'acquisition des langues étrangères cherchent à expliquer et à comprendre les difficultés de prononciation des apprenants. Il a été montré, dès les premières études de ce champ de recherche, que si la maîtrise de la prononciation d'une langue étrangère est difficile, c'est parce que le système de la langue maternelle rend « sourd » aux systèmes des autres langues (Polivanov, 1931 ; Troubetzkoy, 1939). Par conséquent, il est très souvent reconnu que l'expérience linguistique

que l'apprenant a déjà acquise lors de l'apprentissage de sa langue première (L1) constitue un élément essentiel du processus d'acquisition de la phonétique et de la phonologie d'une langue seconde¹ (L2) que ce soit en production ou en perception. Ainsi, en cours d'apprentissage, il est couramment observé que l'apprenant, confronté au système phonético-phonologique d'une autre langue, éprouve des difficultés avec les unités sonores de cette langue qui n'existent pas dans sa langue maternelle (entre autres, Lado, 1957 ; Eckman, 1977, 2004 ; Odlin, 1989 ; Renard, 1989 ; Best, 1995 ; Flege 1995). Ces unités et structures manquantes dans la langue maternelle sont plus ou moins bien perçues suivant les cas et plus ou moins bien reproduites, parce que l'apprenant perçoit ce qu'il a appris à percevoir, c'est-à-dire ce qui est significatif dans sa langue, à travers ce que Troubetzkoy (1939) appelle le *crible phonologique*. Ce filtre, aux mesures de la langue maternelle, ne laisse passer que les éléments sonores de la langue seconde qui sont présents tels quels dans la langue maternelle. De ce fait, l'apprenant d'une langue seconde ne pourra pas produire ou répéter correctement les éléments de la langue seconde qui ne sont pas déjà existants dans sa langue maternelle parce que incapable de les percevoir correctement. C'est le principe de base même des travaux de Guberina (1965/1984) sur la « linguistique de la parole ». Ainsi,

« Apprendre une langue étrangère apparaît à plus d'un titre comme une remise en cause et une réorganisation partielle des conditionnements et des habitudes de perception propres à la langue et à la culture d'origine » (Galisson & Coste, 1976).

De fait, les tentatives de description/compréhension du processus d'acquisition d'une langue seconde se font rarement sans recours à la notion d'*interlangue* (Ioup & Weinberger, 1987). L'idée derrière ce concept, avancé par Selinker (1972), est que chaque apprenant construit sa propre « version » de la L2 (*learning language*). Selinker définit l'interlangue comme l'existence d'un système structurel construit par l'apprenant à tout moment du développement de son apprentissage de la langue cible (voir aussi Ellis, 1994) et précise que la structure de l'interlangue varie par conséquent sur une échelle progressive entre la langue de départ native et la langue cible apprise. L'interlangue est cependant indépendante de l'une et de l'autre, même si elle leur est inévitablement liée. Ainsi, l'interlangue d'un individu peut reproduire des particularités de sa langue maternelle, mais aussi de la langue cible, de même que d'autres caractéristiques qui n'appartiennent ni à l'une, ni à l'autre, mais à des généralisations universelles des langues du monde (Eckman, 1991). Plus l'individu progresse dans sa connaissance de la langue cible et améliore sa compétence communicative, plus son interlangue se détache graduellement des structures de la langue de départ pour se

¹ Le terme *langue seconde* est utilisé ici dans le sens de la « langue dont l'usage est acquis à une étape de la vie ultérieure à l'apprentissage de la *langue première* ou *maternelle* » (Dubois, Giacomo, Guespin, Marcellesi & Mével, 2001, p. 417).

rapprocher de celles de la langue cible. Les interlangues constituent ainsi des systèmes « intermédiaires » d'acquisition de la langue. Plusieurs études ont été menées dans ce champ de recherche avec un objectif principal, celui de comprendre comment les habitudes en perception et production de la parole des non-natifs et/ou les universaux typologiques influencent l'apprentissage d'une langue seconde, notamment ses aspects phonétiques et phonologiques (entre autres Lado, 1957 ; Eckman, 1981 ; Sheldon et Strange, 1982 ; Broselow, 1983 ; Renard, 1989 ; Archibald, 1993 ; Flege, 1995 ; Best, 1995 ; Broselow, Chen et Wang, 1998 ; Hancin-Bhatt, 2000). Selon certains auteurs comme Borrell (1992), il y aurait, chez les apprenants d'une langue seconde, « systématisation des erreurs de perception et donc de production. Pour qu'un son soit correctement (re)produit il faut d'abord qu'il soit bien perçu, c'est-à-dire reconnu, identifié par rapport aux autres » (Borrell, 1992).

Le travail présenté dans ces pages s'inscrit dans ce cadre de recherche sur l'apprentissage des langues étrangères et traite plus particulièrement de l'acquisition des clusters et autres séquences de consonnes en langue seconde en tentant de poser des éléments de la structure phonologique de l'interlangue à partir 1/ de l'analyse acoustico-perceptive des séquences consonantiques de la langue première et 2/ de l'analyse de la perception et production des clusters et autres séquences de consonnes de la langue cible. Les deux langues observées sont le vietnamien (langue source) et le français (langue cible).

Vietnamien et français sont deux langues pour lesquelles de nombreuses dissemblances phonético-phonologiques existent dont celle qui nous intéresse en particulier ici : les groupes consonantiques. Ils peuvent être inter-syllabe, c'est-à-dire répartis de part et d'autre d'une frontière syllabique : c'est le cas en vietnamien et en français. Ils peuvent être intra-syllabe, c'est-à-dire situés dans un constituant sub-syllabique, soit l'attaque (initiale), soit la coda (finale) : c'est le cas en français mais pas en vietnamien.

Les enseignants du Français Langue Étrangère (FLE) relatent souvent que les étudiants vietnamiens font partie de ces apprenants qui éprouvent des difficultés récurrentes dans la réalisation des groupes de consonnes. Les suites de deux ou plusieurs consonnes n'étant pas permises en vietnamien à l'intérieur d'une même syllabe, elles sont la plupart du temps réalisées déformées par rapport à la cible, entraînant chez l'auditeur incompréhension ou malentendu (ex. [ritm] *rythme* prononcé [rim] *rime* ; [taks] *taxe* prononcé [tas] *tasse* ; [filtɚ] *filtre* prononcé [fil] *fil*). Ainsi les multiples variantes de prononciation observées chez les Vietnamiens pour les personnages créés par Peyo, auteur de la bande dessinée belge *Les Schtroumpfs*, sont un clin d'œil à cet inconvénient à ne pas réussir la production de plusieurs consonnes successives. Et ces difficultés persistent quel que soit le nombre d'années d'apprentissage (même après plusieurs années de pratique, même chez des étudiants de niveau déjà confirmé) ou d'exposition à la langue (Nguyễn Thị Bình Minh, 2000). Quelles sont alors les véritables raisons de ces difficultés ? Quels sont les éléments du crible phonologique qui gênent ou empêchent l'acquisition des clusters ? Quels sont leurs

implications dans l'acquisition des percepts phonétiques de la langue cible ? En quoi consistent exactement les erreurs de réalisation ? Quelle est la part des caractéristiques de la langue maternelle et quelle est la part d'autres facteurs tels ceux relevant des principes universels des systèmes phonologiques des langues ? Peut-on trouver dans cette problématique de l'acquisition des clusters des éléments propres à l'interlangue ? Tenter de répondre à ces questions inscrit également cette recherche dans le cadre d'une problématique de l'enseignement et de l'apprentissage de la prononciation à l'usage des apprenants vietnamiens du FLE.

Les approches aujourd'hui retenues pour permettre le développement théorique et de la modélisation des processus d'acquisition en langue seconde sont essentiellement expérimentales. Dans cette voie, la démarche scientifique répondant à la problématique centrale qui fait l'objet de cette thèse se décline en cinq parties faisant chacune l'objet d'un chapitre.

Le premier chapitre « *Acquisition de la phonologie en langue seconde. Le cas des groupes de consonnes* » resitue le cadre théorique de l'apprentissage des langues, présentant les travaux majeurs sur l'acquisition des aspects phonético-phonologiques d'une langue seconde en général et celle des groupes de consonnes en particulier. En partant de l'observation et de l'analyse des difficultés rencontrées par les apprenants, ces travaux proposent des explications et des hypothèses basées sur l'influence de la langue native ou sur certaines propriétés universelles des langues et donc du langage. Ces travaux serviront de base à la discussion des résultats des différentes études expérimentales conduites dans le cadre de cette thèse et qui sont exposées dans les chapitres suivants. Comprendre l'interlangue et les processus d'acquisition des langues étrangères comporte une dimension cognitive, puisqu'il s'agit aussi de comprendre le fonctionnement du langage et de son acquisition, l'émergence et la structuration des systèmes phonologiques et donc d'une manière plus générale de développer des connaissances sur le fonctionnement du cerveau humain en matière de représentation.

En accord avec les méthodes en acquisition des langues secondes, le deuxième chapitre « *Phonologie du vietnamien : description et comparaison avec le français* » est consacré à l'étude de la phonologie du vietnamien, langue source, en comparaison avec le système phonémique du français, langue cible. Cette partie explore particulièrement les données de la littérature dans lesquelles les points de divergences dans les descriptions de la phonologie du vietnamien se manifestent, tant chez les descripteurs que chez les spécialistes du traitement automatique de la langue vietnamienne.

Le troisième chapitre, en lien avec le questionnement central de la thèse, et intitulé « *De la constitution d'un lexique du vietnamien à l'étude comparative des séquences de consonnes et leur distribution en vietnamien, comparaison avec le français* » traite en particulier les différences de

distribution des séquences consonantiques entre les deux langues. Comme peu de données sont disponibles à l'heure actuelle, le choix a été fait, dans le cadre de cette recherche doctorale, de procéder à une étude descriptive et quantitative à partir d'un travail de transcription phonologique et de syllabation d'un lexique de plusieurs milliers de lemmes du vietnamien. Les éléments d'analyse de cette étape permettent d'une part d'établir des corpus et, d'autre part, d'interpréter les résultats des expériences ultérieures. En effet, de cette étape ont été extraites les séquences de consonnes communes aux deux langues qui ont servi de base à l'établissement des corpus de la partie expérimentale.

Le volet expérimental de la thèse est divisée en deux grandes parties correspondant chacune à un chapitre. Le chapitre IV « *Corrélat acoustico-perceptifs des consonnes simples et en séquences du vietnamien* » présente un ensemble de travaux expérimentaux qui visent à rendre compte des paramètres acoustico-perceptifs des consonnes du vietnamien en fonction de leur distribution (position dans le mot, dans la syllabe) afin de rechercher d'éventuels effets de position, de frontière ou d'environnement consonantique. Cette étape sur la recherche de paramètres physiques acoustiques et leurs corrélats perceptifs est nécessaire pour comprendre s'il existe un effet de la L1 sur l'acquisition des clusters du français, fournissant ainsi d'éventuels éléments du crible phonologique et son rôle dans les contraintes d'apprentissage.

Le chapitre V « *Perception et production des groupes de consonnes en langue seconde. Le cas du français par des apprenants vietnamiens* » concerne plus particulièrement la perception et production des groupes de consonnes du français par des apprenants vietnamiens de deux niveaux, intermédiaire et avancé. Il regroupe un ensemble d'études expérimentales sur la perception et la production des consonnes simples et en séquence du français en tenant compte des facteurs distributionnels que sont la position dans le mot, dans la syllabe ainsi que la nature des segments consonantiques constituants ; l'objectif étant de situer et de comprendre les difficultés rencontrées par les étudiants vietnamiens, même de niveau avancé, à réaliser les clusters du français. Les résultats sont analysés par rapport aux facteurs pouvant être lié au crible phonologique de la L1 et à l'influence de la perception sur la production, puis ils sont discutés par rapports aux éléments théoriques et de modélisation des travaux antérieurs, présentés dans le chapitre I.

Enfin, une dernière partie intitulée « *Discussion générale et perspectives* » resitue les éléments théoriques alimentés par l'ensemble des résultats obtenus et propose un prolongement expérimental immédiatement envisageable à des fins théoriques et appliquées dans le domaine de la didactique. Ainsi, au-delà du questionnement théorique sur l'acquisition de la forme orale des langues étrangères, la thèse se situe aussi dans les enjeux de la formation des enseignants et des avancées théoriques et technologiques pour le développement d'outils de phonétique corrective.

Chapitre 1

Acquisition de la phonologie en langue seconde : le cas des groupes de consonnes

Les études en acquisition de la phonologie d'une langue seconde cherchent à expliquer et comprendre les processus d'apprentissage, les éléments favorisant comme les difficultés rencontrées par les apprenants. De nombreux travaux des dernières décennies ont montré que « *L2 learners construct mental grammars that are constrained by general and independently motivated principles* » (Eckman, 2004). A partir du milieu du siècle dernier, de nombreux travaux ont étudié ces contraintes dans deux domaines importants : celui de la langue première des apprenants (L1) et celui des propriétés universelles des langues et donc du langage.

Les premières recherches en acquisition d'une langue seconde se sont centrées sur le rôle de la langue première (Leather, 1999 ; Major, 2001). De nombreux travaux ont suggéré que phonétique et phonologie de la L1 ont une influence importante sur la prononciation des sons de la L2 (Odlin, 1989 ; Best, 1995 ; Flege, 1995 entre autres). Il a fallu du temps pour comprendre que la phonologie interlangue n'était pas simplement le résultat d'un transfert direct de la phonologie de la langue première. Plusieurs études soutiennent l'idée que la L1 n'est pourtant pas la seule source d'erreurs phonologiques (Ioup & Weinberger, 1987; Odlin, 1989; Osburne, 1996 ; Hansen, 2001). Selon le *modèle de l'Ontogenèse* de Major (1987), l'influence de la langue première semble avoir des effets importants dans les premiers stades de l'acquisition d'une langue seconde, mais plus tard, alors que la compétence des apprenants se développe, d'autres facteurs de développement sont susceptibles d'avoir un rôle plus important. Ainsi, les préférences typologiques universelles, le marquage ou encore les facteurs sociolinguistiques et psychologiques peuvent influencer la phonologie de l'interlangue (Greenberg, 1966 ; Eckman, 1977 ; Sato, 1984, 1985 ; Bebee, 1987 ; Broselow, 1987 ; Tarone, 1987 ; Weinberger, 1987 entre autres).

L'objet de ce chapitre est de présenter les travaux importants de ces dernières décennies qui ont cherché à mettre en évidence les contraintes et leurs propriétés qui s'exercent sur l'acquisition de la phonologie d'une L2.

1.1. Influence de la langue native

1.1.1. La notion de « Crible Phonologique »

Le rôle de la langue native dans la perception et la production d'autres langues est reconnu depuis longtemps. Il a beaucoup intéressé les linguistes du Cercle Linguistique de Prague dont Polivanov qui, le premier, a évoqué cette relation dans une publication de 1931 :

« Les phonèmes et les autres représentations phonologiques élémentaires de notre langue maternelle [...] se trouvent si étroitement liés avec notre activité perceptive que même en percevant des mots (ou des phrases) d'une langue avec un système phonologique tout

différent, nous sommes enclins à décomposer ces mots en des représentations phonologiques propres à notre langue maternelle. En entendant un mot inconnu étranger [...] nous tâchons d'y retrouver un complexe de nos représentations phonologiques, de le décomposer en des phonèmes propres à notre langue maternelle, et même en conformité de nos lois de groupement en phonèmes » (Polivanov, 1931, p. 79-80)

Selon Polivanov, les difficultés rencontrées dans la perception et la production des langues étrangères résultent de l'influence de la structure phonologique de la langue maternelle, notamment en raison du fait que les représentations phonémiques d'une langue seconde sont perçues selon le système des représentations de la langue première.

Cette idée a été reprise puis largement défendue et développée par Troubetzkoy (1939). En utilisant l'image d'un filtre, Troubetzkoy conçoit le système phonémique de la langue maternelle comme un « crible » à travers lequel tous les sons des langues étrangères doivent passer pour être classifiés.

« Le système phonologique d'une langue est semblable à un crible à travers lequel passe tout ce qui est dit. Seules restent dans le crible les marques phoniques pertinentes pour individualiser les phonèmes. Tout le reste tombe dans un autre crible où restent les marques phoniques ayant une valeur d'appel ; plus bas se trouve encore un crible où sont triés les traits phoniques caractérisant l'expression du sujet parlant. Chaque homme s'habitue dès l'enfance à analyser ainsi ce qui est dit et cette analyse se fait d'une façon tout à fait automatique et inconsciente. Mais en outre le système des cribles, qui rend cette analyse possible, est construit différemment dans chaque langue. L'homme s'approprie le système de sa langue maternelle. Mais s'il entend parler une autre langue, il emploie involontairement pour l'analyse de ce qu'il entend le « crible phonologique » de sa langue maternelle qui lui est familier. Et comme ce crible ne convient pas pour la langue étrangère entendue, il se produit de nombreuses erreurs et incompréhensions. Les sons de la langue étrangère reçoivent une interprétation phonologiquement inexacte, puisqu'on les fait passer par le « crible phonologique » de sa propre langue. » (Troubetzkoy, 1939, p. 54)

Ce crible conceptuel constitue, dans notre cerveau, une structure qui permet de reconnaître les unités de la langue maternelle. Un nouveau son appartenant à une autre langue sera alors interprété selon ceux qui correspondent à la langue maternelle. Selon cette notion de crible, un apprenant ne perçoit donc pas à travers le système des phonèmes de la langue cible, mais il filtre ce qui est traité par son oreille à travers son propre système sonore. C'est de ce « crible phonologique » que proviendraient de nombreuses erreurs faites par des apprenants débutants quand ils sont confrontés aux unités de la langue seconde et qui n'ont aucun équivalent dans la langue maternelle. Dans cette voie ouverte par Troubetzkoy, plus précisément, Archibald (1998, p. 3) donne un exemple dans lequel « have » [hæv] avoir en

anglais est filtré et interprété différemment par un apprenant français ([æv]) et par un apprenant allemand ([hæf]), chacun « entendant » le mot selon le répertoire phonémique de sa langue. La forme que produit l'apprenant français reflète le fait que sa langue première ne possède pas le phonème glottal /h/ alors que le changement de [v] en [f] dans la prononciation de l'apprenant allemand vient du fait qu'il existe une règle du dévoisement de l'obstruante finale dans sa langue maternelle. Quel est alors l'impact de la langue première dans l'acquisition d'une langue seconde ? De nombreuses études qui se sont attachées à répondre à cette question, sont parvenues à proposer des hypothèses et des facteurs de la L1 impliqués dans le processus d'acquisition des sons de la L2.

1.1.2. Hypothèse de l'analyse contrastive

Une hypothèse forte apparue dans les années 40 suggère que la plupart des difficultés rencontrées par les apprenants d'une L2 sont causées par leur langue première. Dans ce courant de pensées, la question essentielle à résoudre pour l'apprentissage des langues secondes porte sur le *transfert* de la langue maternelle à la langue seconde, c'est-à-dire sur l'influence des règles et des structures de la langue maternelle sur l'acquisition d'une langue étrangère. Cette influence se manifeste de plusieurs façons (Ellis, 1994). Les unités de la L2 qui sont semblables à celles de la L1 sont acquises vite et facilement : il y a « *transfert positif* ». Dans l'autre cas, la différence des unités entre les deux langues rend l'acquisition des unités de la L2 plus difficile. Dans ce cas, le processus jugé responsable est appelé le *transfert négatif* ou l'*interférence* (Ellis, 1985, p. 6).

Afin d'identifier les facteurs à l'œuvre dans le transfert, un procédé appelé *Analyse contrastive* a été élaboré dans les années 50 d'abord par Weinreich (1953) et poursuivi par Lado (1957). Basée sur la conviction qu'un processus de comparaison systématique entre langue maternelle et langue seconde est la source des difficultés rencontrées par les apprenants lors de l'apprentissage, l'analyse contrastive présuppose que les aspects par lesquels les langues sont différentes provoquent plus de problèmes lors de l'apprentissage.

« ... l'apprenant qui entre en contact avec une langue étrangère va trouver certaines caractéristiques de cette langue très faciles et les autres extrêmement difficiles. Les éléments qui sont similaires à sa langue maternelle seront simples pour l'apprenant, et ceux qui sont différents seront difficiles »² (Lado, 1957, p. 2, cité par Ellis, 1994, p. 306, t.p.n)

² « ...the student who comes into contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult »

Les différences constituent, selon les auteurs de l'Hypothèse de l'Analyse Contrastive (HAC), la source majeure des difficultés dans l'apprentissage d'une langue seconde. A titre d'exemple, le mot anglais « school » école /skul/ souvent prononcé [eskul] par des apprenants espagnols peut être expliqué par l'interférence de la L1 (Gerlach, 2004). Etant donné que les clusters de type */sC/ sont illégaux en espagnol en début de mot et que la voyelle /e/ est souvent prédictible en position initiale de mot #esC(C) dans cette langue (cf. « escuela » école), lors de l'apprentissage de l'anglais, des apprenants espagnols ont tendance à insérer [e] devant */sC/.

Les habitudes perceptives de la langue maternelle ne touchent pas seulement les segments phonologiques, mais également la phonotactique et les phénomènes suprasegmentaux. Plusieurs études montrent que les stratégies de segmentation lexicale spécifiques à la langue maternelle et basées sur la phonotactique et la prosodie peuvent être transférées à la langue seconde et ainsi nuire à la reconnaissance des mots de la L2 (Cutler, Mehler, Norris & Segui, 1986 ; Otake, Hatano, Cutler & Mehler, 1993 ; Dupoux, Pailler, Sebastián-Gallés & Mehler, 1997). Ainsi, les auditeurs français de l'étude de Cutler & al. (1986) utilisent la stratégie syllabique, appropriée pour le français, quand ils écoutent l'anglais. Otake & al. (1993) ont montré que les auditeurs français utilisent aussi la stratégie syllabique pour le japonais alors que les auditeurs japonais utilisent les frontières de mores³ comme indices de frontière de mots. Les Japonais, possédant les structures moraiques, sont sensibles, à leur tour, à l'anglais, alors que les auditeurs anglais segmentent le japonais aux frontières des syllabes pleines. Ces études ont mis en évidence que la segmentation lexicale en langue seconde est étroitement liée à la structure rythmique de la langue maternelle et que les auditeurs non-natifs utilisent la stratégie de segmentation à l'œuvre dans leur langue maternelle, même si cette stratégie ne convient pas à la langue seconde.

D'autres types de comportement sont également observés au niveau suprasegmental. Une étude de Dupoux & al. (1997) montre que les Français, à la différence des Espagnols, ont de grandes difficultés à percevoir la différence entre des mots espagnols qui diffèrent seulement quant à la position de l'accent tonique, comme « tópo » *taupe* et « topó » *avoir trouvé*. En français, en prononciation isolée, tous les mots portent l'accent sur la dernière syllabe et l'accent tonique n'est donc jamais utilisé pour des distinctions lexicales. Par conséquent, les locuteurs français n'ont jamais besoin de tenir compte de la position de l'accent tonique pour différencier des mots du français. Leur système perceptif n'est pas

³ « La more constitue l'unité de rythme et de mesure prosodique de la langue japonaise, la seule reconnue dans la réflexion linguistique autochtone. Chaque more articulée occupe une durée perçue comme semblable. Les 103 mores du japonais standard présentent les structures ordinaires CV, CyV (consonne initiale palatalisée) ou V ; ou les mores spéciales correspondant à la nasale /N/, à la première partie d'une géminée /Q/, à la deuxième partie d'une voyelle longue /R/ » (Labrune, 2006 : 143). Par exemple, le mot « hontu » /hoNtoR/ *vrai* est de même longueur prosodique que « origami » /origami/ *pliage de papier*, soit 4 mores.

adapté à ce phénomène linguistique. Comme l'affirme également Peperkamp (2002), leur représentation phonologique des mots n'inclut pas l'accent tonique, ce qui explique leurs problèmes avec l'accent tonique des mots espagnols.

Le système phonologique de la L1 influence donc la perception de la L2 aussi bien en ce qui concerne les sons que l'accentuation et les stratégies de segmentation liées au rythme et à la phonotactique. L'hypothèse originale de l'analyse contrastive semble fonctionner (Lado, 1957 ; Stockwell & Bowen, 1965 ; Suter, 1976 ; Redard, 1973 ; Anan, 1981 ; Ioup, 1984). L'influence de la L1 implique dans ces cas un transfert négatif en L2.

Cependant, toutes les erreurs de l'apprenant ne sont pas décelables à partir des effets de la L1. Au début des années 70, l'HAC a sévèrement été critiquée, en particulier concernant l'acquisition d'autres composantes linguistiques comme la morphologie et la syntaxe. Les résultats des études citées par Ellis (1985) telles que celles de Dulay & Burt (1973, 1974) ont permis d'émettre de sérieux doutes au sujet du transfert négatif comme facteur important au cours de l'acquisition d'une langue seconde. Dans leurs études, seulement 3 % des erreurs grammaticales des apprenants hispanophones de l'anglais étaient dues à la différence entre les deux langues. De nombreuses autres études menées dans les années 80 (voir Ellis, 1985, p. 29) indiquent également une large proportion d'erreurs qui ne peuvent pas être expliquées par l'interférence de la L1. Donc, toutes les erreurs des apprenants ne sont pas dues à l'influence de la L1 et les problèmes que les apprenants rencontrent ne sont donc pas tous prédictibles par l'HAC.

Les erreurs des apprenants ne sont pas toujours explicables par les différences entre les langues et, inversement, les différences ne sont pas forcément source de difficultés. Concernant l'acquisition de la phonétique et de la phonologie en L2, l'HAC a été critiquée car elle n'explique pas pourquoi certains sons sont appris plus facilement que d'autres. Cette hypothèse prévoit de façon inexacte que tous les phonèmes de la L2 qui ne sont pas présents dans l'inventaire phonologique de la L1 ont le même degré de difficulté. Brière (1966) a montré que des apprenants anglais présentent des difficultés à imiter la fricative uvulaire /ʁ/ du français et la constrictive pharyngale /ħ/ de l'arabe mais pas la fricative uvulaire /ɣ/ du vietnamien. Ces trois sons n'existent pourtant pas dans le système consonantique de l'anglais. Ce n'est donc pas parce qu'une unité de la L2 est absente de la L1 que son acquisition est plus difficile qu'une autre. L'étude de Brière montre que les différences linguistiques sont distinctes des difficultés d'apprentissage. Plusieurs études ont même montré qu'il était plus difficile d'acquérir des unités semblables entre les deux langues que des unités bien différentes (entre autres Flege 1987, Major 1996).

Prédire un comportement langagier à partir des comparaisons linguistiques des structures de langues soulève donc plusieurs types de problèmes. « Les résultats de descriptions contrastives pourraient par conséquent s'utiliser pour expliquer les traits

observés dans les productions de l'apprenant mais pas pour prédire ce que l'apprenant est supposé faire » (Lauret, 2007). Klein⁴ (1989) apporte la nuance suivante : « les prédictions sur d'éventuels transferts devraient être basées non sur la comparaison de propriétés structurales, mais sur la façon dont l'apprenant traite ces propriétés ». L'auteur donne l'exemple de la fricative interdentale voisée /ð/ anglais (ex. dans « that ») n'existant pas en français. En général, les francophones apprenant l'anglais se rendent compte qu'il s'agit d'un son inhabituel. Ils cherchent à le remplacer par un son du français. Il y a deux candidats [d] et [z] qui, d'un point de vue purement structural, sont également ressemblants à [ð]. Mais les apprenants francophones choisissent toujours [z], contrairement aux apprenants d'autres langues maternelles et aux locuteurs de certaines variétés régionales d'anglais, qui prononcent [dat]. L'auteur en conclut que cette préférence n'est pas prédictible en comparant uniquement les propriétés phonétiques.

Bien que l'influence de la L1 est démontrée dans des erreurs de prononciation des sons de la L2, les recherches menées dans le cadre de l'Hypothèse de l'Analyse Contrastive ont prouvé que l'influence de la L1 pouvait expliquer seulement une partie des erreurs. D'autres aspects ont fait l'objet d'étude pour expliquer les difficultés des apprenants qui ne peuvent pas être directement liées aux différences entre systèmes linguistiques, notamment les facteurs liés au processus de perception des sons du langage et aux mécanismes d'assimilation des contrastes. Ces études ne se positionnent plus seulement par rapport aux différences entre systèmes phonologiques, mais dans le contexte général de la perception des sons du langage humain.

1.1.3. Perception en langues première et seconde : théories et modèles

Une idée souvent avancée pour expliquer la difficulté rencontrée par des apprenants d'une langue seconde avec les sons de la L2 différents de ceux de la L1 est que ces sons de la L2 en question pourraient ne pas être correctement perçus : « Pour bien produire un son, il faut d'abord l'avoir bien perçu » (Borrell, 1990). Or, parler et écouter ne sont pas des processus simples à décrire et à expliquer parce qu'ils mettent en jeu à la fois variabilité et invariance. Beaucoup des relations entre les signaux physiques perçus et le code linguistique utilisé à des fins de communiquer oralement sont non encore éclaircies. Beaucoup d'hypothèses ont été formulées sur la façon dont un son est perçu et produit. Plusieurs propositions ont été apportées par les tenants des théories auditive, motrice, interactionnelle ou de la théorie des exemplaires ou celle des prototypes. Chacune adopte, pour interpréter le processus de la perception de la parole, une approche particulière qui la singularise. Malgré la diversité des

⁴ Le livre écrit en allemand est traduit en français par Colette Noyau, dont l'ensemble des travaux sur l'acquisition des langues se trouve sur son site personnel à l'adresse : <http://colette.noyau.free.fr>

approches, leur méthode consiste essentiellement à éliminer la variabilité en cherchant des paramètres invariants qui seraient liés aux catégories phonologiques perçues.

Parmi les défenseurs des *théories auditives*, Stevens (1989) propose que l'invariance doit être recherchée au niveau acoustique, le signal contenant toutes les informations nécessaires à la caractérisation perceptive des unités, sans référence nécessaire aux gestes articulatoires sous-jacents, c'est-à-dire, aux gestes produits par de petits déplacements de la langue, des lèvres ou des cordes vocales qui n'ont pas de conséquences *acoustiques* importantes pour l'oreille. Au contraire, selon les auteurs de la *théorie motrice* (Liberman & Mattingly, 1985), l'invariance réside dans le geste articulatoire. Percevoir la parole consisterait à percevoir les gestes articulatoires du conduit vocal. Dans ce cadre, la perception des mots est initiée par la récupération d'une représentation phonétique du signal acoustique et l'auditeur interprète le signal de parole en référence aux mouvements articulatoires dont celui-ci est le produit. La *théorie de la variabilité adaptative* (Lindblom, 1990), soutient, en revanche, que l'invariance procéderait d'un échange constant entre les spécificités articulatoires du locuteur et les nécessités de la perception par l'auditeur. Les corrélats invariants seraient en fait inexistantes, et les éléments de la parole perçus comme tels seraient le résultat de la parole adaptative.

Schwartz et collaborateurs (2002), dans la *théorie de la Perception pour le Contrôle de l'Action*, (PACT) proposent que la perception de la parole est l'ensemble des processus perceptifs (auditifs, visuels, voire tactiles) qui fournissent un ensemble de représentations permettant le contrôle de ses propres actions et le suivi des actions de l'autre. Selon cette théorie, le sujet qui perçoit forme des représentations sensori-motrices à partir de la récupération multimodale des gestes de la parole de son interlocuteur et ces représentations contraignent son système phonologique lui permettant de contrôler ses propres actions. Selon ce schéma, la capacité à retrouver des caractéristiques du contrôle des gestes vocaux passe par la perception de ces gestes. De manière générale, dans les travaux complémentaires de cette théorie (Schwartz, Basirat, Ménard & Sato, 2010 ; Schwartz, Sato & Fadiga, 2011), les auteurs proposent une modélisation qui intègre les relations entre la perception de la parole, le contrôle de l'action et la phonologie. Ils postulent que le domaine d'étude qu'est la parole implique nécessairement la prise en compte des liens entre ces trois niveaux. Cette approche est donc centrée sur la co-structuration des systèmes de perception et de production de la parole en lien avec le niveau phonologique.

La théorie des exemplaires (Kirchner, 1999), s'attache à montrer comment les représentations lexicales sont formées, stockées et traitées dans le lexique mental. Il postule que la trace en mémoire d'un mot serait constituée de l'ensemble des traces stockées lors des différentes rencontres avec ce mot, traces appelées *exemplaires*. La diversité des contextes linguistiques et environnementaux ainsi que la diversité des locuteurs (âge, sexe, ...) font que chaque mot n'est jamais prononcé (ou articulé) de la même façon et donc chaque occurrence perçue d'un mot est caractéristique. Toutes ces variantes de signifiant sont stockées sous

forme d'exemplaires. Deux exemplaires seront distincts si les différences articulatoires sont perceptibles et suffisamment récurrentes. Dans cette optique, la reconnaissance d'un mot revient à estimer la ressemblance entre l'occurrence entendue et les différentes « versions » des exemplaires stockées, puis à retenir celles qui se rapprochent le plus de l'occurrence entendue.

Inspirés par un courant de recherche en psychologie cognitive dont l'objectif est d'étudier les processus généraux employés dans la catégorisation des formes sensorielles, les travaux menés notamment par Rosch, Mervis, Gray, Johnson & Boyes-Braem (1976) et Nosofsky (1986) portent sur la façon dont les catégories sont représentées (sous la forme de prototypes ou de liste d'exemplaires). Cette notion de prototype est au centre de la théorie *Perceptual Magnet Effect* de la perception de la parole proposée par Kuhl (1991) dont l'hypothèse est que tous les membres d'une catégorie ne sont pas perçus par l'auditeur comme étant équivalents. Les « bons » exemplaires, ou prototypes, ont un rôle central à jouer dans la structuration perceptive de l'espace phonétique, dans la mesure où c'est à partir de ces prototypes que s'opère la catégorisation des sons de la parole (voir Nguyen, 2005 pour plus de détails).

Deux théories parmi les plus influentes dans les recherches en L2 des années 90 jusqu'à maintenant sont le Modèle d'Assimilation Perceptive (PAM : *Perceptual Assimilation Model*) de Best (1995) et le Modèle d'Apprentissage de la Parole (SLM : *Speech Learning Model*) de Flege (1995). Ces deux modèles sont issus de travaux qui cherchent à comprendre et à expliquer les difficultés que les apprenants rencontrent lors de l'apprentissage d'une langue étrangère, dues à la similarité entre les catégories phonologiques de la L1 et celles de la L2 ; et notamment ceux qui se sont focalisés sur les distinctions segmentales qui posent des difficultés perceptives pour les apprenants (entre autres Goto, 1971 ; MacKain, Best & Strange, 1981 ; Best & Strange, 1992 ; Tees & Werker, 1984 ; Flege & Eefting, 1987). Les deux modèles (PAM et SLM) sont conçus pour prédire les difficultés initiales à percevoir les contrastes de L2 (Best 1994, 1995 ; Best & al., 2001) ainsi que les difficultés à apprendre à percevoir et à produire les sons de L2, même pour des apprenants très avancés (Flege 1995, 1997).

Une étude antérieure de Best, McRoberts & Sithole (1988) tente d'expliquer pourquoi et comment certains contrastes non-natifs peuvent échapper au « crible phonologique », alors que d'autres distinctions tout aussi peu familières sont répertoriées comme difficiles. L'étude qui porte sur la distinction perceptive entre phonèmes d'une LE et phonèmes de la LM montre que des clicks du zulu, non-assimilables à des phones de l'anglais, ont été très facilement distingués par des adultes et des enfants anglophones natifs. Ce résultat est à la base de l'élaboration du modèle PAM.

En focalisant sur la perception de contrastes non-natifs plutôt que sur la perception de sons non-natifs, le modèle PAM proposé par Best (1995) explique la différence de sensibilité aux contrastes des langues étrangères en faisant appel à la notion d'assimilation perceptive phonologique. D'après ce modèle, des segments non natifs tendent à être perçus en fonction de leur similarité ou dissimilarité par rapport à des segments présents dans l'espace phonologique de la langue native.

Best distingue trois possibilités d'assimilation de segments non natifs (1995, p. 194). Dans le premier, le segment de la langue en cours d'apprentissage est assimilé à une catégorie native : dans ce cas, il peut être perçu soit comme un bon exemplaire de cette catégorie, soit comme un exemplaire acceptable mais non idéal de la catégorie, soit comme un exemplaire notablement déviant de la catégorie. La deuxième possibilité est l'assimilation en tant que son de la parole non catégorisable : dans ce cas, le son de L2 est dans les limites de l'espace phonologique de la langue native, mais ne peut être associé à aucune catégorie phonologique spécifique. Enfin, le son peut ne pas être assimilé du tout à l'espace phonologique de la langue native : c'est le cas des sons perçus comme n'appartenant pas à la parole humaine.

Parmi les trois possibilités d'assimilation de segments non natifs, Best propose de distinguer six catégories d'assimilation perceptive correspondant aux différents cas de figure rencontrés (1995, p. 195).

- Assimilation à deux catégories (TC type) : *chaque segment non-natif est assimilé à une catégorie native différente, et on s'attend à ce que la discrimination soit excellente.*⁵
- Différence de qualité dans la catégorie (CG type) : *les deux sons non-natifs sont assimilés à la même catégorie native, mais ils diffèrent dans l'écart par rapport à « l'idéal » natif (par ex. l'un est acceptable, l'autre est déviant). On s'attend à ce que la discrimination soit de modérée à très bonne, en fonction de la magnitude de la différence dans la qualité de catégorie pour chacun des sons non-natifs.*⁶
- Assimilation à une seule catégorie (SC type) : *les deux sons non-natifs sont assimilés à la même catégorie native, mais ils sont également écartés de "l'idéal" natif, c'est-à-dire que les deux sont également acceptables ou également déviantes. On s'attend à ce que la*

⁵ « Two-Category Assimilation (TC Type) Each non-native segment is assimilated to a different native category, and discrimination is expected to be excellent. »

⁶ « Category-Goodness Difference (CG Type) Both non-native sounds are assimilated to the same native category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable, the other deviant). Discrimination is expected to be moderate to very good, depending on the magnitude of difference in category goodness for each of the non-native sounds. »

*discrimination soit mauvaise (bien qu'elle puisse être légèrement au-dessus du seuil du hasard).*⁷

- Les deux sont non catégorisables (UU type) : les deux sons non-natifs tombent dans l'espace phonétique, mais en dehors d'aucune des catégories natives spécifiques, et ils peuvent varier dans leur discriminabilité en tant que sons de la parole non catégorisables. On s'attend à ce que la discrimination varie de mauvaise à très bonne, selon la proximité entre eux, et la proximité avec les catégories de l'espace phonologique natif.⁸
- Non catégorisé versus catégorisé (UC type) : l'un des deux sons non-natifs est assimilé à une catégorie native, et l'autre tombe dans l'espace phonétique en dehors des catégories natives. On s'attend à ce que la discrimination soit très bonne.⁹
- Non assimilables (NA type) : les deux catégories non-natives tombent en dehors du domaine de la parole, étant perçues comme des sons qui ne relèvent pas de la parole, et la paire peut varier en discriminabilité en tant que sons considérés comme n'appartenant pas à la parole. On s'attend à ce que la discrimination soit bonne à très bonne.¹⁰

La figure 1.1 schématise les différents types d'assimilation possibles lors de la perception de contrastes non natifs (Best, 1995). Parmi les contrastes non-natifs catégorisés perceptivement comme des phonèmes natifs, le modèle prédit que la capacité des auditeurs à distinguer entre différents sons non-natifs est meilleure pour le type « Assimilation à deux catégories – TC type », moins bonne pour le type « Différence de qualité dans la catégorie – CG type » et la plus mauvaise pour le type « Assimilation à une seule catégorie – SC type ».

Le modèle prédit en effet que deux sons de la L2 qui peuvent être assimilés à la même catégorie de la L1 sont difficiles à discriminer. La similarité phonétique entre les catégories phonologiques de la L1 et celles de la L2 prédit des difficultés initiales à percevoir les contrastes existants dans la L2. Les auditeurs ne détectent pas les différences entre des sons natifs et non-natifs s'ils les perçoivent comme similaires en ce qui concerne leurs caractéristiques articulatoires. Dans ce cas-là, les sons non-natifs seront assimilés à la catégorie phonologique native la plus similaire (SC type).

⁷ « Single-Category Assimilation (SC Type) Both non-native sounds are assimilated to the same native category, but are equally discrepant from the native "ideal", that is, both are equally acceptable or both equally deviant. Discrimination is expected to be poor (although it may be somewhat above chance level). »

⁸ « Both Uncategorizable (UU Type) Both non-native sounds fall within phonetic space but outside of any particular native category, and can vary in their discriminability as uncategorizable speech sounds. Discrimination is expected to range from poor to very good, depending upon their proximity to each other and to native categories within native phonological space. »

⁹ « Uncategorized versus Categorized (UC Type) One non-native sound assimilated to a native category, the other falls in phonetic space, outside native categories. Discrimination is expected to be very good. »

¹⁰ « Nonassimilable (NA Type) Both non-native categories fall outside of speech domain being heard as nonspeech sounds, and the pair can vary in their discriminability as nonspeech sounds; discrimination is expected to be good to very good. »

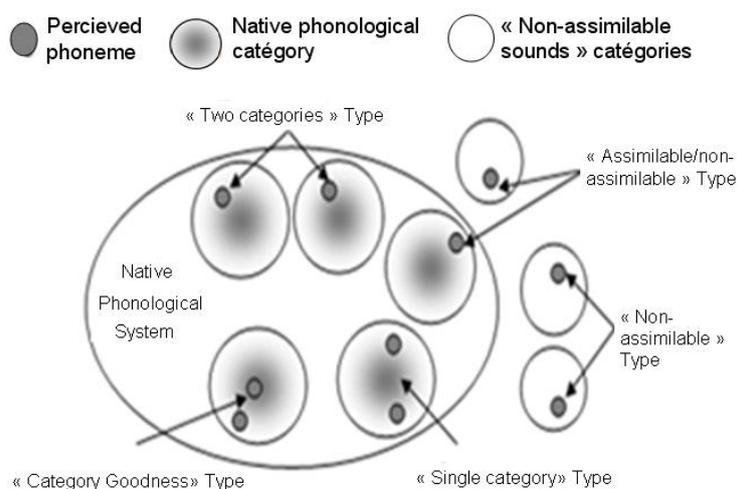


Figure 1.1. Les types d'assimilation perceptive des contrastes non natifs (Best, 1995).

Plusieurs études ont abouti à des résultats en accord avec les prédictions du modèle PAM. Ainsi, les phonèmes /w/ et /j/ de l'anglais américain ont été assimilés à des phonèmes correspondants en japonais /w/ et /j/ (c'est le cas *TC type*), les deux phonèmes /r/ et /l/ de l'anglais ayant été assimilés à un seul phonème japonais /r/ (*SC type*) (Best & Strange, 1992). Burfin, Savariaux, Granjon, Sanchez, Tran, Faraco & Kandel (2011) dans une étude sur la perception audiovisuelle du contraste /f/ ~ /θ/ de l'espagnol par les sujets français montrent que la seule information acoustique de /θ/ n'est pas suffisante pour pouvoir distinguer ce phonème de /f/ (ce dernier existe en français) : /θ/ et /f/ ont été assimilés à un seul phonème /f/ dans la condition auditive (*SC type*). Par contre, l'information visuelle sur les mouvements de la langue a été exploitée pour distinguer les deux phonèmes comme deux catégories différentes (*TC type*). Polka (1992) a examiné, auprès de locuteurs anglais et farsi, la perception du contraste vélaire *vs.* uvulaire dans les contextes voisés /g/ ~ /G/ (existant seulement en farsi) et éjectifs /k'/ ~ /q'/ (n'existant ni en anglais, ni en farsi). Polka observe que, bien au-dessus du seuil du hasard, les auditeurs anglais tendent à percevoir /G/ en tant que « mauvais » exemplaire de /g/ ou en tant qu'aucune consonne claire en anglais, assimilant de ce fait le contraste comme un cas de *CG* ou *UC type*. La plupart des sujets des deux groupes ont eu du mal à percevoir le contraste non natif des éjectives, ce qui correspond à une assimilation de *type SC* ou *UU*.

Le modèle PAM concerne notamment la perception initiale des contrastes non-natifs et non l'acquisition des contrastes d'une L2. Best (1995, p. 198) souligne elle-même qu'il faut développer le modèle PAM aussi du côté des aspects apprentissage et examiner comment la perception de différents types de contrastes non-natifs se modifie lors de l'acquisition d'une L2.

Plus orienté vers les aspects d'apprentissage figure le *Speech Learning Model* (SLM) de Flege (1995, 1997) dont l'objet est d'expliquer les erreurs de production ainsi que l'accent étranger au moyen des erreurs en perception. Basé sur des données expérimentales, le modèle SLM se compose de quatre postulats et sept hypothèses dérivées de ces postulats (Flege, 1995, p. 239).

Le principe fondamental au modèle est le premier postulat qui stipule que « *les processus et les mécanismes utilisés lors de l'apprentissage du système des sons de la L1, y compris la formation des catégories, restent intacts tout au long de la vie, et peuvent être appliqués pour apprendre une langue seconde* »¹¹. Le deuxième postulat explique le concept des catégories phonétiques, qui sont nécessaires à certaines hypothèses ultérieures : « *Les aspects, spécifiques à chaque langue, des sons de parole sont spécifiés dans les représentations de la mémoire à long terme appelées catégories phonétiques* »¹². Le troisième postulat note que « *les catégories phonétiques établies pour les sons de la L1 dans l'enfance évoluent toute la vie afin de refléter les propriétés de tous les phones de la L1 ou de la L2 identifiés comme une réalisation de chaque catégorie* »¹³. Le quatrième postulat concerne les apprenants bilingues qui « *s'efforcent de maintenir le contraste entre les catégories phonétiques de la L1 et de la L2 qui existent dans un espace phonologique commun* »¹⁴.

De ces quatre postulats de base, Flege dérive plusieurs hypothèses présentées ci-dessous et discutées après :

- H1 : *Les sons de la L1 et de la L2 sont perceptivement liés les uns aux autres à un niveau allophonique sensible à la position, plutôt qu'à un niveau phonémique plus abstrait.*¹⁵
- H2 : *Une nouvelle catégorie phonétique peut être établie pour un son de la L2 qui diffère phonétiquement du son de la L1 le plus proche si les bilingues discernent au moins certaines des différences phonétiques entre les sons de la L1 et ceux de la L2.*¹⁶
- H3 : *Plus la dissimilarité phonétique perçue entre un son de la L2 et le son le plus proche de la L1 est grande, plus il est probable que les différences phonétiques entre les sons seront discernées.*¹⁷

¹¹ "P1: The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning"

¹² "P2: Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories".

¹³ P3: Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category

¹⁴ P4: Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

¹⁵ H1: Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

¹⁶ H2: A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

¹⁷ H3: The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

- H4 : La probabilité que les différences phonétiques entre les sons de la L1 et ceux de la L2, et entre les sons de L2 non contrastifs en L1 soient discernées, est inversement proportionnelle à l'âge du début d'apprentissage.¹⁸
- H5 : La formation de catégorie pour un son de L2 peut être bloquée par le mécanisme de classification par équivalence. Dans ce cas, une seule catégorie phonétique sera utilisée pour traiter des sons de la L1 et de la L2 perceptivement liés (diaphones). En définitive, les diaphones se ressembleront en production.¹⁹
- H6 : La catégorie phonétique établie pour les sons de la L2 par un bilingue peut être différente de celle d'un monolingue si : 1) la catégorie chez le bilingue est « déviée » par rapport à une catégorie de la L1 afin de maintenir un contraste phonétique entre des catégories dans un espace phonologique commun à L1 et L2 ; ou 2) la représentation chez le bilingue est fondée sur des traits différents ou sur des pondérations de traits différentes par rapport à celle chez le monolingue.²⁰
- H7 : La production d'un son correspond finalement aux propriétés présentes dans la représentation de sa catégorie phonétique.²¹

Considérant ces sept hypothèses dérivées des quatre postulats, le modèle propose de décrire comment les apprenants adultes acquièrent les sons d'une L2 et établissent des catégories phonétiques nouvelles pour la perception et la production de celles-ci. Les prédictions du modèle prennent en compte l'influence de la structure phonologique de la langue maternelle sur la perception des sons de la L2. De plus, en opposition à d'autres études (voir Rochet, 1995 pour le détail de ces études) selon lesquelles la plupart ou tous les sons de la L2 seront identifiés avec un son de la L1, Flege propose la capacité des apprenants à faire une distinction entre des sons de la L2 relativement semblables à ceux de la L1, contrairement à des sons plus différents ou même « nouveaux » de ceux de la L1.

Par conséquent, au cours de l'apprentissage, l'apprenant classe les sons de la L2 comme identiques, nouveaux ou similaires par rapport aux sons de la L1. Flege (1997a) classe les sons comme identiques, similaires ou nouveaux selon les critères suivants : un son de la L2 qui est *identique* est représenté par le même symbole API utilisé pour un son de la L1. De plus, des analyses acoustiques et perceptives ne montrent pas de différences entre

¹⁸ H4: The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases. (AOL = Age of Learning)

¹⁹ H5: Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

²⁰ H6: The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1 - L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

²¹ H7: The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

les sons. Un son de la L2 qui est *similaire* à un son de la L1 est représenté par le même symbole API que le son de la L1, bien que des analyses acoustiques et perceptives révèlent des différences entre les deux, comme par exemple entre la voyelle /i/ du français et la voyelle /i/ de l'anglais. Un son de la L2 qui est *nouveau* est représenté par un symbole API qui n'est pas utilisé pour un son de la L1 et qui diffère acoustiquement et perceptivement du son le plus proche de la L1, comme par exemple le phonème français /y/ pour un apprenant de langue maternelle anglaise.

Les sons *identiques* ne posent pas de problèmes parce que l'apprenant a déjà établi des catégories phonétiques pertinentes en acquérant sa L1. Les sons *nouveaux* sont ceux qui diffèrent suffisamment des sons de la L1 pour ne pas y être assimilés. Par conséquent, l'apprenant peut construire de nouvelles catégories phonétiques pour les sons nouveaux et, par là, arriver à les prononcer et à les percevoir correctement. Les sons *similaires*, par contre, sont susceptibles d'être classés par l'apprenant comme équivalents des sons de la L1. Les catégories phonétiques de la L1 sont donc susceptibles de bloquer, par le mécanisme de *classification par équivalence*, la formation de nouvelles catégories pour les sons proches de la L2.

La différence des structures phonologiques entre les deux langues a ainsi été intégrée dans le modèle SLM mais interprétée différemment par rapport à l'Hypothèse de l'Analyse Contrastive (HAC). Rappelons que cette hypothèse prédit que « les éléments qui sont similaires à la langue maternelle seront simples pour l'apprenant, et ceux qui sont différents seront difficiles » (Lado, 1957). Selon le SLM, des apprenants auront plus de difficultés en acquérant un son de la L2 qui est *semblable* (mais pas *identique*) à un son de leur langue maternelle. Si l'apprenant ne peut pas discerner les différences phonétiques entre un son de la L2 et le son le plus proche de la L1, aucune nouvelle catégorie phonétique ne sera établie, alors que les *nouveaux* sons présenteront peu de difficultés parce que ces sons se soustraient à la *classification par équivalence*. Dans ce dernier cas, l'input est suffisant pour permettre à des apprenants d'établir de nouvelles catégories phonologiques.

Le mécanisme de *classification par équivalence* a été utilisé par Flege (1987) pour expliquer l'ordre de l'acquisition des unités de la L2, à partir de l'observation de sujets de langue maternelle anglaise, apprenants du français de niveau avancé. Ceux-ci produisaient /y/ d'une manière plus juste que /u/, bien que l'anglais possède /u/ mais pas /y/ dans son système phonologique. Flege a argumenté que ces apprenants s'étaient concentrés sur /y/ soi-disant plus difficile parce que, d'après eux, /u/ en anglais est « équivalent » à /u/ en français. Ces deux /u/ sont pourtant différents en réalité sur le plan acoustique : /u/ de l'anglais possède un F2 plus élevé que /u/ du français.

Plusieurs autres études ont soutenu cette hypothèse de *classification par équivalence*. Bohn et Flege (1992) ont remarqué qu'un groupe d'apprenants allemands expérimentés de

l'anglais a produit la voyelle /æ/ plus authentiquement que le groupe inexpérimenté. Mais peu de différences entre les deux groupes ont été trouvées dans la production de la voyelle semblable /ɛ/ en raison de la *classification par équivalence*. Le même phénomène a été observé par Bradlow, Pisoni, Akahane-Yamada & Tohkura (1997) chez des apprenants japonais de l'anglais lesquels ont montré relativement plus de difficultés à réaliser /l/ que /r/. Les auteurs ont supposé que ce résultat est dû au fait que le japonais possède une unique liquide apico-alvéolaire battue /r/ qui est, sur le plan articulatoire, plus similaire à la latérale /l/ qu'à la vibrante /r/. La liquide du japonais connaît un grand nombre de variantes selon les locuteurs, les registres ou l'environnement phonétique, parmi lesquelles on rencontre principalement la latérale apico-alvéolaire [l] (Labrune, 2006). Ce résultat a été de nouveau confirmé plus récemment par une étude longitudinale d'Aoyama (2004) portant sur la perception et production des liquides /l/, /r/ auprès d'enfants et d'adultes japonais vivant aux États-Unis. La vibrante apico-alvéolaire /r/ en anglais étant perceptivement moins similaire à la liquide du japonais, elle a été acquise plus vite que la liquide /l/ par des apprenants japonais.

Il est à noter que le SLM concerne surtout les apprenants très avancés, ayant parlé la L2 pendant plusieurs années. Selon Flege (1997b), des recherches indiquent que, aux premiers stades de l'apprentissage, les sons similaires sont plus faciles à prononcer que les sons nouveaux. Il cite à titre d'exemple le fait qu'au début de leur apprentissage, les apprenants anglophones substituent souvent la voyelle /u/ de l'anglais à /y/ français. Cependant, aux stades les plus avancés de l'apprentissage, les nouveaux sons sont prononcés plus authentiquement que les sons similaires. Flege pense que pour les débutants, les sons de la L2 pourraient être d'abord identifiés comme des allophones de la L1. Au fur et à mesure, lorsque l'expérience en L2 augmente, les apprenants discerneraient alors la différence phonétique entre des sons de la L2 avec ceux les plus proches de la L1. À ce stade, de nouvelles catégories phonétiques peuvent être établies pour ces nouveaux sons de la L2 qui sont indépendants des représentations déjà établies pour ceux de la L1.

Le modèle SLM converge vers le modèle PAM en ce qu'ils prédisent tous deux que la facilité à percevoir le contraste entre différents phonèmes de la L2 varie en fonction de la similarité phonétique entre les sons de la L1 et ceux de la L2. Ainsi, les sons qui ne ressemblent pas aux sons de la L1 sont plus faciles à percevoir et à prononcer parce qu'il n'y a pas de transfert de la L1, tandis que les sons qui sont similaires aux sons de la L1 sont assimilés à ceux-ci.

La recherche sur les aspects phonologiques de l'interlangue a connu des avancées avec les travaux qui ont considéré le rôle de la L1 à partir du point de vue des différences et des similarités par rapport à la L2 et ceci dans la production comme dans la perception. De ces travaux sont nés des modèles d'explication et de prédiction des erreurs commises par les apprenants d'une L2.

Cependant, si un certain nombre d'erreurs de perception et/ou de prononciation des sons de la L2 sont attribuables à L1 de l'apprenant, d'autres ne le sont pas. Les recherches liées aux propriétés typologiques de la L1 et de la L2 révèlent que les différences de perception et production phonologiques entre des locuteurs natifs et non natifs ne peuvent pas toutes être attribuées au transfert de la L1 et que des principes universels qui structurent les systèmes sonores et l'utilisation de leurs unités dans la chaîne parlée influence considérablement l'apprentissage d'une L2.

1.2. Influence de principes universels

1.2.1. La notion de marquage

Le principe universel de marque, en tant que propriété fondamentale du langage, a été proposé par Troubetzkoy qui l'a ensuite développé avec Jakobson à l'École de Prague en 1930 (voir Gadet, 1994 pour plus de détails). L'idée à la base de ce concept est que des oppositions binaires (présence *vs.* absence d'une caractéristique phonique) sont nécessaires pour classer les phonèmes et leur opposition dans un système. Ce sont des oppositions binaires qui existent entre par exemple obstruents voisés et non voisés, voyelles orales et nasales, syllabes ouvertes et fermées, occlusives et fricatives, consonnes simples et groupes de consonnes. Les comparaisons interlangue ont vite montré que les deux entités d'une opposition binaire n'étaient pas également distribuées dans les langues et que l'une était bien plus fréquente que l'autre (Troubetzkoy, 1939 ; Jakobson, 1941). Ainsi, l'attribution d'une valeur de marquage à une opposition (marqué ~ non marqué) est devenu un moyen de caractériser le statut favorisé *vs.* défavorisé : « *le membre qui est plus largement distribué que l'autre dans une opposition est indiqué en tant que non marqué ; ce qui indique qu'il est, d'une certaine manière définissable, plus simple, plus principal et plus naturel que l'autre, qui est, à son tour, défini en tant que membre marqué* »²² (Eckman, 2008). Dans les exemples cités au-dessus, les obstruents non voisés, les voyelles orales, les syllabes ouvertes, les occlusives, les consonnes simples sont non marquées respectivement aux obstruents voisés, voyelles nasales, syllabes fermées, fricatives et groupes consonantiques. Le terme marquage réfère donc à l'idée que certaines structures linguistiques sont moins fréquentes et « plus complexes » que d'autres. Une tendance générale des langues est de posséder les structures non marquées ou moins marquées.

²² « ... the member of the opposition that was more widely distributed than the other was designated as unmarked, indicating that it was, in some definable way, simpler, more basic and more natural than the other member of the opposition, which was in turn defined as the marked member. » (Eckman, 2008)

Les travaux sur les universaux du langage de Greenberg (1978) ont été utilisés dans de nombreux travaux afin de déterminer les structures linguistiques marquées et non marquées :

« Une structure X est typologiquement marquée par rapport à une autre structure Y (et Y est relativement non marquée à X) si chaque langue qui possède X a aussi Y, mais la présence de Y n'implique pas nécessairement la présence de X »²³ (Gundel, Houlihan & Sanders, 1986, p. 108, t.p.n)

Le marquage sous l'angle de la typologie linguistique est un rapport au niveau des structures ou des représentations linguistiques entre les langues du monde. En tant que principe général du fonctionnement du langage, le marquage influencerait les processus d'acquisition des langues secondes dont la phonétique et la phonologie. Dans les recherches en phonologie de l'interlangue, deux hypothèses sont basées sur ce principe du marquage typologique : l'hypothèse de la différence de marquage (HDM) et l'hypothèse de la conformité structurelle (HCS).

1.2.2. L'Hypothèse de la Différence de Marquage

L'Hypothèse de la Différence de Marquage (HDM) formulée par Eckman en 1977 prédit que les formes de la L2 qui sont différentes et plus marquées que celles de la L1 seront les plus difficiles à produire et que les formes moins marquées sont acquises avant les plus marquées. En anglais par exemple, /θ/ est plus marqué que /s/ parce que /s/ est plus fréquent dans les systèmes phonologiques des langues du monde. De même, un étudiant néerlandais apprenant l'anglais trouvera le contraste de voisement en fin de mot difficile, non seulement parce que le néerlandais ne connaît pas ce contraste dans cette distribution (Hiligsmann, Beheydt, Degand, Godin & Vanderlinden, 2005), mais parce qu'un tel contraste est en général typologiquement marqué.

Bien que les différences entre une langue maternelle et une langue seconde sont nécessaires pour expliquer les difficultés d'apprentissage d'une langue seconde, ces différences ne sont pas suffisantes pour une explication adéquate (cf. section 1.1.3). Selon l'HDM, il faut que les deux aspects suivants soient réunis :

1. La différence entre la langue maternelle (L1) et la langue seconde (L2).
2. L'existence de la relation de marquage entre les formes où il y a des différences.

Selon Eckman (1977, p. 321), les difficultés rencontrées par les apprenants pourraient être prédites par :

²³ « A structure X is typologically marked relative to another structure, Y, (and Y is typologically unmarked relative to X) if every language that has X also has Y, but every language that has Y does not necessarily have X'' (cité dans Eckman, 2008)

- a. Les formes de la L2 qui sont différentes et plus marquées que celles de la L1 seront difficiles à acquérir.²⁴
- b. Le degré relatif de difficulté des formes de différence de la L2 qui sont plus marquées que celle de la L1 correspondra au degré relatif de marquage.²⁵
- c. Les formes de la L2 qui sont différentes mais pas plus marquées que celles de la L1 ne seront pas difficiles.²⁶

Le pouvoir prédictif de l'Hypothèse de la Différence de Marquage a été évalué dans de nombreuses études concernant notamment l'acquisition de structures syllabiques et des groupes consonantiques d'une langue seconde. A titre d'exemple, Moulton (1962) a remarqué que l'acquisition du contraste de voisement en fin de mot est très difficile pour les Allemands apprenant l'anglais, alors que pour les Anglais apprenant l'allemand, l'absence de ce contraste n'était pas difficile à acquérir. Cet exemple a été discuté en contexte de l'HDM où l'auteur a argumenté que, comparé à ce que les Anglais avaient à apprendre, les Allemands devaient acquérir une structure relativement plus marquée : l'opposition de voisement en coda. En allemand, cette opposition est neutralisée en fin de mot en faveur des consonnes non voisées. Alors qu'en anglais, ce contraste existe dans toutes les positions que ce soit en début, au milieu ou en coda de mot.

Larsen-Freeman & Long (1991) ont fourni un autre exemple de l'hypothèse d'Eckman : les anglophones n'auraient pas de problèmes à distinguer /ʃ/ et /ʒ/ en position initiale en français, bien que /ʒ/ soit trouvé en anglais seulement en position médiane et finale. La raison pour laquelle les apprenants anglophones n'auraient pas de problèmes à produire /ʒ/ en position initiale est que cette position est moins marquée que les positions médiane et finale. De même, Major & Faudree (1996) ont constaté que les apprenants coréens de l'anglais ont produit les occlusives en positions initiale et médiane avec un taux de réussite de presque 100 % mais seulement d'environ 50 % en position finale parce que, selon eux, les occlusives en position finale sont plus marquées que celles en positions initiale et médiane.

Anderson (1987) supporte également l'HDM en contestant que la différence de performance des apprenants arabes et chinois, lors de l'apprentissage des clusters en initiale et finale, était étroitement liée au degré de marquage des clusters. Il a aussi trouvé que les clusters marqués en finale causent beaucoup plus d'erreurs que ceux marqués en initiale. Concernant ce contraste de position (initiale *vs.* finale), les résultats de Stockman & Pluut (1992) sont en partie à l'encontre de ceux d'Anderson. En examinant les apprenants

²⁴ *Those areas of the target language which differ from the native language and are more marked than the native language will be difficult;*

²⁵ *The relative degree of difficulty of the areas of difference of target language which are more marked than the native language will correspond to the relative degree of markedness*

²⁶ *Those areas of the target language which are different from the native language, but are not more marked than the native language will not be difficult*

mandarins de l'anglais sur la perception et production des monosyllabes contenant des plosives et des nasales en initiale et finale, ces auteurs ont constaté que la position finale ne pose pas toujours de difficulté. Le cas de la nasale vélaire [ŋ] est un exemple. Cette consonne [ŋ] occasionne plus de difficultés aux sujets mandarins en position initiale que finale. Eckman (2004) relève que les résultats d'Anderson et ceux de Stockman et Pluut sont tous compatibles avec l'HDM, parce que cette nasale vélaire [ŋ] est plus marquée en position initiale que finale et que les langues qui ont [ŋ] en attaque, la possèdent également en coda, mais pas vice versa (Ladefoged et Maddieson, 1996; Ladefoged, 2001; Maddieson, 1984, 1986, 1993).

Eckman a modifié son *Hypothèse de la Différence de Marquage* à partir de 1991, mais a continué à travailler dans le cadre des universaux typologiques. Cette autre hypothèse qui implique aussi le marquage typologique, ou au moins les généralisations qui sont à la base des principes de marquage, est l'*Hypothèse de la Conformité Structurale*.

1.2.3. L'Hypothèse de la Conformité Structurale

Eckman (1991) a proposé l'Hypothèse de la Conformité Structurale (HSC) dans le but d'examiner l'influence d'universaux typologiques sur l'acquisition des groupes de consonnes d'une L2. Elle affirme que « les généralisations universelles qui tiennent pour les langues premières tiennent également pour les interlangues »²⁷ (Eckman, 1991, p. 24, t.p.n).

L'hypothèse HSC d'Eckman formule que la structure d'une interlangue sera conforme à des principes typologiques trouvés dans les langues premières. Les universaux retenus par Eckman pour illustrer l'HSC figurent dans les deux principes :

1. « Le principe fricative-plosive : si une langue a au moins une séquence de consonnes en finale du type plosive + plosive (comme dans *apt*), elle a également au moins une séquence en finale du type fricative + plosive (comme dans *ask*) »²⁸ (cité dans Archibald, 1998, p. 57, t.p.n)
2. « Le principe de résolubilité : si une langue a une séquence de trois consonnes soit en position initiale ou finale, que nous appellerons ici *m*, elle aura également au moins une séquence *m* –. Donc la présence d'une séquence en début de mot tels que *str* par exemple, indique que, la langue contiendra également au moins une des séquences *st* ou *tr* dans la même position »²⁹ (cité dans Archibald, 1998, p. 58, t.p.n.)

²⁷ « the universal generalizations that hold for primary languages hold also for interlanguages »

²⁸ « Fricative-stop principle. If a language has at least one final consonant sequence consisting of stop + stop (as in *apt*), it also has at least one final consequence consisting of fricative + stop (as in *ask*) »

²⁹ « Resolvability principle: If a language has a sequence of three consonants in either initial or final position, which we will here call *m*, it will also have at least one subsequence consisting of *m* –. So the presence of a word-

Les principes peuvent être reformulés de manière implicationnelle comme ci-dessous avec les données de l'anglais (Archibald, 1998) (le signe # indique la frontière de mot) :

1. Principe fricative-plosive :

pt#, kt# > ft#, sp#, st#, sk#

2. Principe de résolubilité :

#spr > #sp, #pr #skr > #sk, #kr sks# > sk#, ks#

kts# > kt#, ts# sps# > sp#, ps# pst# > ps#, st#

#str > #st, #tr spt# > sp#, pt# kst# > ks#, st#

skt# > sk#, kt# pts# > pt#, ts#

Eckman (1991) a cherché à savoir si les interlangues des apprenants d'une L2 suivaient ces principes. Il a ainsi tenté de démontrer que la structure la plus marquée dans une relation implicationnelle n'atteint pas un niveau de référence de l'acquisition si la structure la moins marquée n'est pas encore présente à ce niveau. L'idée de tester l'HSC de cette manière est liée à l'observation que les langues premières ne contiennent pas les structures les plus marquées d'une relation implicationnelle sauf si elles contiennent aussi les moins marquées.

Eckman (1991) a donc collecté les données auprès de onze sujets natifs (quatre japonais, quatre coréens, trois cantonais). Aucune langue native des apprenants ne contient de groupes consonantiques quelle que soit la position initiale ou finale. L'objectif était d'étudier l'acquisition des contraintes sur les groupes de consonnes admissibles en anglais en cherchant si leurs inventaires interlangues violaient les universaux proposés. Les sujets ont effectué plusieurs tâches qui différençaient dans les degrés de contrôle, de la lecture de listes de mots jusqu'à des conversations libres.

Selon l'HSC, si l'interlangue de l'apprenant contient des clusters de type plosive + plosive (pt# ou kt# par exemple), mais ne contient pas de clusters fricative + plosive (ft#, sp#, st# ou sk#), alors l'interlangue viole le principe fricative-plosive. L'application du principe de résolubilité aux interlangues peut être testée de deux manières. Premièrement, le principe peut être confirmé si l'interlangue contient le cluster de trois consonnes et un des deux sous-clusters de deux consonnes (#skr et #sk, par exemple, mais pas #kr) ; Eckman le considère comme un cas faible. Deuxièmement, le principe peut être confirmé si l'interlangue contient le cluster de trois consonnes ainsi que l'ensemble des deux sous-clusters de deux consonnes (#skr et #sk et #kr) ; Eckman le considère comme un cas fort. Les résultats ont montré que sur les 200 cas relevés dans les productions des apprenants, 74 % sont forts, 24 % sont faibles, et seulement 2 % de cas ont violé les universaux implicationnels. Ces résultats suggèrent que

initial sequence such as str, for example, indicates that the language will also contain at least one of the subsequences st or tr in the same position ».

les interlangues sont régies par les mêmes types d'universaux implicationnels trouvés dans les langues premières et soutiennent donc l'Hypothèse de la Conformité Structurale.

Les mêmes résultats sur les clusters ont été trouvés par Carlisle (1997, 1998) qui montre que les clusters anglais plus marqués ont été modifiés plus fréquemment que les clusters moins marqués dans les interlangues des apprenants espagnols. De même, Eckman & Iverson (1994) ont analysé des codas complexes de l'anglais produits par des apprenants japonais, coréens et cantonnais dans les conversations libres. Aucune langue maternelle des apprenants n'autorise les codas complexes. Les résultats montrent que les apprenants ont plus d'erreurs dans la production des codas plus marquées.

Toutes les études menées dans le cadre de l'HCS montrent que les interlangues peuvent contenir des structures qui sont plus complexes que celles autorisées par la L1 mais pas forcément aussi complexes que celles requises pour la L2. Ainsi, selon ces travaux, les interlangues se situent entre la langue de départ et la langue cible, mais toujours en conformité avec les universaux typologiques.

Dans ce champ de recherche, l'acquisition des séquences de consonnes possibles dans la L2 constitue un observatoire de la phonologie de l'interlangue à partir duquel il est possible de confronter théories et modèles qui tentent de décrire et d'expliquer de manière générale les processus d'acquisition des formes orales d'une langue seconde.

1.3. Acquisition des groupes de consonnes en langue seconde

Généralement, les recherches sur les effets de marquage ont donné quelques prédictions utiles sur des zones problématiques pour les apprenants de langues étrangères (Eckman, 1977, 1991 ; Moulton, 1962 ; Anderson, 1987 ; Benson, 1988 ; Carlisle, 1991, 1997, 1998 ; Eckman & Iverson, 1994). Une des hypothèses fréquemment testées en acquisition des clusters de L2 est *l'hypothèse de la Différence de Marquage* d'Eckman (1977) selon laquelle les domaines de la langue cible qui sont différents et plus marqués que ceux de la langue de départ sont plus difficiles à acquérir. Le marquage de groupes consonantiques a reçu beaucoup d'attention dans les recherches en L2, du fait qu'ils ne sont pas présents dans tous les systèmes phonotactiques des langues du monde (pour une analyse des tendances voir Rousset, 2004). Les études en acquisition des clusters convergent en général vers les mêmes résultats : les apprenants d'une langue seconde présentent effectivement des difficultés à acquérir les formes les plus marquées de la langue cible et plus le cluster est marqué, plus les apprenants de la L2 rencontrent des difficultés pour le réaliser (Anderson, 1987 ; Broselow et Finer, 1991 ; Davidson 2002, 2003, 2006 ; Davidson & Stone, 2004 ; Eckman, 1991 ; Eckman & Iverson, 1993 ; Hancin-Bhatt & Bhatt, 1997 ; Kim, 2000). Aussi, il est maintenant bien connu

que les clusters en finale de mot sont plus marqués que ceux en initiale ou position médiane (Anderson, 1987 ; Benson, 1986 ; Major, 1996 ; Sato, 1984 ; Weinberger, 1987 ; Rousset, 2004).

L'acquisition des clusters de l'anglais, langue cible, concerne la majeure partie des études publiées, pour une variété de langues sources telles que cantonais, japonais, coréen, mandarin, espagnol ou arabe. Les langues les plus étudiées de ce point de vue sont celles qui n'autorisent aucun cluster dans toute position d'une syllabe (par exemple, le mandarin, le japonais, le coréen, le vietnamien) ou celles qui autorisent certains clusters en finale (comme l'arabe) ou en initiale seulement (comme le portugais). Un constat commun entre ces études est que les apprenants montrent de grandes difficultés pour acquérir des clusters en position finale par rapport à ceux situés en position initiale de syllabe.

Par ailleurs, les apprenants de différentes origines linguistiques semblent avoir des préférences différentes pour les stratégies de simplification des clusters, allant de l'épenthèse vocalique à la substitution ou la suppression de certains segments du groupe consonantique. Alors que les apprenants mandarins et arabophones semblent préférer la suppression (Anderson, 1987), la stratégie la plus présente chez les lusophones est la substitution du contraste de voisement (Major, 1996). Par ailleurs, Weinberger (1987) a remarqué que des stratégies de simplification de clusters peuvent varier selon le niveau de compétence des apprenants. Aussi, Hansen (2001) a trouvé chez un groupe d'apprenants mandarins de niveau intermédiaire un taux d'épenthèse assez élevé pour les clusters de deux segments (CC) en finale, alors qu'ils ont tendance à n'éliminer que l'un des segments des clusters composés de trois membres (CCC).

Les différents auteurs ont aussi nettement constaté une plus grande facilité de production pour la syllabe canonique universelle³⁰ CV. En effet, les différentes études montrent que les sujets tendent à modifier les groupes consonantiques de manière à produire la structure CV, la moins marquée selon les universaux typologiques. Ils ont également remarqué qu'il existe des différences entre l'acquisition des attaques et des codas complexes³¹, les codas présentant un taux d'erreurs plus élevé en production (Anderson, 1987).

L'influence de la langue première a également été attestée dans l'acquisition des clusters d'une L2. Ainsi, Tarone (1980) qui a examiné l'acquisition des clusters auprès des apprenants coréens, chinois et portugais de l'anglais, a constaté que les apprenants modifient les structures syllabiques de la langue en cours d'apprentissage selon les structures

³⁰ L'universalité de cette structure est expliquée par MacNeilage (1998) dans le cadre de la théorie *Frame, then Content* en la liant au cycle du geste mandibulaire : la consonne est produite lorsque la mandibule est en position haute alors que la voyelle est réalisée lorsque la mandibule est en position basse.

³¹ La complexité d'un constituant syllabique, attaque et/ou coda, est estimée à partir du nombre de segments phonologiques qu'il contient (Rousset, 2004).

syllabiques de la L1. Les coréens et chinois ont préféré la suppression de consonnes, alors que les portugais ont favorisé l'épenthèse comme stratégie de simplification des structures syllabiques complexes. Indépendamment de la L1, ces derniers ont préféré les syllabes ouvertes CV, universellement non marquées et prédominantes en portugais (Carlisle, 2001).

Dans une étude comportant une analyse des stratégies d'épenthèse, Broselow (1983) a remarqué que l'élément inséré dans le cluster en attaque est différent selon les origines dialectales d'apprenants arabophones. Par exemple, les irakiens tendent à prononcer le mot anglais *flow* « courant » comme [iflo], tandis que les égyptiens le produisent comme [filo]. L'explication donnée par Broselow est l'influence des structures syllabiques de la L1, différentes entre les deux dialectes : contrairement au dialecte arabe égyptien, le dialecte arabe irakien connaît un grand nombre de syllabes de structure VC.

Concernant les stratégies par élision, Sato (1984) a observé que les apprenants vietnamiens de l'anglais préféraient la suppression d'un segment lors de la production de clusters de l'anglais en position de coda. Sato a interprété la stratégie des locuteurs vietnamiens comme un résultat du transfert phonologique de la langue première vers la langue cible. Comme le vietnamien possède une large proportion de structures syllabiques CVC, lors de la production des clusters de l'anglais, les locuteurs vietnamiens ont favorisé la suppression d'un segment dans une structure CVCC donnant une réalisation CVC. Ces résultats ont ensuite été confirmés par Osburne (1996) à partir d'une étude de cas d'un apprenant vietnamien de niveau avancé. Cet apprenant a montré des réalisations de plusieurs formes relevant de la phonologie de l'interlangue parmi lesquelles la tendance à omettre les consonnes finales non attestées dans cette position dans sa langue maternelle.

Globalement, il a été démontré que la suppression d'un des segments du cluster à réaliser, ainsi que l'épenthèse vocalique, constituaient les principales stratégies des apprenants (entre autres Davidson, 2002, 2003, 2006 ; Davidson et al, 2004 ; Hancin-Bhatt et Bhatt, 1997 ; Hansen, 2001 ; Kim, 2000 ; Tropic, 1987). Les tendances dans les réalisations de groupes consonantiques d'une langue en cours d'apprentissage pourraient être résumées ainsi : (1) les modifications sont davantage observées lorsque les clusters sont en coda plutôt qu'en attaque ; (2) l'insertion d'une voyelle (épenthèse) est plus fréquente en initiale de mot, alors que l'effacement d'un élément est plus fréquent en finale de mot ; (3) le segment le moins sonore d'un cluster est souvent le plus éliminé ; (4) l'insertion d'une voyelle est plus fréquente après consonne sonore qu'après consonne sourde.

À partir des généralisations universelles des séquences de consonnes de Greenberg (1966), Benson (1986) a trouvé les hiérarchies de marquage des clusters en testant la prononciation des groupes consonantiques de l'anglais par des apprenants vietnamiens. Ces hiérarchies ont été soutenues, plus tard, dans plusieurs études (Benson, 1988 ; Osburne, 1996 ; Sato, 1984) :

1. Des clusters en final qui contiennent une nasale et une obstruente homorganique non voisée sont moins marqués que ceux contenant une nasale et une obstruente hétéroorganique voisée. Par exemple, /nt/ nasale alvéolaire devant plosive alvéolaire non voisée, est moins marqué que /md/, nasale bilabiale suivie d'une plosive alvéolaire voisée.
2. Pour des clusters contenant uniquement des obstruents, les clusters non voisés sont moins marqués que les voisés. Par exemple, /kt/, plosive vélaire devant alvéolaire non voisée, est moins marqué que /bd/, plosive bilabiale suivie d'une alvéolaire voisée.

La longueur des clusters a également été attestée comme une des causes des difficultés rencontrées par les apprenants de L2. Spécifiquement, dans une langue, la complexité des clusters est soumise à des règles implicationnelles. Ainsi, Rousset (2004, p. 117) écrit que « si une structure syllabique avec attaque complexe de n consonnes est attestée dans une langue, elle implique les structures avec attaques de n-i consonnes (i = 1 à n-1) », la même règle étant applicable en coda. Rousset (2004) postule la règle selon laquelle « la fréquence d'apparition d'une structure syllabique diminue lorsque la complexité de son attaque et/ou de sa coda augmente (Rousset, 2004, p. 118). Ainsi, les clusters plus longs en attaque et en coda sont plus marqués que des clusters plus courts. L'étude de Weinberger (1987) montre que chez les locuteurs mandarins, la fréquence de modification des clusters en coda augmente avec la longueur des clusters. Plus précisément, ses analyses montrent que dans les productions des apprenants, 5,5 % des codas mono-consonantiques ; 29,8 % des codas de deux consonnes et 42 % des codas tri-consonantiques ont été modifiés. De même, Anderson (1987) a observé que la performance d'apprenants arabes et chinois de l'anglais baisse significativement avec l'augmentation du nombre de segments dans les clusters. Carlisle (2008) a trouvé la même tendance dans une étude portant sur la production des clusters anglais /sk/, /skr/, /sp/ en attaque chez dix locuteurs espagnols. Ses résultats montrent que de meilleurs scores de réussite sont observés dans la production des clusters à deux segments (CC) que dans celle des clusters tri-consonantiques (CCC).

D'autres études ont montré que les clusters marqués du point de vue du *Principe de Sonorité* comportent une plus grande difficulté d'acquisition. D'après ce principe, les phonèmes sont ordonnés sur une échelle de sonorité qui correspond grosso modo à leur degré d'intensité perçu (Cairns et Feinstein, 1982 ; Clements, 1990 ; Butt, 1992 ; Selfick, 1984 ; Blevins, 1995 entre autres). Suivant cette échelle, les voyelles sont les segments les plus sonores, suivies par les glides (semi-voyelles ou semi-consonnes), les liquides (latérales et vibrantes), les nasales, les fricatives et les plosives.

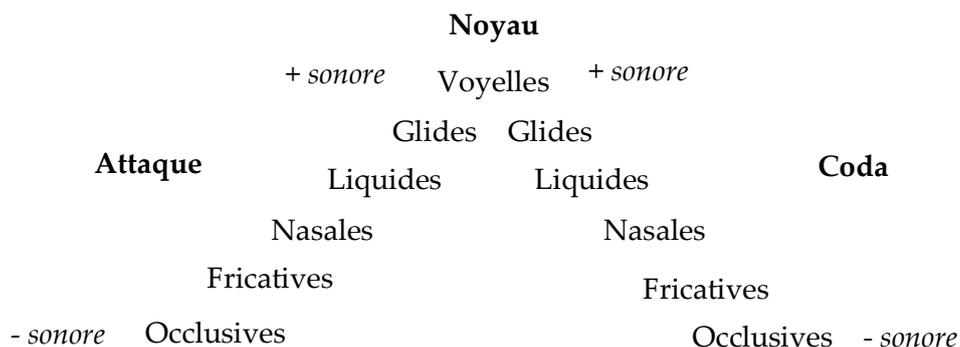


Figure 1.2. Echelle de sonorité (selon Broselow et Finer, 1991).

Selon le Principe de Sonorité, les segments d’une syllabe se succèdent selon un degré de sonorité croissant de l’attaque au noyau de la syllabe et selon un degré de sonorité descendant du noyau vers la coda (Cairns & Feinstein, 1982). Quelques études en acquisition des clusters de leur langue maternelle chez les enfants (Chin, 1996 ; Gierut, 1999 ; Ohala, 1999 ; McLeod, Doorn & Reed, 2001 ; Demuth & McCullough, 2009) se basent sur ce principe pour expliquer les tendances dans la suppression des éléments des clusters. Par exemple, Ohala (1999) a constaté que les enfants réduisent des segments de clusters en position d’attaque en sorte qu’une élévation maximale de sonorité se produit dans cette position. D’autres études sur l’acquisition d’une langue seconde ont montré que les apprenants étaient plus en difficulté pour produire des clusters qui violent le Principe de Sonorité (Carlisle, 1991, 1994 ; Tropic, 1986). Par exemple, en position postvocalique, le cluster [pɸ] dans le mot « câpre » [kapɸ] viole le principe, ce qui n’est pas le cas du cluster [ɸp] dans « carpe » [kaɸp]. [pɸ] étant plus marqué selon le Principe de Sonorité, « câpre » présenterait un niveau de difficulté plus élevé que « carpe ».

Selon le *Paramètre Minimal de la Distance de Sonorité* de Broselow et Finer (1991), le marquage d’un cluster dépend de la distance de sonorité entre ses composants. Les clusters dont les composants sont plus proches sur l’échelle de sonorité sont plus marqués que ceux dont les composants s’y trouvent plus éloignés. Ainsi, plus la distance de sonorité des composants d’un cluster est proche, plus son acquisition pose de problème chez les apprenants. Par exemple, [vɸ] est plus marqué que [pɸ], l’acquisition du mot « vriller » [vɸije] est plus difficile que « prier » [pɸije].

Broselow et Finer ont testé les apprenants coréens et japonais dans leur production de mots anglais contenant les clusters /pj bj fj pr br fr/ en attaque. Selon leur hypothèse, les clusters C+/r/ étant plus proches sur l’échelle de sonorité que C+/j/, ils devraient être plus marqués. Leurs résultats mettent en évidence la relation entre la distance de sonorité des segments dans un cluster et les difficultés de prononciation des apprenants : en effet, ces derniers présentent nettement plus de difficultés pour le cluster /fr/ (fricative et liquide étant

plus proches sur l'échelle de sonorité), alors que le cluster /pj/ leur occasionne le moins de difficultés.

Davidson (2001) fournit un autre regard sur le rôle décisif de la sonorité dans le marquage des clusters. En testant des locuteurs anglais sur leur production de mots polonais contenant des clusters, elle a trouvé que seule la sonorité ne pouvait pas expliquer les résultats. À titre d'exemple, dans son corpus, les clusters /zm/ et /vn/ possèdent la même distance sur l'échelle de sonorité. Pourtant, les locuteurs anglais ont produit /zm/ de manière beaucoup plus performante que /vn/ (63 % *vs.* 11 % de prononciation correcte). Davidson attribue cette différence de performance entre les deux clusters au marquage relatif de /v/ supérieur à celui de /z/ : car /z/ est coronal alors que /v/ labial, est plus marqué.

Des études menées dans le domaine de l'acquisition d'une L2 ont également cherché à savoir si la variation de tâche (lors du recueil de données) pouvait influencer les résultats sur la phonologie de l'interlangue (Labov, 1969 ; Schmidt, 1977 ; Tarone 1979, 1982, 1983, 1985). Tarone (1989) précise que tous les systèmes linguistiques ont une gamme de styles, allant d'un style le plus formel (produit lorsque le locuteur prête le plus d'attention à la forme du langage) à un style « vernaculaire » (dans lequel le locuteur prête le moins d'attention à la forme du langage). Il prédit que plus le style de la tâche est formel, plus l'apprenant prête attention à la forme de la langue, moins il commet d'erreurs. Par exemple, la lecture d'une liste de paires minimales peut être considérée comme le style le plus formel permettant de prédire un degré de performance plus élevé. Tandis qu'une conversation courante devrait être considérée comme la situation la moins formelle susceptible d'engendrer le plus d'erreurs (Dickerson & Dickerson, 1977). Donc, une plus grande performance devrait être obtenue par des apprenants de L2 lorsque le type de tâche est plus formel. Pourtant, les études sur l'acquisition des clusters d'une L2 n'ont pas trouvé cette prédiction si évidente. Ce qui varie selon le degré de formalité des tâches, ce sont les stratégies utilisées par les apprenants plutôt que leur taux d'erreur (Lin, 2001, 2003 ; Major, 1987 ; Weinberger, 1987). Plusieurs études ont en effet montré que la nature des tâches pendant la collecte de données influencent le choix des stratégies de simplification des syllabes chez les apprenants mais pas le nombre d'erreurs (Major, 1987 ; Weinberger, 1987, 1994 ; Bayley, 1996 ; Lin, 2001, 2003 ; Hansen, 2008).

Lin (2001, 2003) a comparé la production des clusters de l'anglais en attaque chez des apprenants chinois dans quatre tâches différentes : paires minimales, liste de mots, phrase porteuse et conversation. Elle a observé que la fréquence des épenthèses augmentait selon le degré de formalité de la tâche (conversation 4,38 % ; phrases porteuses 11,38 % ; liste de mots 27,5 % ; paires minimales 35,13 %), alors que la suppression et les remplacements de segment augmentaient lorsque les tâches devenaient moins formelles. Dans une étude antérieure, Bayley (1996) a également observé que les apprenants chinois de l'anglais suppriment

davantage les plosives /t/ et /d/ en coda dans les conversations informelles que lorsqu'ils lisent des énoncés.

Pour expliquer le lien entre le degré de formalité des tâches et les stratégies de modification de la syllabe, Lin propose le *Principe de Récupérabilité* de Weinberger (1987, 1994), selon lequel « *la proportion d'épenthèse ou de suppression pourrait être plus grande dans les tâches sans contexte linguistique que dans les tâches avec contexte linguistique* »³² (Weinberger, 1987, p. 408, *t.p.n.*). L'hypothèse sous-jacente de ce principe est que dans les tâches sans contexte linguistique telles que les paires minimales ou la lecture de liste de mots, la suppression d'un segment est vraisemblablement moins fréquente car elle peut entraîner une ambiguïté. Sur cette même logique, la suppression peut être trouvée plus fréquemment dans les tâches avec contexte linguistique, telles que la lecture de paragraphe et la conversation, parce que l'interlocuteur peut récupérer les segments supprimés grâce au contexte linguistique. En se basant sur cette notion de récupérabilité, Lin (2001, 2003) propose que l'utilisation de l'épenthèse augmente lorsque la tâche est plus formelle car elle exige plus d'attention à la forme plutôt qu'au contenu. Les apprenants corrigent alors leur « parole » pour rendre mieux audible la consonne cible, en quelque sorte un phénomène d'hypercorrection. Inversement, les taux de suppression et de remplacement sont accrus quand les apprenants prêtent plus d'attention au contenu qu'à la forme.

Les données de la littérature concernant l'influence du type de tâche dans l'acquisition des clusters d'une L2 convergent en montrant que l'utilisation des tâches plus formelles pendant la collecte de données mène à une augmentation d'épenthèses de voyelle, alors que le pourcentage de suppressions et les remplacements deviennent plus élevés dans des tâches moins formelles.

Les résultats et conclusions des études sur l'acquisition des clusters en langue seconde suggèrent que les erreurs des apprenants peuvent être prédites par le marquage, le transfert de L1 ou la variation des tâches. La manifestation de ces facteurs dépend aussi de la maîtrise de la langue par les apprenants, ceux de niveau plus faible ayant des taux d'erreur en production plus élevés. Il convient également de noter que les apprenants de niveau avancé présentent encore des erreurs dans la production des clusters, ce qui n'est pas nouveau, certes, mais cela signifie que la phonologie de l'interlangue est toujours concernée et que les raisons de ces erreurs perdurent (Nguyễn Thị Bình Minh, 2000). En outre, des apprenants de différentes origines linguistiques présentent des stratégies différentes de modification des clusters. L'analyse des productions d'apprenants vietnamiens de l'anglais langue étrangère a montré des stratégies de modification commune telles que la suppression, la substitution, l'épenthèse ou le changement des caractéristiques acoustiques d'un ou de l'ensemble des

³² « *the proportion of epenthesis to deletion should be greater in tasks without linguistic context than in tasks with linguistic context* »

segments des clusters. En ce qui concerne la variation de la tâche, les études n'ont pas toutes conclu que les erreurs sont plus fréquentes lorsque le degré informel des tâches augmente, mais ce sont plutôt les stratégies de modification qui changent.

Les travaux sur l'acquisition des groupes de consonnes en langue seconde sont multiples portant sur une variété de langues sources (arabe, cantonais, japonais, coréen, mandarin, espagnol, etc.) mais pour la plupart sur l'anglais comme la langue cible. Certains travaux étudient l'acquisition des clusters des langues tchèque (Davidson, 2003), polonais (Zydorowicz, 2010), allemand (Teodor, 2001) ou suédois (Johansson, 1973). Il n'existe pas à notre connaissance de recherches sur l'acquisition des clusters du français.

L'établissement d'une hiérarchie absolue de marquage dans une langue ou entre les langues est presque impossible en raison de la diversité des systèmes et des structures. L'influence de la L1 sur la L2 n'est pas évitable, mais elle n'est pas toujours négative et elle ne représente pas la seule source des problèmes. Parce qu'il peut être plus facile d'apprendre certaines formes différentes de la L2 car moins marquées que celles de la L1 selon l'HDM, les similarités entre deux langues ne semblent pas toujours profitables lors de l'apprentissage de la L2. Selon les modèles PAM et SLM, les apprenants tendent à classifier ces sons similaires de la L2 selon les catégories phonémiques de la L1, ce qui serait la cause des accents étrangers. Traditionnellement, une importance considérable est toujours accordée au système phonémique de la langue native dans les études en acquisition d'une langue seconde. C'est pour cette raison que le chapitre suivant présente des analyses phonémiques du système du vietnamien, la langue source, comparé avec le système du français, la langue cible.

Chapitre 2

Phonologie du vietnamien : description et comparaison avec le français

2.1. Éléments de linguistique historique

L'origine de la langue vietnamienne a été un sujet de débat animé parmi les linguistes (Nguyễn Tài Cẩn, 1995) et deux tendances ont dominé les discussions à la fin du XIXe, début du XXe siècle attribuées notamment à Maspéro et Haudricourt. Selon Maspéro (1912), le vietnamien a tiré son origine du thai, tandis qu'Haudricourt (1953, 1954), réfutant cette thèse, confirme que le vietnamien se situe dans la famille austro-asiatique. Cette deuxième opinion est aujourd'hui largement partagée par la plupart des linguistes. Le vietnamien est généralement admis comme appartenant au groupe des langues viet-muong de la branche mon-khmer, famille austro-asiatique. Il existe aujourd'hui un grand nombre de travaux qui attestent cette parenté, notamment ceux en linguistique comparée de Ferlus (1975, 1977, 1982, 1989, 1992, 2001), qui a proposé une correspondance assez régulière entre certains phonèmes du vietnamien et ceux des autres langues austro-asiatiques.

Le vietnamien a connu un développement lié à l'histoire du pays, depuis sa séparation avec le bloc mon-khmer (4 mille ans avant J.C.), puis la division du groupe Viet-Muong en deux langues indépendantes (vietnamien et muong) (du XIVe siècle au XVIe siècle), jusqu'à l'arrivée des missionnaires européens au XVIIe siècle et la naissance du *quốc ngữ*, la langue nationale actuelle.

Dans les domaines de l'ethnologie, de l'archéologie, de la sociologie, des éléments concordent permettant d'affirmer que, dans l'antiquité, le peuple vietnamien a eu des contacts avec des peuples d'origine austro-asiatique (mélanesienne et indonésienne) (Nguyễn Khắc Khâm, 1960). Dès l'époque préhistorique, le Vietnam a connu des vagues d'immigrations, d'abord les Mélanésiens, puis au IIe siècle avant J.C., les Indonésiens, laissant des traces comme par exemple dans les toponymes (noms de lieu, de montagne, de fleuve) et les zoonymes (Nguyễn Khắc Khâm, 1960).

La langue parlée au Vietnam à cette époque, le proto viet-muong, n'était pas une langue tonale. Son système des initiales comportait des consonnes simples, mais aussi des groupes de consonnes tels que */bl/, */tl/, */ml/, */mj/ qui au fil de l'évolution de la langue sont devenus des consonnes simples (Nguyễn Tài Cẩn, 1995). En voici quelques exemples :

<i>Proto Viet-Muong</i>	<i>Vietnamien actuel</i>	<i>Glosse</i>
blai gái / bl aj.ɣaj/	trai gái / t aj.ɣaj/	filles et garçons
con tlâu /kɔn. tl ɔw/	con trâu /kɔn. t ɔw/	le buffle
mlâm lãn / ml ɔm.lɔn/	nhâm lãn / n ɔn.lɔn/	se tromper
hoa mnhài /hwa. m ɔaj/	hoa nhài /hwa. n ɔaj/	fleur de jasmin

Selon Nguyễn Tài Cẩn (1995, p. 243), il existait en position coda de la syllabe, des consonnes comme */l/, */c/, */h/, */s/, */r/, */ʎ/, */ɲ/ qui cohabitaient avec les finales actuelles /p t k m n ɲ j w/. Ainsi, par exemple, la consonne finale */l/ n'existe plus en vietnamien actuel (Vương Lộc, 2006) :

wel (en muong) > về [ve] rentrer (vietnamien actuel)
 quel (en muong) > quê [kwe] village natale (vietnamien actuel)

Alors que */r/ et */h/ se sont estompés et transformés en semi-consonne /j/ :

bir (khmer) ; bar (bahnar) > vài [vaj] quelques (vietnamien actuel)
 kămpor (khmer) > vôi [voj] la chaux (vietnamien actuel)
 tapăh (môn) > bẫy [bǎj] le piège (vietnamien actuel)
 mủh (môn) ; cre mủh (khmer) > mũi [muj] le nez (vietnamien actuel)

Durant les périodes de domination chinoise (de 111 avant J.C. à 938), le vietnamien n'existait qu'en tant que langue parlée. La langue écrite « *chữ nho* » (idéogramme chinois) était exclusivement utilisée dans l'administration et la littérature, et donc, maîtrisée par une infime minorité de la population. Une quantité importante de mots chinois a été introduite dans le vietnamien à cette époque, généralement des termes de morale, de philosophie. Les emprunts au chinois, tout comme ceux d'origine austro-asiatique, ont été ainsi intégrés à la phonologie du vietnamien.

À partir de l'indépendance du Vietnam (en 939), les intellectuels ont commencé à développer leur système d'écriture propre afin de mieux rendre compte de la prononciation « nationale » du vietnamien. Le *chữ nôm* s'est développé alors, en abandonnant définitivement certaines formes résiduelles de l'ancien *chữ nho*, avec des sinogrammes supplémentaires spécifiquement vietnamiens pour l'écriture exclusive du vietnamien. L'évolution de la langue vietnamienne durant cette période et jusqu'au XIXe siècle peut être esquissée avec les étapes suivantes :

- 1) Le vietnamien ancien, d'il y a environ plus d'un millénaire, était probablement encore dissyllabique comme le sont aujourd'hui le sách/rục et le thavung (langues du groupe viet-muong), mais il aurait évolué vers le monosyllabisme sous l'influence du chinois (Ferlus, 1992). Plusieurs linguistes (Haudricourt, 1953 ; Banker, 1964 ; Lee, 1974 cité par Cao Xuân Hạo, 1985) attestent que les mots monosyllabiques du vietnamien ont souvent une origine qui remonte aux formes pluri-syllabiques à présyllabes³³ du mon-khmer.

³³ La présyllabe, représentant souvent « un préfixe ou le produit d'une infixation, est une syllabe auxiliaire (parfois deux) inaccentuée à voyelle réduite ou atone qui précède la syllabe-racine généralement accentuée » (Cao, 1985, p. 88). La présyllabe étant CVCVC s'il s'agit d'un préfixe ou CVCVC s'il s'agit d'un infixe.

Ferlus (1982) propose que le passage au monosyllabisme s'est fait en deux étapes. La première est marquée par la chute des voyelles des présyllabes dans la structure dissyllabique de départ *CVCVC*, qui donne des mots monosyllabiques à structure complexe *CCVC*. Cette voyelle, dans la plupart des cas, était un schwa sans valeur phonémique (Nguyễn Tài Cẩn, 1995, p. 240). L'étape suivante est la compression du mot en gommant la consonne de la présyllabe dans le groupe *CCVC* et donnant des mots de structure *CVC*. L'inventaire des consonnes de la présyllabe était très limité **/p t c k s/* parmi lequel **/k/* est la plus souvent rencontrée (Nguyễn Tài Cẩn, 1995, p. 241).

2) Le système des attaques du vietnamien a subi beaucoup de changements durant la période Xe - XVIIe siècle (Ferlus, 1992 ; Nguyễn Tài Cẩn, 1995 ; Vương Lộc, 2006) :

a. Les préglottalisées **/b/, */d/* ont aujourd'hui évolué en nasales */m/, /n/* avec des tons de la série haute. Les plosives non voisées sont devenues les voisées correspondantes : */p/ > /b/ ; /t/ > /d/ ; /k/ > /g/*. La fricative dentale */s/* s'est fermée en plosive dentale */t/*. Vers le XIe siècle, s'est produit le phénomène de confusion consonantique où les occlusives et fricatives sonores se sont dévoisées et confondues avec leurs homologues sourdes. Ces confusions ont eu des conséquences sur le système tonal (cf. voir l'étape 3 de cette section).

b. Le phénomène de spirantisation a frappé les occlusives et fricatives du vietnamien lorsqu'elles étaient en position médiane, c'est-à-dire à l'initiale de la deuxième syllabe dans les anciens dissyllabes (Ferlus, 1982). Ce phénomène a été établi en comparant le vietnamien aux langues viet-muong méridionales (pong, thavung, sach, ruc ...) ayant conservé un vocabulaire dissyllabique. A la suite de la chute de l'élément présyllabique dans les anciens dissyllabes vietnamiens, ces spirantes se sont phonologisées en enrichissant le système des consonnes initiales. Ci-dessous figurent quelques exemples proposés par Ferlus (1992, p. 113) :

<i>Proto-vietnamien</i>	<i>Spirantisation</i>	<i>Actuel</i>	<i>Exemples</i>	
p - b	ϕ - β	/v/	vôi /voj/ <i>chaux</i>	/kpol/ (thavung)
t - d	θ - ð	/z/	dôi /zoi/ <i>mentir</i>	/pto:j/ (rung)
c - ʝ	ç - j	/z/	giết /zjet/ <i>tuer</i>	/kcit/ (ruc)
tʃ - dʒ	ç - j	/z/	giùm /zum/ <i>aider</i>	/sum/ <i>se réunir</i>
k - g	χ - γ	/ɣ/	gạo /ɣaw/ <i>riz</i>	/əko:/ (thavung)

c. En initiale, le vietnamien a aussi connu la simplification des séquences de consonnes. Ainsi, les groupes **/kl/, */tl/, */bl/, */χl/* disparaissent et font la place aux rétroflexes */t/, /s/*.

klăm, tlăm (XVIIe) [klăm] [tlăm]	>	trăm [tăm] <i>cent, centaine</i>
blăng (XVIIe) [blăŋ]	>	trăng [tăŋ] <i>la lune</i>
không (XVe) [χloŋ]	>	sông [sɔŋ] <i>la rivière</i>

Le groupe */ml/ révélé par le *Dictionarium* d'Alexandre de Rhodes en 1651 où il apparaissait avec ses variantes * [mɲ] ou [ɲ], se transforme actuellement soit en /l/ soit en /j/ selon les dialectes (Mai Ngọc Chừ, Vũ Đức Nghiệu & Hoàng Trọng Phiến, 1997, p. 54)

mlăm [mlăm] <i>se tromper</i>	>	nhâm [ɲăm] (dialecte du Nord)
		lâm [lăm] (dialecte du Sud)
mlõ [mlɔ]	>	nhõ [ɲɔ] (dialecte du Nord)
		lõ [lɔ] (dialecte du Sud)

Le groupe */bl/ noté dans ce même *Dictionarium* avec plusieurs variantes *[bl], [z], apparaît aujourd'hui sous différentes formes en fonction des dialectes.

« blá on » [bla.ɲ]	>	giá on [za.ɲ] (dialecte du Nord)
<i>témoigner sa reconnaissance</i>		trá on [ta.ɲ] (dialecte du Centre et du Sud)

Les groupes avec */r/ du Proto Viet-Muong */pr/, /*br/, */tr/, */dr/, */kr/, */gr/, selon Ferlus (1992), se simplifient d'abord en */k^hr/, puis vers les XVe-XVIe en */kɕ/. La chute du premier élément amène la confusion avec /ɕ/.

- 3) Selon Haudricourt (1954), l'apparition du système tonal du vietnamien date du VIe siècle après J.C. L'influence du chinois a d'abord introduit trois tons (*ngang – ton 1, huyền – ton 2, sắc – ton 5*)³⁴. Le système à six tons actuel est apparu au XIIe siècle.

L'origine des tons en vietnamien a d'abord été discutée par Maspéro (1912) qui a démontré, avec l'aide de mots sino-vietnamiens³⁵, que les six tons pouvaient être analysés en deux séries : *ngang-sắc-hỏi* correspondant à d'anciennes initiales sourdes et *huyền-nặng-ngã* correspondant à d'anciennes initiales sonores.

Les linguistes diachroniciens désignent les trois paires d'inflexions par A, B et C (Michaud, 2004). Alors que les tons *sắc-nặng* dans les syllabes avec finales plosives sourdes /p t k/ sont notés D1 et D2.

³⁴ Nous transcrivons les tons par des chiffres (de 1 à 6) selon le système de notation étymologique qui utilise les chiffres impairs (1, 3, 5) pour les tons de la série haute et pairs (2, 4, 6) pour les tons de la série basse.

³⁵ Par le qualificatif sino-vietnamien sont désignés les mots chinois empruntés par le vietnamien au Xe siècle.

	A	B	C ³⁶
Anciennes initiales sourdes	<i>ngang</i> – A1	<i>sắc</i> – B1	<i>hỏi</i> – C1
Anciennes initiales sonores	<i>huyền</i> – A2	<i>nặng</i> – B2	<i>ngã</i> – C2

Haudricourt (1954) a fourni la première analyse détaillée des tons du vietnamien en démontrant que les trois paires de tons du vietnamien ancien sont originaires des anciennes finales laryngales (cf. table 2.1). En effet, les tons *sắc-nặng* (tons B1 et B2) sont dérivés d’une ancienne constriction glottale en finale (noté */-ʔ/) et les tons *hỏi-ngã* (tons C1, C2) d’une ancienne spirante finale (*/-h/), alors que les tons A1 et A2 *ngang-huyền* se sont développés dans un contexte vocalique ou en raison d’une sonante finale (cf. table 2.1). Les syllabes terminées par des occlusives (/p/, /t/ ou /k/), qui n’avaient pas d’oppositions tonales en proto viet-muong, sont actuellement sous les tons *sắc-nặng* (tons D1, D2).

Table 2.1. L’origine des tons en vietnamien de Haudricourt (1954) (Mai & al., 1997).

Au début de notre ère (absence de tons)	VIe (système à 3 tons)	XIIe (système à 6 tons)	Actuellement
pa	pa	pa	ba <i>trois</i> (A1)
sla, hla	hla	la	la <i>hurler</i> (A1)
ba	ba	pà	bà <i>mamie</i> (A2)
la	la	là	là <i>être</i> (A2)
pas, pah	pà	pả	bả <i>pâturage empoisonnée</i> (C1)
slas, hlah	hlà	lả	lả <i>être prostré</i> (C1)
bas, bah	bà	pã	bã <i>déchets</i> (C2)
las, lah	là	lã	lã <i>eau naturelle</i> (C2)
pax, pa ²	pá	pá	bá <i>titre de compte</i> (B1)
slax, ba ²	hlá	lá	lá <i>feuille</i> (B1)
bax, ba ²	bá	pạ	bạ <i>n’importe</i> (B2)
lax, la ²	lá	lạ	lạ <i>étrange</i> (B2)

³⁶ Il faut noter que phonétiquement, le ton *hỏi* (ton C1) fait partie de la série basse et correspond à des anciennes initiales sourdes comme les tons *ngang* et *sắc*. Alors que le ton *ngã* (ton C2), correspondant à des anciennes initiales sonores avec les tons *huyền* et *nặng*, appartient à la série haute. C’est pour cette raison que le ton *hỏi* est noté par le chiffre pair (ton 4), et le ton *ngã* par le chiffre impair (ton 3).

Série haute	<i>ngang</i> – ton 1 (A1)	<i>ngã</i> – ton 3 (C2)	<i>sắc</i> – ton 5 (B1-D1)
Série basse	<i>huyền</i> – ton 2 (A2)	<i>hỏi</i> – ton 4 (C1)	<i>nặng</i> – ton 6 (B2-D2)

Tout au long de ce manuscrit, les tons seront notés selon le système numérique (notamment dans les transcriptions phonétiques) et alphanumérique comme indiqué ci-dessus.

Haudricourt explique que le changement du système des initiales (perte d'oppositions consonantiques) et des finales (disparition des fricatives) a conduit à la formation du système tonal vietnamien. Le changement s'est produit par un phénomène de compensation : les oppositions lexicales ont été conservées par le biais d'un nouveau type d'opposition (Michaud, 2010). L'évolution s'est faite par transphonologisation d'oppositions consonantiques en oppositions tonales.

4) La rime syllabique a également subi des modifications phonétiques.

a. Parmi elles, l'apparition des diphtongues /a/ > /ɯɤ/ ; /ɔ/ > /uo/ ; /ɛ/ > /ie/ s'est faite à partir des voyelles ouverte /a/ et moins ouvertes /ɔ/, /ɛ/, la plupart du temps dans le contexte des initiales sonantes (nasales, liquides) ou voisées (selon Nguyễn Tài Căn, 1995:302). Ce dernier propose par exemple une comparaison entre le dialecte du Centre³⁷ et le parler de Hanoi.

	<i>Dialecte parlé dans le nord du Centre</i>	<i>Correspondant en vietnamien standard (parler de Hanoi)</i>
/a/ > /ɯɤ/	lả	lừa /lɯɤ/ feu
	ngá	ngứa /ɲɯɤ/ démangeaison
	náng	nướng /nɯɤŋ/ griller
	nác	nước /nɯɤk/ eau
/ɔ/ > /uo/	ló	lúa /luo/ plante de riz
	mọi	muỗi /muoj/ moustique
	rột	ruột /zɯot/ intestin
	rộng	ruộng /zɯoŋ/ rizière
/ɛ/ > /ie/	è	ìa /ie/ aller à la selle

Quelques correspondants trouvés entre l'ancien sino-vietnamien et le sino-vietnamien suggèrent que ce phénomène est dû à l'influence du chinois (Nguyễn Tài Căn, 1995, p. 302)

/u/ > /uo/	phù [fu] > bùa /buo/ amulette
/ɯ/ > /ɯɤ/	lư [lɯ] > lừa /lɯɤ/ tromper
/i/ > /ie/	lì [li] > lìa /lie/ quitter

b. En position finale, les palatales */ɲ/ */c/ ont évolué en alvéolaires respectivement /n/ /t/ (Nguyễn Tài Căn, 1995, p. 313)

*/tɲ/ > đăn /dan/ tricoter
*/mɔc/ > mọt /mɔt/ perce-bois

³⁷ Ce dialecte parlé dans le nord du Centre regroupe six provinces (Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình, Quảng Trị et Thừa Thiên Huế) où le phénomène de diphtongaison n'est pas attesté.

Les finales /ŋ/ /k/ se sont palatalisées après voyelles antérieures /i e ε/ ; mais labialisées après voyelles postérieures arrondies /u o ɔ/. Ce phénomène est dû très probablement à l'influence du chinois (voir Nguyễn Tài Căn, 1995, p. 201)

Les sections suivantes présentent la structure du vietnamien actuel aux niveaux syllabique et segmental. Les caractéristiques prosodiques de la langue (tons, intonation, accent) seront abordées à la suite. Pour les besoins des recherches en FLE dans le cadre du vietnamien langue source, à la fin de ces sections, sont exposées des remarques sur la distribution des unités segmentales et suprasegmentales de la syllabe comparativement avec ces mêmes unités en français.

2.2. Structures des unités lexicales du vietnamien

Le vietnamien est une langue isolante à tons (Trương Văn Chính, 1970 ; Stankevic & Nguyễn, 1975 ; Hoàng Tuệ & Hoàng Minh, 1975 cité par Cao Xuân Hạo, 1985 ; Mai Ngọc Chừ et al., 1997 ; Đoàn Thiện Thuật, 1999). Par définition, une langue isolante est une langue dont « les « mots » sont ou tendent à être invariables et où on ne peut pas, par conséquent, distinguer le radical et les éléments grammaticaux » (Dubois et al., 2001, p. 258). Ainsi, en vietnamien, des phrases comme « tôi nhìn họ » [toj ɲin hɔ] *je les regarde*, « họ nhìn tôi » [hɔ ɲin toj] *ils me regardent* illustrent ce propos. Les unités « tôi », « họ » ne changent pas de forme que leur fonction soit sujet ou objet. De même, le verbe « nhìn » *regarder* est invariable et ne s'accorde pas au sujet quel que soit le nombre, ce qui est différent dans les langues flexionnelles dans lesquelles les unités changent de forme (*je-moi, ils-les-leur-eux* en français par ex.) en fonction de leur fonction grammaticale. Selon Cao Xuân Hạo (1985), « la langue vietnamienne représente l'un des spécimens les plus typiques de langue syllabique isolante qui soit plus ou moins connue en linguistique. En vietnamien, le mot, le morphème et la syllabe semblent coïncider presque entièrement » (Cao, 1985, p. 260). Ainsi, la presque totalité des syllabes vietnamiennes sont des unités signifiantes, dotées selon cet auteur de plus d'un statut lexical : « Si on peut comparer une langue européenne à un mécanisme tournant sur les trois pivots constitués par les unités de base que sont le phonème, le morphème et le mot, le mécanisme de la langue vietnamienne semble réunir ces trois pivots en un seul, la syllabe » (Cao, 1975, p. 99).

2.2.1. Mot et Syllabe

Le vietnamien est caractérisé comme une langue monosyllabique, la description traditionnelle associant tout mot vietnamien à une seule syllabe. Nguyễn Phú Phong (1989) a explicité ce trait particulier de la structure de cette langue à l'oral par le fait que la syllabe du

vietnamien possède des frontières claires, bien délimitées, tant sur le plan phonique que graphique : elle peut être isolée des autres à l’oral par les pauses qui la précèdent ou la suivent et elle est aussi facile à repérer dans sa forme écrite. Ce caractère dominant de la syllabe vietnamienne vient du fait que « *la plupart des syllabes sont des morphèmes et une proportion très importante des morphèmes a le statut de mot, ce qui a conduit nombre d’auteurs à qualifier le vietnamien de « langue monosyllabique »* » (Nguyễn Phú Phong, 1989, p. 27).

Pourtant, dire que le mot vietnamien est dans tous les cas monosyllabique est une erreur, car sur le plan lexical, il existe un nombre non négligeable de mots composés de plusieurs syllabes. La notion de mot est particulièrement complexe en vietnamien et demande de comprendre le processus de formation de composé lexical polysyllabique dans cette langue.

Il existe, dans certains mots composés, des syllabes qui ne sont aucunement des mots, qui sont dépourvues de significations propres et qui ne peuvent être employées indépendamment. Exemple « *đuròi uoi* » /dʊrɔj-ʔurɔj/³⁸ *orang-outan*. Les deux syllabes « *đuròi* » et « *uoi* », prises isolément, n’ont aucune signification et ne sont pas employées telles quelles dans la langue. Seul l’ensemble « *đuròi uoi* » a un contenu sémantique et est employé dans la langue. En d’autres termes, les signifiants [dʊrɔj] et [ʔurɔj] ne sont pas des morphèmes, mais leur réunion constitue un morphème à deux syllabes, appelé traditionnellement mot composé pur (Trương Văn Chỉnh, 1970, p. 7). Rentrant dans cette catégorie les emprunts étrangers transcrits phonétiquement, à l’exclusion des mots sino-vietnamiens (ex. « *ban công* » /ban-kon/ *balcon*, « *bê tông* » /be-ton/ *béton*, « *ô tô* » [o-to] *auto*).

Les mots vietnamiens sont monosyllabiques (mots simples) ou polysyllabiques (mots composés) (Trương Văn Chỉnh, 1970, p. 3). Il existe des mots composés à deux syllabes (ex. « *sạch sẽ* » /sək-ʂe/ *propre*, « *bản thiù* » /bʰn-tʰiw/ *sale*), à trois syllabes (ex. « *sạch sành sanh* » /sək-ʂɛŋ-ʂɛŋ/ *complètement épuisé ou vidé*), à quatre syllabes (ex. « *khập khà khập khiễng* » /χʰp-χa-χʰp-χiɛŋ/ *clopin-clopant*). Les composés à deux syllabes sont les plus nombreux. Il n’existe pas de composés de plus de quatre syllabes.

La description traditionnelle classe les mots composés en deux catégories principales du point de vue sémantique (Mai Ngọc Chừ et al, 1992 ; Nguyễn Thiện Giáp, 1998 ; Diệp Quang Ban, 2004).

- 1) *Les composés coordonnés* sont ceux dont les éléments composants ont une relation égale sur le plan grammatical. La combinaison de sens entre les constituants a pour effet de créer un sens général, global que porte le composé réalisé (Diệp Quang Ban, 2004, p. 43). Dans le processus de création du sens, ces composés coordonnés se divisent en trois sous-catégories :

³⁸ Le tiret figurant entre les syllabes représente la frontière syllabique en cas des mots composés.

- a. Les composés coordonnés itératifs dont la combinaison de sens des éléments composants manifeste un sens commun de mot composé. Dans ce cas, les éléments composants pris isolément ont chacun un sens différent et précis.

Ex. « học » *apprendre* associé à « tập » *exercer, s'entraîner*, forme le mot composé « học tập » qui porte un sens plus général *faire des études*. Le mot composé « xăng dầu » *carburant* est constitué par les deux éléments « xăng » *essence* et « dầu » *pétrole*, « nghe nhìn » *audio-visuel*, « thu phát » *réception et émission de radio*, « tướng tá » *généraux et officiers supérieurs*, « sống chết » *vivre et mourir*.

- b. Les composés coordonnés répétitifs formés par la juxtaposition de deux éléments composants ayant des significations semblables ou voisines. Le vietnamien comprend dans son vocabulaire un grand nombre de mots homophones ou de mots ayant plusieurs significations. Pour « distinguer entre eux les homophones, comme pour distinguer les différentes significations d'un même mot, et afin d'éviter toute équivoque, toute méprise de la part de l'interlocuteur, la langue a recours à la formation des mots composés, soit en juxtaposant deux synonymes, soit en ajoutant au mot simple une syllabe asémantique » (Trương Văn Chính, 1970, p. 4). C'est ainsi qu'à partir des mots simples « dầu » et « già », les composés « dầu vêt » et « già nua » ont été formés dans les exemples suivants :

Le mot « dầu » *trace* a une autre signification plus rare *aimer*. Le composé « dầu vêt » *trace* est formé de deux synonymes « dầu » et « vêt » pour se distinguer de « yêu dầu » *aimer* qui est formé également de deux synonymes « yêu » et « dầu ».

Le mot « già » dont le sens est *âgé, vieux*, s'emploie par extension pour *expérimenté*. « nua », un mot sino-vietnamien qui signifie *courbé sous le poids du temps* s'ajoutant à « già » forment « già nua » pour se distinguer de « già giã » *expérimenté* formé avec un élément asémantique « giã ».

On peut trouver d'autres exemples de mots composés formés par deux synonymes : « tìm kiếm » *chercher*, « đợi chờ » *attendre*, « bởi vì » *parce que*, « thay đổi » *changer, modifier*, núi non *montagne* ...

- c. Les composés coordonnés singulatifs dont la signification s'appuie uniquement sur un seul élément composant, le deuxième élément ayant perdu le trait sémantique compatible du mot composé avec l'évolution de la langue. Celui-ci ne sert qu'à créer le sens « commun » du mot composé et n'est pas employé isolément dans la langue. Sinon, il possède une autre signification dissemblable au sens commun du mot composé. Ex : Dans « chợ búa » *marché*, « chợ » est l'élément constitutif ayant même signification que le mot composé, l'élément « búa » signifie *marteau*. Dans « rộng rãi » *vaste*, « rộng » est le mot simple constitutif signifiant *vaste* ; « rãi » par lui-même, n'a pas de sens. Celui-ci apparaît uniquement dans le mot composé « rộng rãi » et n'est jamais utilisé isolément.

Il en est de même pour les composés de plus de deux syllabes : celles qui sont autres que le mot constitutif sont des éléments asémantiques (Trương Văn Chính, 1970). Aussi « *sạch sành sanh* » est un morphème de trois syllabes qui signifie *complètement épuisé ou vidé*. « *sành* » et « *sanh* » sont des éléments asémantiques.

- 2) *Les composés principo-subordonnés* sont ceux dont les éléments constitutifs ont une relation de subordination. Dans l'association, le premier élément ayant un sens générique est la base, le second ne joue qu'un rôle subordonné du premier. L'adjonction de l'élément subordonné contribue alors à donner au premier élément une précision. Ainsi, dans « *xe máy* » *motocycle*, « *xe đạp* » *vélo*, « *xe ngựa* » *voiture à chevaux*, « *xe xích lô* » *cyclo*, « *xe điện* » *tramway* ; l'élément « *xe* » est le terme générique qui désigne le véhicule ; « *máy* », « *đạp* », « *ngựa* », « *xích lô* », « *điện* » sont des éléments subordonnés qui expriment la discrimination. Grâce à ces éléments subordonnés, on peut distinguer les différentes sortes de véhicules. La combinaison de sens entre eux donne un sens spécifique au mot composé.

Dans la langue vietnamienne, outre la composition, la reduplication est une façon particulière de former des mots et des expressions. Les reduplications forment les mots composés constitués de deux à quatre syllabes qui ont entre elles un lien phonétique, soit au niveau du ton, soit au niveau segmental ou sub-syllabique (initiale et/ou rime). « Les mots composés par reduplication phonétique sont, pour la plupart, des composés purs ou des composés coordonnés singulatifs » (du point de vue sémantique) (Trương Văn Chính, 1970, p. 5). On pourra ainsi classer phonétiquement les composés en distinguant trois types de formation :

- a. Les composés par reduplication intégrale :
Ex : *chuồn chuồn libellule*, *rành rành évident*, *ba ba tortue d'eau douce*
- b. Les composés par reduplication de la consonne initiale :
Ex : *ngỗn ngang en désordre*, *xấu xa laid*, *vắng vè désert*, *buồn bã triste*
- c. Les composés par reduplication de la rime :
Ex : *lơ mơ somnolent*, *bẽn lẽn avoir honte*, *thần lãn margouillat*, *tỉ mỉ minutieux*

Il en résulte que l'unité lexicale en vietnamien, simple ou composée, est mono-morphémique. La différence entre « simple » et « composé » n'existe que par le nombre de syllabes : le mot simple est monosyllabique, le mot composé a de deux à quatre syllabes. Le mot vietnamien, simple ou composé, peut être comparé au mot mono-morphémique français qui est soit monosyllabique (ex. *livre*, *jaune*) soit plurisyllabique (ex. *col-bleu*, *pourboire*). Cette comparaison ne se limite pourtant, du point de vue formel, qu'au nombre de syllabes. Sur le plan sémantique, seul le mot composé pur en vietnamien équivaut au mot simple plurisyllabique en français, ce dernier n'ayant rien de commun avec le mot composé

vietnamien qu'il soit reduplicatif de sens ou à un seul élément sémantique (Truong Văn Chỉnh, 1970, p. 45). Ainsi, les deux syllabes de « stylo » /sti-lo/ prises isolément n'ont aucune signification, tout comme les deux syllabes de « đười ươi » *orang-outan*. Par contre, les mêmes syllabes de « stylo » ne sont pas comparables aux deux syllabes de « rộng rãi » *vaste* qui contient un élément sémantique *rộng*, ni à celles de « dấu vết » *trace* formé de deux mots simples.

La section suivante présente la structure de la syllabe du vietnamien.

2.2.2. Structure syllabique : différents points de vue théoriques

Définie comme une suite de segments sonores dont le groupement est soumis à des contraintes particulières, la notion de la syllabe est fréquemment utilisée dans le traitement de nombreux problèmes linguistiques, comme le souligne Rousset (2004) : « La difficulté des recherches sur la syllabe, en tant qu'objet d'étude propre, semble résider dans le fait qu'on admet généralement sa pertinence mais sans vraiment la définir explicitement. Cet état de fait résulte probablement de la connaissance intuitive de la syllabe » (Rousset, 2004, p. 13).

La structure syllabique du vietnamien est, jusqu'à présent, débattue par les linguistes. Hoàng Tuệ et Hoàng Minh (1975) proposent une description de la structure de la syllabe du vietnamien présentée figure 2.1 ci-après.

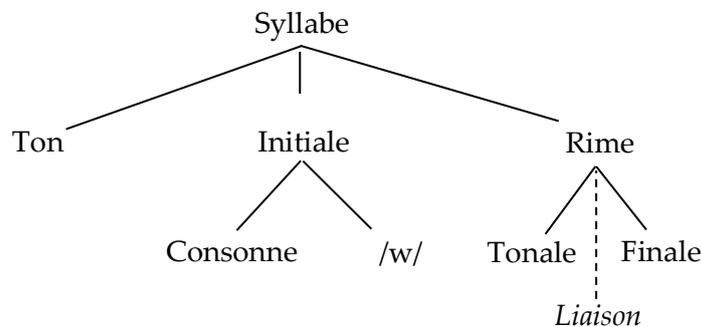


Figure 2.1. Représentation arborescente de l'organisation interne de la syllabe vietnamienne selon Hoàng Tuệ et Hoàng Minh (1975) (d'après Cao Xuân Hạo, 1985, p. 271).

Le modèle de structure syllabique de Hoàng Tuệ et Hoàng Minh (1975) est ainsi commenté par Cao Xuân Hạo (1985, p. 270) : « L'initiale est composée d'un ensemble de traits consonantiques et d'un trait de résonance (labio-vélarisation). La rime se compose d'un ensemble de traits représentant le timbre dominant de la syllabe (la tonale), du mode de terminaison de la syllabe (la finale) et du trait de liaison (ferme ou lâche) qui unit le segment vocalique à la finale quand celle-ci n'est pas nulle ». Cao Xuân Hạo (1985, p. 264) distingue deux types de syllabes : une syllabe dite ferme quand elle est composée de « voyelle brève, de consonne longue et articulée énergiquement ». Une syllabe dite lâche lors qu'elle contient

une « voyelle longue et une consonne brève et qu'elle est articulée faiblement ». « Dans la liaison ferme, la finale est attaquée très tôt, quand la courbe d'intensité de la syllabe atteint à peine son sommet, de sorte que le mouvement fermant est beaucoup plus énergique que dans la liaison lâche, où la finale est articulée quand la syllabe arrive à la fin » (Hoàng Tuệ et Hoàng Minh, 1975, p. 85 cité par Nguyễn Thị Ngân Hà, 2001, p. 28).

Cao Xuân Hạo (1985, p. 264) propose d'analyser la syllabe comme une combinaison de six éléments non segmentaux qui se chevauchent, les trois premiers se réalisant simultanément :

1. Le ton (registre haut ou bas, dessin mélodique plat ou modulant, présence ou absence de constriction glottale)
2. La tonalité dominante (gravité, acuité ou timbre neutre, constance ou décoloration (ou diphtongaison centralisante)).
3. La plénitude vocale (compacité ou degré d'aperture)
4. L'initiale (traits consonantiques d'attaque plus coloration bémolisée ou non bémolisée)
5. La finale (traits consonantiques de fermeture de syllabe)
6. Le contact des deux segments vocaliques et consonantiques (ou semi-vocaliques) qui terminent la syllabe (ferme ou lâche).

Cao Xuân Hạo (1965, p. 50) définit ces éléments comme « des ensembles de traits pertinents groupés d'après leur place et leur extension dans la syllabe (« leur fonction constituante de syllabe »). Cependant, à la lumière de recherches ultérieures, Cao Xuân Hạo trouve que « cette description ne rend pas compte de la division de la syllabe en ses trois éléments constitutifs directs – le ton, la rime et l'initiale –, division réelle et importante que la phonologie ancienne a constatée et que les travaux de Gordina ont confirmée » (Cao, 1980, p. 265).

Hồng Giao (1975) reprend l'idée de Gordina et Andreev (1958), Cù Đình Tú (1977) (cité par Đoàn Thiện Thuật, 1999) selon laquelle la syllabe est divisée en deux parties : initiale et rime (cf. table 2.2). Dans cette structure de la syllabe, le ton couvre seulement la rime, mais pas la partie initiale, et la rime comprend aussi la prétonale, contrairement à la proposition de Hoàng Tuệ et Hoàng Minh (1975), reprise par Cao Xuân Hạo (1985).

Table 2.2. Représentation schématique de l'organisation interne de la syllabe (Hồng Giao, 1975).

Initiale	Ton
	Rime

	Prétonale	Médiane	Finale
--	-----------	---------	--------

Quant à savoir si le ton est porté par toute ou partie de la syllabe, une étude expérimentale de Nguyễn Hàm Dương (1963, cité par Đoàn Thiện Thuật) utilisant l'intonographe montre que le ton est superposé à toute la syllabe, y compris l'initiale. Đoàn Thiện Thuật (1999) confirme cette analyse en étudiant des phénomènes connus dans la littérature vietnamienne tels que le mode de réduplication, le mode de contrepèterie, la formation des mots avec « iêc » [iek] et la loi d'affinité des tons dans la poésie (Đoàn Thiện Thuật, 1999, p. 71). D'après Đoàn, le ton ne peut pas faire partie de la rime seule parce qu'il peut séparer morphologiquement la partie initiale de la rime (Đoàn Thiện Thuật, 1999, p. 94). À titre d'exemple, il donne le mot réduplicatif « khe khẽ » / $\chi\epsilon^1-\chi\epsilon^3$ ³⁹ doucement. La syllabe [$\chi\epsilon$] est répétée au niveau segmental alors que le ton *ngang* (ton 1 ou A1) (qui n'a pas de marque graphique) de la première syllabe a été échangée en ton *ngã* (ton 3 ou C2) (signe tilde) dans la seconde. Il remarque cependant que « bien que la tonalité couvre toute la syllabe ou seulement la rime, cela n'influence guère sa valeur phonologique dans la syllabe » (Đoàn Thiện Thuật, 1999, p. 94). D'autres linguistes vietnamiens partagent la même idée que Đoàn Thiện Thuật en proposant une segmentation de la syllabe en deux niveaux : l'initiale et la rime sous le ton au premier niveau ; au deuxième niveau c'est la division de la rime en segment intermédiaire (la prétonale), segment principal (noyau) et segment final (coda). À partir d'éléments de la littérature, Đoàn Thiện Thuật atteste que le ton est un élément indépendant de la syllabe, couvrant la partie initiale et la rime (cf. table 2.3).

Table 2.3. Représentation schématique de l'organisation interne de la syllabe en vietnamien d'après Đoàn Thiện Thuật (1999).

	Ton		
	Rime		
Initiale	segment	segment	segment
	intermédiaire	principal	final

Dans la dynamique internationale des recherches actuelles en reconnaissance et synthèse de la parole, les études récentes sur le traitement automatique du vietnamien ont livré quelques résultats intéressants sur la réalisation du ton dont la persistance d'un désaccord entre spécialistes concernant les caractéristiques de la fréquence fondamentale (F0) sur toute ou partie de la syllabe :

1. porté par la syllabe entière (Nguyễn Tiến Dũng, Mixdorff, Lương C. M., Ngô H. H. & Vũ K. B., 2004, Đỗ Trọng Tú & Takara, 2003)

³⁹ Rappelons que dans les transcriptions phonétiques, nous utilisons des chiffres (de 1 à 6) situés à droite de la syllabe concernée pour la notation des tons.

2. porté essentiellement par la rime de la syllabe (Nguyễn Quốc Cường, 2002, Trần Đỗ Đạt, Castelli, Serignat, Lê X. H. & Trịnh V. L., 2005)

À noter que dans la représentation proposée par Đoàn Thiện Thuật (1999), certains éléments sont facultatifs. C'est le cas du segment intermédiaire et du segment final. Par exemple, le segment final est vide dans « ra » /zə/ *sortir* mais il est occupé par la consonne /t/ dans « hát » /hat/ *chanter* ou la semi-consonne /j/ dans « hai » /haj/ *deux* ou /w/ dans « đầu » /daw/ *avoir mal*. Dans la syllabe « hoa » /hwa/ *fleur* ou « quà » /kwa/ *cadeau*, le segment intermédiaire /w/ est présent avec l'écriture « o » dans « hoa » et « u » dans « quà ». Il est absent dans les syllabes « hát » /hat/ *chanter* ou « nói » /noj/ *parler*.

Autre point de divergence entre linguistes, la prétonale (ou le segment intermédiaire) qui, selon Cao Xuân Hạo (1975) fait partie de l'initiale alors que pour d'autres (Hoàng Giao, 1975 ; Đoàn Thiện Thuật, 1999), elle appartient à la rime. Lorsque la prétonale est présente, elle ne peut être que le glide /w/ labiovélaire. En étudiant le processus de formation des reduplicatifs en vietnamien, lequel est soumis à certaines contraintes et régit par certaines lois phonétiques, Đoàn Thiện Thuật montre que l'appartenance de la prétonale à la rime est plus logique. Ainsi, certaines reduplications ne concernent que l'initiale : les deux syllabes du reduplicatif ont la même initiale, des rimes différentes et, pour la plupart, les tons appartiennent au même registre (soit série haute, soit série basse), en vertu de la première loi d'affinité⁴⁰. En effet, dans « lấp lòè » /lɤp⁶-lwe²/ *briller et s'éteindre*, la rime /ɤp/ est séparée de l'ensemble pour pouvoir être remplacée par une autre /wɛ/. Il en est de même pour les mots composés par reduplication de la rime dans lesquels les deux syllabes du reduplicatif ont la même rime, des initiales différentes et, pour la plupart, le même ton (ex. « loạng choạng » /lwaŋ⁶-cwaŋ⁶/ *chanceler*). En vietnamien, il est aussi possible de former des reduplicatifs à partir d'un grand nombre de mots simples, par l'addition d'un élément asémantique ayant la même initiale et dont la rime est « iêc » [iek]. L'élément redoublant qui se place toujours après le mot simple constitutif ajoute généralement une nuance de mépris, de plaisanterie ou de grande indifférence. En raison de sa plosive sourde finale, il a soit le ton 5 (ton D1) (-iêc) soit le ton 6 (ton D2) (-iêc). Par exemple : « Nó có **học hiêc** gì đâu !» *Il n'étudie rien!*. Ces faits linguistiques convergent pour une position de la prétonale dans la rime plutôt que dans l'attaque.

Compte tenu de l'étude de Đoàn Thiện Thuật (1999) et de Trần Đỗ Đạt & al. (2005), nous retenons pour notre étude la représentation suivante de la structure syllabique du

⁴⁰ Loi d'affinité des tons : la formation des reduplicatifs est soumise principalement à deux lois d'affinité des tons : affinité des tons appartenant à un même registre, et affinité des tons appartenant à un même groupe. Autrement dit, dans les reduplicatifs, les tons des syllabes constituantes appartiennent soit au même registre, soit au même groupe. (selon Trương Văn Chính, 1970, p. 14)

vietnamien⁴¹ (cf. tab 2.4). Les valeurs quantitatives pour chacun des constituants de la structure syllabique dans un lexique de 40 000 entrées sont données ci-dessous par Lê Việt Bắc (2006, p. 88).

La diversité des structures de syllabe en vietnamien est plutôt limitée. La position d'attaque est toujours occupée par une consonne. Le noyau de la syllabe est toujours un élément vocalique (simple ou diphtongue), tandis que la position finale est facultative et peut être occupée par une consonne. Ainsi, le vietnamien compte quatre types de structures syllabiques (cf. table 2.5).

Table 2.4. Structure syllabique du vietnamien avec entre parenthèses le nombre de phonèmes pouvant apparaître dans chaque position (axe paradigmatique), d'après Lê Việt Bắc (2006).

Syllabe avec ton (6 686)		
Syllabe sans ton (2 376)		
Attaque (22)	Ton (6)	
	Rime (155)	
	Prétonale (1)	Noyau (16) Coda (8)

Table 2.5. Les types de structure syllabique du vietnamien. (C₁ : attaque ; /w/ : prétonale ; V : noyau vocalique ; C₂ : coda)

N ^o	Types de structures syllabiques	Exemples
1	C ₁ + V	ô [ʔo] <i>parapluie</i> , ba [ba] <i>trois</i>
2	C ₁ + V + C ₂	ăn [ʔăn] <i>manger</i> , hát [hat] <i>chanter</i>
3	C ₁ + /w/ + V	oa [ʔwa] <i>vagissement d'un nouveau-né</i> , hoa [hwa] <i>fleur</i>
4	C ₁ + /w/ + V + C ₂	oan [ʔwan] <i>victime d'une injustice</i> , toán [twan] <i>mathématique</i>

2. 3. Inventaire et distribution des phonèmes

2.3.1. L'inventaire des consonnes en attaque

Les consonnes initiales forment un système de 22 phonèmes. La table 2.6 ci-après présente l'inventaire des consonnes en position d'attaque du vietnamien (Đoàn Thiên Thuật, 1999, p. 153).

⁴¹ Dans notre étude, à la place de Initiale, Tonale et Finale, nous utiliserons la terminologie Attaque, Noyau, Coda car c'est celle retenue pour toutes les études conduites à partir d'ULSID (Rousset, 2004 ; Vallée, Rossato, Rousset, 2009).

Table 2.6. Les consonnes du vietnamien en position d'attaque.

	Bilabial	Labio-dental	Alvéolaire	Rétroflexe	Palatal	Vélaire	Uvulaire	Glottal
Plosives	b		t t ^h d	ʈ	c	k		ʔ
Nasales	m		n		ɲ	ŋ		
Fricatives		f v	s z	ʂ ʐ		ɣ	χ	h
Latérale			l					

Les initiales sont au nombre de 19 pour le dialecte du Nord, 22 pour celui du Centre et du Sud : les consonnes rétroflexes /ʈ/, /ʂ/, /ʐ/ sont absentes de la variété du Nord en raison d'assimilation :

- /t/ et /c/ se prononcent [tɕ] selon notre corpus d'analyse acoustique (cf. chapitre IV) du parler du Nord. Ainsi, « **tra**i » /tʰaj¹/ *garçon* ou « **ch**ai » /caj¹/ *bouteille* sont homophones et se réalisent tous les deux [tɕaj¹].
- /s/ et /ʂ/ dans « **ăn** xôi » /ʔăn¹.soj¹/ *manger le riz gluant* et « **nước** sôi » /nưək⁵.ʂoj¹/ *eau bouillie* sont réalisés tous les deux [s].
- /z/ dans « **gi**a » /zə¹/ morphème sino-vietnamien qui signifie *maison*, /z/ dans « **đ**a » /za¹/ *la peau* ou « **r**a » /za¹/ *sortir* (prononcé rarement [ra¹] dans quelques parlers du Nord) : les trois mots se prononcent [za] dans le dialecte du Nord.

Bien que sans correspondance graphique dans l'écriture vietnamienne, le coup de glotte /ʔ/ ou occlusion glottale est un phonème faisant partie des unités distinctives du vietnamien (Cao Xuân Hao, 1985 ; Đoàn Thiện Thuật, 1999). Il n'en a pas toujours été de même puisque certains auteurs, surtout au XVII^e siècle, pensaient que le coup de glotte représentait une attaque normale des mots à initiales vocaliques (Đoàn Thiện Thuật, 1999, p. 155), alors que la plupart des linguistes actuels attestent qu'il fonctionne en tant que consonne à valeur de phonème (Lê Văn Lý, 1948 ; Thompson, 1965 ; Hoàng Tuệ & Hoàng Minh, 1975 ; Đoàn Thiện Thuật, 1999).

2.3.2. La prétonale /w/

Le cas du segment prétonal /w/ est particulier. Certaines descriptions du vietnamien proposent pour ce segment une alternance avec zéro (en l'absence de la partie prétonale) (Đoàn Thiện Thuật, 1999).

En ce qui concerne la distribution de /w/, plusieurs remarques sont à faire (Đoàn Thiện Thuật, 1999, p. 176) :

- /w/ n'apparaît jamais derrière les consonnes labiales /b m f v/ et apicales /n z/ (sauf dans les mots d'emprunt comme par exemple « buýt » [bwit⁵] *bus* du français).
- l'association de la semi-voyelle /w/ avec la consonne initiale /n/ apparaît seulement pour le mot d'origine chinoise « noãn » [nwan³] *ovule*.
- la consonne /z/ apparaît devant la semi-voyelle /w/ uniquement dans le mot « roạt » [zwat⁶] *bruit de faucille coupant les céréales*.
- la semi-voyelle peut apparaître après toutes les consonnes initiales restantes (cf. table 2.7). Le signe (+) signifie que la combinaison est possible, le signe (-) signifie qu'elle est impossible.

Table 2.7. Combinaisons possibles entre la prétonale et les initiales (Nguyễn Quốc Cường, 2002).

	ʔ	t	t ^h	ʈ	c	k	b	d	m	n	ŋ	ɲ	f	s	ʃ	χ	h	v	z	z̥	ɣ	l
w	+	+	+	+	+	+	-	+	-	-	+	+	-	+	+	+	+	-	+	-	+	+

2.3.3. Le système du noyau

Le vietnamien possède 9 phonèmes vocaliques /i e ε a u ɤ u o ɔ/, 4 voyelles brèves /ě ä ɣ ɔ̃/ et 3 diphtongues /ie uɤ uo/.

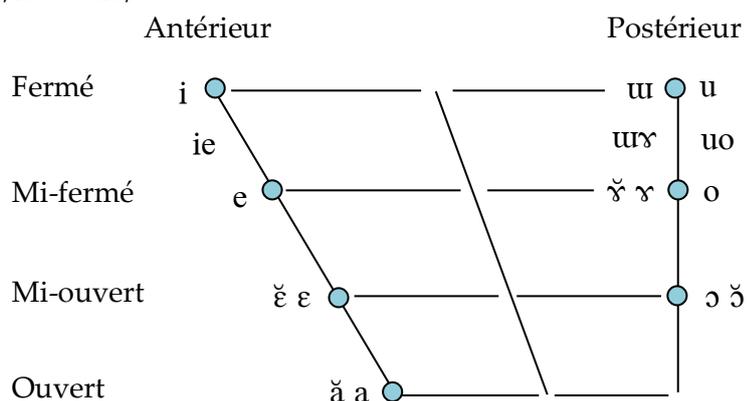


Figure 2.2. Description en traits de lieu et d'aperture du système vocalique du vietnamien.

Une voyelle non brève peut terminer la syllabe alors qu'une voyelle brève est toujours suivie d'une consonne finale. Les 3 diphtongues sont ouvrantes, formées sur le degré d'aperture fermé à mi-fermé, des voyelles antérieures étirées /ie/, des voyelles postérieures non

arrondies /uɤ/ ou arrondies /uo/. Dans le cadre de l’allongement de la syllabe comme par exemple, sous l’accent (cf. section 2.4.4), c’est la première partie de la diphtongue qui s’allonge au détriment de la seconde. Ex : la première syllabe « *viêt* » de mot « *Việt Nam* » [viet⁶.nam¹] ne peut être identifiée que selon cette prononciation [vi:et] mais non [vie:t].

Une controverse sur l’opposition non brève / brève des voyelles monophthongues existe sur le nombre de voyelles (de 16 à 14 phonèmes selon les cas). Ci-dessous, quatre paires minimales illustrent ce trait distinctif :

Phonèmes non brefs : /a ɤ ε ɔ/	Phonèmes brefs : /ă ǎ ɛ ɔ/
san /sɑn ¹ / <i>aplanir</i>	săn /sǎn ¹ / <i>chasser</i>
son /sɤn ¹ / <i>peinture</i>	sân /sǎn ¹ / <i>cour</i>
éc /ʔɛk ⁵ / <i>grognement</i>	ách /ʔɛk ⁵ / <i>joug</i>
xoong /sɔŋ ¹ / <i>casserole</i>	xong /sɔŋ ¹ / <i>finir</i>

Certains linguistes (Lê Văn Lý, 1948 ; Emeneau, 1951 cité par Đoàn Thiện Thuật, 1999) constatent que les paires minimales d’opposition phonologique entre /ɛ/ et /ε/, /ɔ/ et /ɔ/ ont peu de rendement. Des paires comme « *ông* » [ɔŋ¹] vs. « *oong* » [ɔŋ¹], « *óc* » [ɔk⁵] vs. « *oóc* » [ɔk⁵], « *cảnh* » /kǎŋ⁴/ *paysage* vs. « *keng* » /kɛŋ⁴/ *sonnerie*, « *ách* » /ʔɛk⁵/ *boucher* vs. « *éc* » /ʔɛk⁵/ *grognement* sont rares dans la langue, les voyelles [ɔ] et [ɛ] étant parfois considérées comme des variantes des phonèmes /ε/ et /ɔ/. Dans cette proposition, le système vocalique du vietnamien est alors composé de 14 phonèmes : 11 voyelles monophthongues (sans /ɛ/ et /ɔ/) et 3 diphtongues.

Cependant, Haudricourt (1951), Đoàn Thiện Thuật (1999) et d’autres linguistes considèrent en revanche les phonèmes /ɛ/ et /ɔ/ comme des phonèmes distincts et non des variantes de /ε/ et /ɔ/. Selon Đoàn Thiện Thuật (1999, p. 264, *t.p.n*), « les rimes /ɛŋ/, /ɛk/, /ɔŋ/ et /ɔk/ malgré leurs apparitions restreintes dans les mots vietnamiens, ne doivent pas être oubliées lors de la détermination des phonèmes de cette langue ». Đoàn Thiện Thuật (1999, p. 201) atteste que les paires vocaliques /ɛ/~ε/ ; /ɔ/~ɔ/ dans les mots susmentionnés possèdent les mêmes caractéristiques qualitatives (même timbre, même degré d’aperture), elles ne se distinguent que par leur durée. À partir d’analyses acoustiques de la durée de la voyelle en contexte CVC, Nguyễn Việt Sơn (2009) a constaté que les séries de voyelles longues et brèves /a ă/, /ɤ ǎ/, /ɔ ɔ/ sont distinguées par leur durée, les voyelles /a ɤ ɔ/ étant significativement plus longues que leurs homologues brèves. La voyelle brève /ɛ/ se comporte plutôt comme une diphtongue. Dans cette étude, les résultats de tests perceptifs ont confirmé que les Vietnamiens sont parfaitement capables de distinguer les voyelles non brèves des voyelles brèves en se basant seulement sur leur durée relative.

Table 2.8. Combinaisons possibles entre la prétonale et les voyelles (Nguyễn Quốc Cường, 2002).

	i	ɯ	u	e	ɤ	o	ɛ	a	ɔ	ɤ̣	ɛ̣	ǎ	ɔ̣	ie	ɯɤ	uo
w	+	-	-	+	+	-	+	+	-	+	+	+	+	+	-	-

Les voyelles /ɤ̣ ǎ/ sont réalisées devant toutes les consonnes alors que /ɛ̣/ et /ɔ̣/ se trouvent uniquement devant vélaire /k/ ou /ŋ/. Les voyelles postérieures /ɯ u o ɔ/ comme les deux diphtongues /ɯɤ uo/ ne suivent jamais la prétonale /w/ comme le montre la table 2.8.

2.3.4. Les consonnes finales

L'inventaire des consonnes en finale de syllabe est relativement restreint en vietnamien. Il comprend 3 plosives sourdes /p t k/, 3 nasales /m n ŋ/ et 2 semi-consonnes /w j/.

Table 2.9. Description des consonnes finales possibles en vietnamien.

	Bilabial	Alvéolaire	Palatal	Vélaire	Labio-vélaire
Plosive	p	t		k	
Nasale	m	n		ŋ	
Approximante			j		w

Les syllabes en vietnamien peuvent être terminées par une consonne occlusive /p t k m n ŋ/ (ex. « khát » /χat⁵/ avoir soif, « tam » /tam¹/ trois, « ngang » /ŋaŋ¹/ horizontal, etc.), par un glide /w/ ou /j/ (ex. « hai » /haj¹/ deux, « đau » /dǎw¹/ avoir mal). Rappelons que la syllabe peut être ouverte ou fermée. Les proportions de la fréquence des structures ouverte CV et fermée CVC en vietnamien sont respectivement 69,5 % et 24,1 % (voir section 3.2.4 pour plus de détails).

Les finales /k/, /ŋ/ ont chacune deux variantes contextuelles :

- Après des voyelles antérieures /i, e, ɛ̣/, elles connaissent une antériorisation du lieu vélaire et se réalisent dorso-palatales « nh » [ɲ] et « ch » [ç]. « nh » et « ch » sont les diagraphes utilisés respectivement pour les variantes palatalisées de /ŋ/ et /k/. Exemple :

sinh /ɕiŋ¹/ [ɕiŋ¹] *naître*, bệnh /beŋ⁶/ [beŋ⁶] *maladie*, tránh » /tʃɛ̣⁵/ [tʃɛ̣⁵] *éviter*
 dịch /zic⁶/ [zic⁶] *traduire*, lếch /lek⁶/ [lec⁶] *déjeté*, sách » /sʃɛ̣k⁵/ [sʃɛ̣⁵] *livre*

- Dans les autres cas, elles sont articulées vélaire « ng » [ŋ], « c » [k] :

tiếng keng /tieŋ⁵.keŋ⁴/ *signal sonore fait d'une barre de métal*
 xứng đáng /suɯŋ⁵.daŋ⁵/ *mériter* ; trông mong /tɔŋ¹.mɔŋ¹/ *attendre*
 séc siéc /sɛ̣k⁵.ɕiek⁵/ *chéquier* ; xác xươc /sʃk⁵.suɯk⁶/ *impudent*

La semi-consonne /w/ n'apparaît qu'après les voyelles non-arrondies et /j/ uniquement après les postérieures. Donc, /w/ et /j/ ne sont en opposition qu'après les voyelles /ɯ ʏ ɤ a ǎ/ (cf. table 2.10). À la différence de /w/, /j/ ne se rencontre qu'en position finale.

Table 2.10. Combinaisons possibles entre voyelles et semi-consonnes finales.

Monophthongues	Diphthongue	Finale /w/	Finale /j/
i e ε	ie	+	-
u o ɔ	uo	-	+
ɯ ʏ ɤ a ǎ	ɯʏ	+	+

Une des caractéristiques les plus connues du consonantisme vietnamien est que les occlusives en attaque et les occlusives en coda sont réalisées différemment. Si en attaque, elles sont toutes explosées, comme le sont les occlusives françaises ainsi que la majorité des plosives dans les langues du monde (Ladefoged & Maddieson, 1996), en coda, elles ne sont pas relâchées⁴², la phase de relâchement n'étant pas une composante nécessaire de la

⁴² Précisons ici que nous préférons le terme « *non relâché* » de l'API (Alphabet Phonétique International révisé de 2005) au terme « implosif » pour désigner cette caractéristique dans la production des plosives finales vietnamiennes. Dans des langues européennes, le terme « consonne implosive », proposé par Saussure dans son « *Cours de Linguistique générale* » (1916), est réservée à « une consonne qui se trouve après la voyelle ou le noyau syllabique et qui correspond donc à la phase de tension décroissante de la syllabe » (Dubois et al., 2001, p. 241). Dans le mot « rare » [ʁaʁ] du français par exemple, la deuxième consonne est implosive, tandis que la première est explosive. Ce dictionnaire précise également qu'historiquement, les consonnes implosives, d'intensité plus faible que les consonnes explosives, s'affaiblissent et disparaissent plus facilement. Dans l'exemple donné par Saussure [appa], on distingue deux [p] différents car le premier est implosif et correspond à un mouvement de fermeture du conduit vocal, le second est explosif et correspond à un mouvement d'ouverture du conduit vocal. À partir de cette définition, Saussure définit quatre types de combinaisons possibles qui permettent de définir la place de ces éléments dans la syllabe et propose une localisation de la frontière syllabique lors du passage d'une implosive à une explosive. Cao Xuân Hạo (1985) nuance ce propos « *les consonnes finales des langues de ce type ne sont pas proprement « implosives » ... Elles comportent fréquemment (surtout quand il s'agit d'occlusives) une phase de détente essentiellement identique à celle qui caractérise les consonnes initiales, avec un bruit d'explosion nettement audible qui constitue l'indice acoustique le plus important pour l'identification* » (Cao, 1985, p. 78). De plus, par le processus de « resyllabation » (Kasevic, 1977 cité par Cao, 1985, p. 72), les consonnes implosives peuvent devenir explosives en passant de la position de coda à la position d'attaque de la syllabe suivante. En effet, les consonnes implosives qui terminent les mots français *il, aime, une, autre* deviennent explosives en changeant de syllabe dans la phrase [i-le-my-no-tre-lev] (*il aime une autre élève*) (Delattre (1962, p. 598), cité par Cao Xuân Hạo, 2001, p. 25). Il n'en est pas de même dans des langues isolantes syllabiques comme le chinois, le thaï ou le vietnamien. La structure phonologique de la syllabe de ces langues est en effet caractérisée par une faible proportion de l'inventaire phonologique pouvant apparaître en finale. Autre particularité, les consonnes non relâchées sont toujours en position finale de syllabe et appartiennent donc à un autre paradigme que celui des consonnes explosives. Elles remplissent des fonctions différentes dans la syllabe. Ainsi, une consonne non relâchée ne peut jamais devenir consonne explosive en changeant de syllabe donc de morphème.

consonne (Đoàn Thiện Thuật, 1999, p. 229). Elles sont donc produites sans phase de détente, « la phonation cessant pendant l’occlusion » (Nguyễn Thị Bình Minh, 2000, p. 103). Cette particularité dans la production des plosives finales n’est pas phonologique, elle ne change en aucun cas le sens du mot. Nous discutons davantage le cas des plosives non relâchées en vietnamien dans la section 4.1 du chapitre IV.

2. 4. Le système tonal

Comme certaines langues de l’Asie de l’Est et du Sud-Est (le cantonais, le mandarin, le thaï, le birman...), le vietnamien est une langue à tons. Cao Xuân-Hạo, comme André Martinet, définit les tons comme un « ensemble de traits distinctifs caractérisés comme des traits prosodiques, des unités suprasegmentales, par opposition aux traits « inhérents » qui constituent les phonèmes » (Cao Xuân-Hạo, 1985, p. 101). Ce sont des éléments inséparables du noyau syllabique qui se caractérisent essentiellement par la représentation d’une courbe mélodique. Les linguistes considèrent le ton comme une unité discrète au même titre que le phonème, à la seule différence qu’elle n’est pas isolable. Le ton est pertinent en vietnamien et change le sens de la syllabe.

2.4.1. Description des tons du vietnamien (en syllabe isolée)

Le vietnamien possède un système de *tons complexes* définis non seulement par une certaine modulation de la hauteur au cours de la syllabe, mais aussi par des caractéristiques de qualité de voix (mode de vibration des plis vocaux) (Michaud, 2005, 2010 ; Brunelle, 2009).

Les tons du vietnamien sont au nombre de six et portent chacun un nom (cf. table 2.11). Ils sont signalés à l’écrit par des signes diacritiques se plaçant au-dessus ou au-dessous de la voyelle, centre de la syllabe.

Table 2.11. Système des tons du vietnamien (* : B1 et B2 à sonantes finales, D1 et D2 à plosives finales selon Michaud, 2004).

Tons	Appellations		Registres	Signe diacritique	Exemples
	Numérique	Alpha-numérique			
<i>Ngang</i>	1	A1	Haut	(rien)	« ma » [ma ¹] <i>fantôme</i>
<i>Huyền</i>	2	A2	Bas	`	« mà » [ma ²] <i>que</i>
<i>Ngã</i>	3	C2	Haut	~	« mã » [ma ³] <i>cheval</i>
<i>Hỏi</i>	4	C1	Bas	’	« mả » [ma ⁴] <i>tombeau</i>
<i>Sắc</i>	5	B1-D1*	Haut	´	« má » [ma ⁵] <i>joue</i>
<i>Nặng</i>	6	B2-D2*	Bas	.	« mạ » [ma ⁶] <i>semis</i>

Une même syllabe peut être réalisée sous six tons, sauf si elle possède une occlusive sourde (/p/, /t/, /k/) en coda. Dans ce dernier cas, seuls deux tons sont possibles, décrits comme des tons brefs, l'un étant un ton haut et le second un ton bas.

Table 2.12. Compatibilité entre les tons et le type de consonne finale en vietnamien ((-) absent ; (+) présent)

Coda	Ngang (1)	Huyền (2)	Ngã (3)	Hỏi (4)	Sắc (5)	Nặng (6)
Finale sourde	-	-	-	-	+	+
Finale sonore	+	+	+	+	+	+

Le nombre de tons peut cependant varier de six (parler du Nord), à cinq (parler du Sud), voire à quatre pour certains parlers régionaux du Centre (Nguyễn Thị Ngân Hà, 2004). Les recherches instrumentales sur la tonologie du vietnamien du Nord sont assez nombreuses (parmi elles, Gordina, 1957 ; Han, 1966 ; Kim, 1974 ; Brunelle, 2003 ; Phạm, 2003 ; Michaud, 2004 ; Vũ Ngọc Tuấn, d'Alessandre & Michaud, 2005). Parmi les parlers du nord, celui de Hanoi (la capitale) reflète le nombre maximal de tons rencontrés dans un système considéré comme standard. La figure 2.3 montre l'évolution temporelle des tons et leur hauteur relative d'après l'étude d'Andreev et Gordina (1957) (cité par Đoàn Thiện Thuật, 1999, p. 111).

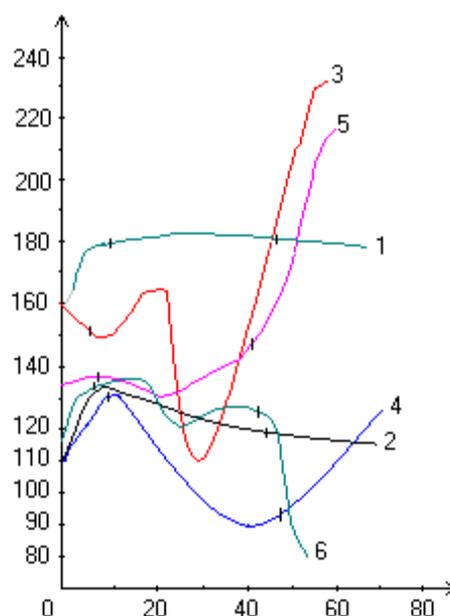


Figure 2.3. Evolution temporelle des tons du vietnamien (d'après Andreev & Gordina, 1957) (en ordonnée la fréquence en hertz (Hz), en abscisse le temps en milliseconde (ms)). La réalisation d'un ton est repérée par le registre tonal dans lequel il évolue et par le mouvement de la courbe de F0. Les mesures des variations de F0 justifient clairement la dénomination des tons : « F0 augmente quand le ton est dit montant, diminue quand le ton est descendant ». (Vũ Ngọc-Tuấn, d'Alessandre & Rosset, 2002).

Les 6 tons vietnamiens se divisent en deux registres. Ceux du registre haut (*ngang* (1), *ngã* (3), *sắc* (5)) évoluent dans la partie supérieure de l'espace tonal du locuteur :

- « *Ngang* » a un contour plat, légèrement descendant en finale. Ce ton est produit en voix modale.
- « *Sắc* » se définit toujours comme un ton montant avec une montée faible au début, parfois même une légère descente, mais une montée très rapide au dernier tiers du parcours.
- « *Ngã* » a aussi un contour montant qui se termine sur une fréquence fondamentale plus élevée qu'en début de la rime (Michaud, 2004), mais une constriction glottale se produit vers le milieu du déroulement. Cette constriction médiane dépend des locuteurs et existe plus fréquemment chez des sujets féminins que masculins (Nguyễn Quốc Cường, 2002).

Les tons du registre bas (*huyền* (2), *hỏi* (4), *nặng* (6)) se trouvent dans la partie inférieure de l'espace tonal du locuteur :

- « *Huyền* » se caractérise par un contour graduellement descendant. Ce ton est produit en voix modale et peut avoir une phonation relâchée ou soufflée (*breathy*) (Nguyễn Văn Lợi & Edmondson, 1998)
- « *Hỏi* » présente une courbe concave qui commence par une descente pendant les deux premiers tiers du parcours puis remonte vers la fin de sa réalisation. Michaud (2004) atteste que ce ton est produit avec un voisement typiquement plus relâché ; il varie (entre locuteurs et parfois chez un même locuteur) entre une réalisation haute-descendante-montante et une réalisation descendante avec relâchement laryngal final (se traduisant par une laryngalisation, ou une forte friction, comme une illustration extrême de la voix soufflée) (Nguyễn Văn Lợi & Edmondson, 1998 ; Vũ Ngọc Tuấn & al., 2005).
- « *Nặng* » a un contour descendant et est en général plus bref que les autres tons en raison d'une forte constriction glottale en finale (Michaud, 2004). Vũ Ngọc Tuấn et al. (2002) remarque que la diminution du quotient ouvert⁴³ de la glotte, qui provoque une diminution brutale de l'énergie du signal EGG (électroglottographie) et du quotient ouvert est un corrélat perceptif de ce ton.

⁴³ Le quotient ouvert de la source glottique est une indication sur la durée d'ouverture de la glotte relative à la durée du cycle glottique. Lié perceptivement à l'axe tendu/relâché de la voix et présente des corrélations avec les mécanismes laryngés, l'intensité et la fréquence fondamentale, ce paramètre permet de déterminer le degré "d'étranglement" des tons, c'est-à-dire la variation de qualité vocale des tons (Vũ Ngọc Tuấn et al., 2002 ; Henrich, 2001).

Une autre représentation schématique des réalisations phonétiques des tons du parler de Hanoi est fournie par Michaud (2004) (cf. figure 2.4).

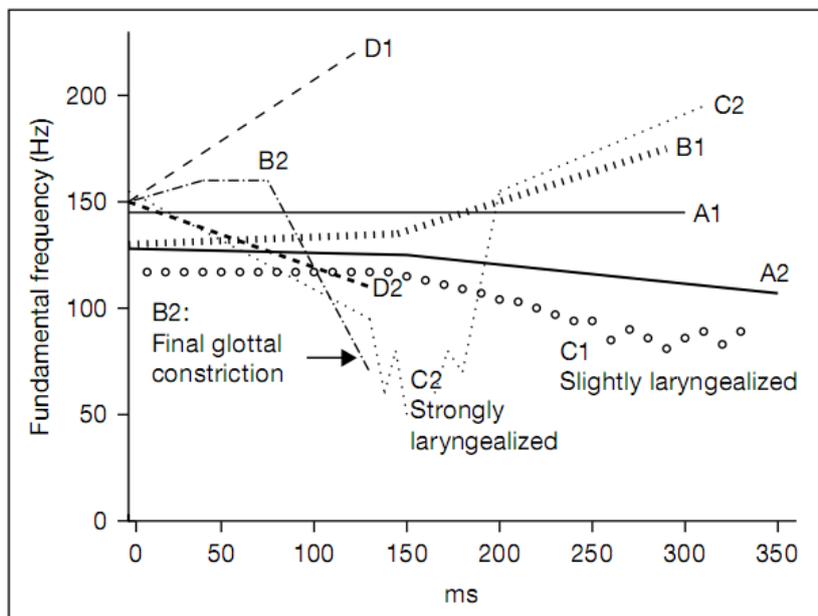


Figure 2.4. Représentation temporelle de l'évolution des tons du vietnamien (parler de Hanoi) (Michaud, 2004).

Michaud (2004) atteste que les tons *ngang* (A1), *huyền* (A2), *sắc* (B1 spécifique des syllabes à finales sonantes) sont produits en voix modale, alors que les tons *nặng* (B2) et *ngã* (C2) sont glottalisés. Un trait distinctif des tons vietnamiens est la présence ou l'absence de glottalisation pendant l'émission. À partir d'études expérimentales, Michaud (2004) est arrivé à la conclusion que les tons D1 et D2 (*sắc* et *nặng*) qui apparaissent sur les syllabes à finales occlusives, ne comportent pas de glottalisation, contrairement à ce qui avait été avancé par d'autres chercheurs (Đoàn Thiện Thuật, 1999). Michaud (2005, p. 226) atteste qu'une glottalisation finale occasionnelle ou fréquente accompagne ces consonnes occlusives finales non relâchées dans certaines langues tonales de l'Asie du Sud-Est et de l'Est (dans le dialecte chinois du Fujian, en chinois cantonais, en thaï et en vietnamien). Il propose également une distinction entre le ton D2 (ton 6 frappant les syllabes avec finale de type occlusive) et le ton B2 (ton 6 à finales sonantes), en dépit du fait que l'un et l'autre ne peuvent apparaître sur les mêmes syllabes. Il démontre ainsi que la réalisation des consonnes finales est liée aux tons (Michaud, 2005, p. 227).

Brunelle, Nguyễn D.D., Nguyễn K.H. (2010) ont récemment réalisé une étude laryngographique et laryngoscopique de la production des tons du parler du Nord. Ils montrent que le contour de F0 des tons du Nord est non seulement attribuable aux changements de la longueur et de la tension des plis vocaux (partiellement aux changements de la taille du larynx), mais que les baisses (*drops*) de F0 sont également en grande partie provoquées par les configurations glottales responsables des qualités de voix contrastives

liées à certains tons. Les auteurs constatent également que les contrastes de qualité de voix sont la plupart du temps dus à la constriction glottale, mais que de temps en temps ils impliquent l'incursion additionnelle des faux plis vocaux et aussi de la constriction épiglottale. Comme l'illustre la figure 2.5, le ton B2 généralement reconnu comme le ton le plus glottalisé peut être accompagné par l'incursion des faux plis vocaux.

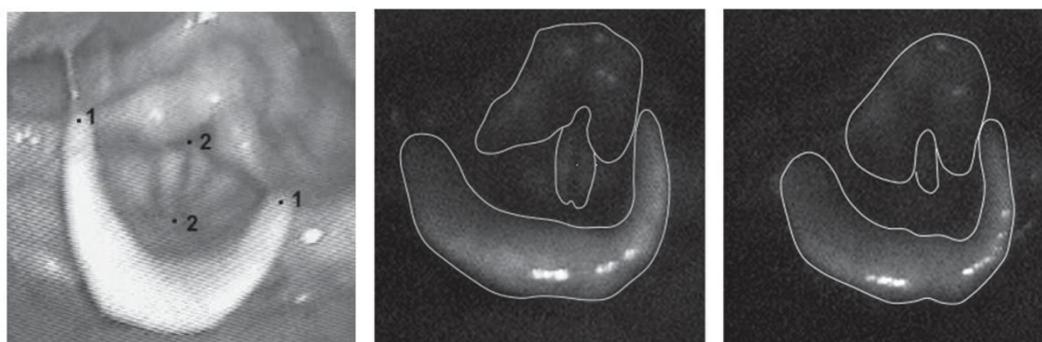


Figure 2.5. La première image à gauche montre les mesures prises sur l'image (1 = largeur de l'épiglotte, 2 = longueur des plis vocaux). Les bandes ventriculaires (faux plis vocaux) se situent des deux côtés des plis vocaux. Les deuxième et troisième images montrent l'adduction de faux plis vocaux pendant la réalisation de la syllabe /ti/ sous le ton B2 (première et dernière image du ton). Le fait des plis vocaux plus raccourcis pourrait être une stratégie pour réduire F0. La taille plus faible de l'épiglotte est une conséquence d'un abaissement léger du larynx qui accompagne la glottalisation (Brunelle et al., 2010).

Concernant le parler du Sud, Brunelle (2009) remarque qu'il y a deux différences importantes dans les systèmes tonaux entre ce parler et celui du Nord : le parler du Nord possède 6 tons tandis que celui du Sud en a seulement 5 (*hỏi* (C1) et *ngã* (C2) sont fusionnés) ; le contour mélodique et la qualité de voix sont combinés dans les tons du Nord alors que le système du Sud se base exclusivement sur la hauteur (cf. figure 2.6).

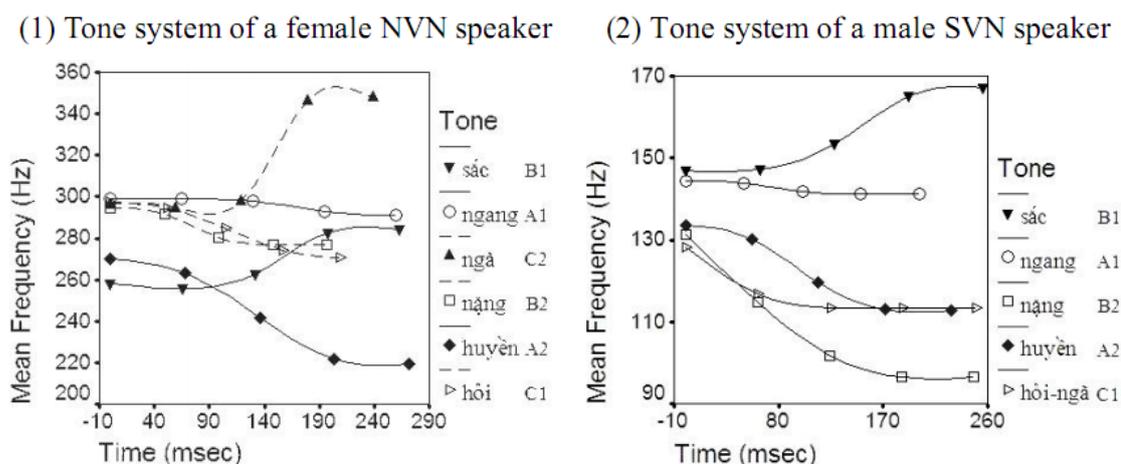


Figure 2.6. Représentation temporelle de l'évolution des tons du dialecte du Nord (à gauche) et du Sud (à droite) (Brunelle, 2009).

Les parlers du Centre très divers, passant d'un système penta-tonal à un système tétra-tonal selon les parlers, rendent la description des tons plus complexe. D'après Seitz (1986) (cité par Nguyễn Thị Ngân Hà, 2001), les tons du parler de Huế (l'ancienne capitale impériale) se répartissent en 3 registres haut, moyen et bas. « *Ngang* » et « *sắc* » évoluent dans le registre haut : le premier a un contour plat, le second un contour légèrement montant. Dans le registre moyen, le ton « *huyền* » dessine une pente descendante ou presque égale et le ton fusionné « *hỏi-ngã* » est considéré comme un ton montant et laryngalisé. Le ton « *nặng* » se trouve dans le registre bas, sa courbe mélodique est très proche de celle de « *huyền* », voire même confondue avec cette dernière.

Quant au contour mélodique, la tradition prosodique vietnamienne classe les six tons en deux types : les tons « plats » (*bằng*) et les tons « non plats » (*trắc*). Les tons « plats » (*ngang* et *huyền*) ont un contour mélodique plutôt simple, à l'inverse des tons dits « non plats » (*ngã*, *hỏi*, *sắc* et *nặng*).

2.4.2. Variation des tons dans la chaîne parlée

Du point de vue phonologique, les tons se réalisent pleinement dans les oppositions d'unités lexicales isolées. Par contre, ils ne se manifestent pas de la même manière dans une chaîne continue de plusieurs syllabes où ils subissent des modifications dues à la présence ou l'absence de l'accent, aux tons environnants, au débit, etc. Il existe, en vietnamien, le phénomène appelé « co-articulation » ou « assimilation tonale » où un ton se trouve modifié par les tons qui l'entourent ; ce qui le différencie du phénomène du sandhi tonal (modification tonale du troisième ton) du mandarin ou phénomène de neutralisation tonale (existence des tons neutres) du thaï (Nguyễn Thị Thanh Hoa, 2004, p. 20).

Les remarques de Gordina et Bystrov (1984) rapportés par Đỗ Thế Dũng (1986) confirment, dans les grandes lignes, l'existence des phénomènes d'assimilation tonale qui se manifestent quand deux syllabes adjacentes ne sont pas séparées par une pause importante, c'est-à-dire quand elles se trouvent au sein d'un même syntagme. Cette assimilation tonale entraîne des modifications de la valeur moyenne de F0 et de son contour mélodique (Nguyễn Thị Thanh Hoa, 2004).

Dans une étude sur les variations des tons (parler du nord) dans les mots et les énoncés dissyllabiques, Han & Kim (1974) (cité par Nguyễn Thị Thanh Hoa, 2004) ont montré que les modifications de la hauteur globale, aussi bien que du contour mélodique, sont souvent considérables. Ces modifications dépendraient de l'environnement tonal immédiat et de la position de la syllabe dans le mot. Le phénomène d'assimilation tonale dans deux combinaisons de tons dont l'une avec deux tons hauts montants *sắc* - *sắc* et l'autre avec deux tons bas descendants *huyền* - *huyền* se manifeste comme suit : Le ton « *sắc* » (ton 5 ou B1-D1) de la seconde syllabe voit sa hauteur globale se trouver à un niveau beaucoup plus élevé

sous l'influence de la hauteur élevée du ton « *sắc* » que porte la syllabe précédente. Tandis que le deuxième ton « *huyền* » (ton 2 ou A2) se trouve à un niveau plus bas sous l'influence de la hauteur peu élevée du premier ton « *huyền* ».

Brunelle (2003) confirme les résultats de Han & Kim (1974) selon lesquels les tons du parler du Nord varient considérablement selon leur environnement tonal immédiat et que la coarticulation tonale progressive est plus forte que la coarticulation anticipée. Brunelle (2009) compare ce phénomène dans les parlers du Nord et du Sud et atteste que la coarticulation tonale progressive est toujours beaucoup plus forte que régressive dans les deux dialectes. Il signale aussi que le dialecte du Nord montre une coarticulation plus forte que celui du Sud, surtout la coarticulation progressive.

2.4.3. Ton et intonation

Les tons d'une langue tonale ne doivent pas être confondus avec l'intonation prosodique. « Le ton est réservé aux variations de hauteur à l'intérieur d'une syllabe tandis que l'intonation indique les variations de hauteur qui affectent un groupe de mots ou une phrase ». « Les tons sont tout à fait autonomes par rapport à la hauteur, au rythme, à l'accent » (Lê Thị Xuyên, 1989).

La langue vietnamienne n'utilise pas les montées et descentes de l'intonation aussi catégoriquement que le français pour distinguer les phrases. Cette langue à 6 tons lexicaux possède tout un système de marqueurs morpho-syntaxiques pour exprimer émotions, attitudes et modalités. Ainsi, on pourrait ajouter à la fin de l'énoncé « Cô ấy xinh thật » (*Elle est vraiment jolie*) un marqueur morpho-syntaxique pour la transmettre en une phrase interrogative comme « Cô ấy xinh thật không ? » (*Est-elle vraiment jolie ?*). L'emploi de marqueurs morpho-syntaxiques est une nécessité quasi-absolue en vietnamien pour exprimer une interrogation, une injonction et même une attitude ou émotion particulière (Mai Ngọc Chừ et al., 1997). L'ordre syntaxique des constituants dans les différents types de phrases est toujours le même, mais on peut différencier facilement la modalité grâce à la présence de ces marqueurs morpho-syntaxiques placé en fin de phrase. Ainsi, l'énoncé « *các anh đi* » suivante (*littéralement* : morphème pluriel – frère (vous) – partir) (Nguyễn Thị Thanh Hoa, 2004, p. 29):

<i>Các anh đi</i> + (rien)	→ <i>Các anh đi.</i>	= Vous partez.
<i>Các anh đi</i> + morphème interrogatif neutre	→ <i>Các anh đi không ?</i>	= Partez-vous ?
<i>Các anh đi</i> + pronom interrogatif	→ <i>Các anh đi đâu ?</i>	= Où partez-vous ?
<i>Các anh đi</i> + morphème interro-expressif	→ <i>Các anh đi à ?</i>	= Vous partez donc ?
<i>Các anh đi</i> + morphème impératif neutre	→ <i>Các anh đi đi !</i>	= Partez !

Les résultats de quelques travaux expérimentaux qui ont étudié la prosodie du vietnamien (Lê Thị Xuyên, 1989 ; Đỗ Thế Dũng, Trần Thiên Hương & Boulakia, 1998 ; Nguyễn Thị

Thanh Hoa, 2004 ; Vũ Minh Quang, 2007), surtout ceux concernant l'évolution des contours intonatifs, ne vont pas dans le même sens. Cependant, presque tous les auteurs partagent la même idée que la distinction des types de phrases se manifeste plutôt par le registre de la phrase (c'est-à-dire la hauteur globale) que par le contour mélodique. Đỗ Thế Dũng et al. (1998) montrent que la déclinaison de la courbe de F0 caractérise en général les phrases déclaratives, alors que les phrases interrogatives et impératives possèdent un contour montant, un niveau de hauteur global plus élevé et un débit plus rapide. Ils attestent aussi que l'évolution de la courbe d'intonation est la même dans la phrase impérative marquée par la particule « *đi* » que dans la phrase interrogative avec particule « *không* ».

Selon Nguyễn Thị Thanh Hoa (2004), chaque type de phrase possède ses caractéristiques prosodiques particulières. Le contour général de la F0 n'est pas considéré comme un facteur essentiel en vietnamien à cause de la présence des tons, le registre de F0, la durée et l'intensité, constituant les paramètres discriminants des types de phrases. Nguyễn Thị Thanh Hoa (2004) précise que les assertives sont prononcées avec un registre bas alors que les questions et les injonctives le sont avec un registre haut. Elle remarque, contrairement à Đỗ Thế Dũng (1998), qu'une pente descendante du contour intonatif de la phrase ne correspond pas toujours à une phrase déclarative. Une différence d'intensité entre énoncés interrogatifs et énoncés des autres modalités est aussi attestée.

Vũ Minh Quang (2007), dans une étude sur la différence prosodique (F0, intensité, durée) entre des paires de phrases déclaratives et interrogatives de type de réponse « oui/non » (dans le but d'une segmentation automatique) trouve un taux global de bonne identification des phrases non-questions et questions (environ 70 %) qui suggère que les paramètres prosodiques de la phrase vietnamienne transportent des informations extralinguistiques qui peuvent permettre à l'auditeur de discriminer le type de phrase. Il relève que les différences essentielles de l'intonation se situent à la fin des énoncés : le contour de la dernière syllabe ou de la deuxième moitié de celle-ci semble être croissant pour les phrases interrogatives, alors que la pente reste descendante pour presque tous les tons, sauf les deux tons faisant partie du registre haut *ngã* (C2) et *sắc* (B1). L'intensité des phrases question est plus forte que celle de phrases non-question dans 87 % des cas. Contrairement à Lê Thị Xuyên (1989) et Nguyễn Thị Thanh Hoa (2004) qui ont suggéré un débit plus rapide pour les phrases interrogatives, l'effet de la durée n'est pas significatif entre les phrases non-question et question pour les locuteurs de son corpus.

Les études sur les expressions émotionnelles dans la parole de Lê Thị Xuyên (1989) montrent que la phrase neutre est caractérisée par un registre moyen et un débit moyen. La colère se traduit par un registre plus élevé, un débit plus rapide et une intensité plus forte alors que la tristesse se manifeste par les caractéristiques inverses (registre plus bas, débit plus lent, intensité moins forte).

2.4.4. Accentuation

En vietnamien, le rôle de l'accent est resté assez obscur en raison de l'absence de beaucoup de recherche particulière sur ce sujet. De plus, les résultats de quelques études relevées sur l'accent du vietnamien (Hoàng Tuệ & Hoàng Minh, 1975 ; Đỗ Thế Dũng, 1986 ; Nguyễn Thị Thanh Hoa, 2004 ; John Ingram, 2006 ; Nguyễn Thị Anh Thư, 2007) sont parfois contradictoires.

Selon Nguyễn Thị Thanh Hoa (2004), l'accent en vietnamien a une fonction démarcative car la syllabe accentuée se trouve *toujours en position finale d'un composé polysyllabique, d'un syntagme ou d'un groupe prosodique*. Sa valeur distinctive peut être observée dans des cas d'homonymie entre un mot composé et un groupe de mots ou encore entre deux séquences qui possèdent les mêmes constituants et dont les structures syntaxiques et prosodiques sont différentes, comme l'illustrent les exemples suivants de paires minimales d'accentuation (Nguyễn Thị Thanh Hoa, 2004, p. 25) :

thấy anh em tôi vui (littéralement : voir – grand frère – frère – 1^{er} personne - content)

→ (a) *thấy anh em tôi 'vui* = de voir mes frères contents

→ (b) *thấy anh 'em / tôi 'vui* = de vous voir (amis), j'en suis content

→ (c) *thấy 'anh / em tôi 'vui* = de te voir (ami), mon frère en est content

Dans (a), l'accent tombe normalement sur « *vui* » qui a valeur d'attribut du complément d'objet et qui occupe la dernière position dans le syntagme verbal ayant pour verbe « *thấy* » (voir). Dans (b), l'accent est sur « *em* » qui fait partie du composé « *anh em* » (qui au sens littéral veut dire « frères ») et « *tôi* » se trouve renvoyé au syntagme suivant où il devient pronom ayant pour fonction sujet de l'adjectif « *vui* ». Enfin dans (c), l'accent est porté sur « *anh* », dernière syllabe du premier syntagme, lequel signifie littéralement « grand frère », alors que « *em* » à son tour rejoint « *tôi* » dans le syntagme suivant pour former le groupe nominal « *em tôi* » qui veut dire « mon frère ».

Hoàng Tuệ et Hoàng Minh (1975) ont signalé l'existence de deux types d'accent, accent de fin de syntagme et accent emphatique. Pour ces auteurs, les particules modales (bien que ces dernières soient placées en fin de phrase) ne sont pas accentuées.

Une étude expérimentale réalisée par Đỗ Thế Dũng (1986) a mis en évidence la présence d'un accent sur la dernière syllabe de mots composés. L'accent de mot aussi bien que l'accent de syntagme se traduit souvent par un accroissement de la durée de syllabe et une élévation du niveau d'intensité et du contour mélodique tonal. Cet auteur affirme l'existence d'une réalisation totale et non altérée du ton en syllabe accentuée et l'a considérée comme une des conséquences naturelles de l'allongement de la syllabe, alors que Hoàng Tuệ et Hoàng Minh (1975) l'ont étudiée comme une des « marques positives » de l'accent. En revanche, tous ces auteurs ont signalé une plus courte durée dans les syllabes inaccentuées,

avec par conséquent un espace de déploiement plus réduit et une instabilité qui entraîne des modifications tonales (cité par Nguyễn Thị Thanh Hoa, 2004, p. 27).

Ingram et Nguyễn Thị Anh Thu (2006), quant à eux, attestent que l'accent lexical existe en vietnamien seulement comme une *tendance phonétique*, mais pas comme un contraste phonologique actif. Ils ont étudié les corrélats phonétiques de l'accent lexical dans les mots composés dissyllabiques (ex. « **hoa hồng** thì đẹp » *la rose est belle*) et les constructions syntagmatiques correspondantes (ex. « **hoa hồng** thì đẹp » *la fleur qui est rose est belle*).

Les auteurs ont trouvé qu'en réalisant une pause entre les syllabes S1 et S2 des syntagmes, la majorité des paires (composé *vs.* syntagme) a été distinguée, mais seulement dans les formes pour lesquelles il a été demandé aux sujets de produire les paires de manière à ce qu'un auditeur parvienne à différencier un composé de sa correspondance syntagmatique. Au contraire, lorsque les sujets ont décrit une image en utilisant une phrase porteuse dans laquelle figuraient les composés ou les syntagmes, tous les indicateurs prosodiques (durée, contour de F0, intensité, pentes F1-F2) permettant de distinguer un composé dissyllabique d'un syntagme n'ont pas été observés. Ces résultats ont été renforcés par l'étude perceptive de Nguyễn Thị Anh Thu (2007) qui montre que les auditeurs ont compté seulement sur la pause entre les deux composants des syntagmes nominaux comme indice pour distinguer entre mots composés et groupes nominaux. Aucune évidence acoustique et perceptive significative n'a été détectée pour prouver l'existence d'un accent faisant contraste entre un composé et un syntagme. Ainsi, les auteurs concluent qu'un tel mécanisme phonologique n'existe pas dans cette langue, ce qui n'est pas le cas en anglais ou en italien. Les Vietnamiens peuvent distinguer entre un composé et un syntagme sur la base de la signification en contexte ou en employant une pause dans la parole.

Devant de telles contradictions entre résultats d'études expérimentales, la décision de ne pas tenir compte de la possibilité d'une unité accentuelle en vietnamien a été retenue pour la suite de l'étude (chapitre IV et V).

2.5. Comparaisons avec le système phonémique du français

Rappelons que l'objectif de cette présente étude se situe dans le cadre de l'apprentissage de la prononciation des sons en langue seconde. La comparaison entre les phonèmes et les structures sonores de deux langues, l'une source, l'autre cible, fondée sur l'analyse contrastive inspire souvent de nombreux manuels d'enseignement de prononciation (ex. la collection de la méthode Assimil dont « Le vietnamien sans peine » pour l'apprentissage du vietnamien langue étrangère de Đỗ Thế Dũng & Lê Thanh Thủy, 1994). À partir de la description des unités distinctives en vietnamien présentée dans les sections précédentes de

ce chapitre, un certain nombre de différences et de similarités entre le système phonémique du vietnamien et du français ont pu être relevées.

2.5.1. Structures syllabiques

Les deux langues présentent des différences dans les structures syllabiques possibles comme l'illustrent les figures 2.7 et 2.8.

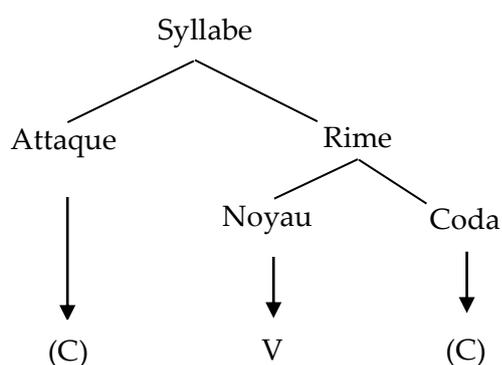


Figure 2.7. Représentation arborescente de l'organisation interne de la syllabe en français (Rousset, 2004) (C = segment(s) consonantique(s) ; V = voyelle), entre parenthèses les constituants facultatifs.

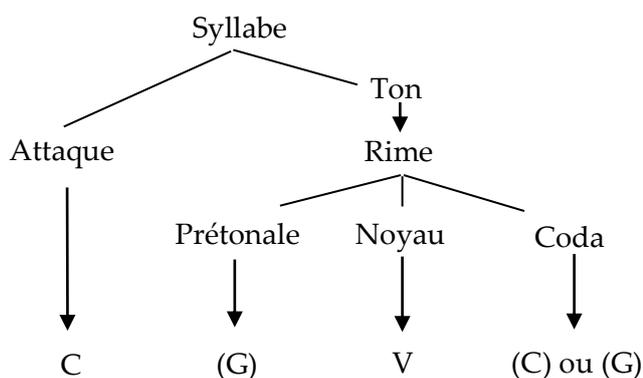


Figure 2.8. Représentation arborescente de l'organisation interne de la syllabe en vietnamien (selon Doãn Thiên Thuật, 1999 et Trần Đỗ Đạt & al., 2005) (C = consonne ; G = glide ; V = voyelle ou diphtongue), entre parenthèses les constituants facultatifs.

En vietnamien, toute syllabe est porteuse de ton, la syllabe fermée par les plosives sourdes /p/, /t/, /k/ porte toujours soit le ton montant 5 (ton D1) (ex. « bác » [bak⁵] oncle), soit le ton descendant 6 (ton D2) (ex. « bạc » [bak⁶] argent).

En français, les frontières syllabiques peuvent être déplacées dans le flux de parole, en raison de phénomènes d'élosion, de liaison ou d'enchaînement. Le phénomène de resyllabation n'est jamais rencontré en vietnamien. Pour la suite de notre étude, rappelons que l'attaque et la coda de la syllabe française peuvent être complexes, c'est-à-dire être occupées par des groupes de consonnes (clusters). En vietnamien, les séquences consonantiques ne se rencontrent qu'en frontière syllabique de mots et appartiennent de fait à deux syllabes différentes.

Dans les deux langues, le noyau de la syllabe est toujours occupé par un segment vocalique : voyelle en français ; voyelle ou diphtongue en vietnamien.

2.5.2. Systèmes consonantiques

Le français compte 21 phonèmes consonantiques /p t k b d g m n ɲ ŋ f v s z ʃ ʒ l ʀ j w ɥ/ présents dans toutes les positions de la syllabe, sauf /w/ et /ɥ/ qui n'apparaissent pas en coda.

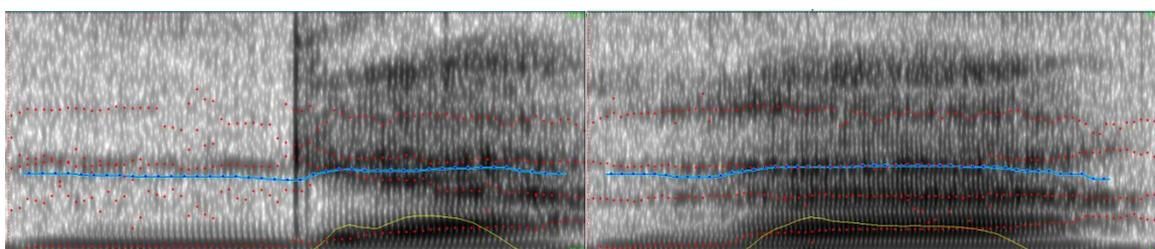
Alors qu'en vietnamien, une restriction du système consonantique entre la position d'attaque et de coda est attestée : 22 consonnes en attaque (/b t t^h d ʈ c k m n ŋ ɲ f v s z ʂ ʎ χ h l ʔ/) contre 8 en coda (/p t k m n ŋ j w/).

Il faut remarquer qu'en vietnamien, l'occlusive bilabiale sourde /p/ n'est présente qu'en finale de syllabe. Ce phonème /p/ du français est habituellement réalisé /b/ par les Vietnamiens : la consonne /b/ occupe la place initiale du mot dans l'écriture ainsi que dans la prononciation (Đoàn Thiện Thuật, 1999, p. 157). De fait, les mots empruntés du français comme « pile », « pâte » sont prononcés [bin] « bin », [bate] « ba tê » en vietnamien.

On peut relever certaines consonnes existant dans un système mais pas l'autre et vice-versa. Le vietnamien ne connaît pas la vibrante /r/ dont l'allophone [ʀ] est très répandu en français. [r], vibrante apicale, est présente dans certaines variétés du français comme en français d'Afrique (Boutin, Lyche & Prignitz, 2007). Il existe une variante vibrante dentale [r] du phonème /z/ présente dans quelques dialectes régionaux du Vietnam mais sa fréquence d'usage est faible. Elle apparaît récemment dans certains mots empruntés (ex. « Paris » est prononcé [pari]).

À l'inverse, /χ/ n'existe pas en français. C'est une fricative sourde produite par l'air qui s'échappe du passage formé par le voile du palais et la partie postérieure du dos de la langue qui se relève. La pointe de la langue est abaissée. En vietnamien, on peut opposer la fricative /χ/ à l'occlusive /k/ dans des mots de type : « kéo » [kew] *ciseaux* et « khéo » [χew] *habile*. /χ/ connaît une variante libre [x] dans le parler du Nord.

/ɣ/ n'existe pas en français non plus. Sur le plan articulatoire, un seul trait distinctif différencie [g] français de [ɣ] vietnamien : le trait « occlusif/ fricatif ». La figure 2.9 illustre ce fait : la barre d'explosion est présente dans le premier spectrogramme pour le cas de [g] occlusif français mais elle n'existe pas dans le second pour /ɣ/ consonne fricative en vietnamien, dans la réalisation de laquelle le trait de friction est visible.



[ga] français

[ɣa¹] vietnamien

Figure 2.9. Sonagrammes des syllabes [ga] en français et [ɣa¹] en vietnamien.

L'opposition entre /g/ et /ɣ/ n'a pas de valeur phonologique ni en français, ni en vietnamien. Les apprenants vietnamiens assimilent souvent /ɣ/ vietnamien à /g/ français lors de l'apprentissage du FLE (Nguyễn Thị Bình Minh, 2000).

Le vietnamien connaît la consonne fricative glottale et sourde [h]. Le français actuel ne connaît pas cette consonne. Dans l'apprentissage de l'oral à partir de l'écrit, l'apprenant vietnamien prononce cette consonne dans les mots français pourvus de « h muet ou aspiré » (Nguyễn Thị Bình Minh, 2000) (ex. « héros » [heʁo], « hôtel » [hotel]).

Les plosives et nasales finales sont articulées de deux façons différentes entre les deux langues : dans l'une, elles ne sont pas relâchées « à la vietnamienne » et dans l'autre, elles sont explosives « à la française ». La figure 2.10 illustre cette différence. Le mot « loupe » français est emprunté en vietnamien « *lúp* » [lup⁵]. La prononciation française montre clairement un brusque relâchement de l'articulation bilabiale de la plosive finale générant un burst (partie gauche) alors que dans la prononciation vietnamienne, la détente brusque est absente (partie droite).

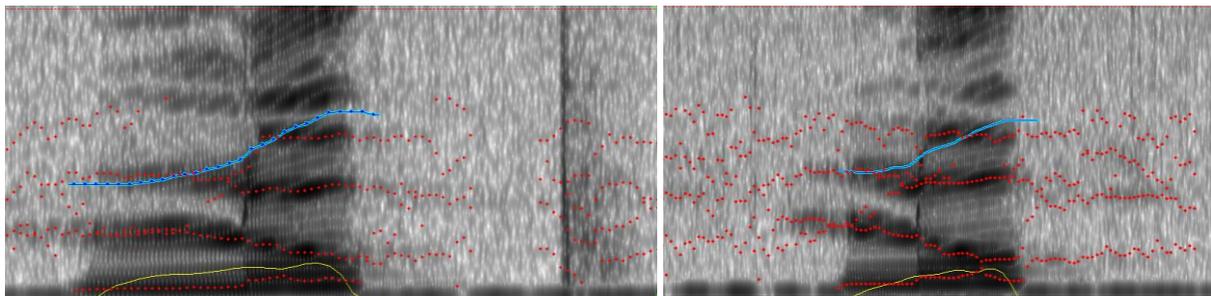


Figure 2.10. Spectrogrammes de « loupe » [lup] français (à gauche) et [lupʔ] vietnamien (à droite).

En français, le coup de glotte fonctionne comme « signe d'expressivité » lorsqu'il s'agit de mettre en valeur un mot commençant par une voyelle comme l'initiale (Ex. « *Encore !* »). La voyelle est précédée notamment une attaque brusque alors que dans la phonation normale française, « les voyelles commencent par une attaque douce » (Léon, 1992). Selon Lê Thị Xuyến (1989), l'existence de la consonne glottale devant les voyelles dans le système des initiales du vietnamien empêche beaucoup d'apprenants de faire l'enchaînement vocalique en français. Elle cite le cas d'un apprenant français qui perçoit distinctivement 3 syllabes dans l'énoncé « *Ái yêu ông* » [ʔaj⁵ ʔiew¹ ʔoŋ¹] (*Ái aime son grand-père*) mais qui prononce au début de son apprentissage du vietnamien l'énoncé avec l'enchaînement vocalique : [ajjewoŋ]. Inversement, l'apprenant vietnamien distinguera nettement [ty-a-y-œ-ʔit], surtout quand on lui montre d'abord la phrase écrite. En raison de la consonne, il réalisera [ty-ʔa-ʔy-ʔœ-ʔit] et ne sera habitué à l'enchaînement vocalique qu'avec l'entraînement oral qui doit précéder la lecture.

2.5.3. Systèmes vocaliques

La figure 2.11 présente un classement des phonèmes vocaliques du français et du vietnamien. Les voyelles en caractères gras sont celles qui n’existent que dans l’une des deux langues.

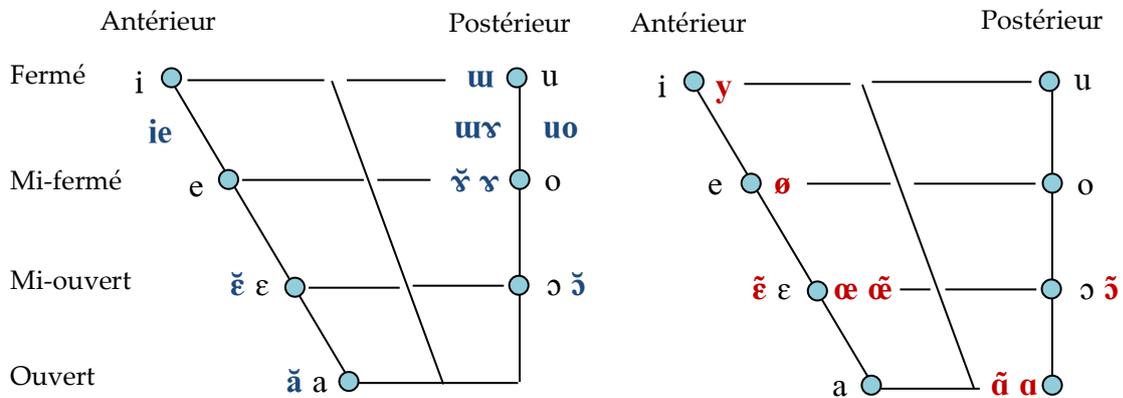


Figure 2.11. Description en traits d'aperture et de lieu des systèmes vocaliques du vietnamien (à gauche) et du français (à droite).

Les deux langues possèdent quatre degrés d'aperture identiques. Le système vocalique du français est caractérisé par une tendance à la réalisation antérieure et à la labialisation, à l'inverse les voyelles vietnamiennes ont une tendance à la réalisation postérieure et à la non-labialisation.

En effet, deux tiers des voyelles du français sont des antérieures et deux tiers des voyelles du vietnamien sont des postérieures. Deux tiers des voyelles françaises sont arrondies et deux tiers des voyelles vietnamiennes sont non arrondies. Le français dispose de quatre voyelles nasales alors que la nasalité vocalique est absente en vietnamien. Inversement, le vietnamien possède trois diphtongues et quatre voyelles brèves absentes dans la langue française.

La majorité des difficultés de prononciation des sons vocaliques du français pour des apprenants vietnamiens reposent sur les voyelles nasales dont la langue ne connaît aucun son vocalique nasal mais elle est dotée de la moitié des consonnes finales en sons nasals (cf. section 2.3.4). Par conséquent, des apprenants vietnamiens du français ont tendance à rajouter une consonne nasale à la voyelle pour y affecter le trait de nasalité (Nguyễn Thị Bình Minh, 2000). Le mot « bon » [bɔ̃] devient ainsi [boŋ¹], « banc » [bã] devient [bãŋ¹] ...

Une description précise et détaillée de la phonologie du vietnamien s'imposait comme préalable indispensable aux études acoustico-perceptives ainsi qu'aux expériences de perception et de production présentées dans les chapitres IV et V suivants.

En effet, il était nécessaire de relever, signaler et comprendre les discordances entre spécialistes dans la description de la phonologie du vietnamien, notamment comme on l'a vu, à propos de la structure syllabique et ses constituants, de la réalisation du ton et de l'accent, du statut de l'occlusive glottale. La comparaison avec la phonologie du français constitue également une étape préliminaire à la constitution des corpora et analyse des résultats de l'ensemble de la partie expérimentale de cette thèse.

L'objet d'étude portant sur l'acquisition des clusters du français par des apprenants vietnamiens, le chapitre suivant présente une analyse descriptive et quantitative des séquences consonantiques relevées en vietnamien, puis comparées à celles du français.

Chapitre 3

**De la constitution d'un lexique du vietnamien
à l'étude comparative des séquences de
consonnes et de leur distribution en
vietnamien - comparaison avec le français**

3.1. Construction d'un lexique du vietnamien

L'objet central de cette étude portant sur les séquences de consonnes dans le cadre de l'apprentissage des sons d'une langue seconde, il est nécessaire d'examiner la distribution des consonnes et des séquences consonantiques présentes dans la langue des apprenants mais également dans la langue cible.

Pour ce faire, il a été décidé de collecter un lexique exploitable dans le cadre d'analyses typologiques sur les données lexicales et syllabiques des langues du monde (Rousset, 2004). L'objectif était d'intégrer le lexique constitué pour le vietnamien à la base G-ULSID (*Grenoble-UCLA Lexical and Syllabic Inventory Database*) développée par Maddieson et Precoda (1992) et depuis une dizaine d'années au sein du Département Parole et Cognition du GIPSA-lab, ancien Institut de la Communication Parlée de Grenoble (ICP). Il s'agit d'une banque de données constituées dans le but de nourrir la recherche d'universaux sur l'organisation des structures syllabiques et lexicales des langues du monde (Vallée, Rossato, Rousset, 2009). Puisque le nombre d'unités lexicales (parmi les lexiques des 16 langues déjà présentes dans G-ULSID (à l'époque de cette étude sur le vietnamien) variait de 1 989 pour le ngizim à 12 181 pour le français et la moyenne se situant à 5 908 items lexicaux (Rousset, 2004, p. 75), il a été décidé de choisir une taille d'environ 5 000 entrées pour le lexique du vietnamien. Le choix des entrées lexicales a été fait sur les mêmes critères que pour les 16 autres langues de G-ULSID : 1) « *Recent loan words, especially those of wide international currency relating to modern technological, political or cultural concepts* (« *telephone* », « *democracy* », « *football* »), *have been excluded wherever recognizable* » (Maddieson, 1993, p. 1 cité par Rousset, 2004, p. 53) ; 2) seuls les lemmes figurent comme entrées ; 3) le lexique est constitué des lemmes les plus utilisés par les locuteurs.

Dans les sections qui suivent figurent le détail de la constitution du lexique pour le vietnamien puis les analyses menées sur la distribution des segments et les résultats quantitatifs obtenus.

3.1.1. Recueil des données lexicales

Dans le cadre du traitement automatique de la parole, Lê Việt Bắc (2006) a proposé deux méthodes de recueil d'un vocabulaire : récupération à partir de ressources lexicales existantes (1), génération automatique à partir de données textuelles (2).

Pour les applications du traitement de la parole, la première méthode de récupération peut être utilisée pour une langue, ou tâche spécifique, qui possède déjà un lexique constitué de bonne qualité, c'est-à-dire avec une bonne couverture lexicale. La couverture lexicale indique le taux des mots du vocabulaire d'une langue présents dans un corpus de textes de

cette même langue. L'avantage de cette méthode réside dans l'assurance de la qualité du vocabulaire car construit par des linguistes. Par ailleurs, pour des langues ou des tâches spécifiques qui ne possèdent pas de vocabulaire ayant une bonne couverture lexicale, il peut être appliqué la seconde méthode pour générer automatiquement un vocabulaire. Cependant la qualité du vocabulaire dépend de la qualité de la source de données textuelles utilisée (Lê Việt Bắc, 2006, p. 34). Étant à la recherche d'un échantillon de 5 000 entrées lexicales vietnamiennes, nous avons été contraints de mener une réflexion sur une bonne méthode de récupération du lexique.

Les besoins s'orientant toujours vers de grandes quantités d'informations lexicales, il n'a pas été possible de sélectionner un dictionnaire de moins de 20 000 entrées, même en format de « poche » (ex. « Từ điển Việt Pháp » *Dictionnaire Vietnamien Français* de Lê Phương Thanh (1996) est de 55 000 entrées).

Avec le développement des technologies de traitement des langues naturelles, de plus en plus de ressources de dictionnaires sont maintenant disponibles sur support informatique. Elles peuvent être de différentes formes telles que fichiers de données dans un logiciel de dictionnaires (ex : *Lac Viet – MTD - Từ điển Pháp Việt (Dictionnaire français - vietnamien)*) ou dictionnaires gratuits publiés sur Internet (ex : *Free Vietnamese Dictionary*⁴⁴, *Wikipédia*⁴⁵...).

Grâce à des communautés en ligne « ouvertes », nous avons pu obtenir des ressources dictionnaires très variées en format, en quantité, en qualité, en information linguistique, etc. Cependant, pour construire une liste de 5 000 unités lexicales, ces dictionnaires ne répondent pas tels quels à notre besoin car trop volumineux. Nous avons donc eu recours à la deuxième méthode « génération automatique d'un vocabulaire à partir de données textuelles » proposé par Lê Việt Bắc (2006) dans le cadre d'un projet au Laboratoire d'Informatique de Grenoble (LIG) sur la reconnaissance des langues peu dotées dont le vietnamien. Le problème de cette méthode résidant dans le fait que la qualité de la source de données textuelles décide la qualité du vocabulaire généré, nous donnons quelques éléments permettant d'apprécier la qualité du lexique source duquel nous avons tiré, pour notre recherche, le lexique du vietnamien d'environ 5 000 items.

Un lexique a été généré automatiquement à partir de la construction d'un corpus de textes récupérés à partir des documents *html* du Web par Lê Việt Bắc. Le point de départ de la méthode consiste à déterminer manuellement des sites Internet riches en ressources et présentant un haut débit de transmission. Par exemple des sites de nouvelles, au fort contenu rédactionnel tel que *VnExpress*⁴⁶ pour le vietnamien ont été sélectionnés. La quantité de pages

⁴⁴ <http://www.informatik.uni-leipzig.de/~duc/Dict/index.html>

⁴⁵ <http://fr.wikipedia.org> ou <http://vi.wikimedia.org>

⁴⁶ <http://www.vnexpress.net>

web collectée par Lê (2006) à partir de ce site de courrier vietnamien est de 2.5 Go. Après avoir traité le corpus de texte, la quantité de données textuelles pouvant servir à l'apprentissage d'un modèle de langage statistique était d'environ 400 Mo (5 millions d'énoncés). La dernière étape est celle du filtrage, en fonction du vocabulaire, des phrases et aussi des blocs de mots. Les modèles de langage sont appris à partir de ce corpus de texte. Pour assurer la qualité du corpus de texte, des filtrages ont été appliqués. Un certain nombre de traitements ont été effectués (voir Lê, 2006, p. 48) afin de rendre les données recueillies exploitables.

À partir de ce corpus de vocabulaire ainsi automatiquement généré, nous avons utilisé l'outil *text2wfreq*⁴⁷ pour sélectionner 5 000 items lexicaux sur le critère de la fréquence lexicale. Nous avons vérifié dans ce lexique la présence de 100 mots « aussi indispensables que possible » de la liste de Swadesh pour le vietnamien, extrait de la thèse de doctorat de Lê Văn Trường⁴⁸ (2004) ainsi que celle des 200 mots les plus fréquents de Hoàng⁵ (1994). Nous les avons enfin comparés également aux 1 000 mots les plus courants de la méthode Assimil (manuel d'apprentissage du vietnamien) de Đỗ Thế Dũng & Lê Thanh Thủy (1994). Quel que soit l'échantillon considéré, il est inclus dans le lexique du vietnamien retenu pour la base G-ULSID.

Concernant les formes lexicales « étrangères » empruntées qui ne sont pas d'origine vietnamiennes bien que souvent utilisées dans la vie quotidienne, comme « Internet », « email », « mobile », « fax », « euro » ... elles ont été exclues de notre liste (cf. Maddieson et Precoda, 1992 ; cité par Rousset, 2004, p. 53).

3.1.2. Intégration du lexique dans la base G-ULSID

Le lexique a été saisi sous forme de liste de mot dans un fichier *langue.txt* avec les conventions suivantes pour l'exploitation automatique des données :

- la police API est SilDoulosIPA 93
- un item lexical par ligne (entrée)
- les syllabes sont séparées par des points graphiques (cas des mots composés) précédés et suivis d'un espace
- les constituants de la syllabe (attaque, noyau, coda) sont séparés entre eux par un blanc graphique (espace)

⁴⁷ Source : http://svr-www.eng.cam.ac.uk/~prc14/toolkit_documentation.html#vocab_file

⁴⁸ Source : <http://ngonngu.net/index.php?p=264> (site sur le langage vietnamien)

À noter que plusieurs entrées lexicales comportent les mêmes segments consonantiques et vocaliques mais sous des tons différents. Illustration ci-dessous avec un extrait de quelques entrées du lexique du vietnamien :

ʔ a
 ʔ a j
 ʔ a j . n ɤ̃ j
 ʔ a n
 ʔ a n . n i ŋ
 ʔ a n . t w a n

Dans cet extrait, nous pouvons identifier trois unités lexicales monosyllabiques et trois dissyllabiques, ces dernières correspondant à des mots composés (cf. section 2.2.1). La figure 3.1 schématise la structure syllabique du cinquième item lexical (avec σ = syllabe, A = attaque, R = rime, N = noyau, C = coda).

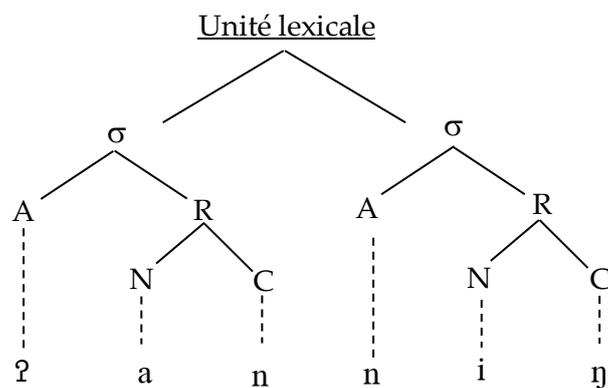


Figure 3.1. Représentation syllabique de l'unité lexicale [ʔ a n . n i ŋ].

3.2. Analyses quantitatives

3.2.1. Nombre de syllabes par item lexical

Parmi l'ensemble des requêtes possibles avec l'interface d'exploitation des données de G-ULSID (Maupeu, 2005), il est possible d'obtenir automatiquement, pour chaque langue, la distribution du lexique en fonction du nombre de syllabes. Rousset (2004) propose la typologie des langues suivante, basée sur la distribution du lexique en fonction du nombre de syllabes par unité lexicale (cf. fig. 3.2).

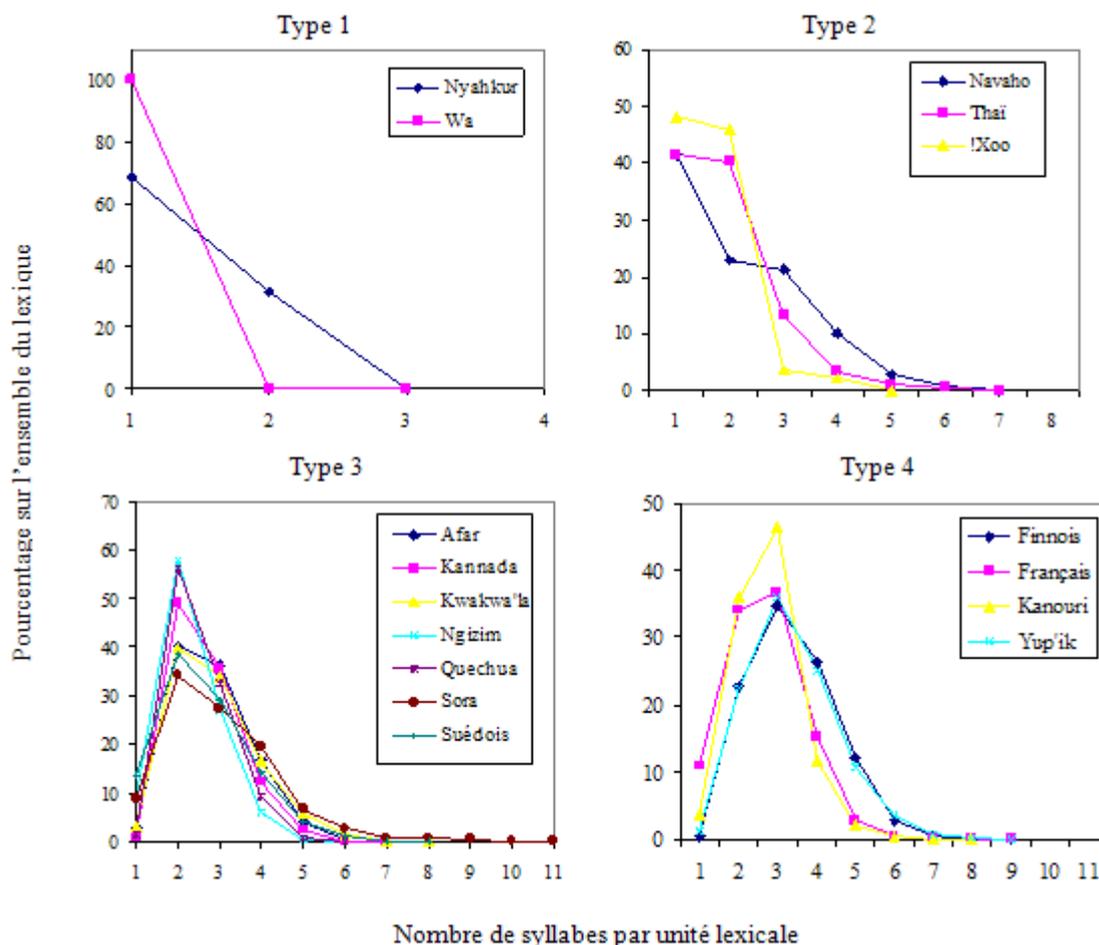


Figure 3.2. Typologie des 16 langues de G-ULSID, élaborée sur la répartition du lexique en fonction du nombre de syllabes par unité lexicale (Rousset, 2004, p. 90).

Nous proposons de situer le vietnamien dans la typologie des langues proposé par Rousset (2004). La figure 3.3 montre que les mots vietnamiens sont soit monosyllabiques (mots simples) soit plurisyllabiques (mots composés). Cette langue appartient au type 2 des langues de la classification de Rousset (2004) qui regroupe les langues présentant « une distribution des unités lexicales telle que 40 % au moins d'entre elles sont monosyllabiques, entre 20 % et 40 % dissyllabiques... » (Rousset, 2004, p. 91). Bien que la langue vietnamienne soit majoritairement monosyllabique et souvent décrite phonologiquement comme telle (Nguyễn Phú Phong, 1989), la figure 3.3 montre que les mots composés, notamment ceux à 2 syllabes représentent la moitié des mots du vocabulaire (2 503 sur 5 000 mots). Le nombre de mots monosyllabiques est proche (2 480 mots). Les items composés de 3 syllabes sont très limités (17 mots), soit 0,68 % du lexique. Les items lexicaux de 4 syllabes et plus pouvant être comptés sur les doigts de la main, aucune entrée de ce type ne figure dans le lexique des mots les plus fréquents en vietnamien.

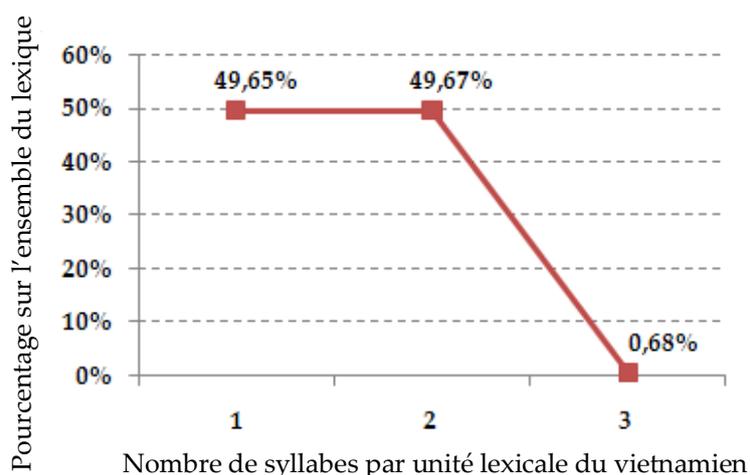


Figure 3.3. Distribution du lexique vietnamien en fonction du nombre de syllabes par unité lexicale.

Dans la typologie proposée par Rousset (2004), le français possède très peu d'unités monosyllabiques (10 %), mais une majorité d'unités trisyllabiques (37 %), 34 % de dissyllabes, 15 % de quadrisyllabes, 2,5 % de quinquasyllabes et 1 % d'unités comptant de 6 à 9 syllabes (cf. figure 3.2). Du fait de cette distribution, le français a été classé par Rousset dans le type 4 qui regroupe les langues majoritairement trisyllabiques. Mentionnons que le type de répartition le plus fréquent rencontré dans l'échantillon de langues observé par Rousset est celui qui favorise le nombre d'unités dissyllabiques : « les structures à deux syllabes représentent le type d'unités lexicales le plus répandu dans les langues du corpus » (Rousset, 2004, p. 95). L'introduction du vietnamien dans G-ULSID ne modifie donc pas la tendance observée par Rousset sur la prédominance des unités dissyllabiques dans toute la base de données en général.

3.2.2. Rendement syllabique

Les 5 000 entrées lexicales du vietnamien transcrites en API livrent un inventaire de 1 433 syllabes phoniques différentes. Rousset (2004) a observé une grande variabilité entre le nombre de syllabes des langues et la taille du lexique : « le nombre de syllabes d'une langue est globalement indépendant de la taille du lexique » (Rousset, 2004, p. 80). Par comparaison au vietnamien, le kannada, langue dravidienne de la base G-ULSID, possède un lexique de taille comparable à celui du vietnamien (4 559 unités lexicales), mais un nombre de syllabes beaucoup plus élevé (12 126).

Le rendement syllabique du vietnamien a été calculé comme le rapport entre « le nombre de syllabes obtenues dans le découpage d'un lexique donné et le nombre de syllabes différentes comptabilisées après regroupement des syllabes identiques au niveau de la nature des segments qui la composent » (Vallée, Boë, Maddieson & Rousset, 2000). Un rendement syllabique (RS) égal à 1 signifie que chaque syllabe n'apparaît qu'une seule et unique fois dans le corpus. Plus le RS

d'une langue est haut, plus la fréquence d'utilisation d'une même syllabe phonique dans les unités lexicales est importante. Le rendement syllabique permet de montrer dans quelle proportion les langues réutilisent leurs syllabes dans la construction des unités lexicales (Rousset, 2004).

Rousset constate un rendement élevé pour les langues ayant une forte proportion d'unités lexicales trisyllabiques (RS de 24 pour le kanouri, 23 pour le finnois). Les rendements les plus faibles sont rencontrés pour les langues de type 1 (cf. fig. 3.2) totalement ou majoritairement monosyllabiques (comme le wa, langue famille parlé en Chine et le nyah kur parlé au centre de la Thaïlande où le RS est inférieur à 2) (Rousset, 2004, p. 121).

Avec un total de 7 510 syllabes dont 1 433 syllabes différentes, le lexique du vietnamien livre un rendement de 5,2. En d'autres termes, une même syllabe, c'est-à-dire identique au niveau de la nature des segments qui la compose, peut apparaître en moyenne 5,2 fois dans l'échantillon des 5 000 items lexicaux.

Vallée et al. (2000), puis Rousset (2004) ont proposé un classement des langues en fonction de leur rendement syllabique (cf. figure 3.4). Le vietnamien se comporte au niveau du RS comme les langues du type 2 de la typologie de Rousset. Sous ce type 2, Rousset classe des langues comme le thaï, le navaho, le vietnamien (RS autour de 5) qui possèdent un taux plus faible de réutilisation des syllabes par rapport aux langues des types 3 et 4 qui connaissent respectivement des unités lexicales majoritairement di- ou trisyllabiques (RS entre 5 et 14). Ce rattachement du vietnamien dans le type 2 est probablement à relier au caractère de langue tonale syllabique et isolante où « le mot, le morphème et la syllabe semblent coïncider presque entièrement » (Cao Xuân Hạo, 1985).

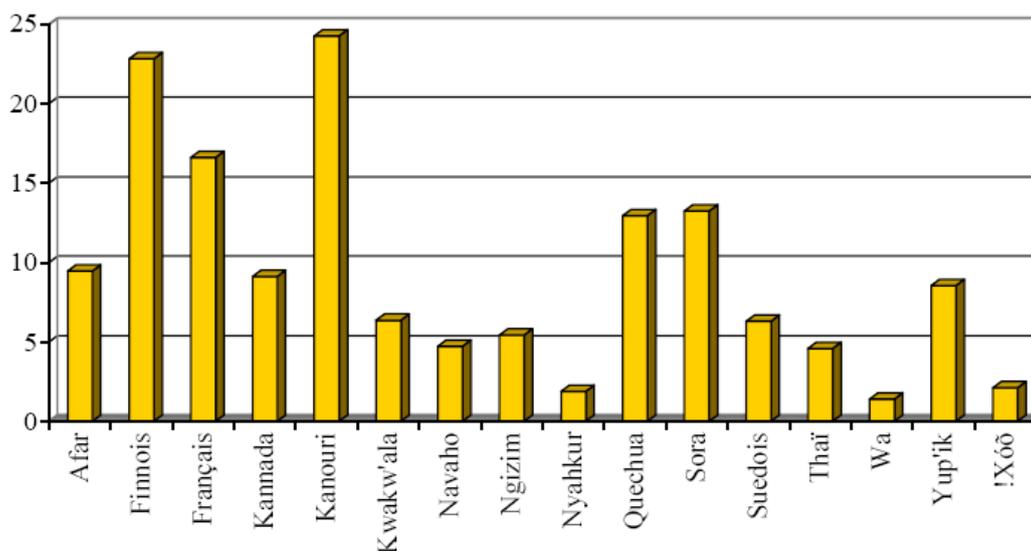


Figure 3.4 Rendements syllabiques des langues de G-ULSID (Rousset, 2004, p. 121) (La valeur de RS pour le vietnamien est de 5,2).

La réutilisation de syllabes dans la construction du lexique permet aux langues de limiter leur inventaire syllabique. Les langues augmentent d'autant plus la réutilisation de syllabes que la longueur de leurs unités lexicales peut être grande.

3.2.3. Gabarits lexicaux du vietnamien

En calculant le nombre des gabarits possibles des items lexicaux pour chaque langue, en fonction du type auquel elle appartient (le type définissant la répartition du lexique selon le nombre de syllabes par unité lexicale), Rousset a montré que les langues majoritairement monosyllabiques ont le nombre de gabarits le plus faible (48 en moyenne). Ceci s'expliquant par des possibilités de combinaisons syllabiques différentes moindres que les langues à tendance di- et trisyllabiques, qui possèdent un potentiel gabaritique plus élevé (392 en moyenne).

Le nombre de gabarit en vietnamien est de 23. La table 3.1 présente, pour la langue vietnamienne, la liste de tous les gabarits représentant plus de 2 % des unités lexicales classées en fonction du nombre de syllabes qu'ils contiennent.

Nous remarquons que le gabarit favori des unités monosyllabiques en vietnamien est comme en français (cf. table 3.2) la structure fermée CVC. C'est aussi une tendance des langues en général : CVC est la structure lexicale (mot) la plus fréquente dans 10 des langues de G-ULSID. Nous remarquons une plus grande variété de gabarits rencontrés en français.

Table 3.1. Structures des gabarits lexicaux et leur fréquence d'apparition dans le lexique du vietnamien (5 000 entrées), à gauche les structures monosyllabiques (CCVC = C/w/VC), au centre les dissyllabiques, à droite 3 syllabes et plus.

CVC	35 %	CVC.CVC	24,4 %		
CV	11 %	CV.CVC	8,1 %	Autres	7,1 %
CCVC	2,6 %	CVC.CV	8,1 %	(16 gabarits)	
		CV.CV	3,7 %		

Table 3.2. Structures représentant les gabarits lexicaux et leur fréquence d'apparition dans le lexique du français (Rousset, 2004, p. 192).

	CV.CVC	6,9 %	CV.CV.CV	6,8 %		
CVC	3,7 %	CV.CV	6,6 %	CV.CV.CVC	4,1 %	CV.CV.CV.CV
	CCV.CV	2,2 %	V.CV.CV	2,5 %	Autres	64,9 %

Table 3.3. Structures représentant plus de 2 % des gabarits de l'ensemble des 16 lexiques de G-ULSID.

Gabarits	Fréquence d'apparition dans l'ensemble des lexiques
CVC	8,80 %
CV.CV	7,79 %
CV.CVC	6,85 %
CV.CV.CV	5,74 %
CVC.CV	5,44 %
CVC.CVC	4,60 %
CV	2,62 %

La table 3.3 présente les gabarits les plus fréquents dans l'ensemble des lexiques de la base (Rousset, 2004, p. 105) : 41 % des gabarits présents dans l'ensemble des lexiques se répartissent en seulement 7 structures. La table 3.3 montre que, dans les langues, le gabarit lexical CVC est plus fréquent que le gabarit dissyllabique CV.CV. Ce résultat est fortement influencé par les langues totalement ou majoritairement monosyllabiques. Nous remarquons la faible présence des séquences de consonnes dans les sept structures gabaritiques les plus fréquentes : seulement deux structures au niveau inter-syllabique. Le vietnamien présente les séquences de consonnes en inter-syllabique (cf. table 3.1) alors qu'en français, les séquences sont rencontrées davantage en intra-syllabe (cf. table 3.2).

3.2.4. Structures syllabiques du vietnamien

Rappelons qu'en vietnamien, l'inventaire des consonnes en coda est restreint (cf. 2.3.4), la proportion de l'inventaire consonantique pouvant apparaître en attaque et en coda se répartit comme suit :

Attaque	Coda
95,65 %	34,8 %

En comparaison avec les langues de la base G-ULSID (voir Rousset, 2004, p. 127), pour ce qui concerne la position d'attaque syllabique, nous constatons qu'il n'y a que peu de disparité entre les langues et que la majorité des consonnes d'une langue donnée peut apparaître dans cette position (toujours plus de 86,67 % de l'inventaire phonologique de la langue). Pour ce qui est des codas, le comportement varie beaucoup plus d'une langue à l'autre. Ainsi, en français, toutes les consonnes, sauf /w/ et /ŋ/, peuvent apparaître en coda, alors qu'en vietnamien, seulement 34,8% d'entre elles (8 sur les 23 segments consonantiques) peuvent occuper cette position.

Table 3.4. Structures syllabiques classées par type et leur proportion dans le lexique du vietnamien (*)
la deuxième consonne en attaque de syllabe correspond à la prétonale /w/ (cf. section 2.3.2)).

N°	Structures syllabiques du vietnamien	Pourcentage
1	CV	24,09 %
2	CCV*	1,79 %
3	CVC	69,47 %
4	CCVC*	4,65 %

Dans cette partie est examinée l'organisation des syllabes du vietnamien au niveau de leurs segments consonantiques et vocaliques. Les syllabes fermées sont plus fréquentes (74,12 %) que les syllabes ouvertes (25,98 %) (cf. table 3.4). Dans une langue de type isolant monosyllabique comme le vietnamien, aucun fait de phonétique syntactique entraînant une resyllabation du fait d'élision, d'enchaînement ou de liaison, à cause de la constitution et la nature de la syllabe, ne peut se produire (Cao Xuân Hạo, 1985).

Les structures syllabiques du français sont plus nombreuses (16 structures) et montrent des degrés différents de complexité (cf. table 3.5), la complexité étant estimée par le nombre de consonnes présentes dans l'attaque et/ou dans la coda : l'attaque peut contenir jusqu'à trois positions consonantiques et la coda jusqu'à quatre. En français, les structures contenant des attaques complexes sont 4 fois plus nombreuses que celles contenant des codas complexes.

Table 3.5. Structures syllabiques classées par type et leur proportion dans le lexique du français (d'après Rousset, 2004, p. 116).

Rang	Structures syllabiques du français	Pourcentage	Rang	Structures syllabiques du français	Pourcentage
1	CV	54,17 %	9	CVCC	2,70 %
2	CCV	10,84 %	10	CVCCC	0,12 %
3	CCCV	0,25 %	11	CVCCCC	0,002 %
4	V	7,94 %	12	CCVC	3,27 %
5	VC	1,92 %	13	CCVCC	0,32 %
6	VCC	0,41 %	14	CCVCCC	0,01 %
7	VCCC	0,005 %	15	CCCVC	0,09 %
8	CVC	17,92 %	16	CCCVCC	0,03 %

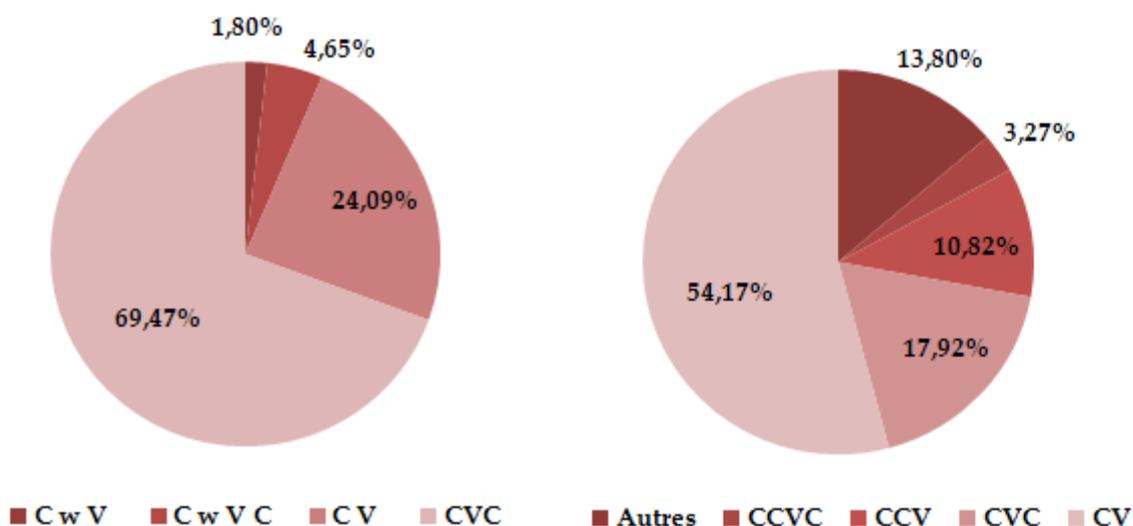


Figure 3.5. Proportions des différents types syllabiques du vietnamien (à gauche) et du français (à droite) (les légendes sont mises en ordre croissant des pourcentages).

Contrairement au vietnamien, les syllabes ouvertes en français sont plus nombreuses (73,18 %) que les syllabes fermées (26,82 %). Les syllabes fermées présentent une diversité beaucoup plus importante : 12 types contre 4 pour les syllabes ouvertes. Cette disparité est explicable puisqu'une syllabe ouverte n'ayant par définition pas de coda, seule la complexité de l'attaque peut varier.

Le calcul du nombre d'occurrences de chaque type syllabique dans l'ensemble des lexiques du vietnamien et du français nous permet de mettre en évidence les structures les plus favorisées dans chaque langue (cf. figure 3.5).

L'observation de la fréquence des types syllabiques permet de faire émerger les deux types dominants du vietnamien et du français : CVC et CV. Ces deux types totalisent à eux seuls 93,56 % des syllabes en vietnamien, et 72,09 % en français. La structure CVC constituant à elle seule presque 70 % des syllabes de la langue vietnamienne, elle représente le type syllabique le plus recruté dans cette langue. Le français est une langue à structure CV dominante (plus de 50 %).

Il est important de repreciser que l'attaque vietnamienne n'est jamais vide. Si on observe ces données en fonction de la présence ou absence de consonnes en coda, on obtient un classement dans lequel les structures du vietnamien avec coda pleine sont largement favorisées (74,12 %) par rapport aux autres types syllabiques sans coda (25,98 %). Ce résultat renforce ce que Rousset a remarqué pour les 5 langues à tons dans l'échantillon de G-ULSID « le wa, le nyah kur, le thaï (toutes les trois issues de la famille austro-asiatique mais de branches différentes), le sora famille parlé en Inde et le suédois possèdent majoritairement des syllabes de type CVC. Ces cinq lexiques favorisent, contrairement aux autres, les syllabes fermées, c'est-à-dire à coda pleine » (Rousset, 2004, p. 113). Cette utilisation de la tonologie

semble influencer le recrutement d'unités syllabiques, comme cela a été montré dans des études précédentes (Hombert, Ohaha & Ewan, 1979, cité par Rousset).

Avec 4 types syllabiques, le vietnamien correspond sur ce point au nombre minimal de structures syllabiques que l'on rencontre dans les langues de G-ULSID. De 4 types (pour le thaï famille et le yup'ik langue eskimo-aléoute), ce nombre peut atteindre 19 pour le suédois. Le nombre moyen est de 9 types par langues, ce qui est très faible quand on le compare au nombre de syllabes différentes répertoriées dans les langues : en moyenne 15 453 syllabes par langue, et une variation comprise entre 3 180 pour le wa et 60 994 pour le français. Les 7 510 syllabes du vietnamien se répartissent donc en 4 types. Rousset précise que « le nombre de types de structures syllabiques d'une langue n'est corrélé ni à la taille des lexiques, ni au nombre de syllabes, ni au nombre de phonèmes de la langue concernée » (Rousset, 2004, p. 109).

Comme remarqué par Rousset (2004), ou encore Blevins (1995), la fréquence des structures syllabiques dans le lexique est inversement liée à la complexité de l'attaque et/ou de la coda. Plus la structure est proche de celle des syllabes de base CVC ou CV, plus elle est présente dans le lexique. Ainsi, les syllabes avec coda vide et les séquences consonantiques sont nettement moins fréquentes dans les langues du monde.

La partie suivante détaille les fréquences d'occurrence des séquences consonantiques du vietnamien en comparaison des groupes consonantiques du français.

3.3. Analyses quantitatives des séquences consonantiques du vietnamien

Avant toute chose, le point terminologique suivant est à rappeler : les clusters consonantiques se distinguent de ce qui sera appelé dans la suite de cette étude séquences consonantiques. Le terme « cluster » désigne une séquence de segments fortement coarticulés réunis dans une position syllabique (attaque ou coda). Alors que « séquence » désigne une succession de segments appartenant à des syllabes différentes, c'est-à-dire qu'une frontière syllabique est présente quelque part dans cette suite de segments consonantiques. Cette terminologie se base donc sur la position du groupement consonantique dans le mot : cluster s'il est intra-syllabe, séquence si inter-syllabe.

Compte tenu du caractère monosyllabique de la langue vietnamienne, aux plans synchronique et phonologique, cette langue ne connaît pas de clusters, autrement dit de groupes consonantiques. Les séquences n'apparaissent donc qu'à l'inter-syllabe. Les groupes de consonnes intra-syllabiques sont inexistant du point de vue des descriptions phonologiques en vietnamien. Cependant, comme déjà indiqué plus haut, on rencontre dans

certaines syllabes, des consonnes initiales séparées de la voyelle, noyau de syllabe, par la prétonale labio-vélaire /w/ (cf. section 2.3.2). Ce segment /w/ est, dans la description phonologique (Đoàn Thiện Thuật, 1999), un élément appartenant à la rime. Vraisemblablement, cette séquence C + /w/ est différente d'une attaque complexe comme on peut en trouver en français : /w/ donne le trait [+labial] à la syllabe (Đoàn Thiện Thuật, 1999). Les séquences CC dans la table 3.6 suivante ne sont donc pas à considérer comme des clusters. Elles sont des suites phonétiques d'unités consonantiques.

Les structures qui contiennent ces suites de segments consonantiques concernent 30 % des structures lexicales en vietnamien. Les suites consonantiques sont présentes à la frontière syllabique des mots composés (82,95 % des suites consonantiques du vietnamien). Par comparaison, les structures CwV sont présentes dans 17,05 % des unités lexicales.

Table 3.6. Types de séquences consonantiques dans le lexique du vietnamien (* : le segment /w/ est un élément de la rime. Le point indique la position de la frontière syllabique).

Séquence de segments consonantiques du vietnamien	Nombre d'occurrences	Pourcentage
V C . C V	1 746	78,33 %
(*) C C V	380	17,05 %
V C . C C V	103	4,62 %
Total	1 849	100 %

Dans le lexique du français, les groupes de consonnes intra-syllabiques (clusters) sont plus importants que les séquences inter-syllabiques (cf. table 3.7). La plupart des clusters occupent la position d'attaque (80 % vs. 20 % pour la position de coda).

Table 3.7. Structures des gabarits lexicaux qui contiennent des groupes consonantiques en français.

Groupes consonantiques	Nombre d'occurrences	Pourcentage
C C	9 886	64,25 %
C . C	4 260	27,69 %
C . C C	719	4,67 %
C C C	285	1,85 %
C C . C	134	0,87 %
C C . C C	81	0,53 %
C . C C C	13	0,08 %
C C . C C C	7	0,05 %
C C C C	1	0,01 %
Total	15 386	100 %

Le nombre de distributions des suites de consonnes est différent dans les deux langues : 2 en vietnamien mais jusqu'à 9 en français (cf. table 3.7), en cohérence avec le fait qu'on ne trouve pas en vietnamien plus de 2 segments consonantiques dans le même constituant syllabique. Ainsi, dans notre corpus, une suite de segments consonantiques intra-syllabique ne peut pas dépasser deux éléments (Cw) et trois en inter-syllabiques, alors qu'en français l'attaque peut contenir jusqu'à trois consonnes et la coda jusqu'à quatre.

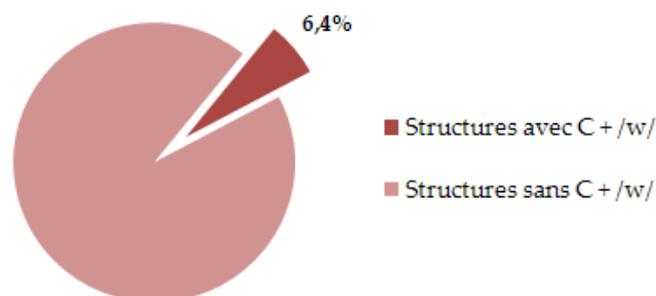


Figure 3.6. Proportions des structures syllabiques avec et sans séquences C + /w/ du vietnamien.

La figure 3.6 présente la proportion de ce type de syllabe CwV(C) dans le lexique de 5 000 entrées du vietnamien dans lequel 7 510 syllabes au total ont été relevées parmi lesquelles 238 combinaisons possibles de consonnes. En français, 499 combinaisons ont été dénombrées.

Si on se penche sur la nature (ou les caractéristiques phonétiques) des segments qui composent les séquences de consonnes, il est possible de pointer ressemblances et divergences entre les deux langues. Pour pouvoir comparer le vietnamien au français tiré des données de Rousset (2004), des regroupements typologiques ont été effectués. Les consonnes ont été regroupées en catégories, en fonction de leur lieu et mode d'articulation (cf. table 3.8). Rappelons que le système du vietnamien contient 25 phonèmes consonantiques : 23 consonnes (22 en attaque + /p/) et 2 approximantes /w j/.

Table 3.8. Regroupement des consonnes du vietnamien selon leurs lieu (à gauche) et mode d'articulation (à droite).

Lieu	Code	Consonnes (API)	Mode	Code	Consonnes (API)
Bilabial	Bi	p b m	Plosif	Pl	p b t t ^h d t c k ?
Labio-dental	Lde	f v	Nasal	Na	m n ŋ ɲ
Coronal	Co	t t ^h d n s z l t̚ s̚ z̚	Fricatif	Fr	f v s z s̚ z̚ ʃ ʒ h
Palatal	Pa	c ɲ j	Latéral	La	l
Vélaire	Ve	k ŋ ʏ ɣ	Approximant	Ap	w j
Glottal	Gl	? h			
Labio-vélaire	Lve	w			

3.3.1. Séquences C + /w/ en vietnamien

Dans différentes études phonologiques (Đoàn Thiện Thuật, 1999 ; Nguyễn Thị Bình Minh, 2000) les séquences C+/w/ ne sont pas considérées comme groupes de consonnes. Rappelons que cette combinaison C+/w/ représente 17,05 % de l'ensemble des séquences de consonnes relevées dans le lexique du vietnamien (cf. table 3.6). 14 sur les 22 consonnes en attaque participent à cette combinaison dans le lexique de 5 000 mots (cf. figure 3.7).

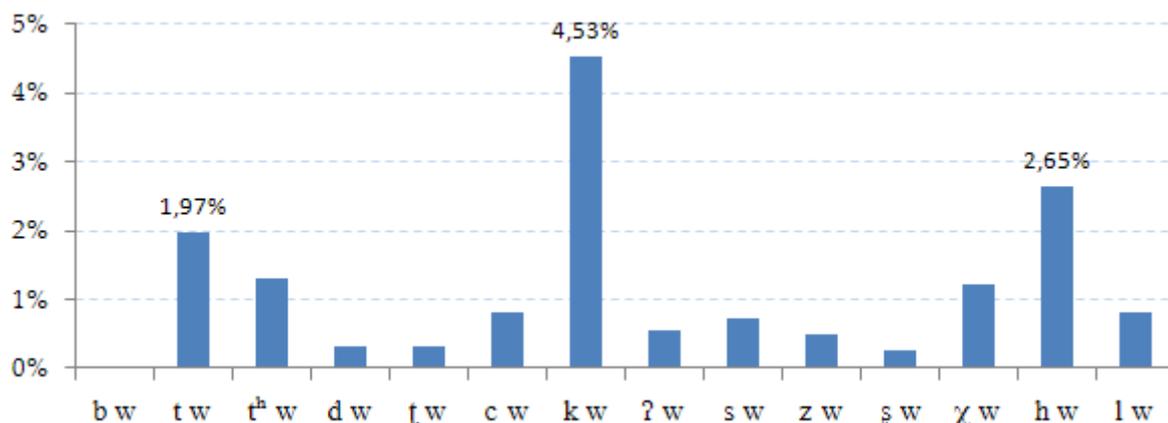


Figure 3.7. Pourcentage d'occurrences des séquences C +/w/ sur l'ensemble des suites de consonnes en vietnamien, classés selon le mode d'articulation.

La figure 3.7 montre que les séquences C+/w/ (C étant consonne initiale) favorisent majoritairement des plosives sourdes vélares et coronales /k t tʰ/. Viennent ensuite la fricative glottale /h/ et la fricative sourde vélaire /ɣ/. Rappelons aussi que la combinaison [bw] du mot « buýt » /bwitʰ/ étymologiquement français *bus* est le seul exemple de cette séquence dans notre lexique.

Les lieux d'articulation vélaire et coronal sont ainsi largement privilégiés en présence de /w/ dans l'initiale (cf. table 3.9).

Table 3.9. Nombre d'occurrences et pourcentage des séquences C+/w/ classées selon le lieu d'articulation.

Séquences	Nombre d'occurrences	Pourcentage sur les syllabes avec Lve	Pourcentage sur l'ensemble de syllabes ayant des suites de consonnes
Ve Lve	149	40,49 %	6,89 %
Co Lve	138	37,50 %	6,38 %
Gl Lve	59	16,03 %	2,73 %
Pa Lve	21	5,71 %	0,97 %
Bi Lve	1	0,27 %	0,05 %

En ce qui concerne le mode d'articulation des consonnes en attaque syllabique, les plosives sont plus recrutées avec 57 % d'occurrences parmi les combinaisons de ce type (cf. table 3.10). La combinaison entre vélaire et labio-vélaire /kw/ est la séquence la plus recrutée parmi les séquences C+/w/ à l'initiale du vietnamien.

Table 3.10. Nombre d'occurrences et pourcentage des séquences C+/w/ classées selon le mode d'articulation.

Séquences	Nombre d'occurrences	Pourcentage sur les syllabes avec approximante	Pourcentage sur l'ensemble de syllabes ayant des suites de consonnes
Pl Ap	219	57,63 %	9,83 %
Fr Ap	119	31,32 %	5,34 %
Na Ap	24	6,32 %	1,08 %
La Ap	18	4,74 %	0,81 %

3.3.2. Séquences consonantiques à l'inter-syllabique des mots composés en vietnamien

En fonction des consonnes qui peuvent apparaître en position de coda en vietnamien (plosives sourdes /p t k/ ; nasales /m n ŋ/ ; approximantes /j w/), huit grands types de séquences consonantiques à l'inter-syllabique peuvent être relevées (le point représente la frontière syllabique) : /p.C/, /t.C/, /k.C/, /m.C/, /n.C/, /ŋ.C/, /j.C/, /w.C/. Les 22 consonnes initiales peuvent se trouver en attaque de la deuxième syllabe du mot composé.

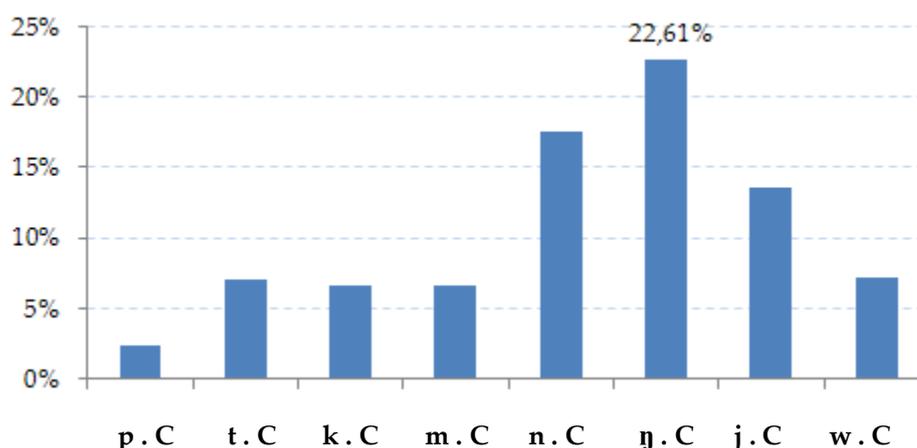


Figure 3.8. Pourcentage des séquences consonantiques à l'inter-syllabique des mots composés sur l'ensemble des suites de consonnes du vietnamien.

La structure syllabique d'un mot composé est de type C₁VC₂.C₃VC₄. La consonne nasale vélaire /ŋ/ est le segment le plus fréquent en coda C₂ de la première syllabe. Les nasales apparaissent en C₂ dans 55 % des cas de mots composés (cf. figure 3.9). Les approximantes /j/

et /w/ occupent à 25 % la position de coda C₂. Quant à /p t k/, elles sont présentes en position de coda de la première syllabe de mots composés dans 19 % des cas de séquences de consonnes inter-syllabiques du vietnamien.

La figure 3.10 montre que les lieux d'articulation vélaire et coronal sont largement privilégiés en position de coda (C₂) des séquences consonantiques inter-syllabiques des mots composés en vietnamien. Les séquences de consonnes Ve.C₃ et Co.C₃ totalisent à elles seules 54 % de l'ensemble des séquences inter-syllabiques.

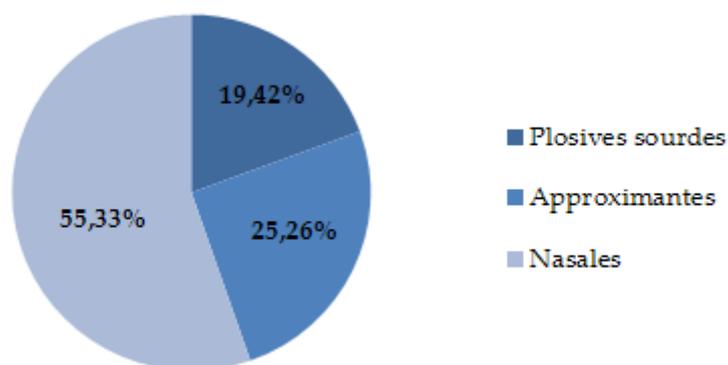


Figure 3.9. Pourcentage d'occurrences des différentes modes consonantiques rencontrées en position de coda C₂ d'un mot composé C₁VC₂.C₃VC₄.



Figure 3.10. Pourcentage d'occurrences des consonnes en coda (C₂) classée selon leur lieu d'articulation sur l'ensemble des suites de consonnes en vietnamien.

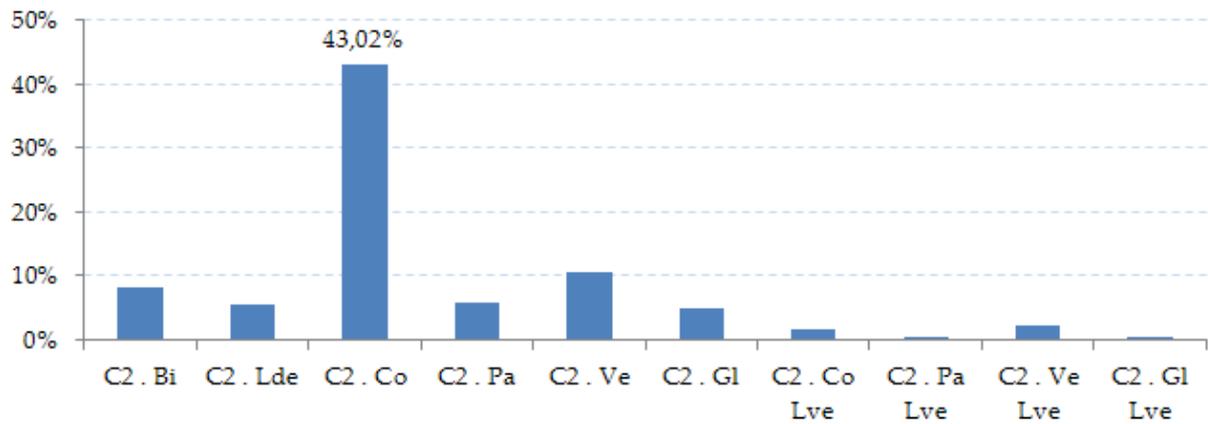


Figure 3.11. Pourcentage d'occurrences des consonnes en position C₃ classées selon leur lieu d'articulation sur l'ensemble des suites de consonnes en vietnamien.

L'observation des occurrences entre les 8 phonèmes en coda C₂ et les phonèmes en attaque C₃ mettent en évidence une forte proportion du trait coronal en C₃ des mots composés. Celle-ci est présente dans plus de 40 % de l'ensemble constitué des suites de consonnes relevées dans les items lexicaux (cf. figure 3.11).

La création de matrice de cooccurrences de lieu d'articulation entre C₂ et C₃ montre que les coronales se trouvent toujours au premier rang des fréquences quelle que soit C₂ (cf. table 3.11).

Table 3.11 : Cooccurrences C₂ et C₃ selon le lieu d'articulation.

C ₃ \ C ₂	C ₂					Total
	Bi	Co	Lve	Pa	Ve	
Bi	18	58	8	33	63	180
Lde	13	27	8	23	49	120
Co	103	274	90	160	304	931
Ve	27	81	19	22	78	227
Pa	14	36	9	20	47	126
Gl	8	27	12	22	39	108
Co Lve	6	13	4	5	8	36
Pa Lve		1	3	1	1	6
Ve Lve	6	15	4	8	15	48
Gl Lve		4		4	6	14
Total	195	536	157	298	610	1796

La table 3.11 montre que les trois séquences les plus favorisées dans des mots composés vietnamiens sont :

- consonne vélaire en C₂ et consonne coronale en C₃ (16,9 % sur l'ensemble des C₂.C₃)
- consonne coronale en C₂ et consonne coronale en C₃ (15,3 % sur l'ensemble des C₂.C₃)
- /j/ en C₂ et consonne coronale en C₃ (8,9 % sur l'ensemble des C₂.C₃)
- Trois séquences de consonnes n'apparaissent jamais dans lexique retenu : il s'agit des séquences de type C.CC. Ainsi :
 - les bilabiales (/p/ ou /m/) ne précèdent jamais les glottales /h/ et /ʔ/ suivies de /w/
 - les bilabiales (/p/ ou /m/) ne sont jamais suivies des palatales en C₃+/w/
 - De même, /w/ en C₂ n'est jamais suivi d'une glottale en C₃, si C₃ est suivie de /w/

De la même façon, les matrices de cooccurrences C₂.C₃ en fonction du mode articulaire ont été calculées. La figure 3.12 montre que les plosives sont les plus favorisées en C₃ dans les mots composés (36,39 %).

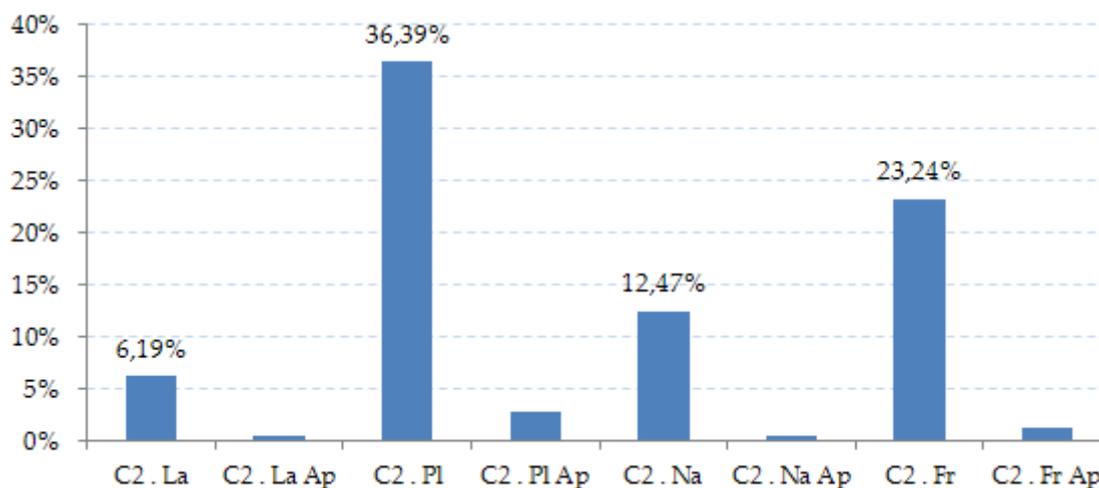


Figure 3.12. Pourcentage d'occurrences des consonnes C₃ classées selon leur mode d'articulation sur l'ensemble des suites de consonnes en vietnamien.

La table 3.12 fait émerger une préférence nette pour les séquences C₂.C₃ de type nasale + plosive. Rappelons que les nasales constituent les segments les plus fréquents en coda de syllabe 1.

Table 3.12. Matrice de cooccurrences entre consonnes dans les séquences C₂.C₃ selon leur mode d'articulation.

C ₃ \ C ₂	Ap	Pl	Na	Total
Pl	193	139	479	811
Fr	143	110	265	518
Na	62	60	156	278
La	40	33	65	138
Pl Ap	22	8	29	59
Fr Ap	4	4	17	25
La Ap	2	4	5	11
Na Ap	1	1	7	9
Total	467	359	1023	1849

Ainsi, les trois séquences les plus fréquentes en fonction de mode d'articulation des consonnes C₂ et C₃ sont :

- nasale en C₂ et plosive en C₃
- nasale en C₂ et fricative en C₃
- approximante en C₂ et plosive en C₃

Les séquences C+/w/ sont peu nombreuses en C₃ (5,62 % des composés lexicaux).

Les analyses des fréquences d'occurrences des consonnes selon leur lieu et mode d'articulation et selon leur position dans la syllabe (attaque ou coda) font apparaître des tendances. Ainsi, en attaque, le lieu et le mode d'articulation le plus répandu est coronal (62,4 %) devant vélaire (14,4 %) et plosif (36,4 %). En coda, les consonnes coronales et vélares sont prédominantes mais changent un peu le classement : vélares (34 %) *vs.* coronales (30 %) ; les modes favorisés sont nasal et plosif. Les consonnes bilabiales, labio-vélares ou palatales sont nettement défavorisées en coda d'une syllabe. Comme souligné par Rousset (2004) pour les autres langues de G-ULSID, les coronales sont très populaires.

Lorsque C₂ et C₃ partagent le même lieu d'articulation, il est intéressant de remarquer que les consonnes vélares, qui sont très majoritaires en coda en vietnamien, sont pourtant défavorisées (4,34 %) dans les séquences inter-syllabiques C₂.C₃ (cf. table 3.13). A l'exception des coronales, la même situation est observée pour les consonnes bilabiales et palatales.

Rousset (2004) a remarqué que les structures syllabiques C₁VC₂ avec consonne de même lieu d'articulation en attaque et en coda étaient défavorisées dans les langues (< 5 %), même pour les coronales (ex. [tat], [pap] *vs.* [pat]). Les résultats sur les séquences inter-syllabiques C₂.C₃ du vietnamien (cf. table 3.13) vont dans le même sens montrant une tendance forte à défavoriser les consonnes de même lieu d'articulation (Ve.Ve, Bi.Bi, Pa.Pa),

alors que la séquence C₂.C₃ avec coda et attaque de même lieu d'articulation coronal (Co.Co) est bien présente avec un fort pourcentage de fréquence d'emploi.

Table 3.13. Cooccurrences des consonnes dans les séquences C₂.C₃ selon leur lieu d'articulation en vietnamien.

	Bi	Co	Pa	Ve
Bi	18	58	33	63
Co	103	274	160	304
Pa	14	36	20	47
Ve	27	81	22	78

3.3.3. Comparaison avec les groupements de consonnes intra- et inter-syllabe du français

L'analyse de la distribution des séquences de consonnes du vietnamien montre de nombreuses dissemblances avec les données du français tirées du lexique figurant dans la base G-ULSID.

Dans le lexique du vietnamien, les suites consonantiques sont très présentes à la frontière syllabique d'un mot composé (82,95 % des suites consonantiques du vietnamien), alors que les structures C+/w/ ne présentent qu'un faible pourcentage (17,05 %) (cf. table 3.6). En français, les groupes consonantiques intra-syllabiques occupent une place plus importante (66 %) que les séquences inter-syllabiques (34 %) (cf. table 3.7). Les séquences C₂.C₃ totalisent 27,69 % du nombre total des groupes consonantiques (cf. table 3.7) où ceux avec les plosives /p t k m n ŋ/ en C₂ occupent seulement 6,65 % du total des clusters du français.

En français, les groupes C₁C₂ intra-syllabiques les plus favorisées sont de type Plosive + Fricative (22,91 % du nombre total des groupements de consonnes en français) (cf. figure 3.13) et de type Coronal+Palatal (10,5 % du total des groupes) (cf. figure 3.14). La consonne la plus recrutée en C₂ des groupes C₁C₂ en attaque en français est la fricative uvulaire /ʁ/. Les clusters /tʁ/, /pʁ/, /gʁ/, /kʁ/, /bʁ/ (classés en ordre décroissant) sont les plus fréquents en attaque, ils totalisent à eux seuls 20,5 % du nombre total des groupements consonantiques en français.

L'analyse des occurrences des consonnes C₂.C₃ en vietnamien nous permet de relever que les séquences Vélaire + Coronale, Nasale + Plosive sont les plus favorisées dans des mots composés vietnamiens. Pour le français, les figures 3.13 et 3.14 montrent que les combinaisons Fricative + Plosive et Uvulaire + Coronale sont les plus fréquentes en inter-syllabique (respectivement 11,27 % et 7,96 % des groupements consonantiques).

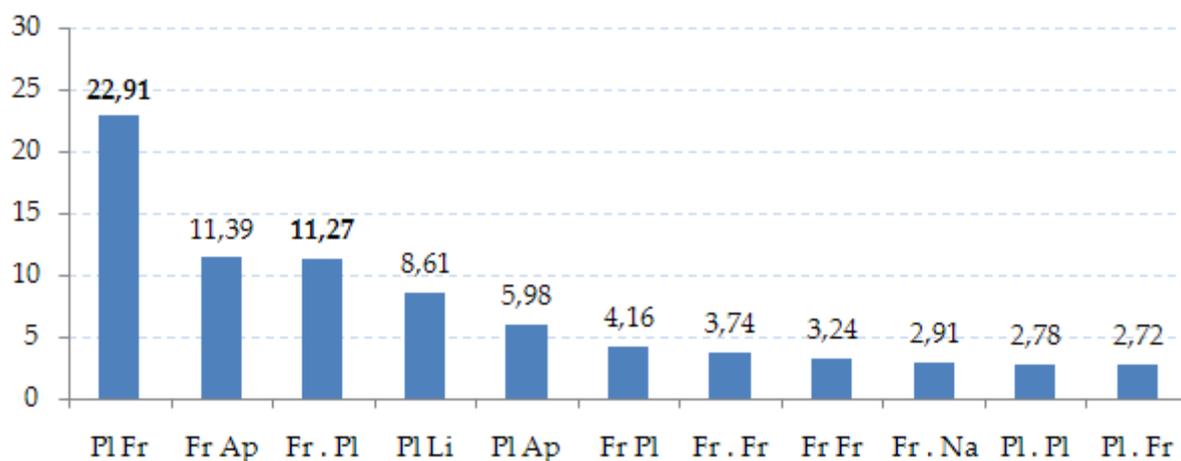


Figure 3.13. Pourcentage d'occurrences des clusters et des séquences de consonnes classés selon leur mode d'articulation (taux de fréquence de plus de 2% sur l'ensemble des groupements de consonnes en français) (le point représente une frontière syllabique intra-mot) (Pl = plosive, Fr = fricative, Ap = approximant, Li = liquide, Na = nasale).

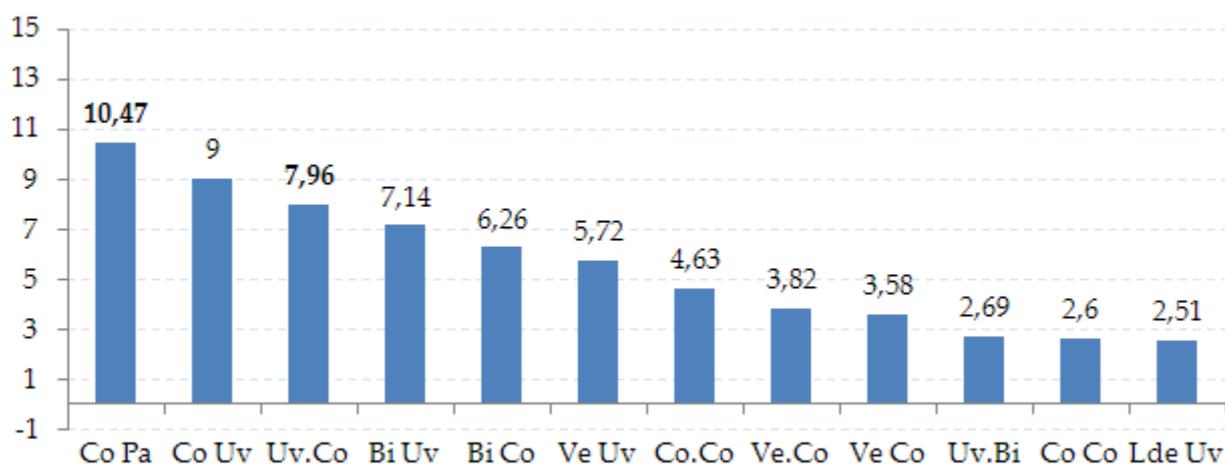


Figure 3.14. Pourcentage d'occurrences des clusters et des séquences de consonnes classés selon leur lieu d'articulation (taux de fréquence de plus de 2% sur l'ensemble des groupements de consonnes en français) (le point représente une frontière syllabique intra-mot) (Co = coronal, Pa = palatal, Uv = uvulaire, Bi = bilabial, Ve = vélaire, Lde = labio-dental).

Après cette étude descriptive et quantitative des séquences de consonnes dans les deux langues, les chapitres suivants sont consacrés aux études expérimentales. Celles-ci sont divisées en deux phases. Dans le chapitre IV figure un ensemble d'études acoustiques et perceptives des séquences de consonnes du vietnamien dans le but de rechercher l'existence ou non de différences de réalisation et/ou de perception en fonction du type de frontière (de mot ou de syllabe). Le chapitre V regroupe un ensemble d'expériences de perception et de production des groupes de consonnes intra- ou inter-syllabiques (clusters *vs.* séquences inter-syllabes) auprès d'apprenants vietnamiens du Français Langue Etrangère.

Chapitre 4

Corrélat acoustico-perceptifs des consonnes simples et en séquences du vietnamien

4.1. Expérience 1 : Étude acoustique des consonnes en séquences du vietnamien

Dans cette section, est présentée l'analyse des paramètres acoustiques des plosives /p t k/ et des nasales /m n ŋ/ du vietnamien (parler du Nord), en comparant leur réalisation, dans un corpus lu, en fonction de leur position dans le mot et la syllabe. Toutes les distributions possibles en vietnamien pour ces consonnes ont donc été prises en compte, l'objectif étant de comprendre les différences dans leurs réalisations qui pourraient exister en fonction de leur distribution et qui pourraient être des éléments du crible phonologique.

Pourquoi ces six consonnes ? Rappelons que l'inventaire des consonnes en coda du vietnamien est restreint. Mises à part les approximantes /w j/, ces six consonnes sont les seules légitimes à occuper la position de coda, et donc à apparaître comme le premier élément d'une séquence inter-syllabique en vietnamien (cf. sections 2.3.4 et 3.3.2).

La parole lue a été choisie afin de répondre à l'exigence des démarches scientifiques imposant le contrôle des conditions expérimentales, nécessaires à la répétabilité des mesures et à la comparaison de plusieurs conditions et de plusieurs sujets.

Trois objectifs sont visés avec cette étude acoustique :

1. L'étude de la réalisation acoustique des plosives et nasales du vietnamien lorsqu'elles apparaissent en coda : soit finale de mot CVC, soit finale de syllabe 1 (S1) de composé lexical CVC.CVC ;
2. La comparaison avec leurs réalisations en attaque de syllabe CVC ;
3. La vérification de l'existence ou non de différences de réalisation d'une même consonne selon qu'elle est coda de S1 dans les dissyllabes CVC.CVC ou qu'elle soit finale de monosyllabes CVC. Il s'agit en fait de comparer les réalisations de consonnes à la frontière de deux syllabes d'un composé lexical (intra-mot) par rapport à une distribution de part et d'autre d'une frontière de mot.

4.1.1. Méthodologie

4.1.1.1. Constitution du corpus

À partir du lexique du vietnamien de 5 000 entrées constitué pour les besoins de l'étude détaillée des structures lexicales et des séquences de consonnes qu'elles contiennent (cf. chapitre III), a été sélectionné un ensemble d'items comportant l'une des 6 consonnes / p t k m n ŋ / en contexte de la voyelle la plus ouverte /a/ dans des structures lexicales

monosyllabiques CV, CVC et dissyllabiques CVC.CVC, de même contexte tonal (ton lexical montant 5 ou B1 à sonantes finales et D1 à plosives finales).

Rappelons que le changement du ton entraînera le changement de sens du mot. Les raisons pour lesquelles nous avons choisi ce contexte tonal résident d'une part au niveau lexical : les syllabes à plosives finales /p t k/ en vietnamien se trouvent uniquement soit sous le ton 5 *sắc* (D1), soit sous le ton 6 *ngặng* (D2). L'examen du lexique de 5 000 entrées montre que les mots à plosive finale qui répondent aux critères fixés par nos objectifs (en syllabe 1, voyelle /a/, pas de prétonale /w/), étaient trois fois plus nombreux sous le ton D1 que D2 (62 à 22 mots respectivement). D'autre part sur le plan acoustique, les tons D1 et D2 ne comportent pas de glottalisation dans le parler du Nord (Michaud, 2004). Le ton B1 est réalisé en voix modale alors que le ton B2 (à sonantes finales) est glottalisé, aucune trace acoustique des segments nasals finals n'étant détectée sous ce ton B2 dans l'étude de Michaud (2006). Pour faciliter la segmentation et les mesures acoustiques, le ton 5 (B1 à sonantes finales et D1 à plosives finales) a donc été retenu pour le corpus.

La voyelle /a/ a été choisie parce qu'elle fournit un meilleur contraste entre le segment vocalique au contact immédiat de l'attaque et/ou de la coda et le segment consonantique qui remplit ces positions.

Pour le cas des mots composés, dans l'inventaire lexical restreint dû à nos différents critères méthodologiques (syllabe 1, voyelle /a/, absence de /w/), le choix du même contexte tonal pour les deuxièmes syllabes (S2) a été envisagé. Ces dernières sont finalement réalisées soit sous le ton 5 *sắc* (B1-D1), soit sous le ton 6 *ngặng* (B2-D2) à cause des critères portant sur S1. Il est à noter que le contexte tonal des S2 pourrait ne pas influencer les mesures de F0 des S1, qui sont les cibles de notre étude acoustique, car selon Brunelle (2003), la coarticulation tonale se fait plutôt dans le sens progressif que régressif dans un mot composé (cf. section 2.4.2).

Dans cet ensemble d'items lexicaux sélectionnés pour l'étude, les six consonnes cibles /p t k m n ŋ/ se trouvent dans l'une des trois distributions suivantes :

- En position initiale C₁ d'unité lexicale monosyllabique de structure ouverte [C₁a] ou fermée [C₁aC₂];
- En position finale C₂ d'unité lexicale de type monosyllabique [C₁aC₂];
- En position finale C₂ de la première syllabe [C₁aC₂] de mots composés lexicaux dissyllabiques [C₁aC₂.C₃VC₄].

Un ensemble de 62 items monosyllabiques et dissyllabiques qui répondent à ces critères ont été extraits de la base lexicale de 5 000 entrées. Ils sont présentés dans la table 4.1 suivante.

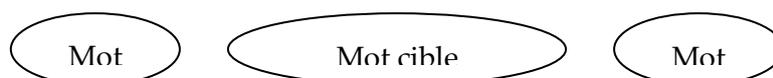
Table 4.1. Liste des 62 items sélectionnés à partir du lexique vietnamien pour l'analyse acoustique des consonnes plosives et nasales. Les consonnes ciblées par l'étude sont mises en évidence en caractère gras.

N°	Stimuli	Transcription	Signification
1	tá	t a	douzaine
2	tác	t a k	bramer
3	tát	t a t	gifler
4	tám	t a m	huit
5	tán	t a n	courtiser
6	táp	t a p	happer
7	táng	t a ŋ	donner un coup de poing (<i>familier</i>)
8	cá	k a	poisson
9	cát	k a t	sable
10	các	k a k	les, des (articles au pluriel)
11	cán	k a n	manche
12	cáp	k a p	câble
13	cám	k a m	son (enveloppe des céréales)
14	cáng	k a ŋ	brancard
15	má	m a	joue
16	mát	m a t	frais
17	mác	m a k	coutelas à long manche
18	mán	m a n	peuple de San Chay (au nord-est)
19	máng	m a ŋ	mangeoire
20	ná	n a	arbalète
21	nát	n a t	écrasé
22	nám	n a m	hâlé
23	ngáng	ŋ a ŋ	barrer
24	ngáp	ŋ a p	bâiller
25	ngác	ŋ a k	désorienté
26	ngát	ŋ a t	pénétrant (<i>parfum</i>)
27	ngán	ŋ a n	être dégoûté
28	pháp lý	f a p - l i	juridique
29	áp bức	? a p - b u k	opprimer
30	bát ngát	b a t - ŋ a t	immense
31	khát nước	x a t - n u y k	avoir soif
32	phát giác	f a t - z a k	révéler
33	hát ví	h a t - v i	air de chanson populaire
34	lác đác	l a k - d a k	clairsemé
35	gác bút	g a k - b u t	abandonner le métier d'écrivain
36	xác đáng	s a k - d a ŋ	pertinent

37	<i>đám cưới</i>	d a m - k u ɤ j	mariage
38	<i>khám phá</i>	x a m - f a	découvrir
39	<i>bán kết</i>	b a n - k e t	demi-finale
40	<i>gián tiếp</i>	z a n - t i e p	indirect
41	<i>đáng kính</i>	d a ŋ - k i ŋ	vénérable
42	<i>đáng tiếc</i>	d a ŋ - t i e k	regrettable
43	<i>sáng kiến</i>	s a ŋ - k i e n	initiative
44	<i>sáng tác</i>	s a ŋ - t a k	composer
45	<i>tháng tám</i>	t ^h a ŋ - t a m	août
46	<i>áp dụng</i>	? a p - z u ŋ	appliquer
47	<i>áp lực</i>	? a p - l u k	pression
48	<i>pháp luật</i>	f a p - l w a t	loi
49	<i>pháp lệnh</i>	f a p - l e ŋ	ordonnance
50	<i>khát vọng</i>	x a t - v ɔ ŋ	aspiration
51	<i>phát động</i>	f a t - d o ŋ	déclencher
52	<i>mát dạ</i>	m a t - z a	satisfait
53	<i>ác liệt</i>	? a k - l i e t	très violent
54	<i>tác dụng</i>	t a k - z u ŋ	utilité
55	<i>xám xỉt</i>	s a m - s i t	couleur de plomb
56	<i>bám trụ</i>	b a m - t ɕ u	tenir pour résister
57	<i>thán phục</i>	t ^h a n - f u k	admirer
58	<i>cán sự</i>	k a n - s u	fonctionnaire du cadre secondaire
59	<i>sáng tạo</i>	s a ŋ - t a w	créer
60	<i>kháng cự</i>	x a ŋ - k u	résister
61	<i>tháng chạp</i>	t ^h a ŋ - t ɕ a p	décembre
62	<i>đáng sợ</i>	d a ŋ - s ɤ	effroyable

Les mots ont été choisis de manière à ce que toute consonne sourde en position de coda, qui est la cible de notre étude acoustique, soit suivie d'une consonne sonore et inversement, pour que toute consonne sonore soit suivie d'une consonne sourde. Ce critère méthodologique a été appliqué aux syllabes 2 des mots composés mais explique aussi le choix des phrases porteuses notamment des mots monosyllabiques.

Deux types de phrase porteuse ont été retenus de manière à ce que l'environnement des phonèmes à étudier soit le plus neutre possible du point de vue articulatoire :



Ainsi, l'unité lexicale A qui précède le mot sélectionné pour l'étude se termine par une voyelle. Ensuite, l'initiale du mot B doit faciliter la segmentation du signal : elle est occupée

soit par une plosive (ex /t/ dans « tiếp » [tiɛp]) si le mot cible est terminé par l'une sonante (voyelle ou /m n ŋ/) (cf. table 4.2a), soit par une consonne sonore (« liên » [liɛn]) si le mot étudié est terminé par une des plosives sourdes /p t k/ (cf. table 4.2b). Les mots B ont été choisis aussi de manière à ce que la signification des deux phrases porteuses ne soit pas trop éloignée.

Pour que les phrases porteuses conservent une prononciation la plus naturelle qui puisse être, les locuteurs du corpus ont été invités à imaginer une situation d'apprentissage du vietnamien (compétence de compréhension écrite), dans laquelle il avait à lire une liste de phrases présentées selon la graphie du vietnamien. Les deux phrases porteuses sont données ci-dessous avec leur traduction (*Trad.*) et leur transcription phonétique (*Tr.Ph*) selon le parler du Nord (cf. table 4.2a & 4.2b).

Le corpus, réparti en deux enregistrements, est constitué de 4 répétitions des 62 mots insérés dans les phrases porteuses et présentées en un ordre aléatoire pour tous les locuteurs. Les 248 phrases du corpus ont été lues à voix haute par 10 sujets, 5 hommes et 5 femmes, âgés de 23 à 28 ans (la moyenne d'âge est de 25 ans), originaires du Nord du Vietnam. Les phrases à lire étaient inscrites sur papier en taille 18 pour le confort du lecteur. Les locuteurs ont été invités à prononcer chaque phrase d'un seul trait sans faire de pause entre les mots de la phrase. L'ensemble du corpus tel que soumis à la lecture aux 10 sujets figure dans l'annexe I.

L'enregistrement s'est déroulé dans la chambre sourde du Département Parole et Cognition de GIPSA-lab avec enregistreur numérique Marantz PMD 670, micro AKG C1000S à directivité cardioïde. Le corpus a été numérisé à 44.1 kHz sous format .wav. Tous les locuteurs confortablement installés devant une table ont été invités à lire le corpus à un débit normal. Les pauses étaient insérées au cours de la tâche de lecture (après toutes les trois pages de lecture, soit environ 60 phrases).

Table 4.2a. Phrase porteuse pour les mots cibles terminés par une sonante. Ici le mot cible est « ma ».

<i>Graphie</i>	Bạn	sẽ	gặp	từ	« má »	tiếp	tục	xuất	hiện	trong	bài	khóa
<i>Tr.Ph</i>	ban	sɛ	ɣăp	tu	ma	tiɛp	tuk	swɔ̣t	hien	tɕɔŋ	baj	xwa
<i>Trad.</i>	Tu vas rencontrer le mot				« joue »	continuer à apparaître dans le texte						

Table 4.2b. Phrase porteuse pour les mots cibles terminés par une plosive sourde. Ici le mot cible est « mát ».

<i>Graphie</i>	Bạn	sẽ	gặp	từ	« mát »	liên	tiếp	xuất	hiện	trong	bài	khóa
<i>Tr.Ph</i>	ban	sɛ	ɣăp	tu	mat	liɛn	tiɛp	swɔ̣t	hien	tɕɔŋ	baj	xwa
<i>Trad.</i>	Tu vas rencontrer le mot				« frais »	apparaître successivement dans le texte						

4.1.1.2. Traitement des données

Les 2 480 phrases enregistrées (3 heures de parole lue) ont été segmentées manuellement sous Praat⁴⁹. Les paramètres acoustiques ont ensuite été analysés de manière semi-automatique sous Praat[®] et Matlab[®]. Une fenêtre d'étiquetage est donnée en exemple figure 4.1. Cette fenêtre présente le signal de parole, le spectrogramme avec la courbe de F0 et le suivi de formants, la transcription orthographique (Tr.Or) de la phrase (Tier 1) ainsi que la transcription phonétique (Tr.Ph) (Tier 2).

Afin d'étudier les caractéristiques acoustiques des consonnes en séquences du vietnamien, un certain nombre de paramètres à mesurer dans le corpus de 2 480 mots de types C₁[a], C₁[a]C₂ et C[a]C₂.C₃VC contenant les 6 occlusives /p t k m n ŋ/ en position C₁, C₂ et C₂ devant C₃ (C₂.C₃) ont été mesurés :

- durée totale de la consonne cible ;
- durée de la voyelle /a/ qui suit ou précède la consonne cible ;
- durée du VOT (Voice Onset Time = le délai d'établissement du voisement) de la consonne cible ;
- durée de l'occlusion ;

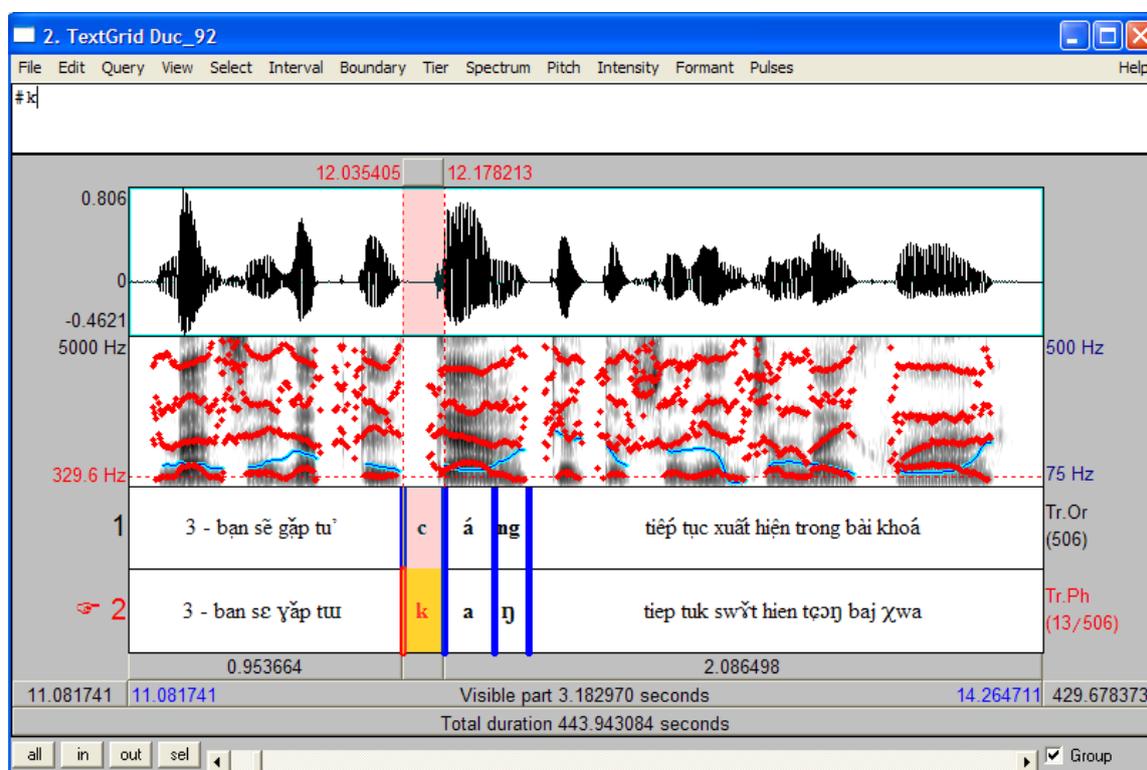


Figure 4.1. Exemple de segmentation (mot et consonne cible) sous Praat.

⁴⁹ <http://www.fon.hum.uva.nl/praat/>.

- durée et intensité du bruit de l'explosion (burst) ;
- transition VC quand C est une plosive : caractéristiques de l'évolution temporelle des 3 premiers formants (F1, F2, F3), de la fréquence fondamentale (F0) et de l'intensité (I) mesurés dans la voyelle à partir de la partie la moins influencée par les consonnes de l'entourage (40 %, 50 %) jusqu'à la partie de transition entre voyelle et consonne finale (à 60 %, 70 %, 80 % et 90 %) de la durée de la voyelle.

L'intérêt d'étudier les transitions VC est qu'en vietnamien, les plosives sont réalisées non relâchées dans cette position. Des études antérieures montrent que les caractéristiques des consonnes en coda sont identifiées par les auditeurs « grâce aux incurvations de formants vocaliques qui se trouvent dans le segment attribué à la voyelle « précédente » (Cao Xuân Hạo, 1985, p. 83). Ainsi, dans notre étude, les variations (Δ) de F0, F1, F2, F3 et de l'intensité ont été obtenues en calculant la différence des valeurs de chaque au temps T2 (de 50 % à 90 %) et au temps T1 (à 40 % partie observée comme la plus stable du noyau vocalique).

La durée totale des plosives en initiale de mot (position C₁) a été mesurée en prenant l'intervalle de temps entre la dernière strie périodique de la voyelle finale /u/ du mot [tu] de la phrase porteuse alignée avec la fin du tracé formantique visible sur le spectrogramme jusqu'à la première strie de périodicité régulière de voyelle /a/ qui suit C₁.

En position finale (C₂), les curseurs ont été positionnés aux endroits qui suivent directement la descente de la dernière crête périodique de la voyelle /a/ jusqu'au début de la vibration des plis vocaux de la consonne /l/ du mot /lien/ qui suit (dans le cas des mots à plosive finale, ou jusqu'à la dernière strie de périodicité régulière du segment consonantique sonant dans le cas des mots à sonante finale).

Dans le cas des plosives, la durée de l'occlusion a été définie dans l'intervalle temporel situé entre le début de l'occlusion et le début du relâchement repéré sur la courbe d'intensité. En position finale, le début de l'occlusion a été déterminé au niveau de la dernière crête périodique de la voyelle /a/ précédente.

Les mesures de VOT et de durée du bruit d'explosion des plosives ont été effectuées à partir des spectrogrammes à larges bandes. Le VOT des consonnes initiales C₁ a été mesuré en prenant l'intervalle de temps entre le début du bruit de détente et la première strie de périodicité régulière de la voyelle suivante. Pour les consonnes en position finale (C₂ devant C₃), le VOT a été mesuré entre le début du burst lorsque celui-ci est présent et le début de la vibration des plis vocaux de la consonne voisée en position C₃.

Les mesures d'intensité du bruit d'explosion sont données en dB relatifs. Les moyennes sont basées sur les 4 répétitions. Rappelons que les mesures d'intensité ne peuvent pas être relevées dans les cas où le burst est absent (cas des plosives non relâchées).

4.1.1.3. Analyses statistiques

Les données des mesures acoustiques ont été statistiquement analysées en utilisant le logiciel SPSS® (*Statistical Package for the Social Sciences*). Des tests d'analyses de variance ANOVA à mesures répétées pour observer des effets intra-sujets ont été effectués.

Deux effets ont été recherchés :

1. L'effet de la position dans la syllabe (C_1 , C_2 , ou $C_2.C_3$)⁵⁰ sur les paramètres acoustiques des consonnes plosives (durée totale du segment, durée de l'occlusion, durée du burst, intensité du burst, durée de VOT). Pour les nasales, un test à part a été mené sur la durée totale de la consonne.
2. L'effet d'interaction entre la position de la consonne en finale de mot C_2 ou en finale de la syllabe 1 de mot composé $C_2.C_3$, le lieu d'articulation (labial, coronal ou vélaire) et les paramètres acoustiques des consonnes cibles (évolution temporelle de l'intensité, de la fréquence fondamentale et des 3 formants de la voyelle mesurés de 40 % jusqu'à 90 % de la durée de la voyelle). Ce test concerne uniquement les consonnes en coda C_2 et $C_2.C_3$.

4.1.2. Résultats

4.1.2.1. Effet de la durée

La figure 4.2 présente les valeurs moyennes de la durée des plosives et des nasales (en milliseconde) selon leur position dans la syllabe. Cette figure montre que, en général, les consonnes les plus longues se trouvent en attaque et en coda d'un monosyllabe (C_1 et C_2), alors que les réalisations les plus brèves, quelle que soit la consonne, sont en coda de syllabe 1 de mot composé ($C_2.C_3$).

L'effet significatif de la différence de durée est globalement observé en intra-sujet pour les plosives [$F(2,18) = 25,634$; $p = 0$] aussi pour les nasales [$F(2,18) = 27,99$; $p = 0$]. Cependant, les effets en intra-sujet⁵¹ de la durée des plosives en C_1 et en C_2 ne sont pas différents significativement [$F(1,9) = 0,004$; $p = 0,951$] pour [t] et [k] (/p/ étant absent en C_1), ce qui n'est pas le cas pour les nasales [$F(1,9) = 5,756$; $p = 0,04$].

⁵⁰ Rappelons que les structures des mots cibles sont [C_1a], [C_1aC_2], [$C_1aC_2.C_3VC_4$] où les consonnes remplissent les positions : attaque (C_1), coda de mot (C_2), coda de syllabe 1 de mot composé ($C_2.C_3$), le point indiquant la frontière syllabique d'un composé lexical.

⁵¹ Pour les tests à mesures répétées, la valeur moyenne de chaque sujet pour chaque condition a été calculée. Etant donné que les valeurs pour /p/ en C_1 sont manquantes car /p/ ne figure jamais en attaque, la comparaison de toutes les plosives /p t k/ confondues n'est possible qu'en coda.

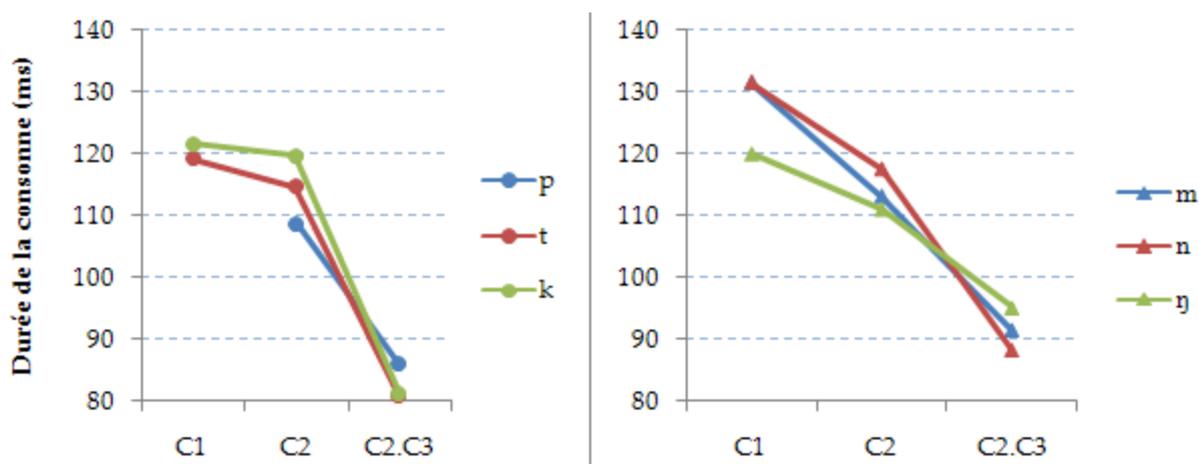


Figure 4.2. Durée moyenne des plosives (à gauche) et des nasales (à droite) en fonction de leurs positions dans la syllabe.

Si on compare C₁ et C₂ de la même structure monosyllabique fermée C₁VC₂, les résultats restent les mêmes. Autrement dit, la durée des plosives en C₁ et C₂ d'une même structure CVC n'est pas différente de manière significative [F(1,9) = 0,192 ; p = 0,671], alors que les nasales sont plus longues en initiale qu'en finale de la même syllabe CVC [F(1,9) = 5,756 ; p = 0,04]. Il est également observé que, de manière générale, la voyelle /a/ est réalisée plus longue dans la structure CVC où elle est précédée par une plosive et suivie d'une nasale (par ex. /tam/) [F(1,9) = 19,188 ; p = 0,002] que dans le patron inverse (par ex. /mat/).

Les nasales ont des durées plus longues que les plosives, mais uniquement en finale de S1 de mot composé C₂.C₃ [F(1,58) = 5,907 ; p = 0,018]. La différence de leurs durées (nasales *vs.* plosives) n'est pas significative en position d'attaque C₁ et de coda C₂ de monosyllabe (p = 0,226 et p = 0,933 respectivement).

L'observation des plosives en position C₁ et la comparaison avec celles en position {C₂, C₂.C₃} montrent des différences entre ces deux séries de position (attaque *vs.* coda), attestées en intra-sujet (échantillons de [t] et [k] seulement, bien entendu) [F(1,9) = 9,746 ; p = 0,012]. Les mêmes effets sont observés pour les nasales [F(1,9) = 25,716 ; p = 0,001].

Aucun effet d'interaction entre la durée et les lieux d'articulation des consonnes n'a été trouvé : les durées de /t/ et /k/ [F(1,9) = 3,142 ; p = 0,11], idem pour les 3 nasales /m n ŋ/ [F(2,18) = 2,038 ; p = 0,159] ne sont pas différentes significativement selon les lieux.

En complémentarité avec l'étude mono-locuteur de Trần et Vallée (2009), ces résultats attestent des différences dans la réalisation de deux types de coda C₂ et C₂.C₃ pour les plosives comme pour les nasales. En effet :

- Des différences significatives de la durée des consonnes existent en fonction des types de structure syllabique (frontière de mot simple C₂ ou frontière de deux syllabes C₂.C₃ à l'intérieur de mot composé) en intra-sujet pour les plosives [F(1,9)

= 61,985 ; $p = 0$] aussi pour les nasales [$F(1,9) = 32,914$; $p = 0$]. Ce résultat montre un effet du type de frontière syllabique sur la durée des consonnes. On peut alors supposer des degrés de coarticulation différents entre des séquences de consonnes se trouvant en position intra-mot par rapport à inter-mot. Aucun effet significatif du lieu d'articulation sur la durée de la consonne n'a été détecté pour les plosives [$F(2,18) = 0,471$; $p = 0,632$], ni pour les nasales [$F(2,18) = 0,046$; $p = 0,955$].

- La voyelle /a/ est plus longue significativement quand elle précède C₂ que C₂.C₃ (cf. figure 4.3) pour les plosives [$F(1,9) = 117,185$; $p = 0$] et les nasales [$F(1,9) = 169,916$; $p = 0$]. Durée qui change aussi de manière significative selon le lieu d'articulation de la plosive finale ([a^k] > [a^l] ($p = 0,006$) ; [a^k] > [a^p] ($p = 0,004$) ; [a^l] ~ [a^p] ($p = 0,129$)) mais pas de la nasale [$F(2,18) = 0,186$; $p = 0,832$].

Ce résultat est intéressant pour l'étude des consonnes finales non relâchées en vietnamien (cf. section 4.2). Il montre que non seulement la partie finale des voyelles porte des indices sur le lieu d'articulation des consonnes qui les suivent, mais que la durée des voyelles contient aussi des informations sur la consonne qui suit (à savoir que sous le ton même montant D1, la voyelle /a/ suivie de la vélaire /k/ est réalisée plus longue que suivie de la coronale /t/ ou de la bilabiale /p/).

- Les plosives en position finale de mot monosyllabique C₂ ont une phase de fermeture plus longue que celles en position finale de S1 de mot composé C₂.C₃ ($p = 0,006$) (cf. figure 4.4). La comparaison entre la position initiale C₁ et finale C₂ d'une syllabe ne montre pas de différence significative ($p = 0,604$). Il n'y a pas non plus d'effet significatif de la durée de l'occlusion des plosives selon leurs lieux d'articulation. La différence, qui semble plus importante pour la vélaire que les deux autres en position C₂.C₃, n'est pas significative en réalité ($p = 0,364$), très probablement en raison de la variabilité des valeurs observées pour la durée de la vélaire.

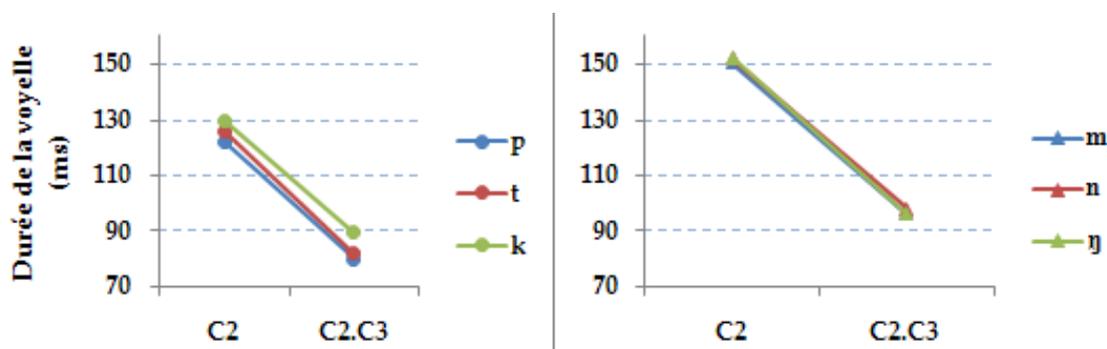


Figure 4.3. Durée moyenne de la voyelle /a/ suivie par des plosives (à gauche) et des nasales (à droite) en fonction de la position finale de mot vs. finale de syllabe 1 de mot composé.

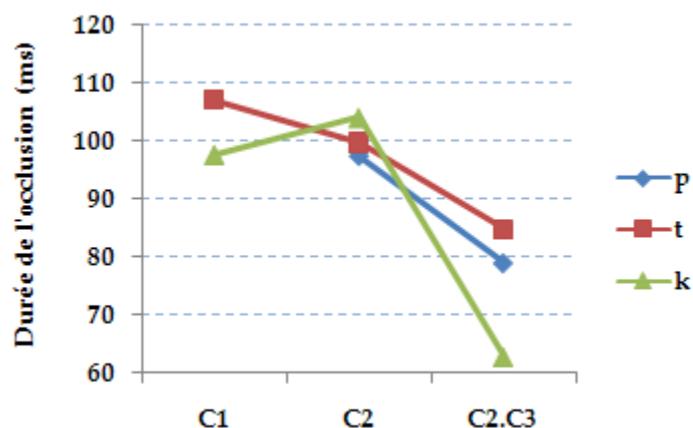


Figure 4.4. Durée moyenne de l'occlusion des plosives selon leur position dans la syllabe.

4.1.2.2. Caractéristiques acoustiques des bruits d'explosion (burst)

Concernant les plosives traditionnellement décrites comme non relâchées en coda (Cao Xuân Hào, 1985 ; Đoàn Thiện Thuật, 1999), l'analyse du corpus multi-locuteurs enregistré pour l'étude montre la présence de bruit d'explosion devant frontière de mot (position C₂) (22,55 %) et de manière plus faible devant frontière syllabique intra-mot (C₂.C₃) (4,72 %). Ces consonnes non relâchées sont plus fréquentes en finale de S1 de mot composé qu'à la frontière de mot. Les bursts sont toujours présents en position d'attaque (C₁). Rappelons que /p/ n'apparaît jamais à l'initiale.

La tendance à un burst plus court en finale est significative. En effet, les bursts des plosives en position de coda C₂ et C₂.C₃ sont nettement moins longs par rapport à ceux des plosives réalisées en initiale (pour /k/ ($p = 0$) aussi bien que pour /t/ ($p = 0,021$)) (cf. figure 4.6). La durée du burst de /k/ est de 12,9 ms en moyenne en C₁ contre 5,6 ms en C₂ et 6 ms en C₂.C₃. Moins long que /k/ mais avec un même effet significatif, le burst de /t/ est de 6,83 ms en C₁, différant de l'ensemble des bursts en C₂ (5,65 ms) et C₂.C₃ (5,8 ms). Les durées des bursts dans les réalisations de /p/, /t/ et /k/ en C₂ et en C₂.C₃ ne sont pas différentes de manière significative ($p = 0,634$).

La force du burst varie également selon la position de la consonne. En effet, si on observe l'effet de la position sur l'intensité du burst de /k/ et de /t/, on trouve que cet effet atteint le seuil de signification pour les deux consonnes (respectivement [$F(2,18) = 19,89$; $p = 0$] et [$F(2,18) = 8,544$; $p = 0,002$]).

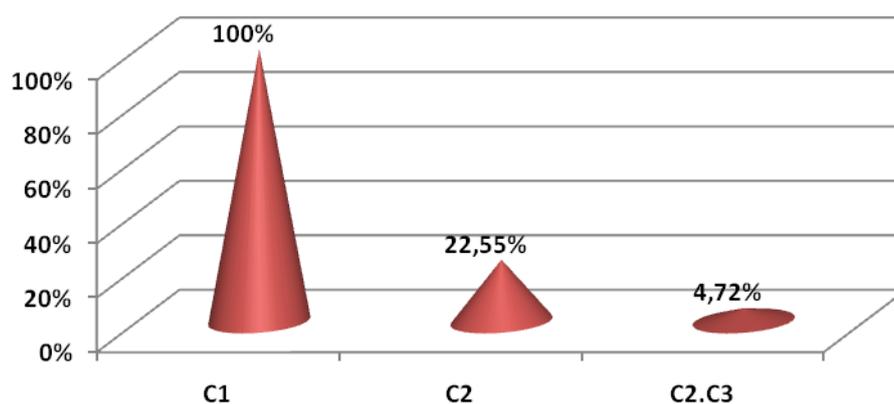


Figure 4.5. Pourcentage de plosives réalisées avec burst en fonction de leurs positions dans la syllabe et dans le mot.

L'analyse de l'intensité des bursts entre C₁ et {C₂, C₂.C₃} de /k/ et /t/ montre un effet significatif (respectivement p = 0 et p = 0,021) indiquant que ces consonnes présentent plus d'énergie dans le bruit de détente en position initiale (65,43 dB de [k] et 61,68 dB de [t]) qu'en position finale du mot (56,95 dB de [k] et 56,38 dB de [t]) ou de la syllabe 1 de mot composé (58 dB de [k] et 57,26 dB de [t]). Autrement dit, les plosives une fois relâchées en finale {C₂, C₂.C₃} comportent un burst significativement plus bref et peu intense par rapport à leur réalisation en initiale C₁ de la syllabe. Mais l'intensité du burst des consonnes de même position finale de mot (C₂) ou de syllabe (C₂.C₃) ne présente pas de différences significatives (p = 0,839).

La consonne vélaire /k/ possède le burst le plus long (12,9 ms) et le plus intense (65,43 dB) en initiale. Dans cette position, /k/ est articulée avec plus d'énergie par rapport aux autres consonnes et aux autres positions.

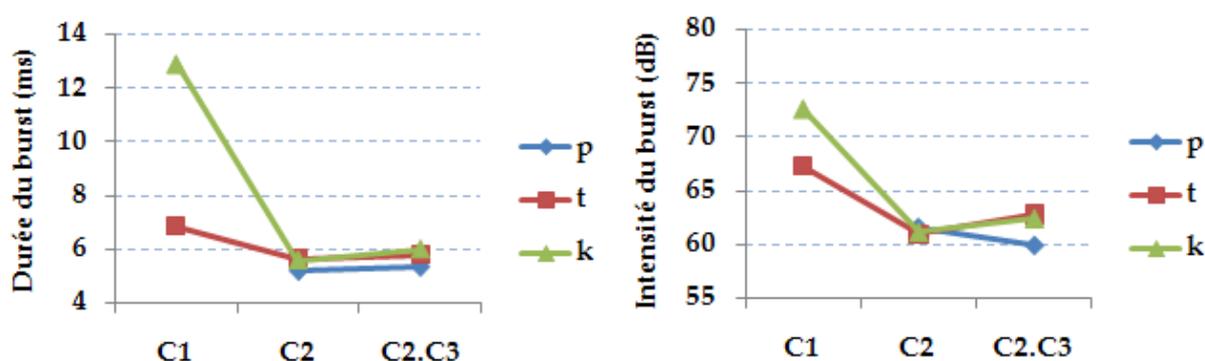


Figure 4.6. Valeurs moyennes de la durée (à gauche) et de l'intensité (à droite) des bursts selon la position des plosives.

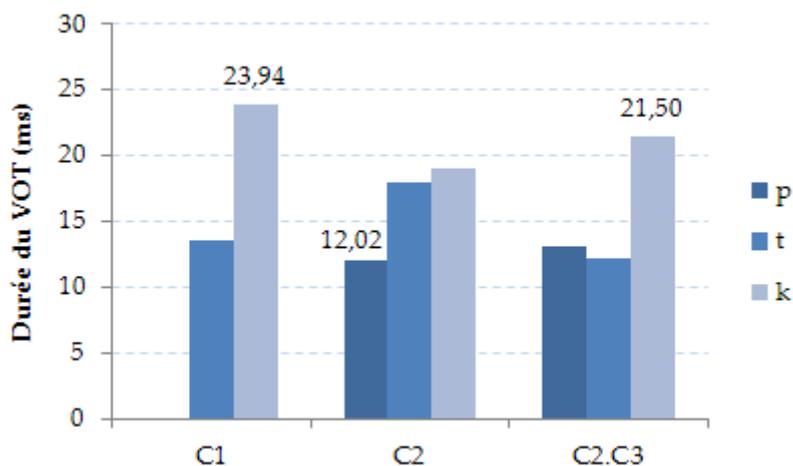


Figure 4.7. Durée moyenne du VOT des différentes plosives selon le type de frontière syllabique (inter-mot vs. intra-mot).

4.1.2.3. Caractéristiques du VOT

Le VOT ou délai d'établissement du voisement (Voice Onset Time) des plosives ne varie pas de manière significative en fonction de leur position dans la syllabe [$F(2,48) = 0,264$; $p = 0,769$]. Par contre, la figure 4.7 montre, comme on pouvait s'y attendre, que le VOT des plosives dépend du lieu d'articulation. En effet, le VOT moyen de la consonne vélaire /k/ a des valeurs plus élevées que le VOT moyen des consonnes articulées aux lieux d'articulation bilabial et coronal et ce, quelle que soit la position dans la syllabe, le mot. Cependant, la différence de VOT entre /k/ et /t/ est estompée en finale C₂ de mot monosyllabique (respectivement 18,9 ms et 17,9 ms). Le VOT de la consonne bilabiale /p/ est significativement le moins allongé par rapport aux autres lieux dans cette position C₂. En position C₂.C₃, la différence de VOT entre /p/ et /t/ n'est plus observée (13 ms vs. 12,1 ms).

4.1.2.4. Transitions voyelle-plosive

Nous constatons dans notre corpus acoustique un grand nombre de cas dans lesquels le bruit d'explosion des consonnes en coda est absent, comme le montre le spectrogramme suivant du mot « gác bút » [ɣak⁵.but⁵] *cesser d'écrire* (cf. figure 4.8).

Le spectrogramme montre que la barre d'explosion (le burst) de /k/ n'existe pas : /k/ final de la première syllabe /ɣak⁵/ du mot composé n'est pas relâché et est prononcé [k^ʔ]. Ce qui a des conséquences particulières : comme la fin de l'occlusion ne produit pas ou presque pas de bruit audible, les finales sont identifiées grâce à leur tenue et à la phase transitoire qui constitue la fin du segment vocalique.

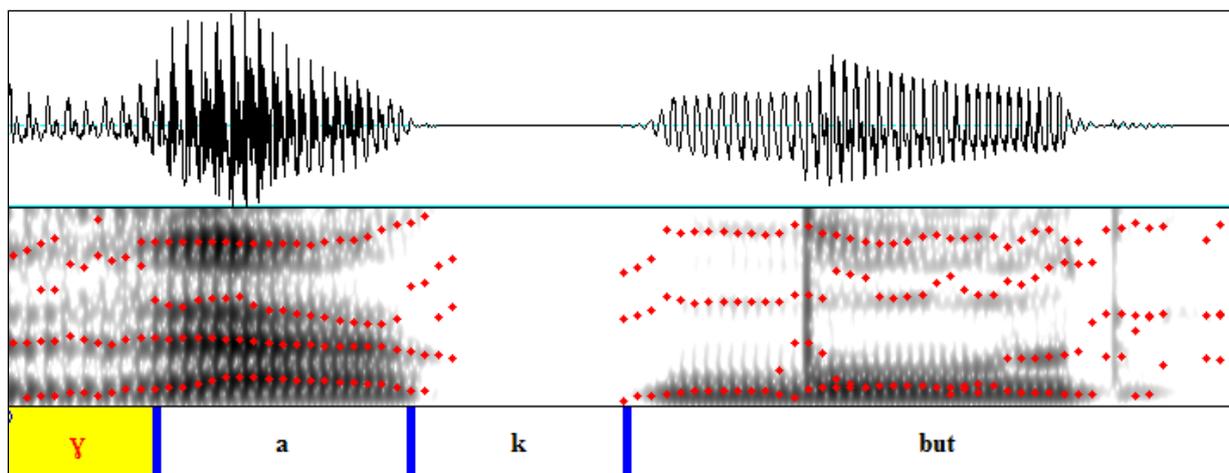


Figure 4.8. Spectrogramme du mot « gác bút » [yak⁵-but⁵].

Comme le précise Cao Xuân Hạo (1985) « la phase correspondant à la tenue de ces consonnes ne donne donc que des informations sur le mode d'articulation ». Les informations sur le lieu d'articulation sont identifiées « grâce aux incurvations de formants vocaliques qui se trouvent dans le segment attribué à la tonale » (Cao Xuân Hạo, 1985, p. 83). Plusieurs études ont aussi considéré les transitions formantiques comme des indices de la perception et de l'identification des consonnes (Lisker, 1957 ; Liberman, 1954 ; Delattre, Liberman, Cooper, 1955 ; Delattre, 1963). En raison de la présence des consonnes non relâchées en position finale C₂ et C₂.C₃, ont été exploités certains indices acoustiques (les trois premiers formants F1, F2, F3, la fréquence fondamentale F0, l'intensité) à partir de 40 % jusqu'à 90 % de la durée de la voyelle dans le but d'étudier les caractéristiques entre types de consonnes en fonction du type de frontière syllabique.

La mesure des formants dans cette partie du signal acoustique est très délicate en raison de la variabilité des formants liée aux phénomènes contextuels de coarticulation. Plusieurs méthodes ont été testées avant de choisir celle qui semblait la plus fiable. La méthode LPC (Linear Predictive Coding) a été retenue car avérée comme la plus sûre dans l'analyse de voyelles très variables (comme celles de l'étude de Ménard (2002) ; Ménard, Chrétien, Lachapelle & Marleau (2010), produites par les enfants).

Les analyses par LPC ont été effectuées afin d'extraire la valeur des trois premiers formants. Les fichiers sonores des 10 sujets ont été ré-échantillonnés à 11 000 Hz. Le signal a été préamplifié à partir de 50 Hz. Les paramètres de l'analyse LPC ont été ajustés pour chaque voyelle analysée. Le premier paramètre ajusté est le nombre de pôles, variant de 12 à 16. Une analyse à cinq fenêtres temporelles d'une largeur de 20 ms pour chaque nombre de pôles (l'écart entre les fenêtres étant de 1 ms) a été conduite afin de suivre l'évolution temporelle de F1, F2, F3 dans la partie de transition VC : une centrée sur le point de repère de la voyelle (c'est-à-dire à 40 %, 50 %, 60 %, 70 %, 80 % et 90 % de sa durée) et 4 autres avec un léger décalage de -2ms, -1ms, +1ms, +2ms. Chaque couple de paramètre (nombre de pôles,

décalage temporel) donne une courbe de réponse spectrale permettant de sélectionner la courbe la plus proche de la courbe moyenne calculée. Les valeurs de référence (moyenne et maximale) pour la voyelle [a] étaient fournies à l’algorithme de détection des formants. Afin de garantir l’exactitude des mesures, les valeurs des formants détectées automatiquement par l’algorithme de LPC ont été vérifiées sur un spectrogramme à large bande. Lorsque la détection automatique ne correspondait pas à la représentation spectrographique, le nombre de pôles était réajusté et la procédure LPC était à nouveau appliquée.

Les valeurs de F0 ont été extraites à l’aide de l’algorithme de détection automatique RAPT⁵² (*A Robust Algorithm for Pitch Tracking*). Les mesures d’intensité ont aussi été détectées de manière automatique à partir de Praat puis Matlab en effectuant des lissages sur les fenêtres d’analyse.

Les mesures de transitions relevées ont ensuite été soumises à des tests ANOVA à mesures répétées pour observer les effets intra-sujet avec les facteurs *lieux* de la consonne post-vocalique (bilabial, coronal, vélaire) et *tons* de la syllabe suivante (neutre A1, montant D1, descendant D2).

4.1.2.4.1. Transitions de F1, F2 et F3

Les valeurs moyennes des trois premiers formants mesurées à six endroits de 40 % à 90 % de la durée de la voyelle avec un pas de 10 ont été analysées en fonction du *lieu* de la plosive située en C₂ vs. C₂.C₃ et du *ton* de la syllabe qui suit. Rappelons que les mots ou parties de mot cibles de l’étude sont tous sous le même contexte tonal *săc* (ton montant B1 à sonantes finales et D1 à plosives finales).

La table 4.3 montre les différences de F1, F2, F3 aux divers endroits en pourcentage de la durée de la voyelle selon le lieu des plosives en coda, quel que soit le type de frontière qui suit immédiatement la consonne cible. Pour une meilleure lisibilité des résultats, les différences non significatives sont présentées avec leur valeur de p ($p > 0,05$), celles significatives avec le symbole *.

Table 4.3. Valeur des seuils du test de significativité pour les valeurs moyennes de F1, F2, F3 selon le lieu des plosives en C₂ vs. C₂.C₃ mesurée de 40 % à 90 % avec un pas de 10 de la durée de la voyelle (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).

Valeurs en moyenne	40 %	50 %	60 %	70 %	80 %	90 %
F1	****	*	p = 0,234	p = 0,49	p = 0,767	p = 0,414
F2	p = 0,13	p = 0,306	p = 0,139	****	***	****
F3	p = 0,88	*	p = 0,139	*	p = 0,17	p = 0,181

⁵² Algorithme extrait de la page Voicebox (Speech Processing Toolbox for Matlab). L’ensemble des routines sont disponibles à l’adresse <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.

Les différences de valeur de F1 en fonction du lieu d'articulation au centre de la voyelle, c'est-à-dire à 40 % et à 50 %, sont trouvés significatives. Les analyses plus fines montrent que c'est le lieu bilabial qui marque cette différence (valeurs de F1 significativement plus élevées que les valeurs de F1 pour les autres lieux). L'effet significatif pour F3 à 50 % est causé par le lieu coronal.

Dans la partie de transition, après 50 % de la durée de la voyelle, des effets significatifs des valeurs moyennes de F2 ont été trouvés à 70 % ($p = 0,001$), à 80 % ($p = 0,002$) et à 90 % ($p = 0$). Les valeurs de F3 sont significativement différentes à 70 % de la durée de la voyelle ($p = 0,025$), alors que les différences de F1 ne sont jamais significatives dans cette zone acoustique de transition entre V et C ($p > 0,05$). Ces résultats suggèrent clairement que la partie finale, et notamment la transition du second formant F2, contient des éléments du lieu d'articulation de la consonne post-vocalique.

Concernant le facteur *ton*, il est observé que le ton de la syllabe suivante influence les valeurs des 2 premiers formants F1 et F2 à partir des 70 % de la durée de la voyelle (cf. table 4.4). Cet effet du ton de la syllabe qui suit la consonne cible est dû au cas des monosyllabes suivies d'une syllabe porteuse d'un ton neutre (ton A1), alors qu'aucune différence significative n'est relevée pour les valeurs des formants F1, F2, F3 de S1 en fonction des tons de S2 dans les mots composés ($p > 0,05$).

Les valeurs moyennes et les écart-types des trois delta F1, F2, F3 de 50 % à 90 % de la durée de la voyelle sont représentées dans l'annexe II. Ces pentes ont été obtenues en calculant la différence entre les valeurs de formants à chaque intervalle (de 50 % à 90 % avec un pas de 10) et les valeurs à 40 %. Par exemple $\Delta F1 \text{ à } 90 \% = F1_{90\%} - F1_{40\%}$; $\Delta F2 \text{ à } 90 \% = F2_{90\%} - F2_{40\%}$; $\Delta F3 \text{ à } 90 \% = F3_{90\%} - F3_{40\%}$. Les $\Delta F0$, ΔI seront obtenus de la même manière. Ces valeurs sont calculées en fonction de la consonne en coda /p/, /t/ ou /k/ et du ton de la syllabe qui suit immédiatement la consonne cible.

Table 4.4. Valeur des seuils du test de significativité pour les valeurs moyennes de F1, F2, F3 selon le ton de la syllabe mesurée de 40 % à 90 % avec un pas de 10 de la durée de la voyelle (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).

Valeurs en moyenne	40 %	50 %	60 %	70 %	80 %	90 %
F1	p = 0,865	p = 0,535	p = 0,112	****	****	**
F2	****	****	****	****	****	****
F3	p = 0,953	p = 0,904	p = 0,954	p = 0,58	p = 0,571	p = 0,857

L'évolution des trois formants dans cette partie finale de la voyelle est présentée par la suite dans les figures respectives 4.9, 4.10 et 4.11. Dans la légende des figures, nous avons nommé les consonnes finales des mots cibles monosyllabiques suivis des syllabes de ton S (« simple » = A1) par p-S, t-S, k-S. Alors que celles suivies des syllabes de ton M (« montant » = D1) ou D (« descendant » = D2) sont les premières syllabes de mots composés, nommés p-M, t-M, k-M et p-D, t-D, k-D. Rappelons qu'aucune différence n'est observée lorsqu'il s'agit d'un mot composé (dissyllabe).

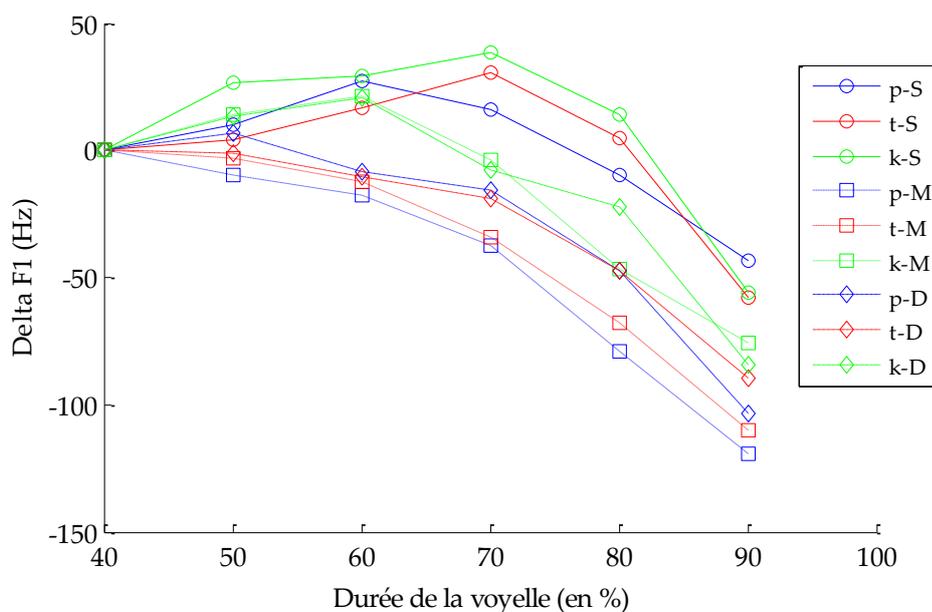


Figure 4.9. Evolution de $\Delta F1$ estimée de 40 % jusqu'à 90 % de la durée de la voyelle en fonction de la plosive en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible (S = A1, M = D1, D = D2).

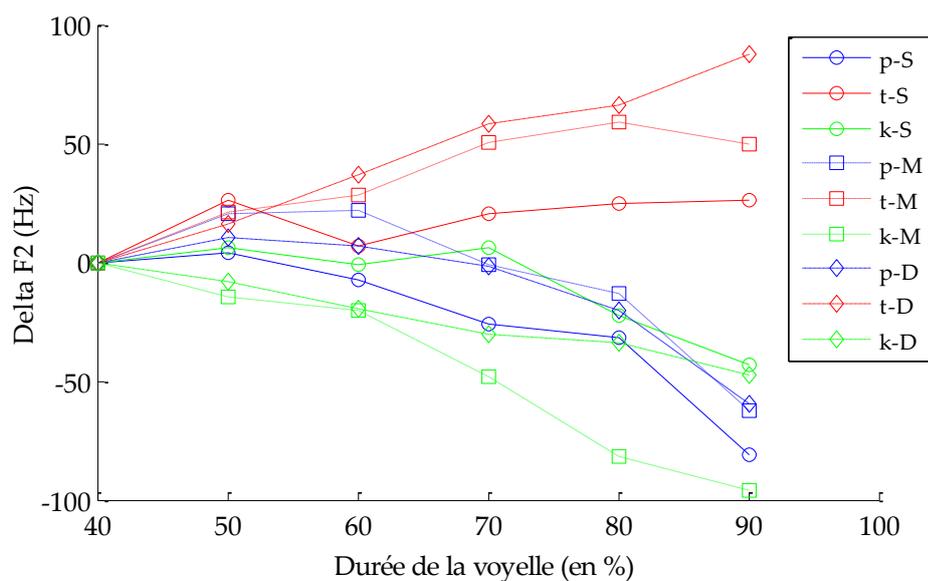


Figure 4.10. Evolution de $\Delta F2$ estimée de 40 % jusqu'à 90 % de la durée de la voyelle en fonction de la consonne en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible (S = A1, M = D1, D = D2).

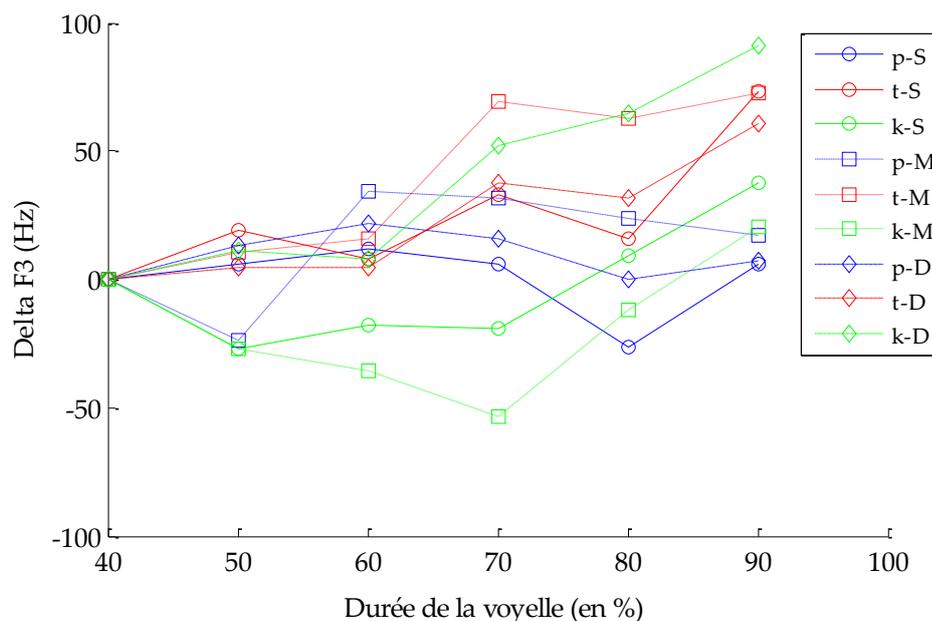


Figure 4.11. Evolution de $\Delta F3$ estimée de 40 % jusqu'à 90 % de la durée de la voyelle en fonction de la consonne en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit immédiatement la consonne cible (S = A1, M = D1, D = D2).

Les résultats des ANOVA à mesures répétées sur les valeurs de $\Delta F1$, $\Delta F2$ et $\Delta F3$ sont présentés dans la table 4.5. Les différences non significatives sont présentées avec leur valeur de p ($p > 0,05$), celles significatives avec le symbole *.

Dès les 50 % de la durée de la voyelle, $\Delta F2$ est déjà significativement différent en fonction du lieu de la coda suivante et le reste jusqu'à la fin de la durée de la voyelle. La figure 4.8 montre qu'à partir des 70 % de la durée de la voyelle, les pentes $F2$ des réalisations en contexte /t/ sont positives, alors que celles en contexte de /p/ et /k/ sont toutes négatives. Dans la transition VC, $\Delta F3$ ne marque pas de différence significative selon le lieu d'articulation de la consonne suivante ($p > 0,05$), très probablement en raison de la variabilité des valeurs observées pour ce formant (cf. figure 4.9). $\Delta F1$ est significativement différent à 60 %, effet peut être dû au hasard de l'échantillonnage.

L'effet du sexe a également été calculé. Que le sujet soit homme ou femme, l'analyse des interactions entre sexe et lieu d'articulation de la coda montre des effets significatifs de $\Delta F2$ et $\Delta F3$ à 80 % (respectivement $p = 0$, $p = 0,039$) mais aussi à 90 % de la durée totale de la voyelle (respectivement $p = 0$, $p = 0,008$). Cela signifie donc que $\Delta F2$ et $\Delta F3$ sont différents selon le lieu d'articulation de la coda, mais que cette différence n'est pas la même pour les hommes que pour les femmes. Les effets du ton de la syllabe qui suit les consonnes cibles en coda sont observables uniquement pour $\Delta F1$ dès 60 % jusqu'aux 90 % de la durée totale de la voyelle ($p < 0,05$) (cf. figure 4.7 et table 4.5). Ce sont les monosyllabes (p-S, t-S, k-S) suivis du ton A1 qui rendent la différence significative.

Table 4.5. Valeur des seuils de significativité des différences de $\Delta F1$, $\Delta F2$, $\Delta F3$ selon les facteurs lieu, et ton de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).

Facteurs	Delta	50 %	60 %	70 %	80 %	90 %
lieux	$\Delta F1$	p = 0,058	*	p = 0,231	p = 0,155	p = 0,459
	$\Delta F2$	**	****	****	****	****
	$\Delta F3$	p = 0,176	p = 0,386	p = 0,151	p = 0,486	p = 0,281
tons	$\Delta F1$	p = 0,075	***	****	***	*
	$\Delta F2$	p = 0,429	p = 0,541	p = 0,554	p = 0,286	p = 0,122
	$\Delta F3$	p = 0,370	p = 0,718	p = 0,142	p = 0,272	p = 0,668

Entre les deux types de syllabe 1 des composés lexicaux (suivis des deux tons différents M et D), aucun effet significatif de $\Delta F1$ n'est détecté selon les tons de S2 ($p = 0,64$). Que le sujet soit homme ou femme, aucune interaction n'est détectée entre les deux facteurs *ton* et *sexe*.

Pour observer les effets du lieu d'articulation des consonnes nasales sur les transitions formantiques de la voyelle vers une coda nasale, la mesure des formants au centre de la voyelle (à 50 % de sa durée) et à 90 % de la durée de la voyelle suivie par des nasales a été effectuée. Compte tenu de leurs caractéristiques acoustiques spécifiques (Rossato, 2000 ; Vaissière, 2007), notamment concernant la présence de formants de nasalité, des ANOVA entre des transitions VC (consonne orale) et des transitions VN (consonne nasale) de même lieu d'articulation ont été effectués. L'objectif était de vérifier l'absence de formant de nasalité dans les valeurs relevées des 3 premiers formants.

Des comparaisons entre les valeurs des $\Delta F1$, $\Delta F2$ et $\Delta F3$ des trois groupes /ap/ vs. /am/, /at/ vs. /an/, /ak/ vs. /aŋ/ ont révélé des différences significatives entre [k] et [ŋ], [p] et [m] (cf. table 4.6). Ce résultat indique que des formants de nasalité figurent sans doute dans les données. Par conséquent, pour les nasales, la même démarche que pour les plosives n'a pas été retenue afin de pouvoir extraire au niveau spectral de manière fiable des informations permettant une classification des réalisations de la voyelle /a/ en fonction du lieu d'articulation des consonnes qui la suivent.

Table 4.6. Comparaison des deltas F1, F2, F3 entre consonne orale et nasale (* différence significative ($p < 0,05$)).

	Delta F1	Delta F2	Delta F3
[p] vs. [m]	p = 0,464	*	p = 0,451
[t] vs. [n]	p = 0,684	p = 0,067	p = 0,161
[k] vs. [ŋ]	*	*	p = 0,62

Une analyse complémentaire basée sur le calcul de douze coefficients MFCC (*Mel-Frequency Cepstrum Coefficients*) et leurs dérivées sur une fenêtre glissante a permis de confirmer l'existence des particularités spectrales dans la partie de transition située à 90 % de la durée de la voyelle. Une analyse en composantes principales (PCA) a été effectuée à partir de ces données.

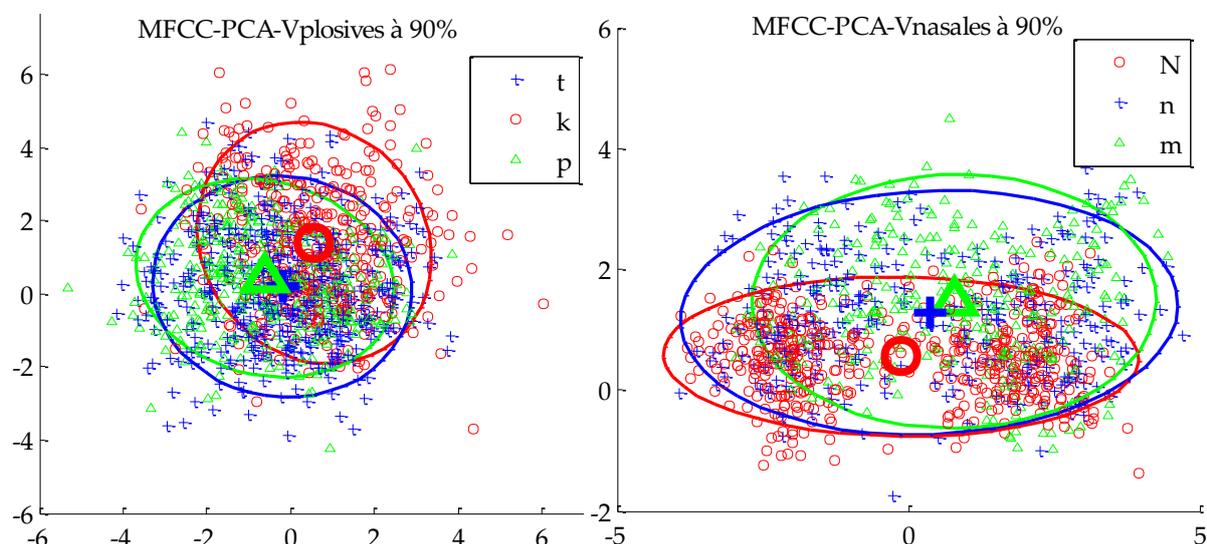


Figure 4.12. Analyse en composantes principales des coefficients MFCC de la transition à 90 % de la durée totale de /a/ en contextes des plosives /t k p/ (à gauche) et des nasales /ŋ n m/ où N = /ŋ/ (à droite).

Cette analyse montre qu'à 90 % de sa durée, la voyelle contient des informations du lieu d'articulation de la consonne post-vocalique : les zones de dispersion sont plus à l'écart pour les vélaires (/k/ et /ŋ/) ; les zones des coronales (/t/ et /n/) sont situées entre celles des vélaires et des bilabiales (cf. figure 4.12). Il y a ici, dans les MFCC, des informations qui séparent les bilabiales des vélaires, ce qui n'était pas visible dans l'examen des $\Delta F2$ (cf. figure 4.10).

Toujours dans le but de révéler les caractéristiques de la consonne en coda, d'autres indices acoustiques dans cette partie de transition VC ont été examinés.

4.1.2.4.2. Transitions de l'intensité (I)

L'intensité a également été mesurée à des trames différentes de la durée totale de la voyelle (de 50 % à 90 %), et comparée ensuite avec les mesures prises à 40 %. Nous obtenons ainsi les valeurs de ΔI dans la seconde partie de la voyelle. Des ANOVA à mesures répétées ont été effectuées pour observer les effets de l'intensité en intra-sujet.

Les valeurs moyennes et les écart-types des pentes de l'intensité de 50 % à 90 % de la durée de la voyelle sont représentées dans l'annexe III. Ces valeurs sont calculées en fonction du type de consonne post vocalique (/p/, /t/, /k/) et du ton de la syllabe qui suit la consonne cible en coda. L'évolution des pentes de l'intensité dans cette partie finale de la voyelle est présentée figure 4.13.

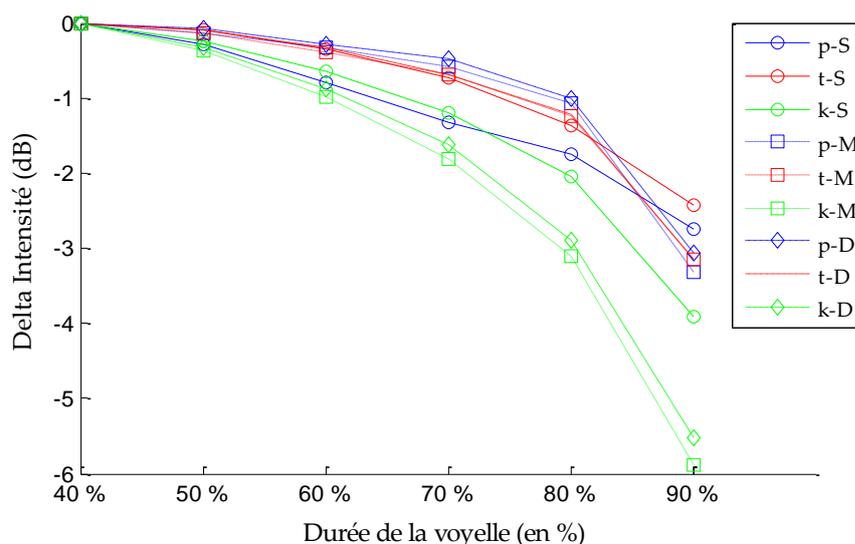


Figure 4.13. Evolution de l'intensité estimée de 40 % jusqu'à 90 % de la durée totale de la voyelle en fonction de la plosive en coda (/p t k/) et du ton de la syllabe qui suit (S = A1, M = D1, D = D2).

A 90 %, des différences de ΔI existent en fonction du ton de la syllabe qui suit la consonne cible [$F(2,18) = 7,196$; $p = 0,005$]. L'intensité d'un mot simple (suivi par un ton neutre A1) est significativement différente d'une syllabe d'un mot composé suivi par un ton montant D1 ($p = 0,018$) ou suivi par un ton descendant D2 ($p = 0,02$). Aucune différence significative entre les deux types de S1 de mots composés qui sont suivies par des tons différents (montant D1 et descendant D2) n'a été détectée ($p = 0,311$). L'effet du ton n'est pas observé à 50 % et jusqu'à 80 %. Les résultats statistiques sur les valeurs de delta intensité figurent dans la table 4.7.

Table 4.7. Valeur des seuils de significativité des différences de ΔI selon les facteurs lieu, ton de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).

Delta I	50 %	60 %	70 %	80 %	90 %
Lieux	**	****	****	****	****
Tons	$p = 0,769$	$p = 0,765$	$p = 0,788$	$p = 0,943$	***

A 90 % de la durée totale de la voyelle, la chute de l'intensité est significativement différente selon le lieu d'articulation de la consonne qui suit [$F(2,18) = 29,344$; $p = 0$]. La chute est plus importante pour les vélaires que pour les bilabiales ($p = 0,001$) et les coronales ($p = 0$) (cf. figure 4.13). La différence d'intensité entre les bilabiales et les coronales n'est pas significative ($p = 0,543$). Le même effet significatif de la chute de l'intensité est observé à 80 %, 70 %, 60 % en fonction du lieu d'articulation de la consonne suivante [$F(2,18) = 28,95$; $p = 0$].

Les vélares ont toujours les pentes négatives les plus importantes par rapport aux bilabiales ($p = 0$) et coronales ($p = 0$). A 50 %, la différence est globalement observée selon le lieu d'articulation ($p = 0,019$), mais la différence calculée entre les réalisations précédées de /p/ et /k/ n'est pas significative ($p = 0,061$). Concernant les nasales, aucune différence significative de l'intensité selon le lieu n'a été observée.

4.1.2.4.3. Transitions de la fréquence fondamentale (F0)

La même démarche pour les $\Delta F0$ a été suivie. Les valeurs moyennes et les écart-types des pentes de F0 de 50 % à 90 % de la durée totale de la voyelle sont présentées dans l'annexe IV. Ces valeurs sont calculées en fonction de la consonne en coda (/p/, /t/, /k/) et du ton de la syllabe qui suit la consonne cible. L'évolution des pentes de F0 dans cette partie finale de la voyelle est présentée figure 4.14.

Les résultats statistiques sur les valeurs de $\Delta F0$ figurent dans la table 4.8. Les transitions analysées sont toutes sous le ton montant *sắc* D1. On remarque que la tendance tonale est montante vers la partie finale de la syllabe dans tous les contextes, d'où les valeurs positives de $\Delta F0$. Des différences significatives de $\Delta F0$ sont observées dès 70 % jusqu'à 90 % de la durée totale de la voyelle ($p = 0,006$; $p = 0,002$; $p = 0$ respectivement) selon le lieu d'articulation de la consonne post-vocalique. La montée de F0 à 90 % de la durée de la voyelle est en général plus importante suivie d'une vélaire que suivie d'une bilabiale ($p = 0,034$) ou d'une coronale ($p = 0,009$), tout comme à 80 % ($p = 0$) et 70 % ($p = 0$).

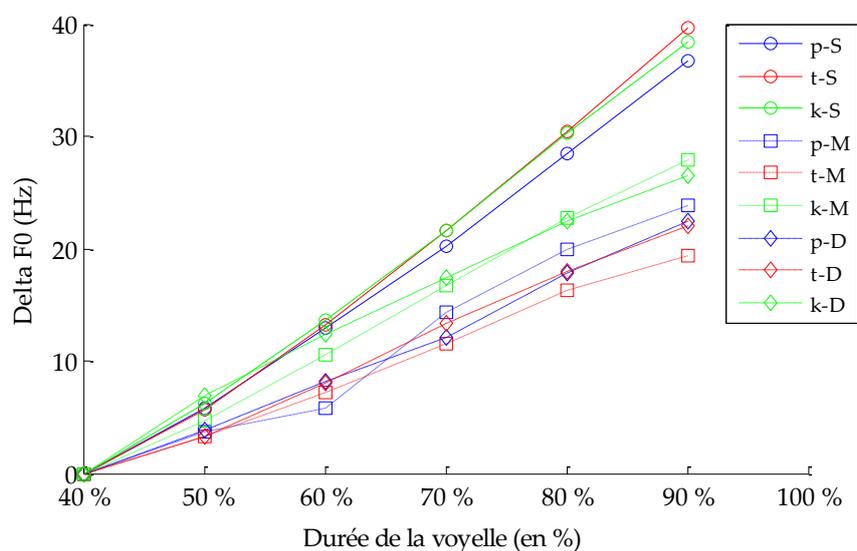


Figure 4.14. Evolution de $\Delta F0$ estimée de 40 % jusqu'à 90 % de la durée de la voyelle en fonction de la plosive en coda (/p/, /t/, /k/) et du ton de la syllabe qui suit (S = A1, M = D1, D = D2).

L'effet du ton de la syllabe suivante sur $\Delta F0$ est observé dès les 60 % de la durée de la voyelle (cf. table 4.8). La figure 4.14 montre que la montée de F0 est significativement plus importante à partir de 60 % de la voyelle lors qu'il s'agit d'un mot simple CVC₂ suivi d'un ton neutre A1 (p-S, t-S, k-S) que de première syllabe d'un mot composé CVC₂.C₃VC (suivie soit par un ton montant D1, soit par un ton descendant D2). Aucune différence significative de $\Delta F0$ entre les deux types de S1 de mots composés ($p = 0,923$). Rappelons que ces deux types de syllabe sont suivis par une syllabe dont le ton appartient à deux registres totalement différents, soit haut soit bas (cf. section 4.1.1.1).

Table 4.8. Valeur des seuils de significativité des différences de $\Delta F0$ selon les facteurs « lieux », « tons » de 50 % à 90 % de la durée totale de la voyelle qui précède la consonne cible (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$).

Delta F0	50%	60%	70%	80%	90%
Lieux	$p = 0,061$	$p = 0,379$	****	***	**
Tons	$p = 0,107$	**	****	****	****

4.1.3. Discussion

Cette expérience multi-locuteur confirme qu'en vietnamien, les nasales sont plus longues en position initiale qu'en position finale d'un item lexical, que celui-ci soit monosyllabique ou dissyllabique, et que les plosives en finale d'un monosyllabe sont plus longues qu'en finale de syllabe devant frontière intra-mot. Ces résultats sont cohérents avec de nombreuses études antérieures qui ont montré dans différentes langues que la prononciation d'un phonème est influencée par la position qu'il occupe dans la syllabe (Lindblom, 1983 ; Keating, 1983 ; Krakow, 1999), dans le mot (Keating, Wringt & Zang, 1999) ou dans le domaine prosodique (Fougeron & Keating, 1997).

L'aspect novateur de cette étude est d'avoir montré l'influence du type de frontière syllabique sur la réalisation des consonnes de coda, et même de la voyelle. En effet, plusieurs résultats pertinents concernant le paramètre de durée (durée de la consonne en coda, durée de l'occlusion, durée de la voyelle précédant la consonne) montrent l'existence de particularités en fonction du type de frontière syllabique et supposent des degrés différents de coarticulation. Une consonne est plus longue en finale de mot (C₂) qu'en finale de syllabe 1 à l'intérieur d'un mot composé (C₂.C₃). Les plosives en C₂ ont une phase de fermeture plus longue que celles en C₂.C₃. La réalisation de la voyelle /a/ est significativement plus longue quand elle précède une consonne finale de mot (C₂) qu'une consonne finale de syllabe 1 de mot composé (C₂.C₃). Ce résultat est valable pour les plosives comme pour les nasales.

L'analyse de la partie transition entre la voyelle et la consonne finale a également mis en évidence l'influence du type de frontière syllabique qui suit immédiatement la consonne.

La trajectoire du premier formant F1 observée entre 40 % et 60 % à 90 % de la durée de la voyelle est significativement différente entre C₂ finale de mot monosyllabe et C₂ finale S1 de mot dissyllabique. La chute de l'intensité à 90 % de la durée de la voyelle des monosyllabes est moins importante lorsque C₂ est devant frontière de mot que devant une frontière syllabique intra-mot. La trajectoire de F0 des monosyllabes est plus positive que celles des S1 de mot composé dès 60 % de la durée de la voyelle. Ces résultats montrent clairement que la transition entre voyelle et consonne se comporte différemment selon que cette consonne se trouve à la frontière de mot ou à la frontière syllabique d'un mot composé.

Concernant le VOT des plosives, il ne varie pas de manière significative en fonction de la position de la consonne dans la syllabe, ni du type de frontière syllabique mais dépend du lieu d'articulation des consonnes. Le VOT moyen de la consonne vélaire /k/ est plus élevé que le VOT des autres plosives de lieux d'articulation bilabial et coronal, et ce quelle que soit la position dans la syllabe, le mot. La comparaison de nos résultats avec ceux de Serniclaes (1987) pour les plosives du français montre que le VOT moyen des consonnes sourdes /p t k/ du vietnamien (/k/ possède la plus grande valeur de VOT en initiale : 23,94 ms) est 1,5 fois plus petit qu'en français (30 et 35 ms). La variabilité des VOT positifs en vietnamien (de 12 à 24 ms) est plus faible qu'en français où la gamme des valeurs s'étend de 10 à 70 ms (Serniclaes, 1987, p. 125).

Aucune différence significative de ΔF_0 n'est observée entre les deux types de S1 de mots composés. Or ces deux types de syllabe sont suivis par une syllabe dont le ton qui appartient à deux registres totalement différents, soit haut soit bas (cf. section 4.1.1.1). Ces résultats confirment les études antérieures sur les effets tonals dans un mot composé dissyllabique. En effet, Han et Kim (1974) ont attesté que l'effet du ton de S1 est en général plus important que celui du ton de S2 et que la variation de F0 du ton sera plus importante dans la seconde syllabe que la première. Brunelle (2003) le reconferme en notant que la coarticulation tonale se fait plutôt dans le sens progressif que régressif. Cela veut dire que même suivies par des tons différents, les syllabes 1 étudiées ne présentent pas de différence significative au niveau de la fréquence fondamentale. Une autre raison qui pourrait expliquer ce résultat, selon nous, est la contrainte physiologique existant dans les mots composés où les premières syllabes sont terminées par des plosives non voisées et suivies tout de suite par des voisées. Par conséquent, même si l'occlusion n'est pas suivie dans tous les cas d'un relâchement audible, les plis vocaux ont pu avoir le temps de se réajuster pour permettre la réalisation d'un autre ton.

La comparaison entre consonnes initiales (C₁) *vs.* finales (C₂ et C₂.C₃) a également montré une tendance significative à un burst plus faible et plus court en finale de syllabe. Les consonnes en position initiale présentent de manière générale plus d'énergie dans le bruit d'explosion qu'en position finale du mot ou de syllabe 1 de composé lexical, la consonne

vélaire /k/ possédant le burst le plus intense. L'intensité du burst des consonnes en position finale de mot ou de syllabe 1 (C₂ et C₂.C₃) ne présente pas de différences significatives.

Les plosives en coda sont traditionnellement décrites comme non relâchées (Đoàn Thiên Thuật, 1999 ; Cao Xuân Hạo, 1985). Nos résultats montrent que toutes les consonnes en position finale ne sont pas des non relâchées. En effet, 27 % des consonnes sont réalisées avec burst dans cette position, avec une durée plus brève et une intensité plus faible du bruit d'explosion que dans les réalisations en initiale. Les consonnes perdent davantage leur burst devant frontière syllabique intra-mot (C₂.C₃) que devant frontière de mot (C₂) et jamais en C₁. La différence considérable entre consonne non relâchée en coda et consonne explosive en attaque en vietnamien consiste « en premier lieu dans la direction des deux mouvements articulatoires et de leurs corrélats acoustiques, qui sont inverses. Ensuite, le bruit caractéristique d'explosion par lequel on reconnaît la consonne en position prévoicallique fait souvent complètement défaut dans les consonnes finales, car la phase de relâchement est rarement obligatoire par ces dernières », atteste Cao Xuân Hạo (1985, p. 66). D'autres linguistes comme Hoàng Tuệ et Hoàng Minh (1975) ont mis en évidence la différence entre les deux catégories de consonnes plosive *vs.* occlusive non relâchée par les paires minimales de constructions bi-morphémiques et bisyllabiques, telles que le mot « phát hành » [fat⁵.hẽŋ²] *diffuser* et « phá thành » [fa⁵.t^hẽŋ²] *démolir le rempart* (cf. figure 4.15 et 4.16). On remarque que [t^h] final dans « phát hành » n'est pas relâché. À l'inverse, on remarque nettement le bruit d'explosion de [t^h] initial et l'incurvation des formants qui fait défaut dans la voyelle [a] du groupe « phá thành ».

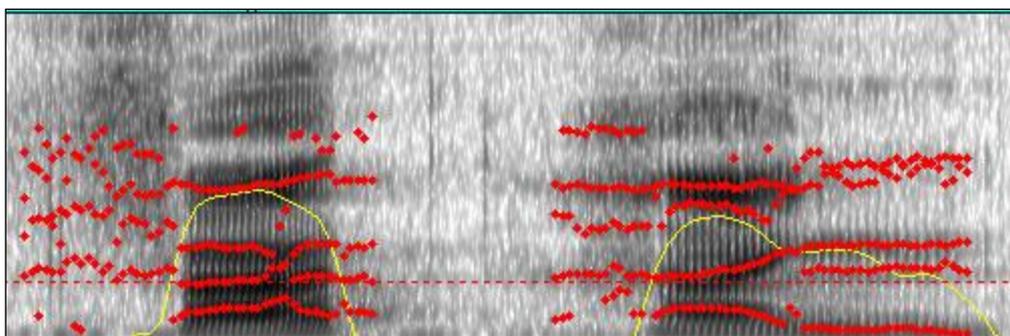


Figure 4.15. Spectrogramme du mot « phát hành » [fat⁵.hẽŋ²] (*diffuser*).

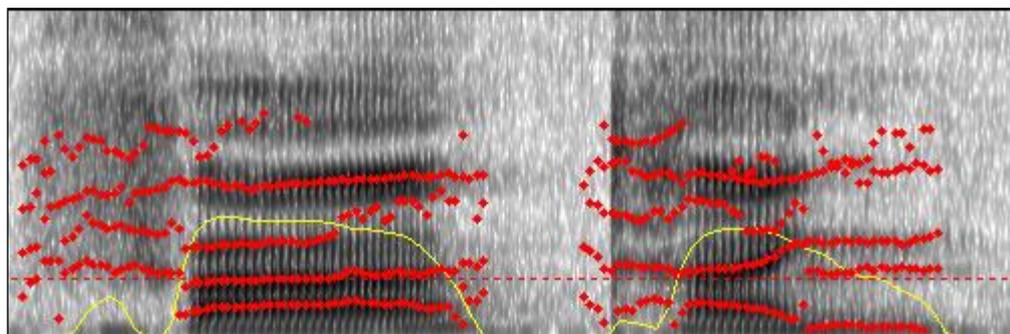


Figure 4.16. Spectrogramme du mot « phá thành » [fa⁵.t^hẽŋ²] (*démolir le rempart*).

L'analyse des trajectoires des formants, de F0 et de l'intensité menées dans la partie de transition entre voyelle et consonnes de coda montre que la structure acoustique de la voyelle contient des informations sur le lieu d'articulation de la consonne en coda, dans la plupart des cas, non relâchée. En effet, à 70 % de la durée totale de la voyelle, les valeurs moyennes de F2 et F3 sont significativement différentes selon le lieu d'articulation de la consonne finale. A 80 % et 90 %, celles de F2 sont significatives, mais ce n'est plus le cas pour celles de F3. Quant à l'évolution des trois formants depuis le centre de la voyelle, seules les pentes de F2 sont différentes de manière significative selon le lieu d'articulation de la consonne suivante. On n'obtient pourtant pas l'allure montante (perceptivement discriminante) dans la partie de transition VC pour la consonne vélaire /k/ comme proposé par Delattre dans ses travaux de 1955 et 1963 sur la parole synthétique. Les informations spectrales obtenues avec MFCC à 90 % de la voyelle montrent par contre les zones acoustiques de dispersion plus à l'écart des zones de réalisation des labiales et des coronales pour les vélares.

Les résultats obtenus montrent également que $\Delta F2$, ΔI et $\Delta F0$ sont différents de manière significative selon le lieu d'articulation de la plosive post-vocalique. Concrètement, la chute de l'intensité et la montée de F0 montrent des effets significatifs en fonction du lieu d'articulation de la consonne en coda. Les vélares obtiennent toujours des tendances (chute de l'intensité et montée de F0) plus importantes que les bilabiales et les coronales.

Ces résultats sont conformes et précisent le propos de Cao Xuân Hạo (1985) selon lequel l'identification des lieux d'articulation des plosives finales non relâchées est possible grâce à des informations récupérables dans la partie de transition avec la voyelle précédente. Notre étude acoustique montre qu'il s'agit des pentes de F2, de l'intensité et de F0.

Rappelons que l'objectif de cette étude acoustique est de comprendre les différences de réalisations des plosives /p t k/ et des nasales /m n ŋ/ qui peuvent exister en fonction de leur distribution. La différence de réalisation en fonction de leur distribution pourrait être des éléments du crible phonologique. Dans la littérature, peu d'études expérimentales prenant en compte la frontière syllabique ont été menées sur les données du vietnamien. Les résultats livrés récemment par des études sur le traitement automatique de la langue concernent principalement le niveau suprasegmental du vietnamien (Nguyễn Quốc Cường, 2002 ; Nguyễn Tiến Dũng, 2004 ; Trần Đỗ Đạt, 2005 ; Lê Việt Bắc, 2006 ; Vũ Minh Quang, 2007). L'ensemble des résultats présentés dans cette étude acoustique constitue une étape préliminaire à la constitution d'un paradigme expérimental rendant compte de la perception des séquences de consonnes que l'on peut trouver en vietnamien et en français, à la différence près de la présence d'une frontière syllabique ou frontière de mot.

La section 4.2 suivante de ce chapitre consiste à mettre en lumière les corrélats perceptifs des consonnes qui apparaissent dans des séquences, et notamment lorsqu'elles sont articulatoirement non relâchées.

4.2. Expérience 2 : Identification des consonnes finales du vietnamien

Un des résultats importants obtenus avec les analyses acoustiques montre que les consonnes finales du vietnamien comportent des caractéristiques différentes en fonction du type de frontière syllabique qu'elles précèdent : d'une part, les plosives et les nasales sont significativement plus longues devant frontière de mot que devant frontière de syllabe à l'intérieur d'un mot composé lexical ; d'autre part, les plosives sont plus souvent non relâchées en finale à l'intérieur de mot qu'en finale de mot. Les bursts en finale, si relâchée, sont significativement plus brefs et de plus faible amplitude par rapport à leur réalisation en initiale.

Dans le prolongement de ce travail, l'étude que nous présentons dans cette section 4.2 a pour objectif d'identifier les corrélats acoustico-perceptifs des consonnes finales du vietnamien en lien avec un éventuel effet de l'influence du type de frontière syllabique sur la perception des consonnes. En d'autres termes, il s'agit de répondre aux questions suivantes : quels indices permettent d'identifier le lieu d'articulation d'une consonne du vietnamien lorsque celle-ci est non relâchée ? Existe-il des différences dans la perception des consonnes en fonction de leur position dans la syllabe ou dans le mot ? C'est-à-dire des différences de perception lorsque la consonne fait partie ou non d'une séquence consonantique intra-mot (dissyllabe) ou inter-mot (monosyllabe).

4.2.1. Méthodologie

4.2.1.1. Stimuli

Un ensemble de 55 items sonores naturels de type CVC correspondant soit à un mot monosyllabique, soit à la première syllabe d'un mot composé dissyllabique ont été sélectionnés pour l'étude (cf. table 4.11). Ils comportent l'une des 6 occlusives / p t k m n ŋ / en contexte de la voyelle la plus ouverte /a/ et de même contexte tonal (ton montant D1). Plus concrètement, dans cet ensemble d'items lexicaux, ces six consonnes se trouvent :

- Soit en position finale C₂ de mots monosyllabiques [C₁aC₂] (exemples : [kak] « *các* » *marque pluriel* ; [maŋ] « *máng* » *mangeoire*) ;

- Soit en position finale C₂ de première syllabe [C₁aC₂] de mots composés lexicaux dissyllabiques [C₁aC₂.C₃V(C₄)] (exemples : [mat] dans [mat.za] « mát dạ » *content* ; [saŋ] dans [saŋ.taw] « sáng tạo » *créer*)

Ces stimuli sont extraits des phrases porteuses du corpus acoustique (cf. section 4.1), lu par un sujet natif masculin, âgé de 27 ans, originaire du Nord du Vietnam. Ce sujet a été choisi parmi les dix participants en raison de son articulation plus distincte et son timbre de voix plus élevé. Les stimuli ont été choisis de manière à ce que leur environnement dans la phrase porteuse soit le plus neutre possible du point de vue articulatoire afin de faciliter l'extraction de la syllabe : toute consonne sourde en coda est suivie d'une consonne sonore, et inversement. Au total, 29 items auditifs CVC terminés par une plosive et 26 items terminés par une nasale ont pu être retenus pour le test.

Table 4.9. Liste des 55 stimuli du vietnamien pour le test de perception des consonnes finales tirés du corpus acoustique (cf. section 4.1).

N°	Stimuli	API	Mot d'origine	N°	Stimuli	API	Mot d'origine
1	áp	ap	áp bức	30	bám	bam	bám trụ
2	áp	ap	áp dụng	31	đám	dam	đám cưới
3	áp	ap	áp lực	32	cám	kam	cám
4	pháp	fap	pháp lệnh	33	khám	xam	khám phá
5	pháp	fap	pháp lý	34	nám	nam	nám
6	cáp	kap	cáp	35	tám	tam	tám
7	ngáp	ŋap	ngáp	36	xám	sam	xám xịt
8	táp	tap	táp	37	bán	ban	bán kết
9	bát	bat	bát ngát	38	cán	kan	cán
10	phát	fat	phát động	39	cán	kan	cán sự
11	phát	fat	phát giác	40	mán	man	mán
12	hát	hat	hát ví	41	ngán	ŋan	ngán
13	cát	kat	cát	42	tán	tan	tán
14	khát	xat	khát nước	43	thán	t ^h an	thán phục
15	khát	xat	khát vọng	44	gián	zan	gián tiếp
16	mát	mat	mát	45	đáng	daŋ	đáng kính
17	mát	mat	mát dạ	46	đáng	daŋ	đáng sợ
18	nát	nat	nát	47	cáng	kaŋ	cáng
19	ngát	ŋat	ngát	48	kháng	xaŋ	kháng cự
20	tát	tat	tát	49	máng	maŋ	máng
21	ác	ak	ác liệt	50	ngáng	ŋaŋ	ngáng
22	gác	ɣak	gác bút	51	sáng	saŋ	sáng kiến

23	<i>các</i>	kak	<i>các</i>	52	<i>sáng</i>	saŋ	<i>sáng tạo</i>
24	<i>lác</i>	lak	<i>lác đác</i>	53	<i>táng</i>	taŋ	<i>táng</i>
25	<i>mác</i>	mak	<i>mác</i>	54	<i>tháng</i>	t^haŋ	<i>tháng chạp</i>
26	<i>ngác</i>	ŋak	<i>ngác</i>	55	<i>tháng</i>	t^haŋ	<i>tháng tám</i>
27	<i>xác</i>	sak	<i>xác đàng</i>				
28	<i>tác</i>	tak	<i>tác</i>				
29	<i>tác</i>	tak	<i>tác dụng</i>				

4.2.1.2. Déroulement du test

Vingt sujets (10 hommes et 10 femmes) originaires du Nord du Vietnam, âgés de 24 à 29 ans (moyenne d'âge de 24 ans pour les femmes, et de 25 ans pour les hommes), ont participé au test. Ils sont en France pour des études supérieures depuis 1 à 4 ans. Ils continuent à parler vietnamien tous les jours. Aucun d'entre eux ne possédait de formation universitaire en phonétique-phonologie. La plupart avaient eu connaissance de notions non approfondies de phonétique lors de leur apprentissage en français langue étrangère au Vietnam. Ceux qui avaient des problèmes de vue devaient venir avec leurs verres correcteurs.

Nous avons choisi de tester les plosives à part des nasales en raison de structures acoustiques très différentes. Le programme utilisé pour le test est *PerceptA*⁵³ de R. Carré et E. Samain. Les sujets avaient pour consigne d'écouter au casque, puis d'identifier la consonne finale du stimulus présenté parmi trois solutions proposées (choix forcé) correspondant aux trois plosives /p t k/ pour le test concernant les plosives, et aux trois nasales /m n ŋ/ pour le test portant sur les nasales, en cliquant la réponse à l'écran avec l'aide de la souris (cf. figures 4.17 et 4.18). La consigne était donnée au sujet de répondre le plus rapidement possible même si l'interface fournit la possibilité de réécouter les stimuli.

⁵³ Le programme *PerceptA*, utilisé pour les tests d'identification de signaux de parole (.wav), est distribué gratuitement sur le site personnel de René Carré, accessible à l'adresse suivante <http://ren.carre.pagesperso-orange.fr/Programmes.html>.

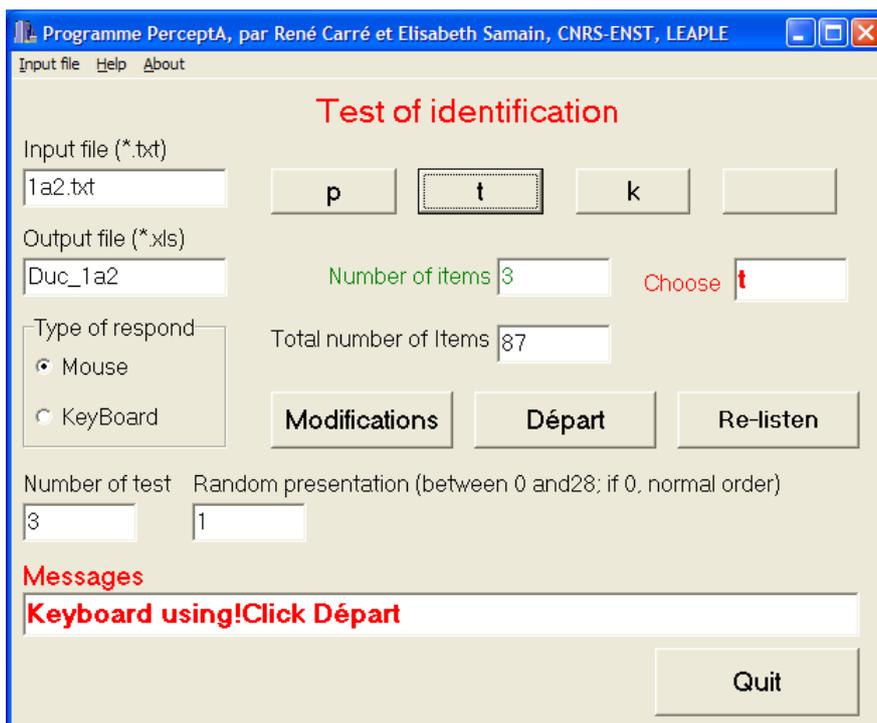


Figure 4.17. Interface du test perceptif utilisé pour les plosives.

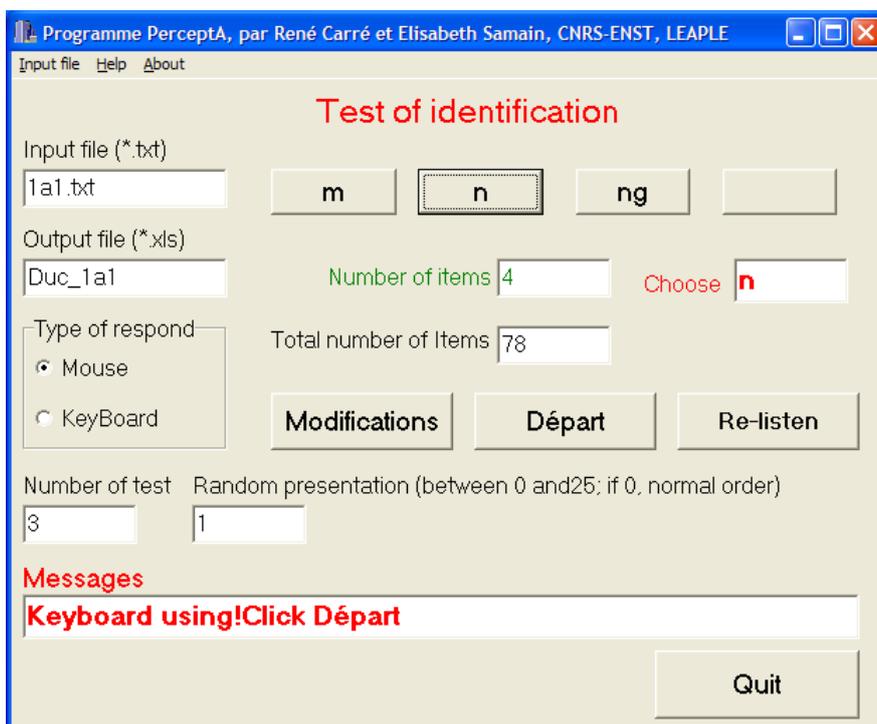


Figure 4.18. Interface du test perceptif utilisé pour les nasales.

Une phase d'entraînement à l'interface et au type de stimuli leur avait préalablement proposée, avec des stimuli non-présents dans le corpus du test lui-même. Les deux tests ont été passés à la suite. Chaque test a duré environ cinq minutes et une petite pause d'une

minute a eu lieu entre les deux. Le premier test concernait les plosives et était constitué de 3 répétitions des 29 items présentés dans un ordre aléatoire pour chaque sujet.

Les sujets n'étant pas familiers des symboles de l'alphabet phonétique international, ceux-ci leur ont été expliqués par l'expérimentateur avant le test. L'interface du programme ne permettant pas d'afficher le caractère spécifique de la nasale vélaire /ŋ/, celle-ci apparaissait sous sa forme orthographique « ng » (cf. figure 4.18).

Le test s'est déroulé dans une salle calme du Département Parole et Cognition de GIPSA-lab en utilisant un casque binaural de haute qualité, Sennheiser HD-25-13.

Le second test comportait 3 répétitions des 26 items sélectionnés pour les nasales présentées également de manière aléatoire.

4.2.2. Résultats

Les données ont été analysées avec le logiciel SPSS[®]. Des tests d'analyses de variance (ANOVA) ont été effectués.

Les analyses montrent que pour les plosives en coda, le meilleur score d'identification est obtenu par /t/ et /k/ (cf. table 4.10). La différence de score de bonnes réponses est significative entre /p/ et /t/ ($p = 0$) ; entre /p/ et /k/ ($p = 0$), mais pas entre /t/ et /k/ ($p = 0,79$).

Pour les nasales, la vélaire /ŋ/ obtient le meilleur score d'identification par rapport à /m/ et /n/ (cf. table 4.11). La différence de score de bonnes réponses est significative entre /m/ et /ŋ/ ($p = 0$) ainsi que entre /n/ et /ŋ/ ($p = 0$), mais pas entre /m/ et /n/ ($p = 0,34$).

Table 4.10. Scores d'identification des plosives.

Stimuli \ Choix	Choix		
	p	t	k
p	74 %	19 %	7 %
t	8 %	84 %	9 %
k	11 %	6 %	83 %

Table 4.11. Scores d'identification des nasales.

Stimuli \ Choix	Choix		
	m	n	ŋ
m	85 %	10 %	6 %
n	9 %	86 %	4 %
ŋ	1 %	5 %	95 %

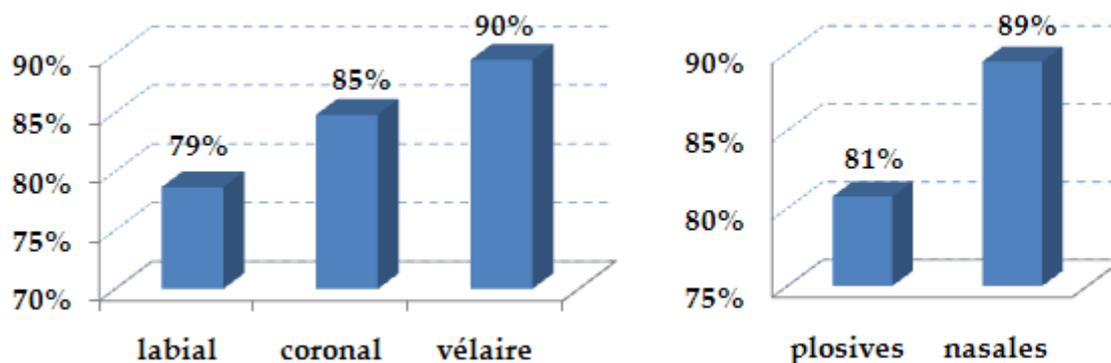


Figure 4.19. Scores d'identification des consonnes finales par lieu et mode d'articulation.

Regroupées par lieu d'articulation, les vélares obtiennent un meilleur score d'identification par rapport aux labiales et coronales (cf. figure 4.19). Par mode d'articulation, les nasales génèrent moins de confusion chez les sujets par rapport aux plosives (89 % contre 81 %).

Concernant l'influence du type de frontière syllabique sur le score de bonnes réponses, la figure 4.20 montre davantage de confusion pour les plosives et les nasales lorsque la coda C_2 précède une frontière intra-mot (devant C_3 de mot composé) et cette différence est significative dans les deux cas (respectivement $p = 0,011$; $p = 0,008$).

Les meilleurs scores d'identification des plosives ainsi que des nasales sont obtenus par les femmes (cf. figure 4.21). Cependant, le test d'égalité des variances des erreurs de Levene confirme l'hypothèse nulle selon laquelle la variance des erreurs de la variable dépendante (scores de bonnes réponses) est égale sur les deux groupes (hommes et femmes), pour les plosives ($p = 0$) comme pour les nasales ($p = 0$). Nous ne pouvons pas, par conséquent, affirmer que la différence de scores est différente significativement en fonction du sexe des sujets pour les plosives quoique la valeur p soit inférieure de 0,05 ($p = 0,003$). La différence de scores entre hommes et femmes n'est pas significative pour les nasales ($p = 0,34$).

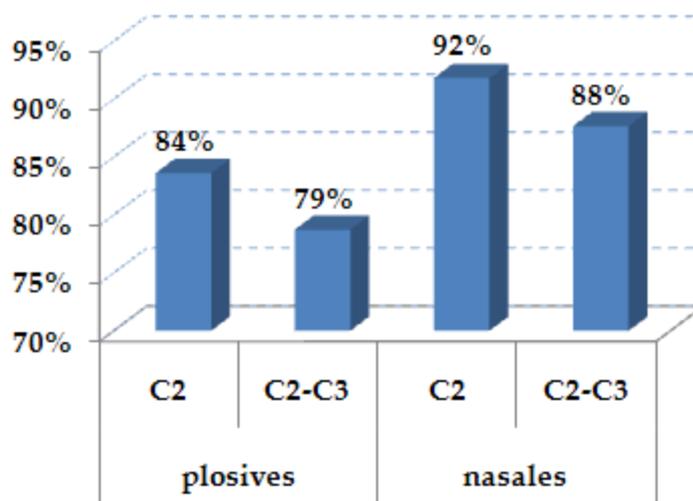


Figure 4.20. Scores d'identification des consonnes finales selon leur position dans la syllabe.

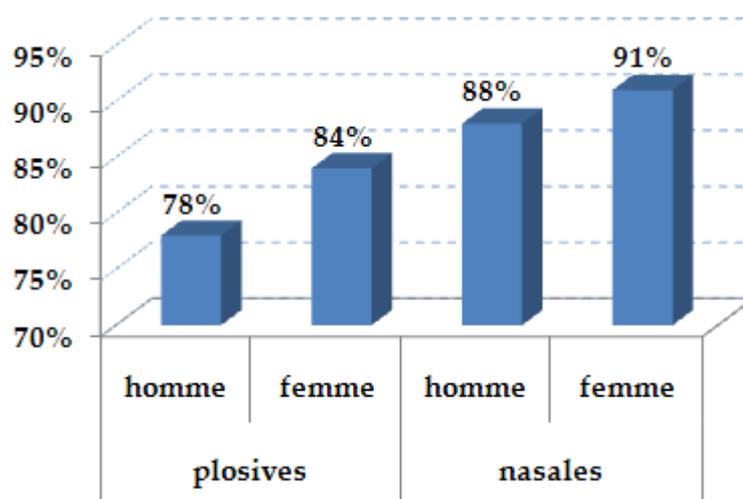


Figure 4.21. Score d'identification des consonnes finales selon le sexe des sujets.

4.2.3. Discussion

Ces résultats montrent que les locuteurs natifs arrivent à identifier les consonnes finales, plus souvent non relâchées, avec des scores bien supérieurs au seuil du hasard. Ce qui nous amène à des questions intéressantes : quels indices acoustiques jouent un rôle d'indicateur ? Quels sont les indices acoustiques parmi ceux mis en évidence dans l'étude acoustique qui précède (cf. section 4.1.2.4 de ce chapitre) qui permettent d'identifier le lieu d'articulation d'une consonne lorsque celle-ci n'est pas relâchée ? Pourquoi les nasales ont été mieux perçues en coda par rapport aux plosives ? Est-ce que ces résultats sont liés à un effet de la fréquence de ces consonnes dans le lexique ?

Pour répondre à ces questions, trois pistes d'explication ont été explorées : la fréquence lexicale, le burst et les transitions formantiques.

4.2.3.1. Fréquence lexicale

A partir du lexique vietnamien de 5 000 entrées intégré à la base de données G-ULSID (cf. chapitre III), il a été observé que, en général, les consonnes /t/, /k/ et /ŋ/ les mieux identifiées par les sujets possèdent les meilleures fréquences d'occurrences dans leur catégorie (plosives et nasales) toutes distributions confondues (cf. figure 4.22).

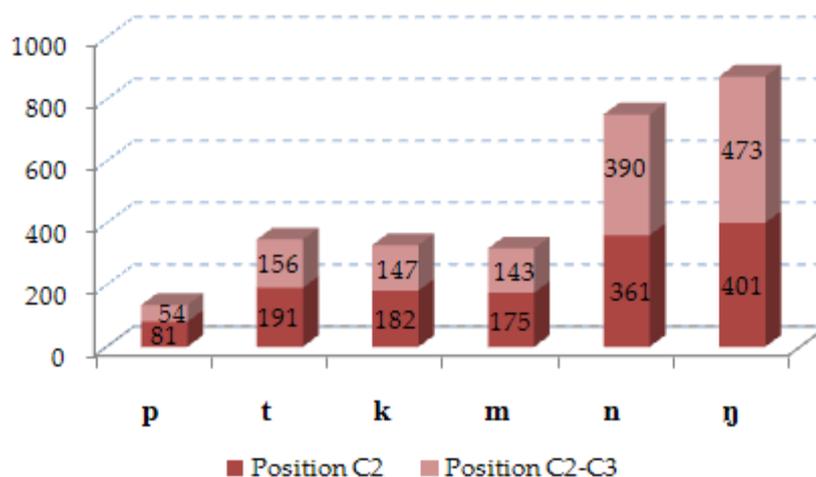


Figure 4.22. Calcul des occurrences des consonnes en coda dans le lexique du vietnamien de 5 000 entrées.

Une tendance entre le score de bonnes réponses et la fréquence d’occurrence des consonnes en position de coda dans le lexique a été effectivement observée. Les coefficients calculés (R^2) indiquent une corrélation importante entre fréquence d’occurrences et score d’identification des plosives comme des nasales, respectivement 0,76 et 0,89 (cf. figure 4.23).

La comparaison des fréquences d’occurrences de ces consonnes entre position initiale *vs.* position finale (cf. figure 4.24) montre que les nasales sont particulièrement présentes en position coda (77 %), soit de mots monosyllabiques C₂, soit de première syllabe de mots composés C₂.C₃, contre 23 % en position initiale. Cette observation pourrait expliquer un meilleur résultat d’identification des nasales par rapport aux plosives en position de coda.

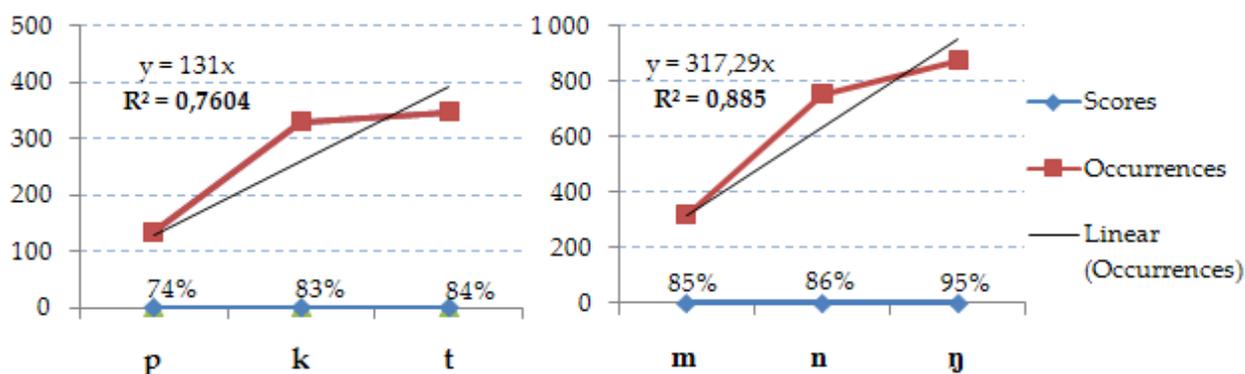


Figure 4.23. Corrélation entre scores de bonnes réponses et nombre d’occurrences des plosives (à gauche) et des nasales (à droite) en coda.

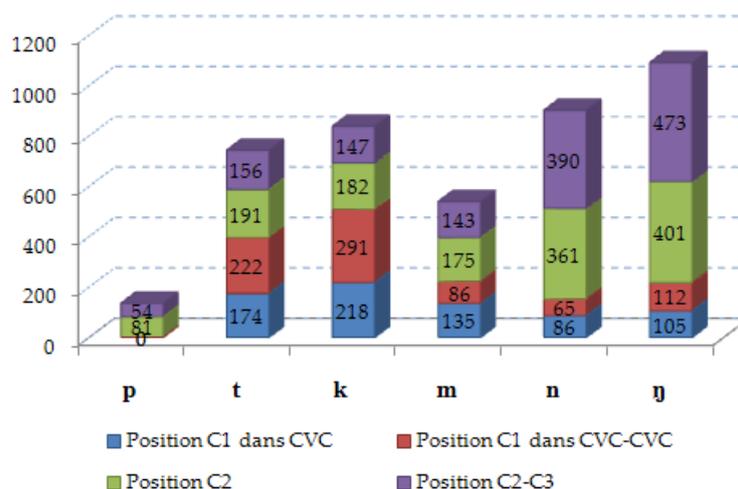


Figure 4.24. Nombre d'occurrences des consonnes en coda en comparaison avec la position d'attaque dans les unités lexicales du vietnamien.

Cependant, les nasales /m/ et /n/ très fréquentes en coda par rapport à la position d'attaque obtiennent des scores d'identification (respectivement 85 % et 86 %) proches de ceux des plosives /t/ et /k/ (respectivement 84 % et 83 %) plus fréquentes en position d'attaque que de coda. Il apparaît donc que la fréquence lexicale n'explique pas tout, même si globalement une certaine corrélation entre le score d'identification et la fréquence dans le lexique a pu être observée.

Le même type d'analyse de fréquence dans le lexique a été effectué pour les consonnes en contexte /a/ et sous le ton montant D1 qui est celui de stimuli du test perceptif. Le résultat montre toujours une occurrence plus forte pour les vélaire qui obtiennent les meilleurs scores d'identification (cf. figure 4.25).

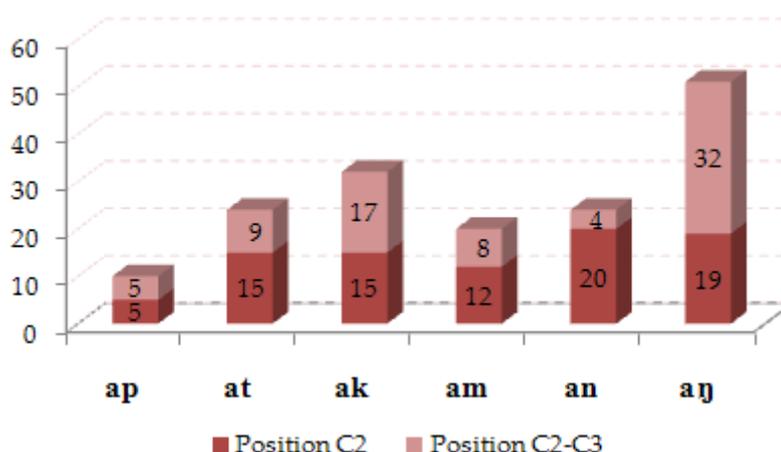


Figure 4.25. Fréquence lexicale des consonnes en coda en contexte de la voyelle /a/ et le ton montant D1.

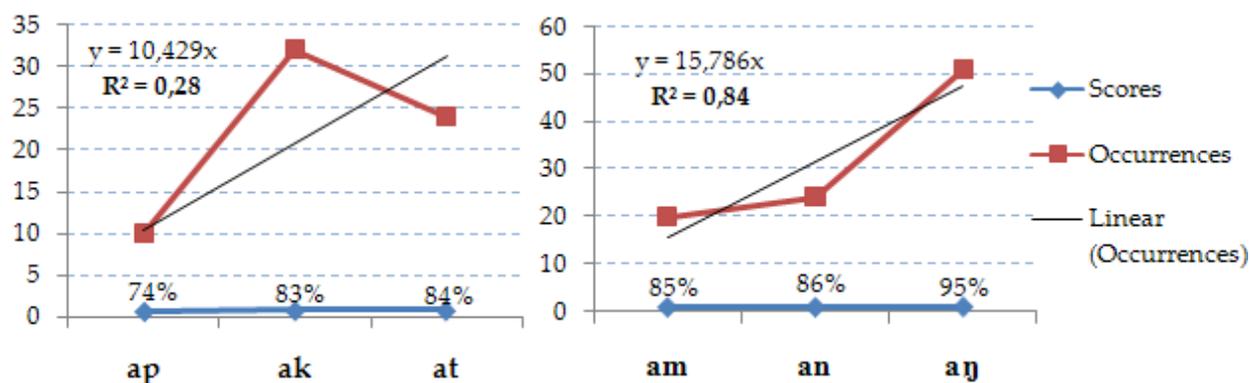


Figure 4.26. Corrélation entre scores de bonnes réponses et fréquences d'occurrences des plosives (à gauche) et des nasales (à droite) en coda, en contexte de la voyelle /a/ et du ton montant D1.

Cependant, la figure 4.26 montre que la corrélation observée de manière générale entre le score de bonnes réponses et la fréquence d'occurrences des consonnes en coda n'est plus vraie pour les plosives en contexte de la voyelle /a/ et du ton montant D1 ($R^2 = 0,28$).

Par conséquent, la fréquence lexicale ne fournit pas toutes les explications nécessaires aux meilleurs scores réalisés par les vélaires et par les nasales. De plus, /p/ toujours en finale a pu biaiser l'opposition /p/ ~ /t/ et orientés les sujets vers le choix /p/ plutôt que /t/ en final. Or /t/ est mieux perçu que /p/. Donc, la fréquence lexicale n'est vraisemblablement pas le facteur qui a guidé les sujets dans leur choix de réponse.

4.2.3.2. Indice acoustique : le burst

Malgré la caractéristique non relâchée des plosives finales en vietnamien, la présence de burst a pu être constatée dans 34,5 % des réalisations (19 des 55 stimuli du test perceptif sont réalisés avec un relâchement de la plosive finale de syllabe). Dans ces cas, les bursts mesurés sont de durée brève (en moyenne 4 ms) et d'intensité plus faible (60 dB en moyenne) comparés à ceux produits en position initiale (10 ms et 68 dB en moyenne).

En considérant la présence *vs.* absence de burst dans la réalisation de la consonne en coda, la figure 4.27 montre des scores significativement meilleurs lorsque le burst est présent et ce quelle que soit la frontière qui suit la consonne (frontière de mot [$F(1,718) = 34,581$; $p = 0$] ou de syllabe intra-mots [$F(1,1018) = 25,66$; $p = 0$]). Cette différence de frontière (de mot ou de syllabe intra-mot) a d'ailleurs moins d'impact sur l'identification des nasales (scores respectifs de 92 % et de 88 %). La tendance est identique lorsque sont considérés les lieux d'articulation.

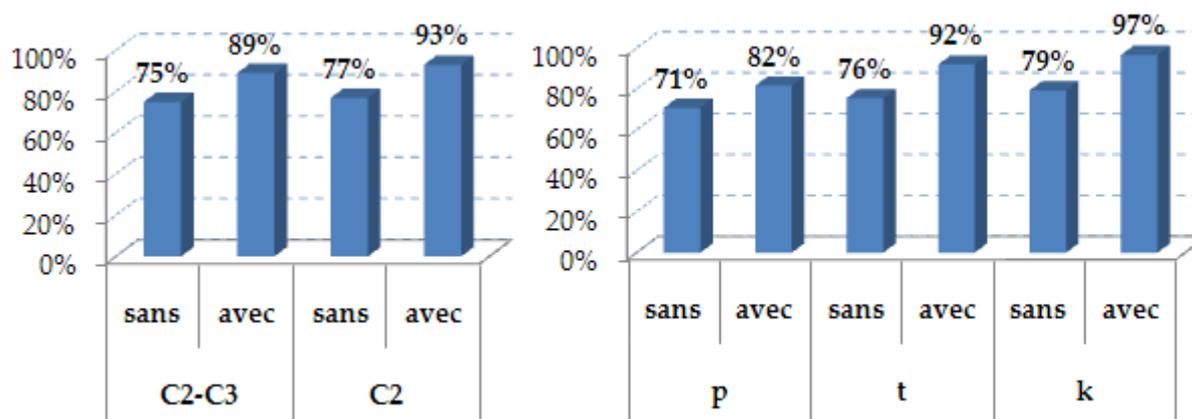


Figure 4.27. Score de bonnes réponses en fonction de la présence vs. absence du burst et du type de frontière : intra-mot (C₂.C₃), inter-mot (C₂) (à gauche), et des lieux d'articulation (à droite).

Ces résultats montrent clairement que le burst améliore nettement l'identification du lieu d'articulation de la consonne. Pourtant, le nombre de plosives finales /p/ /t/ /k/ relâchées (avec présence de burst) dans les stimuli du test est beaucoup plus restreint par rapport au nombre de plosives non relâchées (cf. table 4.12). Ces plosives comportent vraisemblablement d'autres indices acoustiques en dehors du burst porteurs d'informations sur leurs lieux d'articulation.

Table 4.12. Score d'identification en fonction de la présence/absence d'un burst et de la fréquence d'occurrences des plosives dans les stimuli du test.

Stimuli	Burst	Occurrences	Score
p	non	360	71%
	oui	120	82%
t	non	360	76%
	oui	360	92%
k	non	420	79%
	oui	120	97%

4.2.3.3. Indice acoustique : durée de la voyelle

Les analyses du corpus acoustique ont montré que le lieu articulaire de la consonne pouvait laisser des impacts sur la durée de la voyelle (voir section 4.1.2.1). À savoir que la voyelle est plus longue quand la consonne suivante est vélaire que quand la consonne est labiale ou coronale. Une vérification de ce résultat obtenu en moyenne pour les 10 locuteurs du corpus acoustique a été conduite afin de s'assurer des mêmes caractéristiques pour les productions du locuteur masculin retenues pour le test de perception.

Les analyses de variance à une variable dépendante montrent un effet non significatif, sur l'identification de la consonne qui suit, de la durée de la voyelle selon le lieu de la consonne suivante [$F(2,26) = 0,179$; $p = 0,837$]. La voyelle est effectivement plus longue quand la consonne est vélaire (117 ms), mais cette différence n'est plus significative par rapport à d'autres lieux labial (112 ms) ou coronal (109 ms) (la valeur de p est de 0,74 et de 0,55 respectivement).

La durée de la voyelle n'est donc pas un indice qui contribue à l'identification de la consonne post-vocalique.

4.2.3.4. Indice acoustique : la transition VC

L'importance de la transition formantique pour la perception des consonnes dans des séquences bi-segmentales CV a été mise en évidence dans la littérature depuis Delattre (1955). Pour l'analyse des transitions VC des plosives de l'étude acoustique qui a précédé cette étude perceptive, a été observé l'évolution des 3 premiers formants $F1$, $F2$, $F3$, la trajectoire de la fréquence fondamentale et de l'intensité, estimées à partir de la différence des valeurs mesurées à 50 % (partie la moins influencée par les consonnes de l'entourage) et 90% (partie de transition entre voyelle et consonne finale) de la durée de la voyelle. La même analyse a été menée sur les productions du locuteur qui ont fait l'objet du test perceptif. Ainsi, des analyses statistiques ont été effectuées sous SPSS pour repérer d'éventuels effets d'interaction entre lieux d'articulation et paramètres acoustiques dans cette partie de la transition voyelle-consonne.

Les valeurs moyennes et les écart-types des $\Delta F1$, $\Delta F2$, $\Delta F3$, $\Delta F0$ et ΔI sont représentés dans la table 4.13. Ces valeurs sont calculées en fonction de la consonne en coda (/p/, /t/ ou /k/).

Table 4.13. Valeurs moyennes et écarts-types des pentes de $F1$, $F2$, $F3$, $F0$ (Hz) et de l'intensité (dB) en fonction de la consonne en coda (/p t k/) calculés pour les productions du locuteur retenu pour établir les stimuli du test perceptif.

Consonne		p	t	k
$\Delta F1$	Moy.	-144,13	-48,75	-139,33
	Ecart.	109,78	78,99	81,02
$\Delta F2$	Moy.	-214,88	-12,92	-138,44
	Ecart.	72,94	111,54	229,40
$\Delta F3$	Moy.	-90,38	368,75	38,11
	Ecart.	369,05	658,24	496,66
$\Delta F0$	Moy.	12,25	13,08	9,56
	Ecart.	5,09	7,88	7,35
ΔI	Moy.	-2,53	-2,12	-5,70
	Ecart.	1,47	2,22	2,24

Les résultats statistiques sur les valeurs de $\Delta F1$, $\Delta F2$, $\Delta F3$, $\Delta F0$, ΔI figurent dans la table 4.14. Les différences entre lieu d'articulation non significatives sont présentées avec leur valeur de p ($p > 0,05$), celles significatives avec le symbole $*$.

Table 4.14. Seuils de significativité des différences des deltas $F1$, $F2$, $F3$, $F0$, I dans la partie de transition VC (noyau et coda) (**** si $p \leq 0,001$, *** si $p \leq 0,005$, ** si $p \leq 0,01$, * si $p \leq 0,05$)

	$\Delta F1$	$\Delta F2$	$\Delta F3$	$\Delta F0$	ΔI
Lieux	*	*	$p = 0,164$	$p = 0,521$	***

Les ANOVA montrent globalement un effet non significatif de la transition de $F0$ entre les plosives [$F(2,26) = 0,668$; $p = 0,521$]. Que le lieu soit labial, coronal ou vélaire, les analyses indiquent également un effet non significatif des valeurs de $\Delta F3$ pour les plosives finales [$F(2,26) = 1,94$; $p = 0,16$].

Les mêmes tests effectués pour les valeurs de $\Delta F1$ [$F(2,26) = 3,842$; $p = 0,035$], $\Delta F2$ [$F(2,28) = 4,55$; $p = 0,02$] et ΔI [$F(2,25) = 8,107$; $p = 0,002$] montrent des différences significatives globalement de ces transitions en fonction du lieu d'articulation des plosives. Le détail de la dynamique des formants dans cette partie de transition entre voyelle et consonne présente des pentes négatives de $F1$ et $F2$ (cf. figure 4.28).

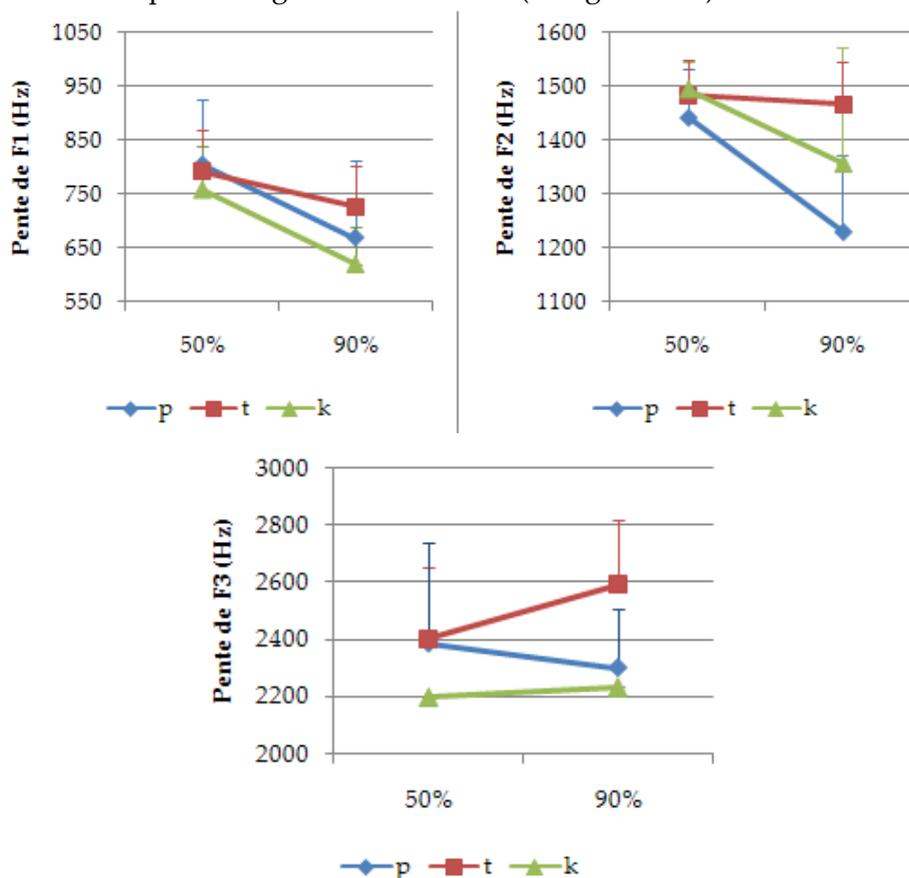


Figure 4.28. Pentes (et écart-types) des transitions formantiques de $F1$, $F2$, $F3$ (Hz) mesurées entre 50 % et 90 % de la durée de la voyelle (de gauche à droite, de haut en bas) en fonction de la consonne en coda (/p t k/).

Pour les pentes de F1, nous retrouvons les mêmes tendances par rapport aux résultats obtenus pour les 10 locuteurs (cf. section 4.1.2.4.1), celles suivies /t/ et /k/ sont descendantes de manière significative ($p = 0,042$ et $p = 0,001$ respectivement). La descente de la pente de F1 suivi de /p/ n'est pas significative, dû à la dispersion des valeurs. Un effet significatif global est observé pour les pentes de F1 selon le lieu d'articulation des plosives ($p = 0,035$). Pourtant, quand on regarde plus en détail, aucune différence entre chaque paire de plosive n'est significative : /p/ ~ /k/ ($0,56$) ; /p/ ~ /t/ ($p = 0,98$) ; /t/ ~ /k/ ($p = 0,56$). Les pentes de F1 ne sont donc pas vraiment un indice fiable permettant d'identifier les lieux d'articulation des plosives.

Pour les pentes de F2, comme les résultats des 10 locuteurs, les trajectoires devant plosives /p k/ sont plus descendantes par rapport à celles devant /t/. Mais seules celles suivies /p/ sont descendantes de manière significative ($p = 0,003$). Les valeurs prises à la fin de la voyelle sont trop dispersées pour les vélaires, ce qui fait que la descente de la pente devant /k/ n'est pas significative [$F(1,16) = 3,498$; $p = 0,08$]. La différence entre des valeurs de F2 prises à 50 % de la durée de la voyelle et celles à 90 % n'est pas significative avec /t/ [$F(1,22) = 0,271$; $p = 0,608$]. La pente F2 de la voyelle lorsqu'elle est suivie de /p/ se distingue significativement de celles suivies de /t/ ($p = 0,005$), aussi de /k/ ($p = 0,004$), ce qui ne nous permet pas d'expliquer la confusion davantage de /p/ avec /t/ (19 %) qu'avec /k/ (7 %). Les allures de la transition du deuxième formant entre la voyelle et l'occlusion de la consonne /k/ et /t/ ne présentent pas de différence significative ($p = 0,91$), pourtant /k/ est perceptivement confondue plus avec /p/ (11 %) qu'avec /t/ (6 %). Cet indice de pente de F2 ne fournit donc pas d'explication pour cette confusion.

La trajectoire de F3 dans les réalisations de /a/ devant /t/ est montante alors que celles de /p/ ou /k/ sont constantes. Que la plosive finale soit labiale, coronale ou vélaire, aucune différence de F3 entre la partie noyau et la partie coda n'a été détectée en fonction de leurs lieux d'articulation ($p = 0,56$; $p = 0,07$ et $p = 0,83$ respectivement).

Concernant la transition de l'intensité dans cette partie de la voyelle, la chute de l'intensité est différente significativement selon le lieu d'articulation de la consonne suivante ($p = 0,002$). Concrètement, la chute est plus importante de manière significative pour les voyelles suivies de vélaire que pour celles suivies de bilabiale ($p = 0,005$) ou coronale ($p = 0,001$). Il n'y a pas de différence significative de l'effet de l'intensité entre bilabiale et coronale ($p = 0,665$).

Donc, les critères pertinents en fonction des lieux d'articulation sont les transitions de F2 et de l'intensité. Même en l'absence de burst, les consonnes contiennent dans leur transition avec la voyelle précédente des informations (F2, I) sur leur lieu d'articulation qui ont guidé les sujets dans leur choix de réponse. Ces deux corrélats perceptifs (deuxième formant et intensité) font partie des indices acoustiques déjà validés par notre étude précédente (cf. section 4.2). Ces résultats sont en conformité avec les études antérieures

(Lieberman, Delattre, Cooper & Gerstman, 1954 ; Agrawal & Wen, 1975 ; Kishon-Rabin, Dayan & Michaeli, 2003 ; Nguyễn Việt Sơn, 2009) sur le rôle des transitions formantiques dans la perception du lieu d'articulation des consonnes.

4.2.3.5. Indice acoustique : les coefficients MFCC

Comme pour l'analyse acoustique des productions des 10 sujets (cf. section 4.1.2.4.1), douze coefficients MFCC (*Mel-Frequency Cepstrum Coefficients*) et leurs dérivées ont été également calculés pour les productions du locuteur retenu pour le test, sur une fenêtre glissante pour caractériser les particularités spectrales dans la partie de transition située à partir des 90 % de la durée de la voyelle. L'idée principale était d'extraire, au niveau spectral, des informations permettant une classification des voyelles /a/ en fonction du lieu d'articulation des consonnes qui la suivent. L'analyse en composantes principales (PCA) a été effectuée à partir de ces données (cf. figure 4.29).

L'analyse MFCC sur ce sujet est dans la même tendance que les 10 sujets de l'étude acoustique. En effet, cette analyse montre qu'à 90 %, la voyelle contient des informations concernant le lieu d'articulation de la consonne finale qui diffèrent par les caractéristiques spectrales mises en place pour les former : les zones de dispersion sont plus à l'écart pour les exemplaires des vélares /k ɲ/ qui obtiennent les meilleurs scores d'identification ; les zones des coronales situées entre celles des vélares et des labiales expliquent probablement les confusions de réponses dans l'identification de /p/ et /t/, de /m/ et /n/ par les sujets vietnamiens. Cette analyse ne nous permet cependant pas d'expliquer pourquoi les sujets confondent davantage les réalisations de /k/ avec celle de /p/ (11 %) qu'avec celles de /t/ (6 %). L'explication pour cette confusion reste à discuter.

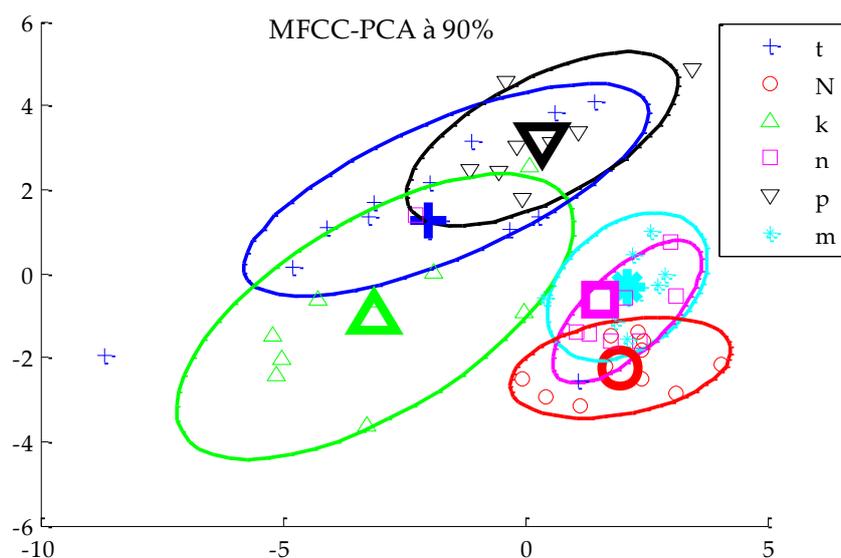


Figure 4.29. Analyse en composantes principales des coefficients MFCC-PCA de la transition à 90% de /a/ dans les contextes consonantiques /p t k m n ɲ/ pour les productions du locuteur natif qui font l'objet des stimuli du test perceptif.

4.2.4. Conclusion

Notre étude montre que des sujets vietnamiens identifient mieux les nasales que les plosives de leur langue maternelle lorsqu'elles sont en position de coda. La confusion entre plosives et entre nasales est plus importante lorsque la coda C_2 précède une frontière intra-mot (devant C_3 de mots composés) qu'une frontière inter-mot (C_2). Ce résultat montre l'influence de la frontière sur la perception des plosives, comme des nasales en fin de syllabe et de mot. L'analyse de la fréquence lexicale a montré que les nasales sont particulièrement présentes en position finale de syllabe (finale de mots monosyllabiques et finale de première syllabe de mots composés) par rapport à la position initiale, ce qui pourrait expliquer un meilleur résultat d'identification des nasales par rapport aux plosives en position de coda.

Bien que les plosives soient généralement non relâchées dans cette position, nos résultats montrent que la présence d'un burst même faible augmente le taux d'identification correcte. Dans le cas des plosives non relâchées, les scores restent supérieurs à 70 %. Par ailleurs, des analyses de la fréquence lexicale de ces consonnes montrent que celle-ci n'explique pas totalement ces scores. L'analyse de la durée de la voyelle montre que celle-ci ne possède pas de corrélat sur le lieu de la consonne suivante. Une analyse acoustique des transitions des fréquences formantiques, de la fréquence fondamentale et de l'intensité montre que $\Delta F1$, $\Delta F2$ et ΔI sont, chez le locuteur qui a produit les stimuli du test perceptif, significativement différents en fonction du lieu d'articulation des plosives. Ces résultats sont en conformité avec les études antérieures sur le rôle des transitions formantiques dans la perception du lieu d'articulation des consonnes. Une nette tendance d'allure de pente de $F2$ (montante pour /t/ et /k/) n'a pas cependant été observée comme dans la littérature, à part le cas descendant significatif de /p/. L'analyse MFCC-PCA a montré, en complémentarité, des zones distinctives pour la catégorie des vélaires.

En l'absence de burst, les consonnes contiennent dans leur transition ($F2$, I) avec la voyelle précédente des informations sur leur lieu d'articulation qui ont probablement guidé les sujets dans leur choix de réponse.

L'analyse acoustique ayant montré des différences dans la réalisation des segments en fonction du type de frontière syllabique (intra-mot vs. inter-mot), l'expérience suivante vise à rechercher si ces différences impactent la perception d'une frontière voire d'un type de frontière.

4.3. Expérience 3 : Identification du type de frontière syllabique en vietnamien

Un des résultats importants de l'étude acoustique (cf. section 4.1) a montré qu'il existe des différences dans la réalisation d'une consonne en fonction de sa position en finale (C₂ finale de mot ou C₂.C₃ finale d'une syllabe intra-mot). Il a aussi été montré dans l'étude précédente que la frontière syllabique influence la perception des consonnes en coda : les consonnes coda intra-mot engendrent plus de confusion que les consonnes en coda inter-mot (cf. section 4.2.2). Des sujets natifs arrivent-ils alors à la simple écoute d'une syllabe à reconnaître s'il s'agit d'une syllabe constituante d'un mot simple ou d'une syllabe d'un mot composé ?

L'objectif de cette expérience est de chercher les corrélats perceptifs du type de frontière syllabique en vietnamien. Selon les résultats de l'étude acoustique (cf. section 4.1.2.1 de ce chapitre), les consonnes en fin de mot devraient être mieux perçues que celles en fin de S1 de mot composé.

4.3.1. Méthodologie

Les mêmes sujets de l'expérience précédente (10 hommes et 10 femmes), originaires du Nord du Vietnam, ont passé cette troisième expérience. Les mêmes corpora et programme *PerceptA* ainsi que la même interface que le test d'identification des consonnes finales ont été utilisés pour ce test.

Les sujets avaient pour consigne d'écouter au casque, puis de choisir s'il s'agissait d'un mot simple (« don ») ou d'une syllabe de mot composé (« ghep ») en cliquant la réponse à l'écran à l'aide de la souris (cf. figure 4.30). La consigne était donnée au sujet de répondre le plus rapidement possible. Une phase d'entraînement à l'interface et au type de stimuli leur avait été préalablement proposée avec des stimuli ne figurant pas dans le test. Ce test a eu lieu quelques minutes après les tests de l'expérience précédente. Les plosives étaient testées à part des nasales. Le test, constitué de 3 répétitions des 55 items (cf. section 4.2.1.1) présentés dans un ordre aléatoire pour chaque sujet, durait environ 10 minutes.

Les données ont été traitées avec le logiciel SPSS®. Des tests d'analyses de variance (ANOVA) ont été effectués.

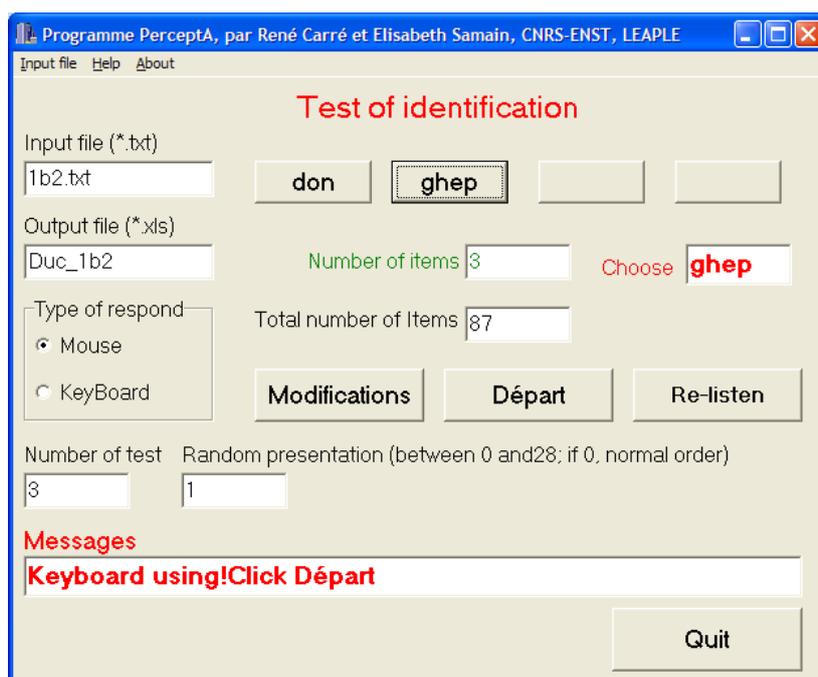


Figure 4.30. Interface du test perceptif utilisé pour les plosives.

4.3.2. Résultats

Les résultats montrent que les sujets arrivent mieux à reconnaître la nature d'une syllabe (c'est-à-dire syllabe de mot monosyllabique ou syllabe 1 de mot composé dissyllabique) lorsque la syllabe se termine par une consonne nasale, plutôt qu'une plosive (85 % contre 69 % respectivement). La figure 4.31 montre que pour les syllabes terminées par une plosive, les S1 des mots composés ont des scores significativement inférieurs (65 % vs. 75 %) [$F(1,1738) = 19,948$; $p = 0$]. Une monosyllabe terminée par une nasale parvient au même score d'identification du type de frontière (85 %) qu'une syllabe S1 à nasale finale précédant une frontière intra-mot (85 %).

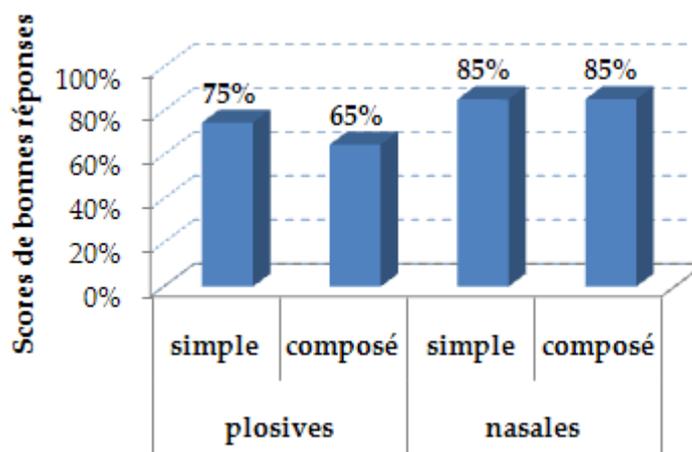


Figure 4.31. Scores d'identification du type de frontière syllabique pour des syllabes à plosives finales (à gauche) et des syllabes à nasales finales (à droite) (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

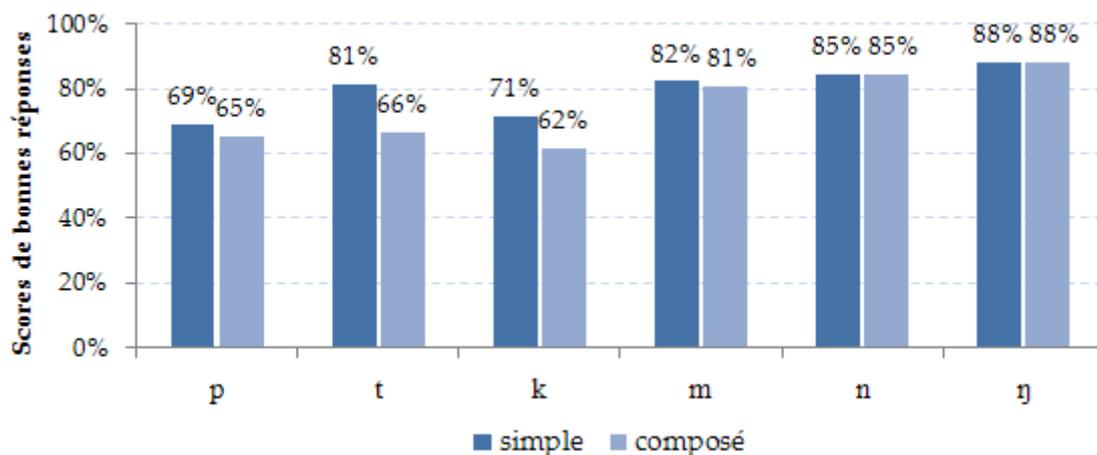


Figure 4.32. Scores d'identification du type de frontière syllabique pour des syllabes à consonnes finales /p t k m n ŋ/ (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

La distinction du type de frontière est mieux reconnue quand il s'agit des syllabes à consonne /t/ ou /k/ final ($p = 0$, $p = 0,02$ respectivement) que des syllabes à /p/ final ($p = 0,425$) (cf. figure 4.32). Aucune différence significative de scores de bonnes réponses n'a été détectée pour les stimuli avec nasales finales. Les scores d'identification restent presque identiques pour les deux types de frontière testés (mot ou syllabe) quel que soit le lieu des nasales.

Que le locuteur soit homme ou femme, il n'y a pas de différence significative de scores de bonnes réponses pour les stimuli à plosive finale ($p = 0,301$), idem pour ceux à nasale finale ($p = 0,342$).

Avec des scores de bonnes réponses supérieures au hasard, les résultats obtenus pour les plosives et pour les nasales montrent que la syllabe comporte des indices acoustiques porteurs d'information sur le type de frontière.

4.3.3. Discussion

Rappelons que le processus de resyllabation n'étant pas permis en vietnamien, les plosives souvent non relâchées en coda ne peuvent jamais passer de cette position à la position d'attaque de la syllabe suivante. Quelques exemples de la distinction entre les groupes dissyllabiques et les monosyllabes homophones (c'est à dire constitués apparemment des mêmes sons) ou les syntagmes dissyllabiques homophones qui se distinguent par la place de la frontière syllabique ont été étudiés par les linguistes vietnamiens Hoàng Tuệ et Hoàng Minh (1975) :

- | | | |
|--|-----|---|
| khoa Y [χwa ¹ ʔi ¹] <i>département de Médecine</i> | vs. | khoai [χwaj ¹] <i>patate</i> |
| xem ô tô [sɛm ¹ ʔo ¹ -to ¹] <i>regarder les voitures</i> | vs. | xe mô tô [sɛ ¹ -mo ¹ -to ¹] <i>motocyclette</i> |
| ô mai [ʔo ¹ -maj ¹] <i>abricot confit</i> | vs. | ôm ai [ʔom ¹ ʔaj ¹] <i>embrasser quelqu'un</i> |

Langue isolante, phonologiquement monosyllabique, les syllabes en vietnamien, dans la plupart des cas, pourraient être des unités signifiantes. Distinguer à l'oreille, sans contexte, une syllabe isolée, soit appartenant à un mot composé, soit constituant un mot monosyllabique, n'était pas tâche facile pour les locuteurs. Elle leur paraissait même un peu « étrange ».

Comment ont-ils pu faire l'expérience avec des scores au dessus du seuil du hasard ? Quels indices leur ont permis de reconnaître le type de frontière ? Pourquoi les syllabes à nasales finales sont moins difficiles à identifier « simple » ou « composé » que ceux terminés par des plosives ? Pourquoi les monosyllabes à plosives finales sont mieux reconnus que les syllabes intra-mots ?

Le degré de coarticulation devrait être plus important entre syllabes d'un mot composé qu'entre mots monosyllabiques. Si c'est le cas, la durée devrait être un indice considérable sur lequel les locuteurs se basent pour effectuer cette tâche. Quelques résultats acoustiques obtenus avec le corpus de l'expérience 1 concernant la durée ont été exploré pour tenter de répondre à ces questions.

4.3.3.1. Durée de la syllabe

Les analyses montrent que les syllabes à nasale finale sont plus longues que celles à plosive finale (en moyenne 332 ms contre 285 ms). Cette différence de la durée de la syllabe est significative [$F(1,53) = 4,548$; $p = 0,038$]. Ce résultat pourrait expliquer le fait que les stimuli à nasale finale ont obtenu un meilleur score d'identification que ceux à plosive finale.

La figure 4.33 montre globalement que la syllabe est significativement plus longue quand il s'agit d'un mot simple que d'une syllabe à l'intérieur d'un mot composé, pour les plosives [$F(1,27) = 31,269$; $p = 0$] comme pour les nasales [$F(1,24) = 87,815$; $p = 0$]. Ce résultat pourrait expliquer le fait que les sujets identifient mieux un mot qu'une syllabe intra-mot.

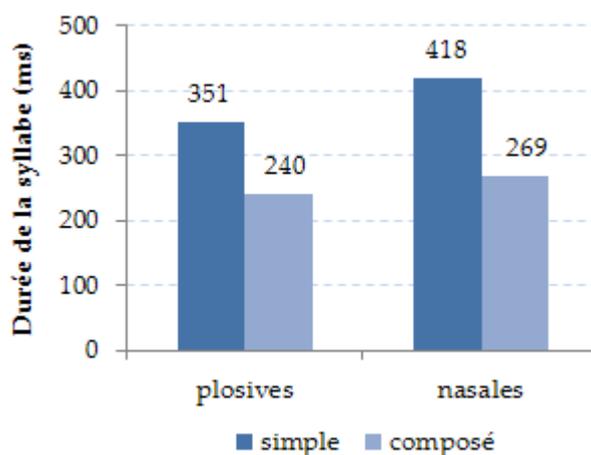


Figure 4.33. Durée des syllabes à plosive finale (à gauche) et à nasale finale (à droite) (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

Cependant, les mots simples à nasale finale sont 1,5 fois plus longs que les syllabes intra-mot à nasale finale (cf. figure 4.33) alors que dans les deux cas, les mêmes scores d'identification (85 %) ont été obtenus (c'est-à-dire quel que soit le type de frontière syllabique). L'explication pour cette performance des nasales reste à discuter.

La figure 4.34 présente une analyse plus fine de la durée de la syllabe en fonction de la consonne finale. Les syllabes à /p/ final ne présentent pas de différence significative de durée en fonction du type de frontière syllabique [$F(1,6) = 3,226$; $p = 0,123$], dû très probablement à la variabilité des valeurs. Ce qui n'est pas le cas pour les consonnes /t k m n ŋ/ où une syllabe à l'intérieur d'un mot composé est significativement moins longue qu'un mot monosyllabique.

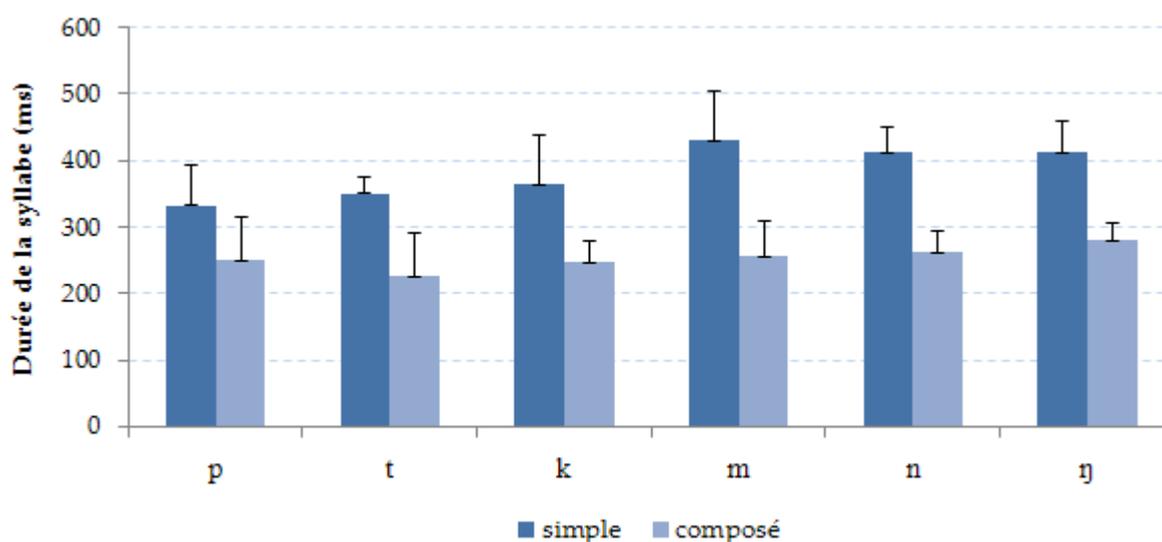


Figure 4.34. Durée des syllabes à plosive finale (à gauche) et à nasale finale (à droite) (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

Ce résultat pourrait expliquer pourquoi malgré le fait que /p/ n'existe qu'en finale de mot ou de syllabe, les sujets reconnaissent moins bien le type de frontière syllabique lorsque celle-ci suit une plosive bilabiale.

4.3.3.2. Durée de la consonne finale

Une distinction de la durée de la consonne a été globalement observée selon sa position dans la syllabe (cf. figure 4.35). Comme les résultats de l'étude acoustique auprès de 10 locuteurs (cf. section 4.1.2.1), les consonnes sont plus longues en C₂ qu'en C₂.C₃. La différence de la durée de la consonne en fonction du type de frontière qui la suit est significative pour les plosives [$F(1,22) = 6,903$; $p = 0,015$], aussi pour les nasales [$F(1,24) = 41,832$; $p = 0$].

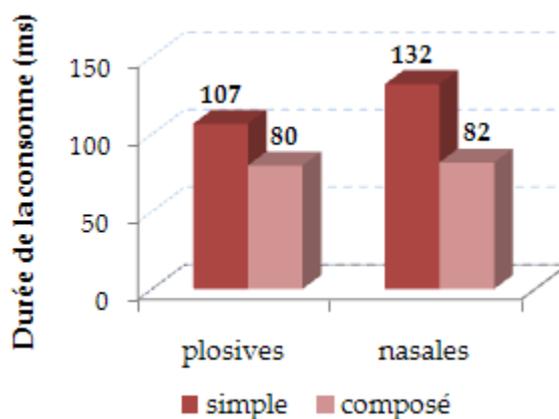


Figure 4.35. Durée moyenne des plosives (à gauche) et des nasales (à droite) en finale (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

Le détail par rapport à chaque consonne (cf. figure 4.36) montre qu'une différence significative de la durée de la consonne n'est observée que pour les nasales /m n ŋ/ ($p = 0,014$; $p = 0,04$; $p = 0,002$ respectivement). La durée des plosives /p t k/ n'est plus différente de manière significative en fonction de la structure syllabique ($p = 0,67$; $p = 0,10$; $p = 0,062$ respectivement) : moins de différences avec /p/ et /t/ ; trop de dispersion des valeurs pour /k/. L'allongement des plosives antérieures est pourtant plus facilement contrôlable que celle de la plosive vélaire à cause de l'extension possible des parois de la cavité buccale pour les antérieures (Vallée et al, 2002). La diminution de la durée est probablement le phénomène induit par la structure plus longue du mot.

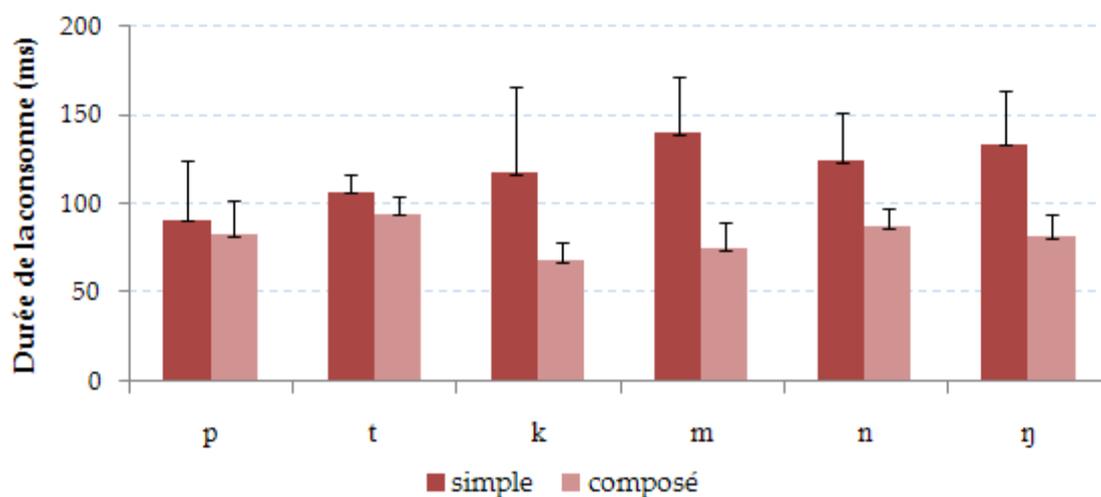


Figure 4.36. Durée moyenne des consonnes en finale en fonction de la structure syllabique (simple = monosyllabe, composé = syllabe 1 de dissyllabe)

4.3.3.3. Durée de la voyelle

La même distinction de durée de la voyelle a été observée selon la structure syllabique, comme les résultats obtenus auprès de 10 locuteurs de l'étude acoustique (cf. section 4.1.2.1), les voyelles sont plus longues devant C₂ que C₂.C₃ (cf. figure 4.37).

Cette différence est significative dans tous les contextes ($p < 0,05$), quel que soit le lieu d'articulation des consonnes (cf. figure 4.38). Donc, la voyelle est toujours porteuse d'information sur la nature de la syllabe quelle que soit la consonne qui suit.

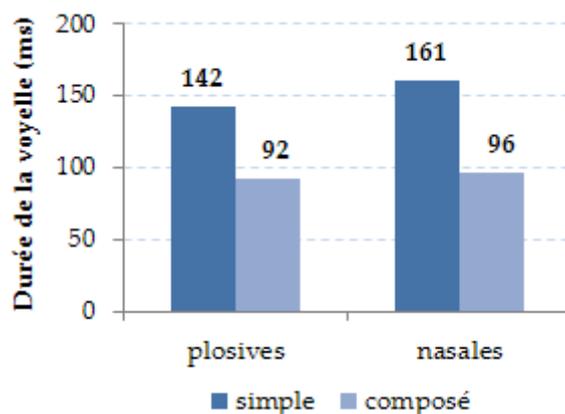


Figure 4.37. Durée moyenne de la voyelle suivie d'une plosive finale (à gauche) et d'une nasale finale (à droite) (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

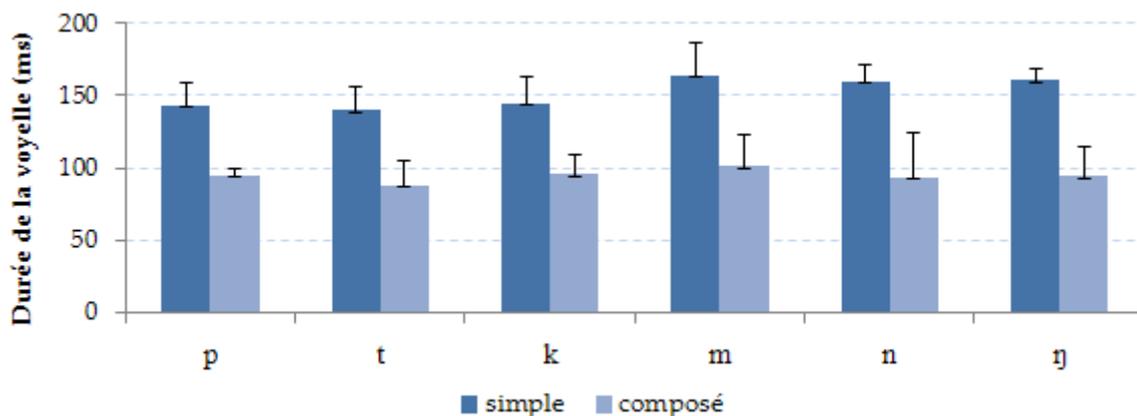


Figure 4.38. Durée moyenne de la voyelle en fonction de la consonne finale de syllabe (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

4.3.4. Conclusion

Cette troisième expérience montre que les sujets vietnamiens sont capables à partir de l'écoute d'une syllabe de reconnaître s'il s'agit de la syllabe 1 d'un mot composé ou d'une syllabe correspondant à un mot simple. Les stimuli à nasale finale engendrent moins de confusion que ceux à plosive finale. Les sujets arrivent mieux à identifier les mots simples que les syllabes à l'intérieur d'un mot composé, si ces mots ou ces syllabes sont terminés par des plosives. Les résultats montrent que la durée de la syllabe, de la consonne finale et de la voyelle sont les indices acoustiques permettant aux locuteurs d'effectuer la tâche bien qu'elle leur a parue difficile.

Les scores d'identification des deux types de frontière pour les syllabes terminées par des nasales sont identiques malgré le fait que les consonnes nasales et la voyelle /a/ qui les précèdent sont plus longues en fin de mot qu'en fin de S1 d'un mot composé. Les performances dans l'identification des frontières de syllabes intra-mots à nasale finale restent à comprendre.

Une expérience complémentaire a été menée, cette fois sur l'identification de la syllabe elle-même, afin de s'assurer que le type de frontière syllabique impacte la réalisation des segments tauto-syllabiques et de confirmer que des indices acoustiques de frontière sont bien présents dans le signal et qu'ils sont récupérables par les auditeurs pour la segmentation des unités lexicales et syllabiques.

4.4. Expérience 4 : Perception des constituants de la syllabe

La différence de durée de la syllabe, et notamment celle de la voyelle sont attestées par les résultats de l'étude acoustique (cf. section 4.1). Cette différence constitue d'ailleurs un indice majeur pour reconnaître le type de frontière syllabique qui suit une structure CVC entendue (cf. section 4.3). Si la voyelle est acoustiquement plus brève quand elle précède une frontière syllabique intra-mot, les locuteurs devraient percevoir cette voyelle /a/ en structure CaC.CVC comme brève /ă/. L'objectif de cette quatrième expérience est de vérifier cette hypothèse.

4.4.1. Méthodologie

Un groupe de vingt sujets (10 hommes et 10 femmes) différents de ceux des expériences précédentes, âgés de 24 à 29 ans (moyenne d'âge de 22 ans pour les femmes, de 25 ans pour les hommes), originaires du Nord du Vietnam, ont participé à l'expérience. Ils sont en France pour suivre des études supérieures et ce depuis 2 à 4 ans. Ils continuent à parler quotidiennement vietnamien depuis leur arrivée en France.

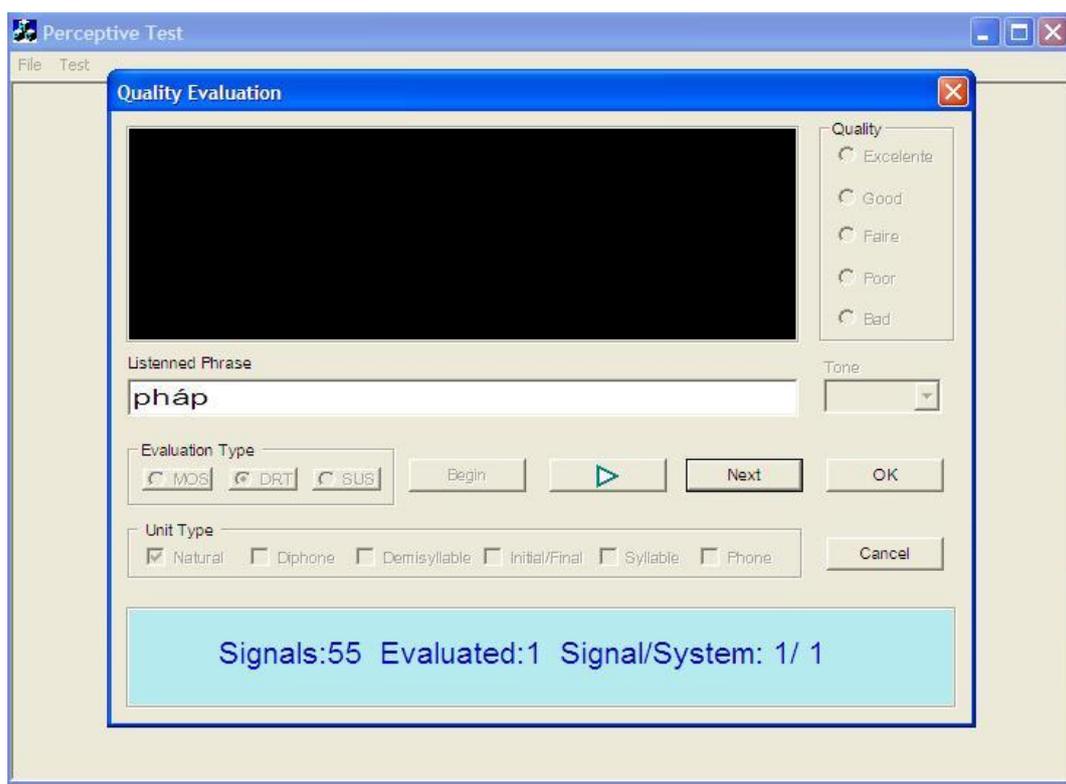


Figure 4.39. Interface du test perceptif de la syllabe.

Le programme *Perceptive Test* réalisé par Trần Đỗ Đạt (MICA, Hanoi) a été utilisé pour ce test. Les sujets avaient pour consigne d’écouter, puis de taper le mot entendu en code orthographique le plus vite possible dans la case « Listened Phrase » en utilisant le clavier (cf. figure 4.39). La réécoute est possible avec l’interface. Après avoir saisi le mot, les sujets appuyaient sur la touche « Next » pour passer au stimulus suivant. Une phase d’entraînement à l’interface et au type de stimuli leur avait préalablement proposée. Ce test a duré en moyenne cinq minutes pour chaque participant.

Les mêmes 55 stimuli (contexte tonal *sắc* B1-D1 et vocalique /a/) des expériences précédentes (cf. section 4.2.1.1), extraits du corpus acoustique, ont été utilisés pour ce test. Ce sont des monosyllabes ou des premières syllabes de mots composés. Ces stimuli ont été présentés en ordre aléatoire pour chaque sujet.

Le test s’est déroulé dans une salle calme du Département Parole et Cognition de GIPSA-lab en utilisant un casque de haute qualité binaural Sennheiser HD-25-13.

Les données ont été traitées avec SPSS[®]. Des tests d’analyses de variance (ANOVA) ont été effectués.

4.4.2. Résultats

Les résultats montrent que les syllabes présentées auditivement (monosyllabes ou syllabes 1 de mot composé) ont été correctement identifiées dans seulement 57,5 % des cas. Les erreurs proviennent de tous les constituants de la syllabe. Nous remarquons que la consonne initiale a été très bien perçue (taux d'erreur de 4,4 %) par rapport à d'autres constituants. C'est aussi le cas du ton *sắc* (B1 à nasales finales et D1 à plosives finales) qui a été également très bien identifié (taux d'erreur de 5,9 %). Dans les cas de confusion du ton, les sujets se trompent le plus souvent avec le ton *ngang* A1. Rappelons que le ton *ngang* A1 et le ton *sắc* B1-D1 appartiennent tous les deux au même registre haut.

Les syllabes 1 de mot composé ont donné le plus de confusion dans les réponses. Ci-dessous sont présentés en détail les types d'erreurs concernant la perception des voyelles et des consonnes finales.

4.4.2.1. Erreur d'identification de la voyelle

Dans les 1 100 cas de syllabe observés (640 monosyllabes et 460 syllabes intra-mot), la voyelle a été correctement identifiée comme longue dans 67,9 % des cas. Concernant les erreurs, 30 % des voyelles /a/ ont été perçues comme /ă/ et 2 % comme la voyelle brève postérieure non arrondie mi-fermée (/ɤ/) (cf. figure 4.40).

Si on regarde l'influence de la structure syllabique sur la perception de la voyelle, la figure 4.37 montre que la voyelle a été perçue comme brève davantage dans la structure CVC.CVC sur l'ensemble des stimuli testés. La différence est significative quel que soit le mode d'articulation des consonnes ($p = 0$). Ce résultat est en accord avec notre étude acoustique selon laquelle la voyelle est plus brève quand elle précède une frontière intra-mot C₂.C₃ qu'une frontière de mot C₂.

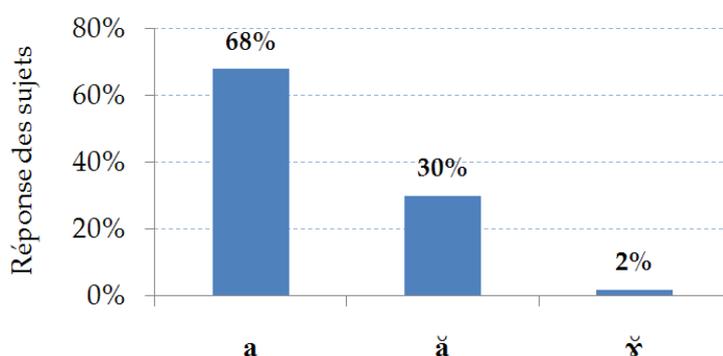


Figure 4.40. Résultats pour l'identification de la voyelle.

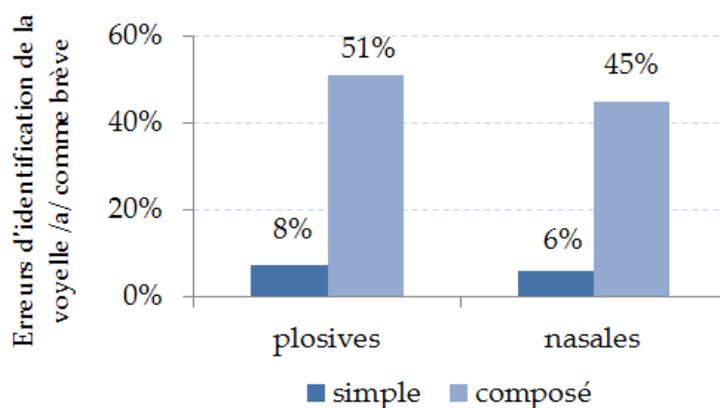


Figure 4.41. Identification de la voyelle /a/ comme brève en fonction du mode et du type de structure syllabique (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

Une analyse plus fine tenant compte de la nature des consonnes post-vocaliques montre que la tendance est indépendante du type de consonne (cf. figure 4.42). La différence est significative en fonction de la structure syllabique pour chacune des six consonnes ($p = 0$). La voyelle de S1 de structure CaC.CVC est perçue plus souvent comme brève lorsqu'elle précède la plosive vélaire /k/ (71 %). Ce résultat corrobore ce qui a été montré dans l'expérience 4.3 où la distinction du type de frontière est bien reconnue, et de manière significative, quand il s'agit des syllabes à /k/ final.

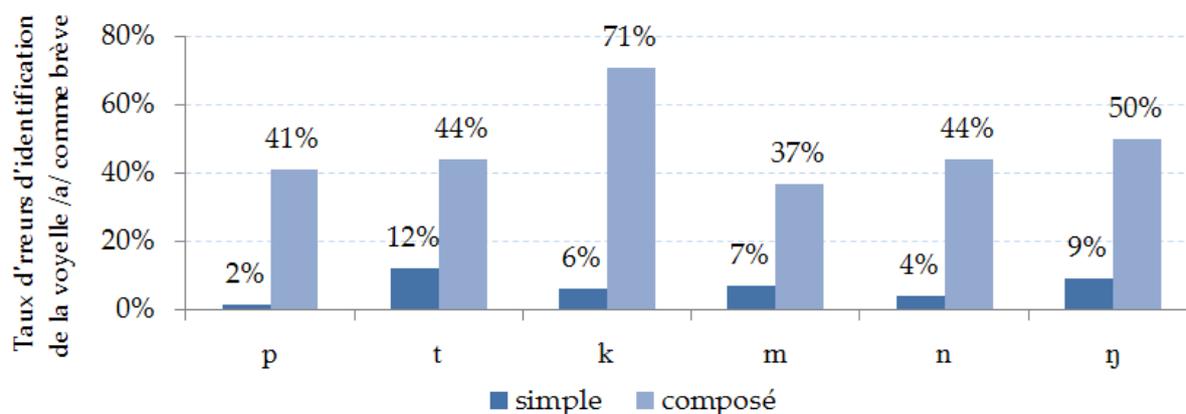


Figure 4.42. Perception de la voyelle /a/ comme brève en fonction de la structure syllabique pour chaque consonne (simple = monosyllabe, composé = syllabe 1 de dissyllabe).

4.4.2.2. Erreur d'identification de la consonne finale

Le taux d'erreur de perception des consonnes finales est faible et s'élève en moyenne à 13 % pour les syllabes à plosives finales et 6 % pour celles à nasales finales.

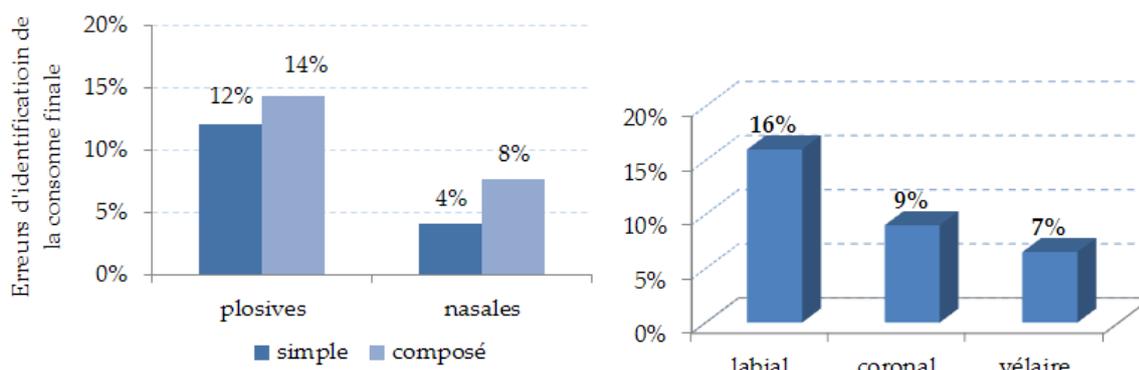


Figure 4.43. Taux d'erreurs dans l'identification de la consonne finale selon le mode et la structure syllabique (simple = monosyllabe, composé = syllabe 1 de dissyllabe) (à gauche) et selon le lieu (à droite).

Les résultats de cette expérience sont en conformité avec ceux de l'expérience 4.2 sur l'identification de la consonne finale. En effet, la figure 4.43 montre que les sujets se trompent davantage dans la perception des plosives finales que des nasales finales [$F(1,1095) = 16,241$; $p = 0$]. Regroupées par lieu d'articulation, les vélares et les coronales obtiennent de meilleurs scores d'identification par rapport aux labiales.

Le type de frontière syllabique n'influence pas de manière significative les scores quel que soit le mode d'articulation (plosives, $p = 0,419$; nasales, $p = 0,092$) (cf. figure 4.44), alors que l'expérience 4.2 montre significativement davantage de confusion pour les plosives comme pour les nasales lorsque la coda précède une frontière intra-mot.

L'analyse des confusions est en accord avec les résultats obtenus dans l'expérience 4.2. Pour les plosives, c'est la bilabiale /p/ qui engendre le plus de confusion (19 % d'erreur de perception). Les sujets la confondent davantage avec /t/ (24 sur 30 réponses) qu'avec /k/ (6 sur 30 réponses). Le meilleur score d'identification est obtenu par /t/ et /k/ (taux d'erreur respectifs de 10 % et 13 %). La différence d'erreurs entre ces deux consonnes n'est pas significative ($p = 0,321$).

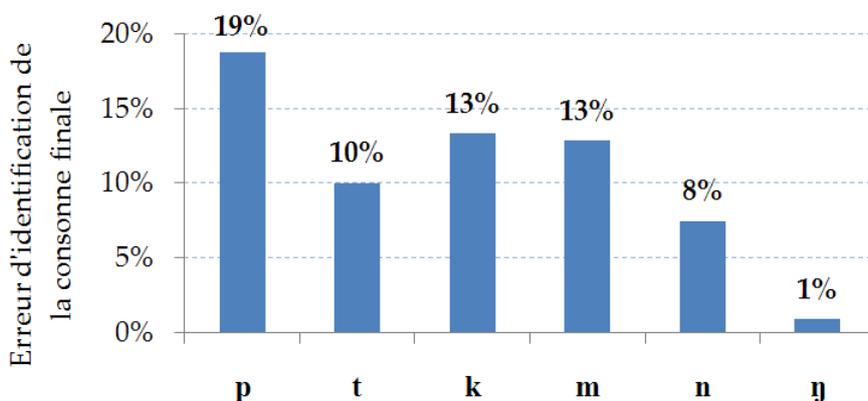


Figure 4.44. Taux d'erreurs dans l'identification de la coda en fonction du type de consonne.

Pour les nasales, la vélaire /ŋ/ obtient le meilleur score d'identification par rapport à /m/ et /n/ (taux d'erreur de 1 %). Les sujets identifient mieux les stimuli à /n/ finale qu'à /m/ finale ($p = 0,05$), alors que dans l'expérience 4.2 précédente, la différence de score de bonnes réponses n'est pas significative entre /m/ et /n/ ($p = 0,34$).

4.4.3. Conclusion

Cette expérience montre que le score d'identification d'une syllabe d'unité monosyllabique ou d'unité dissyllabique dépend notamment de la perception de la voyelle (32 % des erreurs), ensuite de la consonne finale (10 % des erreurs), du ton (6 % des erreurs) et de la consonne initiale (4,4 %).

Les résultats obtenus confirment l'hypothèse formulée selon laquelle la voyelle /a/ acoustiquement plus brève lorsqu'elle précède une frontière intra-mot est perçue effectivement comme le phonème /ă/. Rappelons que, sur le plan phonologique, la paire minimale /a/ ~ /ă/ se distingue uniquement par la durée (Nguyễn Việt Sơn, 2009). Ce résultat renforce l'étude 4.3 sur la perception du type de frontière syllabique dont les résultats montrent que la syllabe contient des indices sur la frontière qui la suit : non seulement la durée de la consonne finale, mais aussi la durée de la voyelle est un indice porteur d'information sur la nature de la syllabe.

Concernant l'identification de la consonne finale, cette expérience confirme les résultats obtenus dans l'expérience 4.2 selon laquelle les nasales engendrent moins de confusion que les plosives, les coronales et les vélaire étant mieux identifiées que les bilabiales : /t/, /k/ et /ŋ/ obtiennent toujours les meilleurs scores d'identification, /p/ est la consonne la moins correctement identifiable par rapport aux autres finales. Cette consonne est confondue avec /t/.

Les études acoustico-perceptives sur les consonnes simples et en séquence du vietnamien, exposées dans ce chapitre, permettent de montrer chez des sujets natifs l'existence de frontière syllabique et lexicale en perception et production. Une frontière syllabique de mot est différente acoustiquement et perceptivement d'une frontière de syllabe à l'intérieur d'un mot composé. Les erreurs en perception sont plus fréquentes devant frontière intra-mot (composé lexical dissyllabique) que devant frontière inter-mot. Il existe des indices acoustiques de ces frontières récupérables en perception : trajectoire du deuxième formant, de l'intensité et de F0 dans la transition VC précédant la frontière, durées de la consonne et de la voyelle. Face à l'objectif principal de cette thèse qui concerne la production et la perception des groupes de consonnes du français par les apprenants vietnamiens, les résultats sur la production et perception des consonnes en séquences du vietnamien peuvent constituer des éléments du crible phonologique de la langue source capables de rendre

compte des difficultés persistantes pour des apprenants vietnamiens à réaliser des clusters du français.

La connaissance des caractéristiques acoustiques et perceptives des consonnes du vietnamien en tenant compte de leur environnement vocalique et consonantique est un préalable à la connaissance des processus d'acquisition des groupes consonantiques inter-syllabe et intra-syllabe auxquels se trouvent confrontés les Vietnamiens apprenant d'une langue étrangère comme le français.

Dans le chapitre suivant, est présenté un ensemble d'études en perception et production de consonnes simples et en séquence en prenant en compte le facteur position du groupe de consonnes : inter-syllabe (séquence consonantique) *vs.* intra-syllabe (cluster consonantique) du français par des apprenants vietnamiens du Français Langue Etrangère. Les analyses sont discutées en regard des résultats des expériences de ce chapitre IV et des éléments en théorie et modélisation présentés dans le chapitre I.

Chapitre 5

Perception et production des groupes de consonnes en langue seconde : le cas du français par des apprenants vietnamiens

L'étude exposée dans ce chapitre porte sur l'acquisition des clusters du français en perception et production, auprès d'apprenants vietnamiens de niveaux intermédiaire ou avancé. *L'Hypothèse de la Différence de Marquage* (Eckman, 1977) prédit que les formes du français différentes et plus marquées que celles du vietnamien sont plus difficiles à acquérir. Bien que les séquences de consonnes peuvent être constituées de segments proches voire identiques dans les deux langues, elles sont différentes en ce que, en vietnamien, elles ne sont rencontrées que de part et d'autre d'une frontière syllabique inter- ou intra-mot et que les consonnes post-vocaliques, premier constituant des séquences consonantiques, présentent des particularités de réalisation phonétique.

L'étude vise les séquences de consonnes que l'on peut trouver dans les deux langues à la frontière de deux syllabes consécutives en français ou en vietnamien, et les clusters qui sont intra-syllabiques et qui existent uniquement en français, l'objectif premier de cette étude étant d'estimer l'impact de la syllabe et de ses frontières dans l'acquisition des groupes de consonnes, dont les clusters d'une langue seconde. Il s'agit également de comprendre l'influence des structures syllabiques de la L1 sur l'apprentissage d'une langue étrangère présentant des structures différentes de syllabe. Plus généralement, cette partie de la thèse cherche à montrer la présence ou l'absence d'un effet de la syllabe et de ses frontières dans le système phonologique interlangue.

À partir des résultats de l'étude acoustique présentée chapitre IV et des éléments théoriques discutés plus avant (cf. section 1.1 et 1.2) comme le crible phonologique (Troubetzkoy, 1939), le marquage (Eckman, 1977), les modèles de processus d'acquisition (Best, 1995, Flege, 1995) et en considérant les résultats des études antérieures sur l'acquisition des clusters de l'anglais par des apprenants vietnamiens (Benson, 1988 ; Osburne, 1996 ; Sato, 1984, 1985), plusieurs hypothèses peuvent être formulées :

- Hypothèse 1 : La structure syllabique C_1VC_2 du vietnamien impacte l'apprentissage des structures syllabiques complexes du français.
- Hypothèse 2 : Les clusters du français sont plus difficiles à récupérer et à produire que les séquences inter-syllabiques.
- Hypothèse 3 : Les apprenants vietnamiens modifient plus fréquemment les codas complexes du français que les attaques complexes, ces dernières étant moins marquées.
- Hypothèse 4 : Les caractéristiques phonétiques des consonnes en coda du vietnamien (consonnes non relâchées) ont des conséquences sur les stratégies de simplification des groupes consonantiques du français.
- Hypothèse 5 : Les combinaisons de phonèmes consonantiques plus marquées selon les universaux du langage sont plus difficiles à produire que les

combinaisons moins marquées. Et dans ce cadre, selon que les clusters respectent ou violent le Principe de Sonorité, leur niveau de difficulté d'apprentissage doit être différent.

5.1. La perception des consonnes simples et en séquences du français par des apprenants vietnamiens

Avant de conduire une étude sur la production des clusters du français, un ensemble de tests portant sur les indices acoustico-perceptifs des consonnes produites en séquences en fonction de leur position dans la syllabe et dans le mot a été effectué. Il s'agit de tester la perception des séquences de consonnes communes aux deux langues chez plusieurs groupes de sujets, apprenants du français langue étrangère à l'Université en fonction de leur nombre d'années d'apprentissage, afin d'estimer l'influence des structures syllabiques de la L1 sur la perception d'une langue seconde présentant des structures différentes de syllabe.

5.1.1. *Expérience 1 : Perception des consonnes simples du français*

Bien que les langues présentent des phonèmes identiques, plusieurs études ont montré que leurs réalisations peuvent différer d'une langue à l'autre (Ladefoged et Maddieson, 1996 ; Ladefoged, 2001). C'est le cas par exemple des plosives finales /p t k/ non relâchées en vietnamien et réalisées avec un bruit d'explosion en général audible en français. Avant d'étudier la perception et production des séquences de consonnes du français par des apprenants vietnamiens, une étape préliminaire s'impose : l'analyse de la perception et production des consonnes simples en fonction de leur position dans la syllabe.

L'objectif de cette étude est double :

- Tester chez des apprenants vietnamiens, la récupération des consonnes du français et leurs caractéristiques articulatoires à partir du signal acoustique ;
- Analyser typologiquement et quantitativement les stratégies de production des consonnes simples du français dans une tâche de répétition, et plus particulièrement des consonnes /p t k m n ŋ/, les seules permises en coda en vietnamien et les seules à initier, dans cette langue, les séquences de plusieurs consonnes. Rappelons que l'étude vise la perception et la production des consonnes présentes dans les deux langues afin de dégager un éventuel effet de la syllabe.

5.1.1.1. Méthodologie

5.1.1.1.1. Constitution du corpus

Dix-huit consonnes du français /p t k b d g m n ɲ f s ʃ v z ʒ l ʁ/ ont été sélectionnées pour l'élaboration du corpus. Des pseudo-mots et mots contenant ces consonnes en contexte de la voyelle /a/ ont été construits en prenant en compte les différentes distributions : position initiale, inter-vocalique et finale dans la syllabe (cf. table 5.1). L'utilisation des pseudo-mots vise à éviter un éventuel effet de la fréquence lexicale des stimuli à produire et à percevoir. Pour un contraste CV maximal facilitant le traitement des données, le contexte de la voyelle /a/ a été choisi. Etant donné que /ɲ/ ne figure jamais ni en initiale ni en inter-vocalique dans le lexique du français, un seul pseudo-mot contenant cette consonne en position finale figure par conséquent dans la liste des items construits.

Table 5.1. Liste des mots et pseudo-mots utilisés pour le test de perception et production des consonnes simples du français.

Initiale	Inter-vocalique	Finale	Initiale	Inter-vocalique	Finale
p a	a p a	a p	f a	a f a	a f
t a	a t a	a t	s a	a s a	a s
k a	a k a	a k	ʃ a	a ʃ a	a ʃ
b a	a b a	a b	v a	a v a	a v
d a	a d a	a d	l a	a l a	a l
g a	a g a	a g	z a	a z a	a z
m a	a m a	a m	ʒ a	a ʒ a	a ʒ
n a	a n a	a n	ʁ a	a ʁ a	a ʁ
ɲ a	a ɲ a	a ɲ			a ɲ

Ces pseudo-mots ont été insérés dans la phrase porteuse : « *Tu prononces ... trois fois* ». L'ensemble du corpus est présenté dans l'Annexe V. Le corpus est constitué de 3 répétitions des 52 pseudo-mots insérés dans la phrase porteuse et mises en ordre aléatoire pour le locuteur. Les 156 phrases ont été lues à voix haute par un sujet masculin⁵⁴, 40 ans, de langue maternelle française, origine des Deux-Sèvres, habitant depuis 20 ans à Grenoble. L'enregistrement s'est déroulé dans la chambre sourde du Département Parole et Cognition de GIPSA-lab, Grenoble avec un enregistreur Marantz PMD 670, micro AKG C1000S à directivité cardioïde. Le corpus a été numérisé à 44.1 kHz sous format .wav. Le locuteur a été

⁵⁴ Le choix du locuteur a été fait à partir de l'enregistrement de plusieurs sujets de langue maternelle française (deux hommes et une femme, âgés de 25 à 40 ans). Les enregistrements ont été soumis à l'appréciation de cinq sujets de langue maternelle vietnamienne qui ont voté à l'unanimité pour le locuteur C retenu pour l'expérience.

invité à lire le corpus à un débit normal. Deux pauses ont été faites au cours de la lecture après chaque tiers du corpus. Les 156 pseudo-mots ont ensuite été segmentés manuellement sous Praat et extraits sous forme de liste de sons. Le matériel expérimental est donc de 3 répétitions de 17 pseudo-mots de structure CV, 17 de structure VCV et 18 de type syllabique VC.

5.1.1.1.2. Participants

26 sujets vietnamiens ont participé au test (5 hommes et 21 femmes), locuteurs du dialecte du Nord, âgés de 18 à 20 ans (moyenne d'âge 19 ans pour les hommes et les femmes). Les participants sont des étudiants de première, deuxième et troisième année du Département de Langue et de Civilisation Françaises de l'École Supérieure des Langues Étrangères de l'Université Nationale de Hanoi. L'équilibre des sujets quant au paramètre sexe n'a pas pu être obtenu car la majorité des inscrits dans ces formations sont de sexe féminin. Leur entourage parle le même dialecte et les régions où ils résident pendant plus de 6 mois au Vietnam se situent dans le Nord du pays. Ils ont tous l'anglais comme deuxième langue étrangère.

Le nombre d'années à l'Université ne reflète pas le niveau réel d'expérience en français des sujets. Seul le concours d'entrée à l'Université exige un certain niveau de français. Cependant, les étudiants en première année ont, à leur actif, au moins trois années d'apprentissage du français au lycée. Certains ont reçu un enseignement de FLE depuis l'école primaire, ce qui peut correspondre jusqu'à 12 ans d'apprentissage en milieu scolaire. Les sujets sont donc de niveau intermédiaire ou avancé. À noter que ces étudiants n'ont passé aucun tests de niveau linguistique français selon le Cadre Européen Commun de Référence (TCF ou DALF), puisque ces tests, qui leur servent uniquement pour la poursuite d'étude en France, coûtent très chers au Vietnam.

Pour les besoins de l'expérience, deux groupes de sujets ont été constitués selon leur nombre d'années d'expérience de la langue française avec un seuil de 5 ans : le niveau (+) regroupe 11 étudiants plus avancés qui apprennent le français depuis plus de 5 ans, la plupart d'entre eux ont suivi des formations bilingues depuis l'école primaire, alors que le niveau (-) contient 15 étudiants qui ont moins de 5 ans d'apprentissage.

5.1.1.1.3. Déroulement du test

Avant de passer le test, chaque étudiant devait remplir un questionnaire rédigé en vietnamien. Ce questionnaire (cf. annexe VI) contenait les informations sur leur identité et leur expérience en français. Les instructions à suivre pour les tests de perception leur ont également été donnés.

L'enregistrement pour chaque participant s'est déroulé dans un studio calme de l'Université Nationale de Hanoi. Chaque participant était assis confortablement. L'expérimentateur restait à côté. L'écoute se faisait au casque binaural Head Set HS-95. Les 156 stimuli ont été présentés dans un ordre aléatoire unique pour tous les participants qui étaient informés au préalable qu'un signal de parole, pas nécessairement un mot, allait être présenté dans le casque et qu'ils devaient répéter, sans attendre, ce qu'ils avaient entendu de manière claire et à un débit normal. Un entraînement avec cinq pseudo-mots différents de ceux du test leur était préalablement proposé et la séance de test durait environ 10 minutes pour chaque participant, avec une pause à mi-parcours. L'enregistrement a été numérisé à 16 kHz sous format .wav.

5.1.1.2. Résultats

Les 156 stimuli sonores produits par chacun des 26 étudiants ont été segmentés manuellement et analysés acoustiquement avec Praat[®]. Les répétitions des étudiants, comparées avec les modèles acoustiques en entrée, ont été codées en 1, si la consonne était identifiée correctement car reproduite correctement ou en 0, dans le cas contraire. Ce codage permet de calculer le score de réussite par sujet et par item en moyennant les réponses codées. Les données ainsi obtenues pour tous les items et pour tous les participants ont été analysées statistiquement avec le logiciel SPSS[®].

5.1.1.2.1. Effet de la distribution

Une ANOVA à mesures répétées a été effectuée pour comparer les moyennes de scores de bonnes réponses obtenues pour chacune des positions des consonnes dans le pseudo-mot (initiale, inter-vocalique, finale). Le test met en évidence un effet de position sur l'identification des consonnes [$F(2,48) = 21,218$; $p = 0$]. Les résultats montrent en effet que les sujets perçoivent mieux les consonnes en position inter-vocalique (81,8 %) et initiale (76,2 %) qu'en position finale (67,6 %), quel que soit leur niveau de français (cf. figure 5.1). La différence est significative quand on compare les contrastes en intra-sujet entre la position initiale et finale [$F(1,24) = 14,199$; $p = 0,001$], entre la position inter-vocalique et finale [$F(1,24) = 44,931$; $p = 0$], entre la position initiale et inter-vocalique [$F(1,24) = 6,534$; $p = 0,017$]. Aucune interaction entre la position et le niveau des sujets n'est détectée [$F(2,48) = 0,349$; $p = 0,705$].

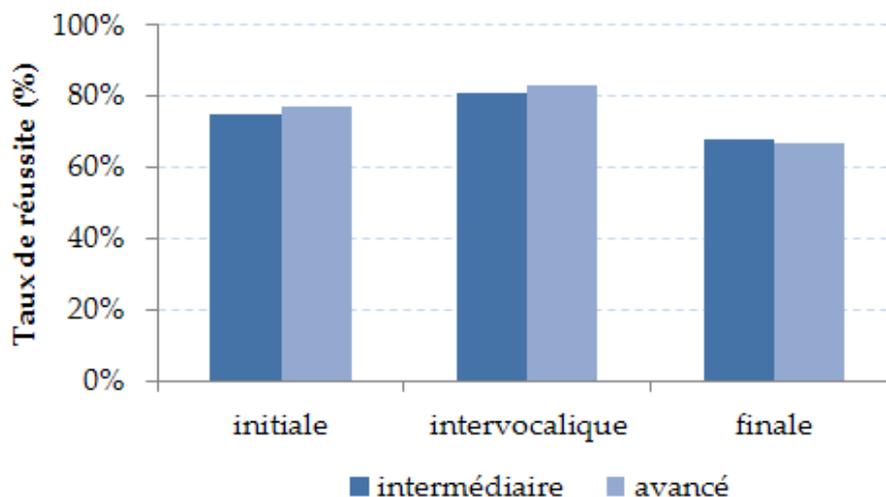


Figure 5.1. Taux de production correcte des consonnes selon leur position dans les pseudo-mots et selon le niveau des étudiants en FLE.

5.1.1.2.2. Effet du mode d'articulation

Le taux de réussite moyen pour chaque consonne est présenté dans les figures 5.2 et 5.3. Les consonnes /ʁ g ʒ/ sont les moins bien reproduites par tous les participants, quel que soit leur niveau, alors que les trois nasales /m n ɲ/ sont les mieux réussies, avec des scores proches de 100 % par tous les étudiants des deux niveaux.

Le score moyen obtenu pour /ɲ/ n'est pas très bon (57,6 % par les apprenants de niveau intermédiaire et 48,5 % par ceux de niveau avancé), peut-être dû à ce qu'il n'apparaît qu'en position finale, la moins performante.

Si on regroupe les consonnes selon leur mode d'articulation, les résultats statistiques montrent un effet significatif global sur le taux de réussite selon le mode [$F(3,61) = 45,09$; $p = 0$]. Les étudiants avancés ne sont pas plus performants que les intermédiaires ($p = 0,614$).

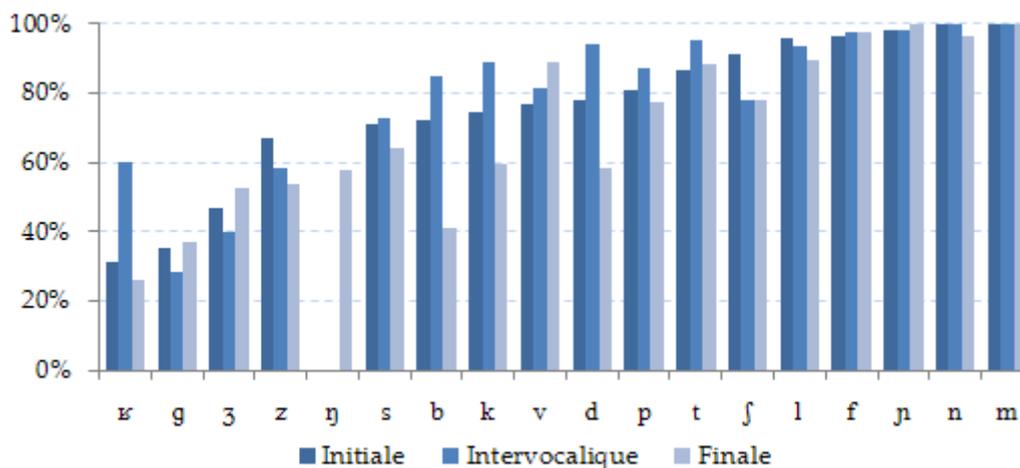


Figure 5.2. Taux de bonnes productions des consonnes simples du français selon leur position dans les pseudo-mots par les étudiants du niveau intermédiaire.

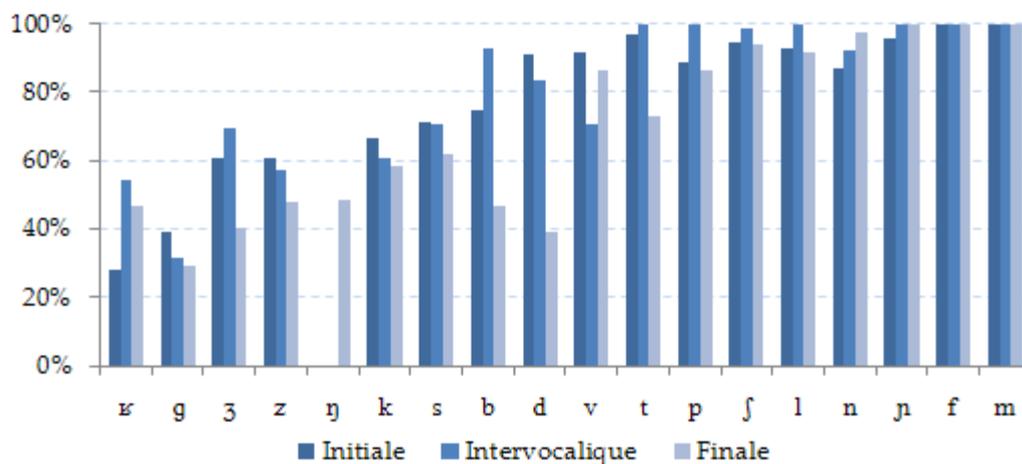


Figure 5.3. Taux de bonnes productions des consonnes simples du français selon leur position dans les pseudo-mots par les étudiants du niveau avancé.

Quel que soit leur niveau, de manière générale, les étudiants identifient mieux les nasales (93,6 %) et les latérales (93,2 %) que les plosives (68,5 %) et les fricatives (69,2 %) (cf. figure 5.4). La différence est significative entre fricatives et latérales ($p = 0$), fricatives et nasales ($p = 0$), plosives et nasales ($p = 0$), plosives et latérales ($p = 0$); mais ce n'est plus le cas entre plosives et fricatives ($p = 0,656$), entre nasales et latérales ($p = 0,976$).

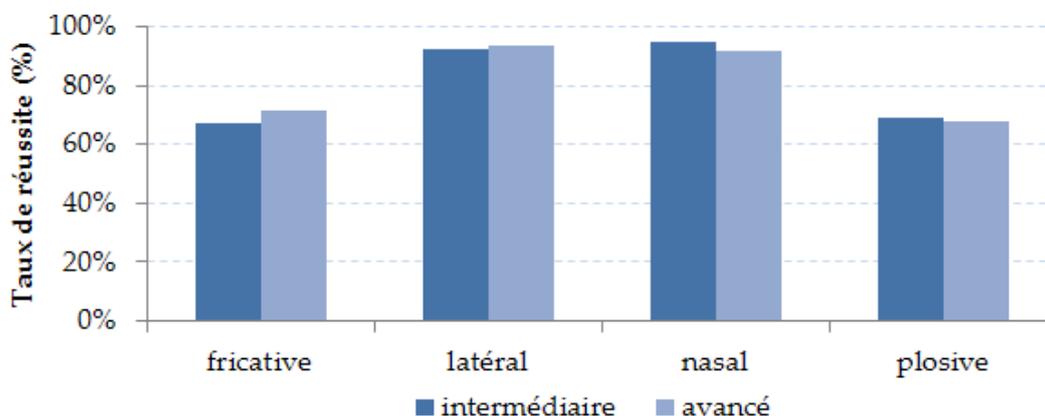


Figure 5.4. Taux de bonnes productions des consonnes selon leur mode d'articulation et le niveau des apprenants.

5.1.1.2.3. Effet du lieu d'articulation

Un effet significatif du score d'identification a été également trouvé selon le lieu des consonnes [$F(3,68) = 51,57$; $p = 0$]. Aucune interaction entre le lieu et le niveau des apprenants en français n'a été détectée ($p = 0,303$). La table 5.2 et la figure 5.5 montrent que, quel que soit leur niveau, les étudiants réussissent mieux à identifier les palatales, devant les labio-dentales et les labiales. Ils rencontrent plus de difficultés à identifier les vélares et uvulaire.

Table 5.2. Taux de bonnes productions des consonnes (du plus fort au plus faible) selon leur lieu d'articulation, tout niveau d'apprentissage confondu.

Lieu	Palatal	Labio-dental	Bilabial	Alvéolaire	Post-alvéolaire	Vélaire	Uvulaire
Cons.	ɲ	f v	p b m	t d n s z l	ʃ ʒ	k g ŋ	ʁ
Taux	98,3 %	90 %	82,4 %	79,3 %	69 %	51,6 %	40,6 %

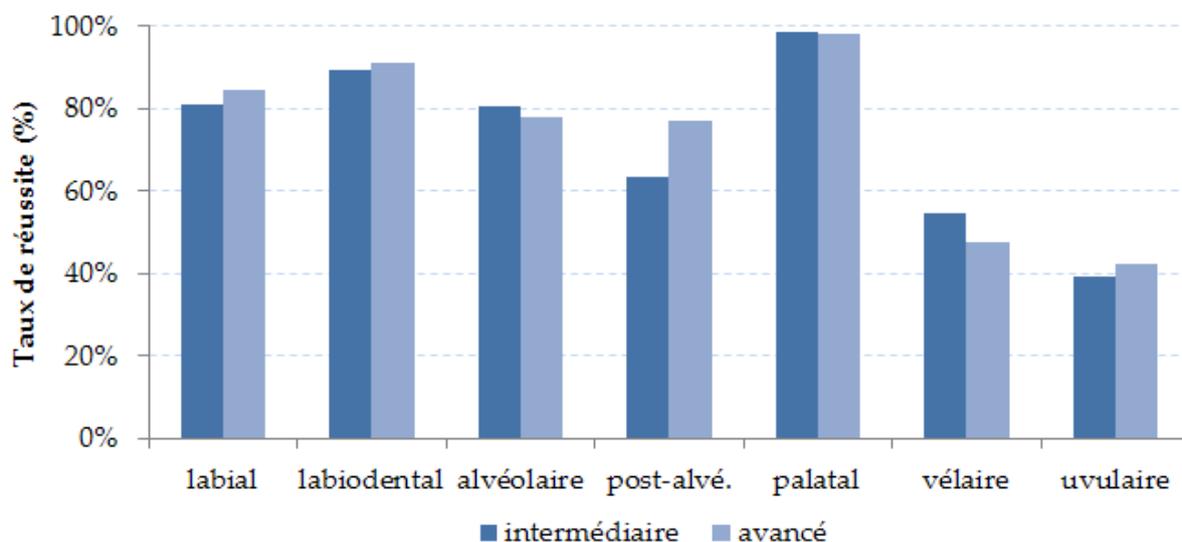


Figure 5.5. Taux de bonnes productions des consonnes selon leur lieu d'articulation et le niveau des étudiants.

5.1.1.3. Effet de similarités entre les deux langues : le cas des consonnes /p t k m n ŋ/

Le cadre de cette étude concerne la perception des consonnes faisant partie des séquences communes aux deux langues. Le vietnamien possède des séquences de consonnes uniquement aux frontières syllabiques de mot ou à l'intérieur de mots composés. Ces séquences se composent donc obligatoirement d'une des six plosives /p t k m n ŋ/, qui peuvent apparaître en coda de syllabe.

Nous remarquons que globalement, les taux de production correcte de ces six consonnes sont au-dessus de 50 %, voire jusqu'à 100 % pour /p/, /t/ dans le groupe de niveau avancé ou /m/, /n/ pour les deux niveaux (cf. figure 5.6). Le niveau d'apprentissage du français des étudiants n'influence pas le taux de répétition correcte des plosives /p t k/ [$F(1,24) = 0,035$; $p = 0,852$], ni des nasales /m n/ [$F(1,24) = 1,17$; $p = 0,29$] dans toutes les positions, ni des nasales /m n ŋ/ en position finale [$F(1,24) = 0,303$; $p = 0,587$]. Aucune interaction entre le niveau des apprenants et leur performance à restituer des consonnes

entendues n'est trouvée, quels que soient les facteurs pris en considération (position, mode ou lieu).

5.1.1.3.1. Effet de la position

Les ANOVA à mesures répétées pour les plosives /p t k/ selon la position montrent le même effet déjà observé pour les dix-huit consonnes testées [$F(2,48) = 7,49$; $p = 0,001$]. Les sujets, quel que soit leur niveau, perçoivent mieux, de manière significative, les plosives /p t k/ en initiale (82 % en moyenne) qu'en finale (72 %) ($p = 0,031$) ; mieux en position inter-vocalique (87 %) qu'en finale (72 %) ($p = 0,001$). Il n'y a pas de différence significative du taux d'identification entre les consonnes en position initiale et inter-vocalique ($p = 0,148$).

L'effet de position n'est pas observé pour les nasales /m n/ : les étudiants des deux niveaux les perçoivent aussi bien en initiale qu'en inter-vocalique ou en finale (97 %, 98 % et 98 % respectivement). Aucune différence significative de score d'identification n'est donc observée en fonction de leur position [$F(2,48) = 0,376$; $p = 0,59$].

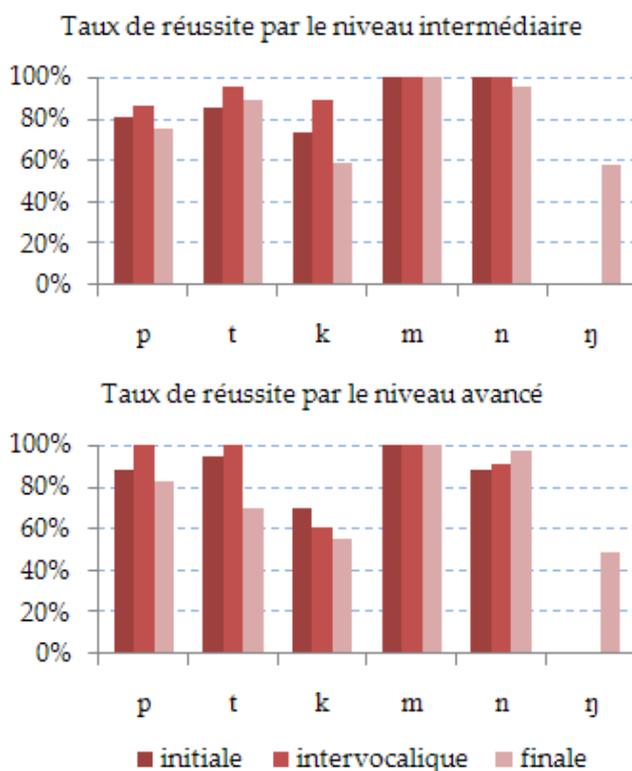


Figure 5.6. Taux de production correcte des consonnes /p t k m n η/ pour les trois positions dans les pseudo-mots par les étudiants des niveaux intermédiaire (en haut) et avancé (en bas).

5.1.1.3.2. Effet du lieu d'articulation

Les bilabiales et alvéolaires sont mieux produites (92 % et 93 % respectivement) que les vélares (65 %) toutes positions confondues. En effet, /k/ est reproduit correctement

seulement dans 57 % des cas, par rapport à /p/ (78 %) et à /t/ (80,8 %). La différence est significative dans les deux cas ($p = 0,015$ et $p = 0,005$ respectivement). De même, la production de /ŋ/ est beaucoup moins performante (53,8 %) par rapport à /m/ (100 %) et /n/ (96 %) ($p = 0$ pour les deux cas). Quelle que soit sa position dans le pseudo-mot, le taux d'identification de la vélaire /k/ est significativement moins fort que celui de la bilabiale /p/ ($p = 0,005$) et l'alvéolaire /t/ ($p = 0$).

5.1.1.3.3. Analyse des erreurs

Une analyse quantitative des erreurs n'étant pas possible du fait du faible nombre d'erreurs pour certaines consonnes, une analyse qualitative des erreurs est présentée table 5.3.

Table 5.3. Traits d'erreurs dans la production des consonnes /p t k m n ŋ/, niveaux d'apprentissage confondus. À gauche, dans la colonne « Réussi », le nombre de réalisations correctes.

Cons.	Réussi	Mode	Lieu	Mode + Lieu	Épen- thèse	Non Relâche- ment	Nasalité	Aspira- tion	Total
p	197		17	3	5	3		8	233
t	209	11	5		3	2		4	234
k	161		10	53	3	3		4	234
m	233								233
n	224	10							234
ŋ	42	2	16		16		2		78
Total	1066	23	48	56	27	8	2	16	1246

Les erreurs concernent principalement la confusion du mode et/ou du lieu d'articulation. Par exemple, les étudiants (avancés comme intermédiaires) confondent le plus souvent [k] avec [tɕ] (63 % des erreurs observées dans la production de [k]) ou avec [t] (12 %). [t] est aussi confondue le plus souvent avec [tɕ] (dans 46 % des erreurs de [t]) ou avec la rétroflexe [ʈ] (23 % des erreurs de [t]). Alors que [p] en initiale comme en finale est confondue le plus souvent avec [t] (47 % des erreurs de [p]).

La nasale [m] est perçue correctement dans toutes les positions, par tous les étudiants des deux niveaux. La confusion entre /n/ et une autre consonne est relativement peu fréquente (seulement 10 cas de confusion sur 234 avec la latérale de même lieu d'articulation [l]). [ŋ] en finale est confondue avec la nasale palatale [ɲ] (33 % des erreurs de [ŋ]). Sinon, dans 44 % des erreurs de [ŋ], les étudiants produisent un bruit d'explosion vélaire très audible de type [k] ou [g] à la fin de ce segment répété [ŋk], [ŋg].

Les erreurs sont présentées plus en détail dans la table 5.4 selon la position des consonnes. À signaler quelques cas dans lesquels les formes répétées des consonnes sont accompagnées d'une forte aspiration, davantage en inter-vocalique que dans les autres positions. Quelques cas de non relâchement ont aussi été observés mais uniquement pour les

plosives en coda. Le taux d’omission concernant les plosives en coda représente 6 % des erreurs concernant les plosives.

Table 5.4. Traits d’erreurs dans les productions des consonnes /p t k m n ŋ/ en fonction de leur position dans les pseudo-mots (totalisés pour les deux niveaux confondus).

Cons.	Position	Réussi	Mode	Lieu	Mode et Lieu	Épen- thèse	Non Relâche- ment	Nasalité	Aspira- tion	Total
p	initiale	68		6	1				3	78
	intervoc.	71		1		2			3	77
	finale	58		10	2	3	3		2	78
t	initiale	70	1	5					2	78
	intervoc.	76							2	78
	finale	63	10			3	2			78
k	initiale	56		3	19					78
	intervoc.	60		3	11				4	78
	finale	45		4	23	3	3			78
m	initiale	78								78
	intervoc.	77								77
	finale	78								78
n	initiale	74	4							78
	intervoc.	75	3							78
	finale	75	3							78
ŋ	finale	42	2	16		16		2		78
Total		1066	23	48	56	27	8	2	16	1246

5.1.1.4. Discussion

L’expérience 1 a mis en évidence des effets de la position sur les performances de production des consonnes simples du français. Des analyses globales ont été menées prenant en compte toutes les consonnes avant d’approfondir les analyses des plosives orales et nasales /p t k m n ŋ/ qui entrent dans la composition des séquences de consonnes en vietnamien.

Les analyses sur l’ensemble des consonnes attestent que les étudiants montrent de grandes difficultés à restituer les consonnes entendues en position finale par rapport aux positions initiale ou inter-vocalique. Ce résultat, en conformité avec les résultats d’études antérieures (Major & Faudree, 1996 ; Anderson, 1987 ; Stockman & Pluut, 1992 ; Benson, 1986 ; Major, 1996 ; Sato, 1984 ; Weinberger, 1987), valide l’Hypothèse de la Différence de Marquage (Eckman, 1977) et montre que l’acquisition des consonnes finales d’une langue seconde reste relativement problématique par rapport à d’autres positions. La même observation a été faite

dans des études développementales qui ont révélé que chez les enfants, les consonnes en coda sont acquises plus tardivement que les consonnes en attaque (Stoel-Gammon, 1985 ; Demuth & McCullough, 2009). Les analyses par mode et lieu d'articulation montrent que les nasales et les latérales sont mieux récupérées et produites que les obstruents par les apprenants des deux niveaux. Cependant, ces apprenants, quel que soit leur niveau, rencontrent plus de difficultés pour les uvulaires et les vélaires (/ʁ g ʒ/ sont les moins bien produites). Ces consonnes, n'existant pas en vietnamien, sont la plupart du temps réalisées sans voisement et produites comme leur homologue sourd (respectivement /x/, /k/, /ʃ/), soit comme un phonème proche faisant partie du répertoire phonologique de la langue de départ (/ʁ/ > [x], /g/ > [ɣ], /ʒ/ > [z]). La majorité des erreurs dans les productions des étudiants, quel que soit leur niveau, trouvent leurs sources dans le système phonologique de la langue de départ. Il semblerait que ces nouveaux sons sont filtrés par le crible phonologique (Troubetzkoy, 1939), lequel perturbe la répétition correcte chez les apprenants, même après plusieurs années de pratique. Il se pourrait que ces nouveaux phonèmes du français ne soient pas « suffisamment » différents des sons de la langue première pour être assimilés en tant que nouvelle catégorie phonologique. Ce phénomène peut être décrit comme le mécanisme de *classification par équivalence* par lequel les catégories phonétiques de la L1 sont susceptibles de bloquer la formation de nouvelles catégories pour les sons proches de la L2, selon le modèle *Speech Learning Model* de Flege (1995). Ces nouveaux sons ont été le plus souvent assimilés à des catégories natives et reproduits comme un bon exemplaire de ces catégories, figurant comme une des possibilités d'assimilation de segments non natifs décrites dans le *modèle d'Assimilation Perceptive* de Best (1995).

L'analyse des données de production des six consonnes /p t k m n ŋ/ qui correspondent aux finales du vietnamien et qui constituent toujours le premier segment des séquences de consonnes dans cette langue, montre un effet de la position pour les orales /p t k/. Les meilleures performances sont toujours observées en position initiale et inter-vocalique, alors qu'en finale, les plosives sont moins bien récupérées et répétées. Par contre, le taux de répétition correcte est identique pour les nasales /m n/ quelle que soit la position. /ŋ/ apparaissant en coda dans les pseudo-mots du corpus présente des scores bien plus faibles par rapport à /m/ et /n/. La raison pour laquelle, dans les répétitions des apprenants, la vélaire /ŋ/ est plus souvent remplacée par la palatale /ɲ/ ou accompagnée par un burst final, semble résider dans l'input (cf. figure 5.7).

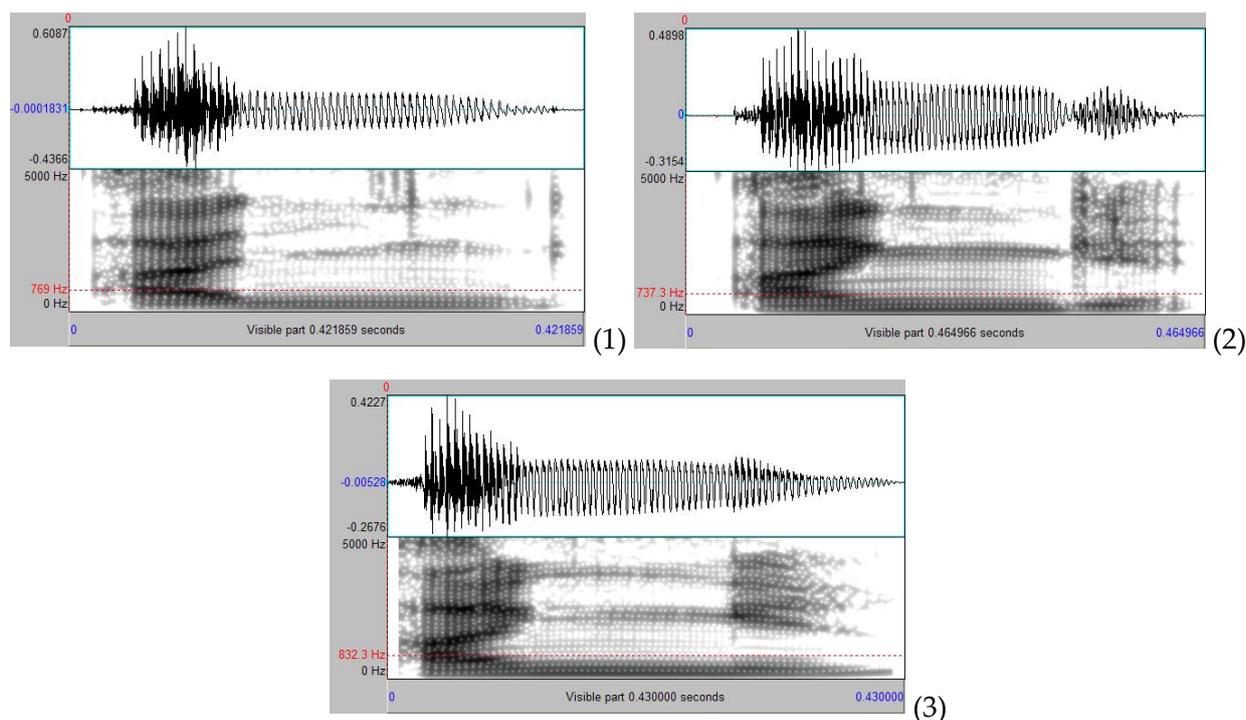


Figure 5.7. Les trois échantillons [aŋ] du locuteur français intégrés dans la liste des stimuli.

Les transitions convergentes de F2 et F3 dans les échantillons (2) et (3) nous indiquent que le segment post-vocalique est de type vélaire, alors que les trajectoires montantes de F2 et F3 dans le premier échantillon (1) montrent que le lieu d'articulation du segment qui suit la voyelle est plus antérieur que vélaire. Ce qui pourrait expliquer le fait que les apprenants ont produit souvent [aŋ] à l'écoute de ce stimulus (1) (cf. table 5.4). En outre, le bruit observé à l'échantillon (2) montre que le locuteur a prononcé une double consonne post-vocalique [ŋŋ], expliquant un taux d'épenthèse consonantique en final à la répétition de ce stimuli (2) par les apprenants des deux niveaux (cf. table 5.4).

Un effet du lieu d'articulation sur le taux de bonne répétition a été également observé : les bilabiales et alvéolaires présentent les meilleurs scores (92 % et 92 %) par rapport aux vélares (61 %) et ce quel que soit le niveau d'apprentissage.

L'analyse qualitative des erreurs en fonction des positions des consonnes montre que celles-ci sont souvent liées à une confusion du mode et/ou du lieu d'articulation (plosif [t], [k] > affriquée [tʃ] ; nasal [n] > latéral [l]). Certains apprenants venant des provinces Hải Phòng, Thái Bình confondent [n] avec [l] dans les répétitions. Cette confusion [n] et [l] est en effet très fréquemment rencontrée dans les parlers de ces provinces (ex. « nón » [nɔn⁵] *chapeau* > [lɔn⁵] *pas de signification* ; « lá » [la⁵] *feuille* > [na⁵] *arbalète*). Cette confusion entraîne des ambiguïtés, la compréhension doit se faire en contexte. Parmi les types d'erreurs de production, les apprenants utilisent un trait phonétique caractéristique des codas de la langue première : les codas non relâchées. Ces erreurs constituent, selon les tenants de

L'Hypothèse de l'Analyse Contrastive (Lado, 1957), un transfert négatif de la langue maternelle. L'influence de la L1 semble importante dans ces erreurs.

Cette première expérience met en évidence l'importance de la position (initiale, inter-vocalique, finale) dans la récupération et la production des consonnes simples en français par des apprenants vietnamiens dans une tâche de répétition. Dans le prolongement de cette première expérience, d'autres tests ont été effectués sur les séquences de consonnes (pseudo-mots et mots) communes aux deux langues.

5.1.2. Expériences 2 : Perception des groupements biconsonantiques en intra- et inter-syllabe du français

Rappelons que les clusters intra-syllabe n'existent pas en vietnamien et que les séquences de consonnes s'y trouvent uniquement en frontière de mot ou frontière de syllabe intra-mot (cf. section 3.3). Les analyses de l'étude acoustico-perceptive des séquences de consonnes du vietnamien traitées au chapitre IV montrent des effets de frontière syllabique et lexicale en perception et production chez des sujets natifs : une frontière de syllabe intra-mot entraîne des différences acoustiques et perceptives par rapport à une frontière syllabique de mot. Plusieurs indices acoustiques de ces frontières sont récupérables en perception. Les résultats du chapitre IV indiquent que des indices sont présents dans la transition VC et dans la coda. Ces résultats de perception des consonnes en séquences en vietnamien pourraient être des éléments du crible phonologique lors de la perception des consonnes en séquences du français.

L'objectif principal de cette expérience est donc d'identifier la prégnance des consonnes du français en fonction de leur position dans les séquences pseudo-mots (expérience 2A) et dans des unités lexicales du français (expérience 2B). Autrement dit, il s'agit de vérifier certaines des hypothèses suivantes : les clusters intra-syllabe sont plus difficiles à identifier que les séquences de consonnes inter-syllabes (Hypothèse 2) ; les clusters en attaque sont plus faciles à identifier que ceux en coda (Hypothèse 3) ; les combinaisons de consonnes plus marquées selon les tendances universelles des langues sont plus difficiles à récupérer (Hypothèse 5).

Expérience 2A : Perception des groupes de consonnes dans les pseudo-mots

5.1.2.1. Méthodologie

5.1.2.1.1. Constitution du corpus

À partir des lexiques syllabés du français et du vietnamien intégrés à la base G-ULSID développée au sein de l'équipe SLD du GIPSA-lab (cf. chapitre III), les séquences de consonnes communes aux deux langues qui possèdent les plus fortes fréquences d'occurrences⁵⁵ ont pu être collectées. Étant donné que les structures syllabiques possibles en vietnamien sont CV et CVC et que par conséquent, les séquences de consonnes se trouvent uniquement à la frontière de mots ou à la frontière syllabique à l'intérieur de mots composés, les séquences communes d'abord sélectionnées sont celles qu'on retrouve de part et d'autre d'une frontière syllabique dans les deux langues. Une fois ces séquences repérées, leurs fréquences respectives ont été calculées en initiale et finale par rapport aux constituants des syllabes du français. Au final, 33 séquences biconsonantiques contenant des segments communs aux deux langues ont été sélectionnées dont 5 existent à la fois en initiale, inter-syllabique et finale ; 11 en inter-syllabique mais aussi rencontrées en français soit à l'initiale, soit en finale de syllabe ; 17 apparaissent seulement en position inter-syllabique. Le corpus contient 54 séquences réparties en trois positions (cf. table 5.5), en contexte de la voyelle /a/.

Table 5.5. Séquences de consonnes communes aux deux langues sélectionnées pour le test, classées selon leur position (en initiale (#CC) et en finale (CC#) uniquement pour le français, en inter-syllabique (C.C) pour les deux langues, le point indique la frontière syllabique attendue en français.

C1	#CC	C.C	CC#
p	pt, pk, pn, ps	p.t, p.k, p.n, p.s	pt, ps
t		t.b, t.m, t.n, t.s, t.l	tm, ts
k	kt, kn, kl, ks	k.t, k.m, k.n, k.f, k.s, k.l	kt, km, kl, ks
m		m.t, m.k, m.b, m.d, m.n, m.v, m.s, m.l	mn
n		n.t, n.k, n.d, n.m, n.f, n.v, n.s, n.l	nt, nd, ns
ŋ		ŋ.b, ŋ.s	ŋs

Les pseudo-mots ont été insérés dans la phrase porteuse : « *Tu prononces ... trois fois* ». L'ensemble du corpus se trouve Annexe 8. Le corpus est constitué de 4 répétitions de 54 phrases et mises en ordre aléatoire pour le locuteur natif du français. Les phrases ont été lues à voix haute par le même sujet de l'expérience précédente. Le locuteur avait pour consigne

⁵⁵ L'ensemble des séquences de consonnes choisies pour le corpus, présenté avec leur nombre d'occurrences dans les deux langues, se trouve dans l'Annexe IV.

de lire les séquences C.C de sorte que les consonnes soient réparties de part et d'autres de frontière syllabique (ex. [at.la] ne doit pas être prononcé de même manière que [a.tla]). Après l'entraînement du locuteur sur plusieurs items, vérifié par l'expérimentateur, l'enregistrement s'est déroulé dans la chambre sourde du Département Parole et Cognition de GIPSA-lab, Grenoble avec l'enregistreur Marantz PMD 670, le micro AKG C1000S à directivité cardioïde. Le corpus a été numérisé à 44.1 kHz sous format .wav. Le locuteur a été invité à lire le corpus à un débit normal. Deux pauses ont été faites au cours de la lecture après chaque tiers du corpus.

De ces données enregistrées ont été segmentées et extraites toutes les séquences de pseudo-mots. Les 54 meilleures réalisations ont été choisies parmi ces séquences, en se basant à la fois sur l'observation du signal acoustique et du spectrogramme. Les séquences réalisées avec hésitation ou contenant soit une friction après une plosive (ex. [ʃ] après /k/) soit un bruit d'inspiration ont été retirées. Le matériel expérimental se compose donc des séquences de consonnes (pseudo-mots) de structure #CCa, aC.Ca et aCC#.

5.1.2.1.2. Participants

Trente-neuf sujets (sept hommes, trente-deux femmes) ont participé au test dont vingt-six de l'expérience précédente. Tous les participants sont des étudiants du Département de Langue et de Civilisation Françaises de l'École Supérieure des Langues Étrangères de l'Université Nationale de Hanoi. Ils sont locuteurs natifs du dialecte du Nord, âgés de 18 à 20 ans (moyenne d'âge de 19 ans pour les hommes comme pour les femmes). Leur entourage parle le même dialecte et les régions où ils résident pendant plus de 6 mois au Vietnam se situent toutes dans le Nord du pays. L'anglais est leur deuxième langue étrangère à l'Université.

Comme pour l'expérience 1, ces étudiants ont été classés en deux groupes selon leur nombre d'années d'apprentissage du français langue étrangère : le niveau avancé contient 20 étudiants qui apprennent le français depuis plus de 5 ans alors que le niveau intermédiaire contient 19 étudiants qui ont moins de 5 ans d'apprentissage.

5.1.2.1.3. Déroulement du test

L'expérience s'est déroulée dans la salle multimédia de l'Université Nationale de Hanoi dans laquelle se trouvent dix ordinateurs fonctionnant sous Windows 1998. Chaque participant était assis face à un ordinateur et équipé d'un des casques binauraux Head Set HS-95, Somic CD-750, Sony MDR-V150.

Le programme utilisé pour le test a été réalisé par R. Carré. Les sujets avaient pour consigne d'écouter un signal, pas nécessairement un mot, puis de choisir le plus rapidement possible, avec possibilité de réécoute des stimuli, la séquence ou la consonne entendue en cliquant avec la souris sur le bouton correspondant à leur choix (cf. figure 5.8). Il s'agit d'un

test à choix fermé. Pour une séquence de pseudo-mot donnée, les choix possibles sont soit l'une des consonnes de la séquence, soit la séquence, soit la séquence biconsonantique inverse. L'ordre de présentation des boutons à l'écran (C_1 , C_2 , C_1C_2 , C_2C_1) est fixé pour tous les stimuli.

Bien que la réécoute de stimuli sonore était possible, les temps de réaction ont été mesurés à partir du début du signal sonore émis. Un entraînement à l'interface et au type de stimuli leur était préalablement proposé avec des stimuli n'appartenant pas au test lui-même. La séance du test durait environ 10 minutes pour chaque participant.

Le test était constitué de 3 répétitions des 54 items présentés dans un ordre aléatoire pour chaque sujet.

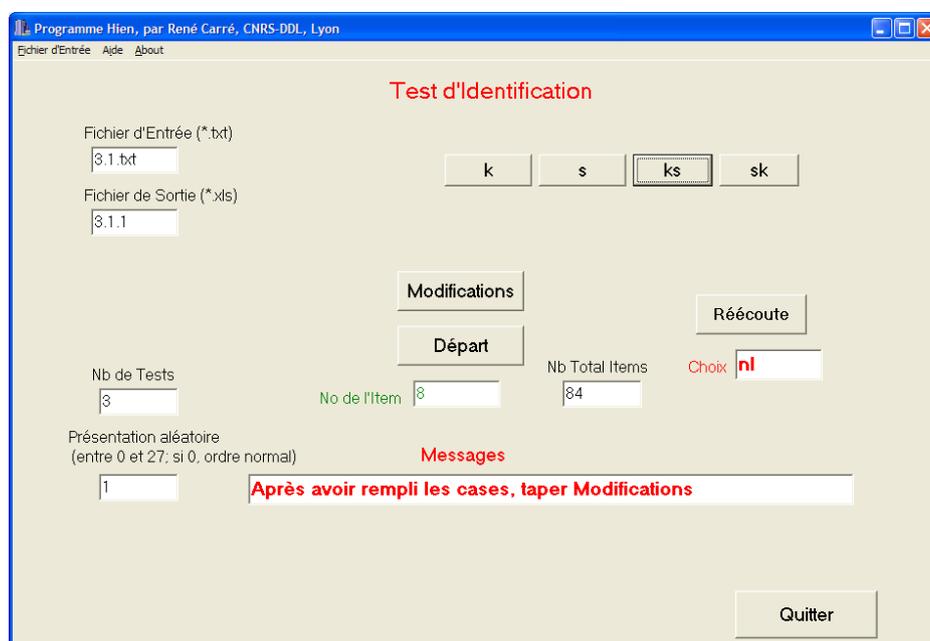


Figure 5.8. Interface du test utilisé pour l'expérience perceptive des séquences de consonnes dans les pseudo-mots.

5.1.2.2. Résultats

Les fichiers de sortie de chaque sujet du programme de test sont sous format Excel. Ces fichiers incluent les paramètres de départ, la succession des choix avec les temps de réponse (TR) correspondants (moyennes et écart-types). Toutes les réponses ont ensuite été rassemblées à l'aide de Matlab avant de pouvoir être traitées statistiquement sous SPSS. Des ANOVA à mesures répétées ont été effectuées avec les facteurs suivants : position dans le pseudo-mot, mode d'articulation, lieu d'articulation, combinaisons de consonnes, niveau des apprenants et temps de réponse.

5.1.2.2.1. Effet de la position

Des analyses ANOVA à mesures répétées ont été effectuées en considérant, comme facteur intra-sujet, la position des séquences et comme facteur inter-sujet le niveau des étudiants ; la variable dépendante étant le score d'identification correcte pour tous les sujets et tous les items. Les séquences consonantiques C₁C₂ ont été observées en fonction de C₁. Cette consonne initiale de séquence appartient à l'ensemble des phonèmes possibles en coda en vietnamien. Rappelons que les séquences de consonnes étant obligatoirement inter-syllabiques dans cette langue, les codas initient forcément les séquences en vietnamien. L'ensemble des résultats est présenté figure 5.9.

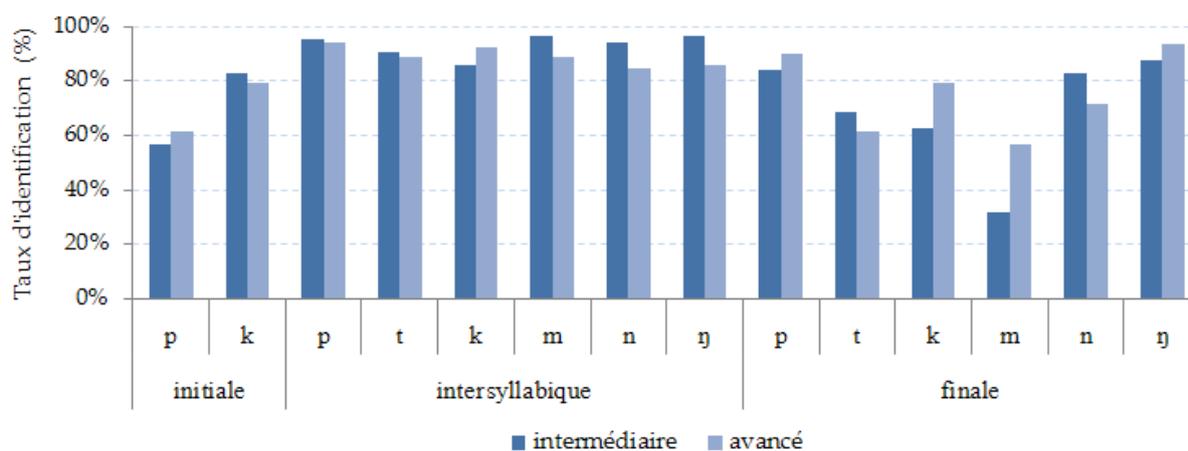


Figure 5.9. Scores d'identification correcte des séquences biconsonantiques en fonction de C₁, de la position dans le pseudo-mot entendu et le niveau des apprenants.

Les résultats statistiques montrent que les apprenants, quel que soit leur niveau, identifient mieux les séquences de consonnes pseudo-mots en position inter-syllabique qu'en finale. La différence de score d'identification a été trouvée significative entre ces deux positions [$F(1,37) = 75,92 ; p = 0$]. Si on considère les consonnes /p k/ car les séquences qu'elles initient sont trouvées dans les trois positions en français et donc dans les pseudo-mots du corpus, un effet significatif entre les positions est effectivement trouvé [$F(2,74) = 27,89 ; p = 0$], ceci quel que soit le niveau des apprenants. Ce résultat est détaillé ci-après. Aucune interaction entre les facteurs position et niveau n'est détectée ($p = 0,166$).

Les étudiants ont mieux reconnu les séquences pseudo-mots /k/+C en position inter-syllabique (89 %) qu'en initiale (81 %) et finale (71 %). Les séquences /p/+C sont aussi mieux identifiées en inter-syllabique (95 %), alors que les apprenants montrent plus de difficultés à reconnaître les séquences en initiale que celles en finale de pseudo-mots (respectivement 59 % vs. 87 %). Des analyses plus fines des contrastes intra-sujet montrent que les différences sont significatives dans tous les cas considérés (cf. table 5.6).

Table 5.6. Résultats statistiques des contrastes intra-sujet en fonction des positions des séquences /p/+C et /k/+C dans les pseudo-mots.

Séquences	Initiale	Inter-syllabique
/p/+C	Initiale	p = 0
	Finale	p = 0,023
/k/+C	Initiale	p = 0,037
	Finale	p = 0,028

Si on considère les séquences selon le mode d'articulation de leur première consonne, les résultats statistiques montrent que les scores des plosives ne sont pas différents significativement de ceux des nasales, quelle que soit la position dans le pseudo-mot, inter-syllabique [F(1,37) = 0,017 ; p = 0,89] ou finale [F(1,37) = 0,014 ; p = 0,907] de la séquence.

Des analyses des contrastes en intra-sujet ont montré qu'en position inter-syllabique, les meilleurs scores des séquences de plosives ont été obtenus par /p/+C (94 % en moyenne des deux niveaux) par rapport à /t/+C (89,7 %) et /k/+C (89 %). La différence entre /p/+C et les autres séquences est significative [F(1,37) = 7,52 ; p = 0,009]. La même tendance des plosives est observée en position finale, les séquences /p/+C sont toujours les mieux identifiées parmi les séquences de plosives [F(1,37) = 27,78 ; p = 0].

Pour les nasales en inter-syllabique, toutes les séquences /m/+C, /n/+C et /ŋ/+C sont bien identifiées par les apprenants des deux niveaux (92,7 %, 89,5 % et 91 % respectivement). En position finale, seules /ŋ/+C restent les mieux identifiées (91 %) par rapport aux autres séquences {/m/+C et /n/+C} [F(1,37) = 39,9 ; p = 0]. /m/+C étant la moins bien reconnue en finale (44 %) parmi les séquences initiées par une nasale [F(1,37) = 42,9 ; p = 0].

Aucun effet significatif de la différence de scores en fonction des lieux des consonnes initiales de clusters (labial, coronal, vélaire) n'a été trouvé entre les deux positions inter-syllabique et finale [F(2,74) = 0,472 ; p = 0,626].

5.1.2.2.2. Temps de réaction

Des analyses statistiques à mesures répétées ont été effectuées pour observer un éventuel effet entre le temps de réaction des bonnes réponses (variable dépendante pour tous les sujets et tous les items), la position des séquences, le mode et le lieu des premières consonnes des séquences. Le facteur intra-sujet est la position des séquences dans les pseudo-mots, le facteur inter-sujet est le niveau des apprenants.

Le résultat montre que le temps de réaction des bonnes réponses n'est pas différent significativement selon la position inter-syllabique et la position finale des séquences testées [F(1,23) = 3,651 ; p = 0,069]. Aucune interaction entre ces deux positions et le niveau des apprenants en FLE n'a été détectée [F(1,23) = 0,068 ; p = 0,79].

Si on considère les consonnes /p k/ car les séquences se trouvent dans les trois positions en français (cf. table 5.5), un effet significatif du temps de réaction entre les trois positions a été trouvé [$F(2,68) = 40,03$; $p = 0$], ceci quel que soit le niveau des étudiants. Le temps de réaction est plus court quand il s'agit des séquences en initiale que celles en inter-syllabique ($p = 0$) et en finale ($p = 0$). Il n'y a pas de différence de temps de réponse entre inter-syllabique et finale ($p = 0,21$). L'interaction entre les facteurs position et niveau n'est pas significative ($p = 0,182$).

Aucune différence significative du temps de réaction selon le mode des premières consonnes des séquences n'a pu être mise en évidence [$F(1,37) = 0,075$; $p = 0,786$]. Par contre, les résultats statistiques montrent un effet significatif du lieu des consonnes sur le temps de réaction [$F(2,74) = 9,986$; $p = 0$]. En effet, les étudiants, quel que soit leur niveau, ont répondu juste dans un délai plus court pour des séquences commençant par les vélaire. La différence des temps de réaction est significative entre les séquences biconsonantiques initiées par une vélaire et celles initiées par une coronale ($p = 0,008$), idem entre vélaire et labiale ($p = 0$), mais pas entre coronale et labiale ($p = 0,055$).

5.1.2.2.3. Types de consonnes dans les combinaisons

La performance d'identification des séquences de consonnes varie de manière significative selon la nature des phonèmes impliqués [$F(3,111) = 39,5$; $p = 0$]. La figure 5.10 montre qu'en début de syllabe, la suite « Muta Cum Liquida » de type Plosive+Latérale présente un taux d'identification très élevé (96,6 % en moyenne pour les deux niveaux), suivie par la séquence Plosive+Plosive (77,2 %) et Plosive+Nasale (67,9 %). La combinaison Plosive+Fricative semble poser plus de problèmes pour les étudiants (48,7 %), quel que soit leur niveau d'apprentissage du français. La différence est significative entre les quatre types de combinaisons en initiale de pseudo-mots. Aucune interaction n'est trouvée entre les types de combinaisons et le niveau des étudiants [$F(3,111) = 0,412$; $p = 0,683$].

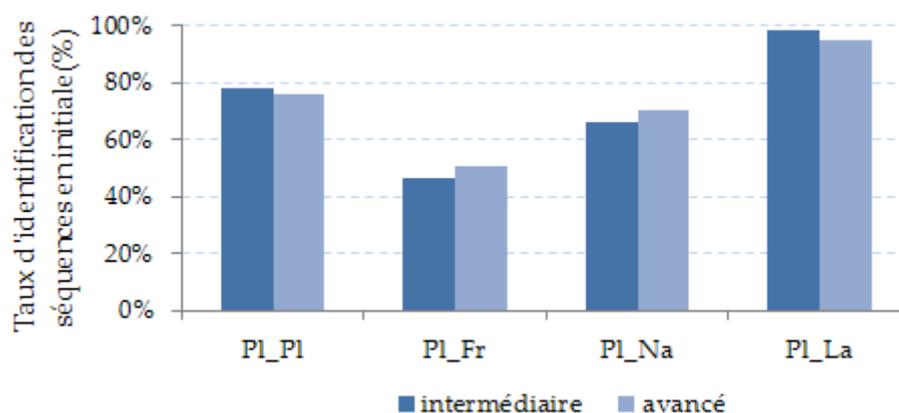


Figure 5.10. Taux d'identification correcte des initiales complexes de pseudo-mot selon le type de combinaison (Pl = plosive, Fr = fricative, Na = nasale, La = latérale).

Concernant les résultats obtenus pour la position finale, les analyses statistiques montrent un effet significatif global pour les séquences impliquant une plosive [$F(3,111) = 3,099$; $p = 0,03$]. On retrouve ici la même tendance observée en initiale : les apprenants identifient mieux la combinaison impliquant une latérale en position C₂ des séquences initiées par une plosive que si la plosive est suivie par une fricative. Le test des contrastes intra-sujets indique une différence significative entre Plosive+Latérale (82 %) et Plosive+Fricative (68 %) ($p = 0,007$). Une interaction entre le type de combinaison et le niveau des étudiants n'est pas détectée pour cette position ($p = 0,42$).

Les scores sont aussi différents de manière significative si on considère les combinaisons impliquant une nasale [$F(2,74) = 27,2$; $p = 0$]. La figure 5.11 montre que les séquences impliquant deux nasales présentent un score bien plus faible de bonne identification (44 %) par rapport à d'autres combinaisons. La suite Nasale+Fricative est la mieux reconnue, parmi les séquences initiées par une nasale ($p < 0,001$ dans tous les cas considérés), quel que soit le niveau de français des apprenants (88,6 % et 89 %).

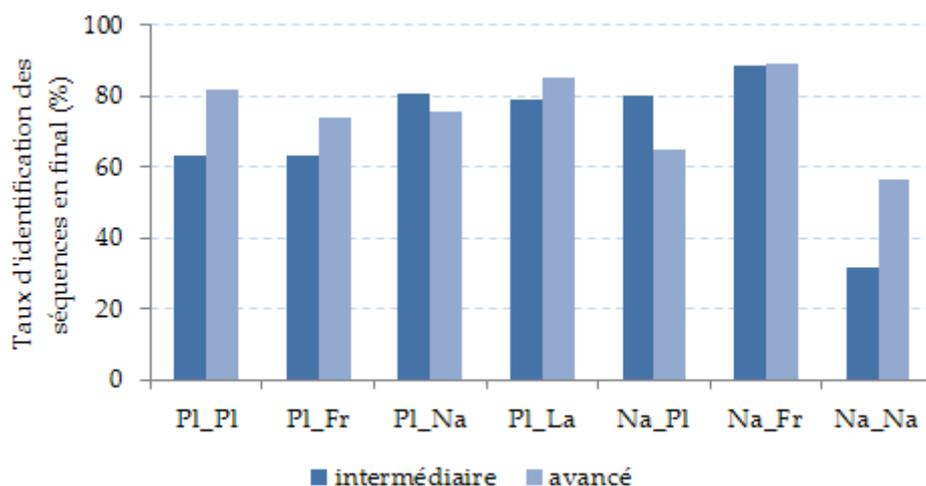


Figure 5.11. Taux d'identification correcte des codas complexes selon le type de combinaison (Pl = plosive, Fr = fricative, Na = nasale, La = latérale).

La figure 5.12 présente le taux d'identification correcte pour la position inter-syllabique selon les types de combinaison. Les résultats statistiques montrent un effet significatif entre les types de combinaisons [$F(3,111) = 3,314$; $p = 0,023$]. La séquence Plosive+Fricative est toujours la moins bien identifiée parmi les combinaisons impliquant une plosive (86 % en moyenne pour les deux niveaux d'apprenants). La différence de scores d'identification correcte entre la séquence Plosive+Fricative et la combinaison la mieux identifiée Plosive+Plosive (93 %) est significative ($p = 0,003$). Il n'y a pas d'interaction entre le niveau des étudiants et le type de combinaison impliquant une plosive ($p = 0,232$).

Bien que la figure 5.12 affiche des scores moyens d'identification correcte des séquences biconsonantiques initiés par une nasale nettement supérieure chez les étudiants de niveaux intermédiaires, aucune différence significative n'est trouvée entre les types de combinaisons impliquant une nasale [$F(3,111) = 1,55$; $p = 0,204$]. Dû à la variabilité des réponses des étudiants avancés (valeur d'écart-type élevée), aucune interaction entre le niveau des étudiants et le type de combinaison des nasales n'est trouvée [$F(3,111) = 0,913$; $p = 0,437$].

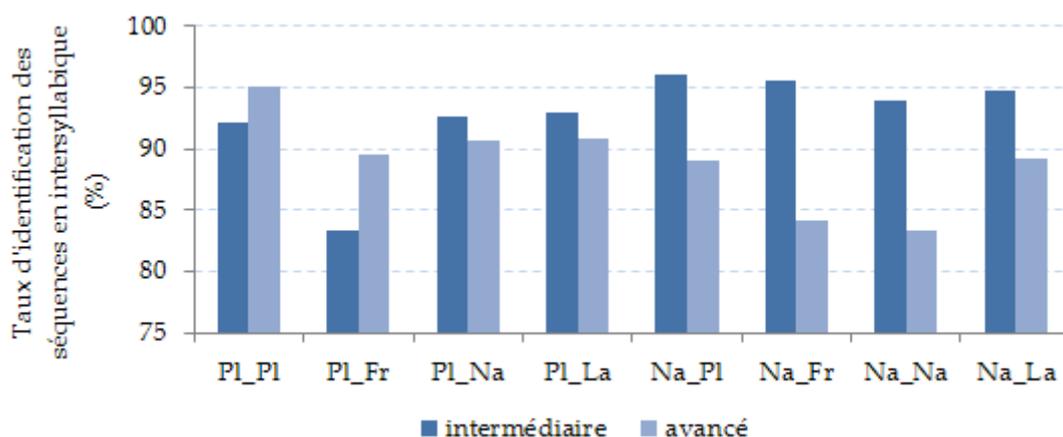


Figure 5.12. Taux d'identification correcte des séquences inter-syllabe selon le type de combinaison (Pl = plosive, Fr = fricative, Na = nasale, La = latérale).

5.1.2.3. Discussion

Cette expérience teste la perception des séquences de deux consonnes communes aux deux langues, par des apprenants vietnamiens de plusieurs niveaux, dans le but de rechercher une éventuelle influence de la position des séquences sur leur identification. Les résultats confirment notre deuxième hypothèse formulée plus haut (cf. page 148) selon laquelle les clusters intra-syllabique (en initiale ou finale) sont plus difficiles à identifier que les séquences inter-syllabique.

Ce résultat correspond aux prédictions de l'*Hypothèse de la Différence de Marquage* d'Eckman (1977), selon laquelle les formes d'une langue seconde qui sont différentes et plus marquées que celles de la langue maternelle sont plus difficiles à acquérir. Conformément à cette hypothèse, nous avons constaté que les groupes de consonnes en finale et en initiale des pseudo-mots correspondant à des schémas phonotactiques du français mais absentes en vietnamien, sont beaucoup moins bien identifiées que celles en position inter-syllabique (score moyen de /p t k m n ŋ/ en finale de 72,5 % vs. 91,2 % en position inter-syllabique ; score moyen de /p k/ en initiale de 70 %, en finale de 79 % vs. 91,9 % en position inter-syllabique), et ce quel que ce soit le niveau des apprenants, intermédiaire ou avancé.

Selon Greenberg (1978), les apprenants d'une langue seconde devraient avoir une plus grande difficulté à acquérir les codas complexes que les attaques complexes, ces dernières étant moins marquées (cf. hypothèse 3). Si on considère les séquences /p/+C et /k/+C qui se trouvent dans les trois positions (cf. table 5.5), nos résultats pour les séquences /k/+C confirment cette hypothèse alors que celle-ci n'est pas validée pour les résultats de /p/+C. Ceci pourrait être expliqué par le fait que /pn/ en attaque a été moins bien identifié et fait chuter le score de /p/+C. La combinaison /pn/ est d'ailleurs défavorisée dans cette position dans les langues du monde, les nasales étant peu rencontrées dans les lexiques au contact immédiat de consonnes plosives comme /p/ (Vallée et collègues, 2009). /pn/ en attaque est en effet la séquence la moins réussie en perception des séquences /p/+C et totalise à elle seule trois fois plus d'erreur par rapport aux erreurs des séquences /p/+C en inter-syllabique et deux fois plus que des erreurs en coda (69 erreurs comptabilisées de /pn/ vs. 24 erreurs des séquences /p/+C en inter-syllabique et 30 en coda). La raison semble résider dans l'input : le burst de [p] étant de durée très brève (8 ms) et d'intensité faible (53,5 dB) pour une plosive en initiale (cf. figure 5.13), dans 98,6 % des erreurs de [pna], ce stimulus est identifié comme [na]. Cette séquence /p/+C est pourtant la mieux identifiée parmi les séquences de plosives en inter-syllabe et en finale de pseudo-mots.

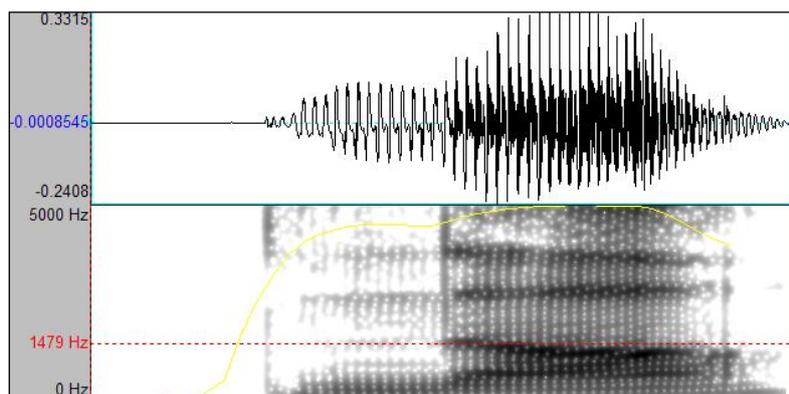


Figure 5.13. L'échantillon [pna] du locuteur français intégré dans la liste des stimuli.

Un effet significatif entre les temps de réaction (seulement pour les cas des bonnes réponses) et le lieu d'articulation des consonnes a été mis en évidence, quelle que soit la position des séquences et quel que soit le niveau des apprenants du français. Les apprenants répondent plus vite quand il s'agit des séquences commençant par les vélares que par d'autres lieux d'articulation. Le mode n'a pas d'impact sur le temps de réponse. À l'aide de la plateforme d'exploitation des lexiques syllabés de G-ULSID, l'interrogation du nombre d'occurrences des séquences vélaire+C dans le lexique des deux langues, toutes positions confondues, a livré que les séquences commençant par les vélares /k/ ou /ŋ/ sont les plus fréquentes des séquences testées dans le lexique du vietnamien, ainsi que /k/ en français (respectivement 45 % et 56 % des séquences de /p t k m n ŋ/+C). Ce qui pourrait expliquer une réaction plus rapide des sujets. Ce résultat reconferme les études antérieures de Vitevich et Luce (1999,

2005) et de Vitevich, Luce, Charles-Luce et Kemmerer (1997) qui ont aussi montré que, à la répétition de pseudo-mots, les locuteurs répondent plus vite lorsque les items contiennent des séquences de phonèmes plus fréquents dans le lexique. Pourtant, le temps de réaction le plus rapide est observé pour /k/+C alors que son taux d'identification est moins élevé que /p/+C en inter-syllabique (88,9 % vs. 94,9 %) et en coda (71% vs. 87 %). Ce résultat reste à discuter.

L'analyse des combinaisons de consonnes les moins bien reconnues par les apprenants des deux niveaux montre que Plosive+Fricative (/ps/, /ks/, /ts/) pose plus de problème d'identification quelle que soit la position initiale, inter-syllabique ou finale. Selon Greenberg (1978), toute langue qui possède en début de syllabe une suite Plosive+Plosive, possède aussi une suite Plosive+Fricative. L'*Hypothèse de la Différence de Marquage* prédit qu'une structure X est marquée par rapport à Y si la présence de X dans une langue implique la présence de Y. La séquence Plosive+Plosive est donc une combinaison marquée par rapport à la suite Plosive+Fricative. Plosive+Plosive devrait être plus difficile à acquérir selon cette hypothèse que Plosive+Fricative. Or le score d'identification des suites Plosive+Plosive est significativement meilleur que celui des combinaisons Plosive+Fricative (77,2 % vs. 48,7 %). L'aspect prédictif de l'HDM ne semble pas marcher dans ce cas.

En position finale, la combinaison Nasale+Nasale /mn/ possède le moins bon score d'identification correcte parmi les séquences initiées par une nasale. Cette suite de nasales, très peu fréquente en français, existe seulement dans 4 des quelques 135 000 mots français de la base *Lexique*⁵⁶. Greenberg (1978) atteste que la présence de cette séquence Nasale+Nasale dans les langues implique la présence d'autres séquences Nasale+Obstruente. /mn/ est donc plus marquée et plus difficile à acquérir en conformité avec l'*Hypothèse de la Différence de Marquage*, par rapport à /nd/, /ns/, /nt/ ou /ŋs/ testées dans notre étude.

La combinaison que les étudiants ont le mieux réussi à identifier en initiale et finale est la séquence Plosive+Latérale, ceci quel que soit le niveau des étudiants. Kühnert et Hoole (2006), dans une étude sur la cohésion temporelle des groupes C+/l/ initiaux en français, expliquent que « la production de la liquide [l] n'implique pas une constriction complète dans le conduit vocal et ne masque pas les informations perceptives éventuelles de la consonne précédente, de la façon dont une plosive le ferait. Ainsi, s'il y en a, les problèmes de récupérabilité perceptuelle sont plus faibles dans le cas d'une production Plosive+Liquide que dans une production Plosive+Plosive ». Dans notre étude, en position inter-syllabique, cette séquence de consonnes Plosive+Latérale reste une des meilleures séquences identifiées par les apprenants des deux niveaux. À noter que la séquence Plosive+Plosive, étant la combinaison la plus fréquente des séquences initiées par une plosive en inter-syllabique du vietnamien (cf. section 3.3.2), récolte aussi un très bon score d'identification par les étudiants.

⁵⁶ Lexique est une base de données lexicales libre, accessible à l'adresse <http://www.lexique.org/>

L'influence de la L1 pourrait donc expliquer ce meilleur score de la séquence Plosive+Plosive par rapport à Plosive+Fricative en inter-syllabique. Concernant les séquences impliquant une nasale, la combinaison Nasale+Fricative est la plus réussie par les étudiants en position finale. Il n'y a pas une préférence significative entre des séquences de nasales en position inter-syllabique.

Expérience 2B : Perception des groupements biconsonantiques dans des mots du français

À la suite de l'expérience 2A portant sur l'identification des consonnes en séquence du français dans des pseudo-mots, l'objectif de cette expérience 2B est de prolonger l'étude sur des items lexicaux du français. Plus précisément, il s'agit avec cette expérience de tester perceptivement les effets de la position dans le mot sur la récupération des consonnes par des apprenants vietnamiens du FLE.

Table 5.7. Unités lexicales du français sélectionnées pour l'expérience (cf. table 5.5) et classées selon la position des groupes de consonnes dans le mot et dans la syllabe.

#CC	Graphique	Transcription API	C . C	Graphique	Transcription API	CC#	Graphique	Transcription API
p n	<i>pneu</i>	p n ø	p . n	<i>hypnose</i>	i p . n ɔ z			
p s	<i>psychose</i>	p s i . k ɔ z	p . s	<i>capsule</i>	k a p . s y l	p s	<i>ellipse</i>	e . l i p s
	<i>psychique</i>	p s i . ʃ i k		<i>apside</i>	a p . s i d		<i>laps</i>	l a p s
p t	<i>ptose</i>	p t ɔ z	p . t	<i>capture</i>	k a p . t y ʁ	p t	<i>apte</i>	a p t
				<i>optique</i>	ɔ p . t i k		<i>adepte</i>	a . d e p t
			p . k	<i>chapka</i>	ʃ a p . k a			
			t . m	<i>wattman</i>	w a t . m a n	t m	<i>rythme</i>	ʁ i t m
			t . l	<i>atlas</i>	a t . l a s			
				<i>atteler</i>	a t . l e			
			t . n	<i>ethnique</i>	ɛ t . n i k			
				<i>ethnie</i>	ɛ t . n i			
k l	<i>climat</i>	k l i . m a	k . l	<i>check-list</i>	ʃ e k . l i s t	k l	<i>cycle</i>	s i k l
	<i>cloche</i>	k l ɔ ʃ	k . b	<i>milk-bar</i>	m i l k . b a ʁ		<i>boucle</i>	b u k l
		<i>snack-bar</i>		s n a k . b a ʁ	k m	<i>drachme</i>	d ʁ a k m	
k n	<i>knout</i>	k n u t	k . n	<i>acné</i>	a k . n e			
				<i>bloc-notes</i>	b l o k . n ɔ t			
			k . s	<i>taxi</i>	t a k . s i	k s	<i>phénix</i>	f e . n i k s
				<i>vexer</i>	v ɛ k . s e		<i>taxe</i>	t a k s
			k . t	<i>acteur</i>	a k . t œ ʁ	k t	<i>acte</i>	a k t
				<i>dictée</i>	d i k . t e		<i>insecte</i>	ẽ . s e k t
			k . f	<i>bec-fin</i>	b ɛ k . f ẽ			
				<i>stockfisch</i>	s t o k . f i ʃ			
			m . n	<i>gymnase</i>	ʒ i m . n a z	m n	<i>hymne</i>	i m n
			m . b	<i>rumba</i>	ʁ u m . b a			
			n . t	<i>sprinter</i>	s p ʁ i n . t e	n t	<i>sprint</i>	s p ʁ i n t
						n d	<i>weekend</i>	w i . k ɛ n d
						n s	<i>jeans</i>	d ʒ i n s

5.1.2.4. Méthodologie

5.1.2.4.1. Corpus

À partir des séquences de consonnes communes aux deux langues établies pour les besoins de l'expérience précédente, a été établie une liste d'unités lexicales du français comportant les séquences biconsonantiques. Des mots possédant plusieurs groupes de consonnes ont été écartés afin de ne pas perturber l'identification des clusters ciblés. Les mots retenus sont monosyllabiques ou dissyllabiques. Au final, 48 mots du français ont été sélectionnés pour le test (cf. table 5.7). Ces mots totalisent ensemble 19 clusters répartis en trois positions (#CC initiale, C.C inter-syllabique et CC# finale). La consonne /ŋ/ ne fait pas partie du corpus de cette expérience 2B en raison du nombre très restreint de clusters /ŋ/+C dans le lexique du français. En effet, la seule entrée trouvée dans le lexique syllabé du français de la base G-ULSID contenant /ŋs/ en final est « leggings ». Cette séquence a été écartée en raison de sa faible occurrence. Les mots d'emprunt récents ont été exclus de l'analyse (ex. come-back).

Chaque unité lexicale a été insérée dans la phrase porteuse « *Tu prononces ... trois fois* ». Le corpus est constitué de trois répétitions de quarante-huit phrases mises en ordre aléatoire. L'ensemble du corpus se trouve annexe IX. Les phrases ont été lues à voix haute par le même sujet masculin des expériences précédentes. L'enregistrement s'est déroulé dans la chambre sourde du Département Parole et Cognition du GIPSA-lab, Grenoble avec l'enregistreur Marantz PMD 670, micro AKG C1000S à directivité cardioïde. Le corpus a été numérisé à 44.1 kHz sous format .wav. Le locuteur a été invité à lire le corpus à un débit normal. Deux pauses ont été faites au cours de la lecture après chaque tiers du corpus.

À partir du corpus enregistré, les mots cibles ainsi que les groupements de consonnes cibles ont été segmentés. Les quarante-huit meilleures versions (produites sans hésitation, sans bruit de friction après plosive) ont été choisies en nous basant à la fois sur l'observation du signal acoustique et du spectrogramme. Le matériel expérimental se compose donc des clusters et des séquences biconsonantiques (mots) de structure #CCV, VC.CV et VCC#.

5.1.2.4.2. Participants

Ce sont les mêmes sujets de l'expérience 2A qui ont passé ce test, 19 sujets de niveau intermédiaire et 20 sujets de niveau avancé.

5.1.2.4.3. Déroulement du test

L'expérience 2B s'est déroulée dans la même salle multimédia de l'Université Nationale de Hanoi que l'expérience précédente. Après la passation du test d'identification des clusters pseudo-mots, une pause de 10 minutes entre les deux expériences a été imposée à tous les

participants. Le protocole de test est identique à celui de l'expérience 2A. Les sujets avaient pour consigne d'écouter, puis choisir la séquence ou la consonne entendue en cliquant sur le bouton correspondant. La réécoute du stimulus était possible. Le temps de réaction a été calculé à partir du début du signal sonore émis. Le test était constitué de 3 répétitions de 48 items présentés dans un ordre aléatoire pour chaque sujet.

5.1.2.5. Résultats

Comme pour l'expérience précédente, les résultats sur la perception des clusters et des séquences de consonnes sont présentés en fonction des facteurs suivants : selon leur position dans le mot et dans la syllabe, selon le mode et le lieu d'articulation des premières consonnes impliquées dans les séquences, le niveau des apprenants, le temps de réaction, le type de consonnes C₁ et C₂ impliquées dans la combinaison.

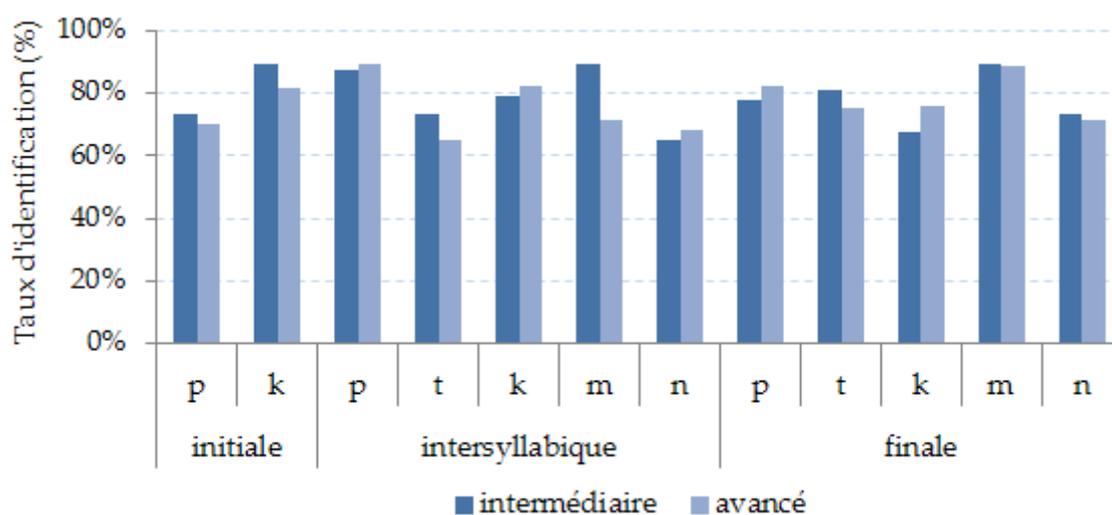


Figure 5.14. Scores d'identification correcte des clusters et des séquences de consonnes du français en fonction de la position de C₁ et du niveau des apprenants.

5.1.2.5.1. Effet de la position

Des analyses ANOVA à mesures répétées ont été effectuées en considérant, comme facteur intra-sujet, la position des séquences et comme facteur inter-sujet, le niveau des apprenants, la variable dépendante étant le score d'identification pour tous les sujets et tous les items. L'ensemble des résultats d'identification des clusters est présenté figure 5.14.

Les résultats statistiques ne présentent pas de différence significative de score d'identification des séquences de consonnes entre les positions inter-syllabique et finale [$F(1,37) = 0,32$; $p = 0,575$].

Si on considère les consonnes /p/ et /k/ qui sont présentes en français dans les trois positions initiale, inter-syllabique et finale, un effet significatif entre les positions est trouvé [$F(2,74) = 5,175$; $p = 0,008$], ceci quel que soit le niveau des apprenants (aucune interaction

entre les facteurs position et niveau [$F(2,74) = 2,414$; $p = 0,096$]. Les étudiants ont mieux identifié significativement les séquences /p/+C et /k/+C en position inter-syllabique (84,5 %) qu'en finale (75,9 %) [$F(1,37) = 22,2$; $p = 0$]. Il n'y a pas de différence significative entre les positions initiale et finale ($p = 0,393$), ni entre initiale et inter-syllabique ($p = 0,054$).

Si on considère les séquences biconsonantiques selon le mode d'articulation des consonnes qui les initient, les résultats statistiques montrent que les scores de plosives ne sont pas différents significativement de ceux des nasales, quelle que soit leur position dans le mot : inter-syllabique [$F(1,37) = 2,001$; $p = 0,166$] ou finale [$F(1,37) = 0,187$; $p = 0,668$].

Un effet significatif de différence de scores en fonction du lieu d'articulation a été trouvé [$F(2,74) = 10,984$; $p = 0$]. En effet, les clusters contenant une labiale (84 %), quel que soit leur position, sont mieux identifiés par rapport à d'autres types de lieux [$F(1,37) = 26,6$; $p = 0$].

5.1.2.5.2. Temps de réaction

Des analyses statistiques à mesures répétées ont été effectuées pour observer un éventuel effet entre le temps de réaction et la position des séquences dans l'unité lexicale, le mode et le lieu des premières consonnes des séquences. La variable dépendante est le temps de réaction pour tous les sujets et tous les items, le facteur intra-sujet est la position des séquences, le facteur inter-sujet est le niveau des apprenants en français.

Le temps de réaction des réponses correctes est plus court pour les clusters en coda finale de mot (2,4 secondes en moyenne quelle que soit la consonne) que pour ceux en inter-syllabique (2,7 secondes). Cette différence est significative [$F(1,29) = 10,354$; $p = 0,003$]. L'interaction entre le niveau des étudiants et le temps de réaction n'est pas significative ($p = 0,812$).

L'effet de position est observé également quand on considère /p/ et /k/ [$F(2,76) = 11,59$; $p = 0$]. Comme les résultats obtenus dans l'expérience 2A, les apprenants, quel que soit leur niveau, répondent significativement plus vite lorsque ces consonnes initient des clusters en initiale [$F(1,37) = 15,94$; $p = 0$] et moins vite quand il s'agit des séquences inter-syllabiques [$F(1,37) = 6,8$; $p = 0,013$].

Il n'y a pas de différence significative du temps de réaction selon le mode des consonnes [$F(1,37) = 0,103$; $p = 0,75$]. Par contre, les résultats statistiques ont présenté un effet significatif si on considère le temps de réaction selon le lieu des consonnes [$F(2,76) = 9,8$; $p = 0$]. En effet, les étudiants, quel que soit leur niveau, ont donné des réponses exactes en un temps plus court quand il s'agit des séquences commençant par une labiale [$F(1,38) = 23,37$; $p = 0$]. Alors que dans l'expérience 2A, les vélaires avaient le meilleur score.

5.1.2.5.3. Types de consonnes dans les combinaisons

La performance de l'identification des clusters en initiale varie de manière significative selon les phonèmes impliqués [$F(3,111) = 7,16$; $p = 0$]. En effet, la figure 5.15 montre que la combinaison Plosive+Fricative est moins bien identifiée par les apprenants des deux niveaux (64,9 % en moyenne) [$F(1,37) = 13,15$; $p = 0,001$]. La séquence Plosive+Latérale présente, comme dans l'expérience 2A portant sur des pseudo-mots, un taux d'identification élevé (82,5 % en moyenne pour les deux niveaux). Elle ne présente pas de différence significative avec la combinaison Plosive+Nasale qui est perçue à 88,8 % en moyenne [$F(1,37) = 2,36$; $p = 0,133$]. Le taux d'identification correcte n'est pas différent significativement entre les deux types de combinaisons Plosive+Plosive et Plosive+Fricative [$F(1,37) = 0,952$, $p = 0,335$]. Aucune interaction n'est trouvée entre le niveau des apprenants et les types de combinaisons ($p = 0,356$), bien que la figure affiche de meilleurs scores d'identification des combinaisons Plosive+Nasale, Plosive+Latérale en initiale par des apprenants de niveau intermédiaire.

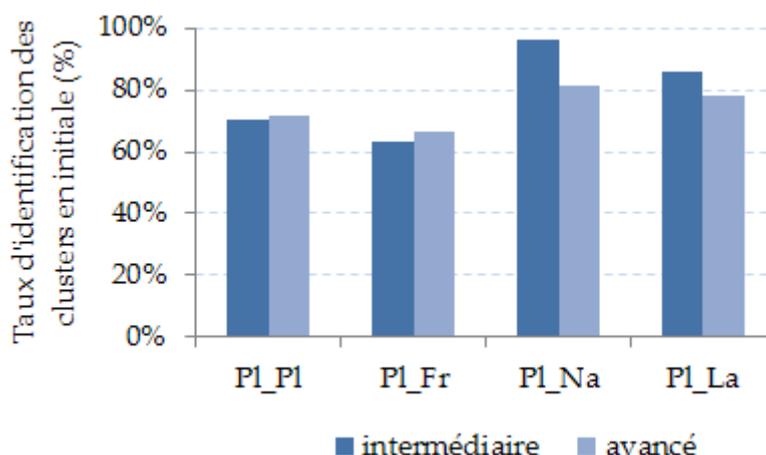


Figure 5.15. Taux d'identification correcte des clusters en initiale selon le type de combinaison (Pl = plosive, Fr = fricative, Na = nasale, La = latérale).

Pour la position finale, les analyses statistiques ne présentent pas de différence significative entre les clusters impliquant une plosive initiale [$F(3,111) = 1,795$; $p = 0,152$]. Une interaction entre le type de combinaisons des plosives et le niveau français des apprenants n'est pas trouvée dans cette position ($p = 0,195$).

Si on considère les combinaisons impliquant une nasale, un effet significatif des scores de bonnes réponses est mis en évidence [$F(2,74) = 6,125$; $p = 0,003$]. La figure 5.16 montre que le cluster impliquant deux nasales présente significativement un score plus élevé (88,9 %) que d'autres combinaisons [$F(1,37) = 13,468$; $p = 0,001$]. Le score d'identification n'est pas différent significativement entre les combinaisons Nasale+Plosive et Nasale+Fricative ($p = 0,542$).

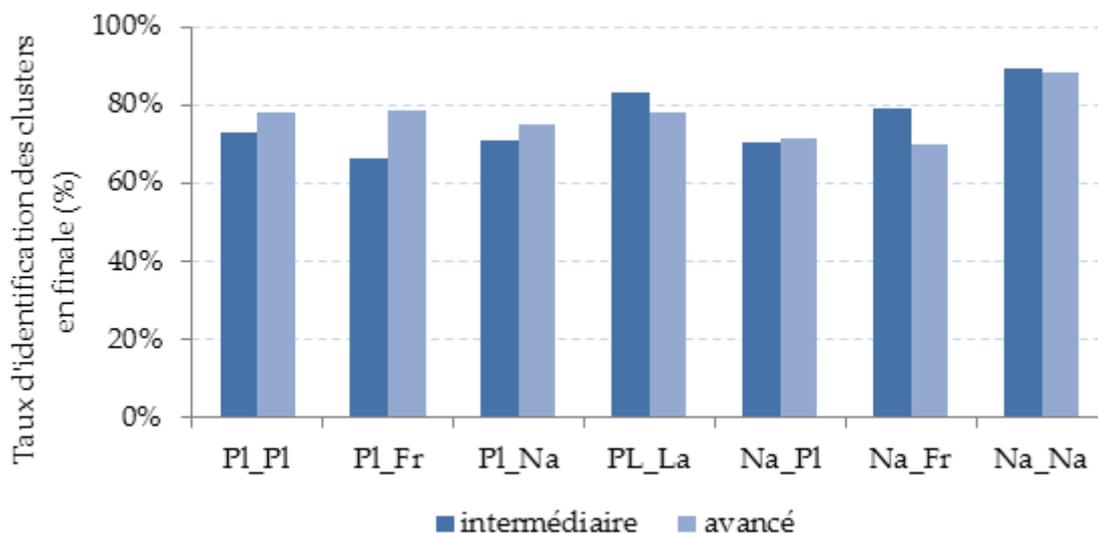


Figure 5.16. Taux d'identification correcte des clusters en finale selon le type de combinaison (P1 = plosive, Fr = fricative, Na = nasale, La = latérale).

La figure 5.17 présente le taux d'identification des séquences consonantiques inter-syllabe selon les types de consonne impliquée.

Les résultats statistiques montrent globalement un effet significatif de la différence de scores en fonction des types de combinaisons impliquant une plosive [$F(3,111) = 4,128$; $p = 0,008$]. Des analyses en contraste intra-sujet indiquent que la combinaison Plosive+Plosive est significativement la mieux identifiée (86,7 %) dans cette position [$F(1,37) = 18,18$; $p = 0$]. Les autres combinaisons Plosive+Fricative (identifiée à 78,9 %), Plosive+Nasale (74,8 %), Plosive+Latérale (77,5 %) ne présentent pas de différence significative au niveau des scores de réussite ($p > 0,05$ dans tous les cas considérés). Il n'y a pas interaction entre le niveau des apprenants en français et les types de combinaison testés dans cette position ($p = 0,142$).

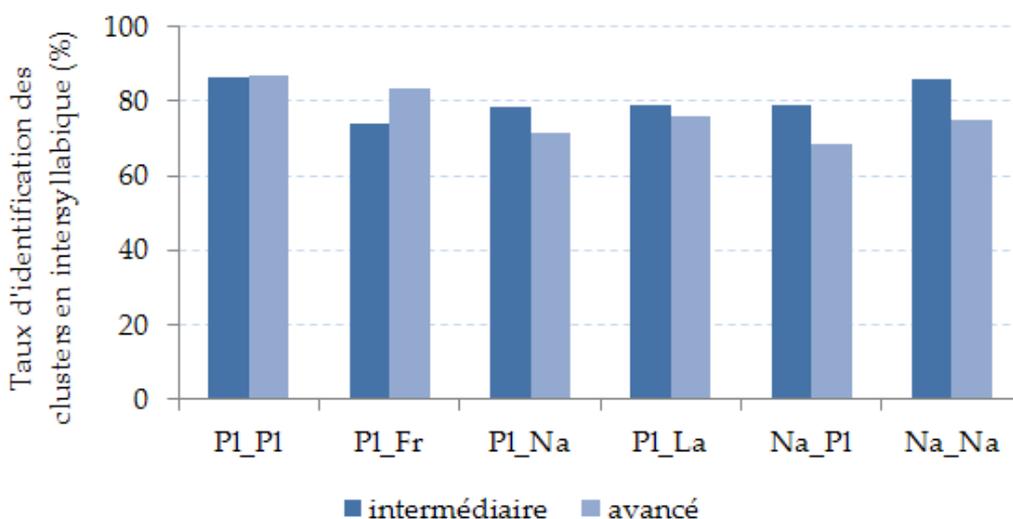


Figure 5.17. Taux d'identification correcte des séquences consonantiques inter-syllabe selon leur mode articulatoire (P1 = plosive, Fr = fricative, Na = nasale, La = latérale).

Aucune différence significative n'est trouvée entre les deux types de combinaisons impliquant une nasale (Nasale+Plosive, Nasale+Nasale) [$F(1,37) = 1,007$; $p = 0,322$].

5.1.2.6. Discussion

Rappelons que l'objectif de cette expérience 2B est de chercher l'effet de la position dans la syllabe et à la frontière intra-mot. Cette expérience a consisté à tester la perception de clusters et de séquences de consonnes du français par des apprenants vietnamiens de deux niveaux. Contrairement à l'expérience 2A sur les séquences pseudo-mots, aucun effet significatif n'a été observé dans la différence de scores d'identification entre les séquences en position inter-syllabique et les clusters en coda (le score moyen de toutes les consonnes et des deux niveaux d'apprentissage confondus est respectivement de 77 % et 78 %). Cependant, si on considère les clusters contenant /p/ et /k/ en C₁, et qui en français existent dans les trois positions, les résultats en contrastes intra-sujet montrent que les apprenants ont identifié significativement mieux (mais avec un temps de réaction plus long) ces séquences en position inter-syllabique qu'en finale (84,5 % *vs.* 75,8 %). La différence de score n'est plus significative entre initiale et finale (78,5 % *vs.* 75,8 %) contrairement à ce qui a été trouvé avec les pseudo-mots.

Ce résultat correspond donc en partie aux prédictions de l'*Hypothèse de la Différence de Marquage* (Eckman, 1977) selon laquelle les clusters en coda, inexistants en vietnamien et plus marqués dans les langues du monde (Rousset, 2004), sont plus difficiles à récupérer à partir du signal audio que les séquences réparties de part et d'autre d'une frontière syllabique. Rappelons que ce dernier cas de figure est celui rencontré en vietnamien. Les apprenants, quel que soit leur niveau, rencontrent plus de difficultés à percevoir les clusters dans les mots tels que « *adepte* » [adept] ou « *insecte* » [ɛ̃sɛkt] que dans les mots comme « *optique* » [op.tik] ou « *dictée* » [dik.te]. Toujours selon l'HDM, les apprenants d'une langue seconde devraient avoir une plus grande difficulté à acquérir les codas complexes que les attaques complexes, ces dernières étant moins marquées (Greenberg, 1978). Cette hypothèse se voit rejetée par nos résultats en perception des clusters /p/+C et /k/+C qui ne montrent pas de différence significative de scores d'identification entre les attaques complexes et les codas complexes. Rappelons que le résultat obtenu sur les séquences pseudo-mots /p/+C de l'expérience 2A montre même que ces séquences en position coda sont mieux récupérées que celles en attaque (87 % *vs.* 59 %).

Le mode d'articulation des premières consonnes des séquences et des clusters n'a pas d'impact sur la performance de leur perception. Que le cluster ou la séquence commence par une nasale ou une plosive, le score ne présente pas de différence significative. Par contre, concernant le facteur lieu d'articulation, les étudiants identifient significativement mieux et plus vite un cluster commençant par une labiale que par une vélaire ou une coronale. Alors

que les apprenants ont répondu juste dans un délai plus court pour les séquences de consonnes pseudo-mots commençant par une vélaire.

Le degré de difficulté de perception des groupes de consonnes du français semble varier selon les phonèmes impliqués dans les séquences biconsonantiques. En attaque de syllabe, comme le résultat obtenu sur les pseudo-mots de l'expérience 2A, un cluster Plosive+Latérale dans le mot « cloche » [kloʃ] ou « climat » [klima] est plus facilement identifié que les clusters Plosive+Fricative ou Plosive+Plosive tels que « psychique » [psiʃik], « psychose » [psikɔz] ou « ptose » [ptɔz] et ce, quel que soit le niveau des apprenants. Ce résultat confirme les propos de Kühnert et Hoole (2006) concernant le meilleur taux de réussite des séquences impliquant une liquide que les séquences Plosive+Plosive dans cette position.

Les combinaisons Plosive+Nasale en initiale et Nasale+Nasale en finale qui ont été moins bien perçues à l'intérieur des pseudo-mots (cf. expérience 2A), ont obtenu de meilleurs scores d'identification correcte lorsque insérées dans de vrais mots. En effet, les clusters « pneu » [pnø] et « knout » [knut] comportant une séquence Plosive+Nasale ont été identifiés à 88,8 % (meilleur score des combinaisons en initiale) par rapport à la perception de la même séquence dans les pseudo-mots à 67,9 % (le moins bon score à l'initiale). De même, en finale, le cluster Nasale+Nasale /mn/ a récolté le plus de réponses correctes parmi ceux initiés par une nasale. Contrairement aux résultats dans les pseudo-mots (le moins bon score de 44 % en finale), ce cluster dans un mot tel que « hymne » [imn] est mieux reconnu que les clusters Nasale+Obstruente dans les mots comme « weekend » [wikend], « jeans » [dʒins] ou « sprint » [sprɛ̃nt].

Tous les clusters en final ne respectent pas le principe de sonorité (Plosive+Plosive, Plosive+Fricative, Plosive+Nasale, Plosive+Latérale), car C₁ n'a pas une sonorité strictement supérieure à C₂. Ceux-ci présentent à peu près le même score d'identification, aucune différence significative entre ces clusters n'a été observée.

En position inter-syllabique, la séquence Plosive+Plosive étant la combinaison la plus fréquente des séquences initiées par une plosive dans cette position en vietnamien, elle établit le meilleur score d'identification correcte par les étudiants des deux niveaux. Ainsi, les groupes consonantiques /kt/, /pt/ dans « acteur » [ak.tœʁ], « capture » [cap.tyʁ] ont été nettement mieux perçus que /kn/, /ks/, /kf/, /ps/ dans « acné » [ak.ne], « taxi » [tak.si], « vexer » [vek.se], « bec-fin » [bɛk.fɛ̃], « capsule » [kap.syl], « apside » [ap.sid]. Il n'y a pas une préférence significative entre des séquences initiées par une nasale en position inter-syllabique, malgré le fait que la combinaison Nasale+Plosive soit la plus fréquente de toutes les séquences confondues en fonction du mode d'articulation des consonnes du vietnamien dans cette position.

Les résultats des deux tests de perception des séquences de consonnes pseudo-mots (expérience 2A) et mots (expérience 2B) du français montrent que les étudiants vietnamiens, de niveau intermédiaire ou avancé, ont réussi à identifier les clusters intra-syllabe et les séquences de consonnes inter-syllabe avec un taux moyen de plus de 70 %. Les erreurs d'identification des séquences biconsonantiques sont globalement observées davantage en coda qu'à la frontière syllabique intra-mot (dans les pseudo-mots, score moyen de 72,5 % en coda *vs.* 91 % en inter-syllabique ; dans les mots, score moyen des clusters /p/+C et /k/+C en coda de 78,9 % *vs.* 85,8 % en inter-syllabique). L'effet de la différence entre les deux positions est significatif quel que soit le niveau des étudiants [$F(1,37) = 53,3$; $p = 0$]. Ce résultat permet de valider l'*Hypothèse de la Différence de Marquage* selon laquelle, une forme dans une langue seconde qui est différente et plus marquée que celle de la langue maternelle sera plus difficile à acquérir. Il reste à démontrer et valider cette hypothèse sous l'angle de la production (cf. expérience 3).

Dans les deux expériences 2A et 2B, les apprenants présentent un taux de réussite plus élevé (respectivement 96,6 % et 82,2 %) quand une plosive est suivie d'une latérale en position d'attaque. La suite Plosive+Fricative en attaque pose par contre plus de difficultés aux étudiants (taux d'identification de 48,6 % et 64,9 % respectivement).

En position de coda, la combinaison Plosive+Latérale qui ne respecte pas le Principe de Sonorité reste le type de cluster le mieux identifié par les étudiants dans les deux tests 2A et 2B (taux de réussite de 82 % et 80,9 %). La différence est significative par rapport à l'ensemble des séquences [$F(1,38) = 6,179$; $p = 0,017$]. La suite Nasale+Nasale est significativement la moins bien récupérée dans les pseudo-mots par les apprenants par rapport aux autres combinaisons (taux de réussite de 44 %). Pourtant, lors du test 2B, ce cluster a été correctement identifié à 88,9 %. L'étendue du lexique des apprenants semble jouer un rôle ici dans la capacité à récupérer un groupe biconsonnique de façon précise.

En inter-syllabe, la suite Plosive+Plosive reste la mieux perçue parmi les séquences impliquant une plosive, que cette combinaison soit insérée dans des pseudo-mots ou des mots du français (93,6 % et 86,7 % respectivement) [$F(1,38) = 15,78$; $p = 0$]. La séquence qui établit le moins bon score dans cette position dans les deux tests 2A et 2B est Plosive+Fricative (86 % et 78 %), bien que ce score reste quand même d'un bon niveau de réussite par rapport à d'autres positions.

Pour récapituler, les groupements biconsonantiques les mieux récupérés par les apprenants des deux niveaux se trouvent à la frontière syllabique des pseudo-mots quelle que soit C_1 (test 2A) et des mots initiés par /p/ et /k/ (test 2B). Malgré le crible phonétique, les plosives du français en coda bien que relâchées et différentes de celles du vietnamien sont bien reconnues. Ce résultat est en accord avec le test perceptif des consonnes en séquences du vietnamien (cf. section 4.2.3.2) qui montre que la présence du burst améliore

l'identification des consonnes. Que ce soit des pseudo-mots ou mots, le meilleur cluster biconsonantique identifié en attaque comme en coda est Plosive+Latérale, la séquence la mieux perçue en inter-syllabe est Plosive+Plosive. Cette séquence semble être perçue en fonction de sa similarité par rapport à des segments présents comme *bons exemplaires* dans l'espace phonologique de la langue native (Best, 1995). Les apprenants semblent utiliser pour cette séquence les catégories *identiques* déjà présentes dans leur système de la L1. Selon le *Speech Learning Model* de Flege (1995), ces catégories sont plus faciles à acquérir. Rappelons que la combinaison Plosive+Plosive, qui est la plus fréquente parmi les séquences initiées par une plosive en inter-syllabique du vietnamien (cf. section 3.3.2), pose le moins de problèmes pour les apprenants des deux niveaux dans cette position. La séquence Nasale+Plosive qui possède la meilleure occurrence de toutes les séquences (initiées soit par une nasale, soit par une plosive) pouvant exister à la frontière syllabique du vietnamien (cf. section 3.3.2) est bien récupérée mais pas mieux que d'autres séquences initiées par une nasale dans les pseudo-mots et dans les mots du français.

Les consonnes bilabiales simples qui ont été les mieux récupérées, toutes positions confondues lors de l'expérience 1 (cf. section 5.1.1.3.2), établissent effectivement les meilleurs scores d'identification dans les séquences pseudo-mots en position inter-syllabique et en coda. Dans les mots, un cluster intra-syllabe ou une séquence inter-syllabe initié par une labiale est mieux identifié et avec un temps de réaction plus rapide que d'autres séquences commençant par une vélaire ou une coronale.

Afin d'estimer un éventuel impact de difficultés de perception sur la production, l'expérience suivante vise à évaluer la production des clusters et des séquences consonantiques inter-syllabe chez des locuteurs natifs du vietnamien.

5.2. Expérience 3 : Production de clusters et de séquences biconsonantiques du français par des apprenants vietnamiens

Au cours de l'apprentissage d'une langue seconde, les unités manquantes de la langue première sont plus ou moins bien perçues suivant les cas et plus ou moins bien reproduites, parce que l'apprenant perçoit ce qu'il a appris à percevoir, c'est-à-dire ce qui est significatif dans sa langue, à travers ce que Troubetzkoy (1939) appelle le *crible phonologique*. Ce filtre aux mesures de la langue maternelle, ne laisse passer que les éléments sonores de la langue seconde qui sont présents tels quels dans la langue maternelle. De ce fait, l'apprenant d'une langue seconde ne pourra pas produire ou répéter correctement ce qu'il ne perçoit pas correctement. Il y aurait, chez les apprenants d'une langue seconde, « systématisation des erreurs de perception et donc de production. Pour qu'un son soit correctement (re)produit il

faut d'abord qu'il soit bien perçu, c'est-à-dire reconnu, identifié par rapport aux autres » (Borrell, 1992).

Les séquences de consonnes en vietnamien n'existent qu'en frontière de mot ou qu'en frontière de syllabe intra-mot. Les combinaisons de consonnes successives en français peuvent figurer non seulement à la frontière syllabique, mais aussi dans un constituant de la syllabe. Les deux langues relèvent de nombreuses dissemblances à propos des gabarits lexicaux et patrons syllabiques (cf. chapitre III). Quels sont alors les impacts de la structure syllabique d'une langue monosyllabique à CVC dominante sur la production des structures syllabiques complexes du français ? De plus, les plosives, premier élément d'une séquence inter-syllabique en vietnamien, dans la plupart des cas ne sont pas relâchées en coda (cf. chapitre IV). Cette caractéristique a-t-elle des conséquences dans la réalisation des clusters du français ?

Les expériences examinant la perception des consonnes simples et en séquences du français dans des pseudo-mots et des mots montrent de meilleures performances d'identification des consonnes simples et séquences de consonnes du français en inter-syllabe que des clusters en positions initiale et finale. Dans la suite de cette étude, nous nous intéressons à la production des clusters et séquences de consonnes du français en fonction de leur position dans la syllabe et dans le mot. Notre objectif est de cerner et d'expliquer les difficultés de prononciation de ces groupes chez les apprenants vietnamiens de niveau intermédiaire et avancé, en analysant du point de vue typologique et quantitative les stratégies de simplification des clusters.

Il s'agit de vérifier les hypothèses posées au début de cette étude :

- **Hypothèse 1** : La structure syllabique C₁VC₂ du vietnamien impacte la production des structures syllabiques complexes du français.
- **Hypothèse 2** : Les clusters du français sont plus difficiles à récupérer donc à produire que les séquences inter-syllabiques.
- **Hypothèse 3** : Les apprenants vietnamiens modifient plus fréquemment les codas complexes du français que les attaques complexes, ces dernières étant moins marquées.
- **Hypothèse 4** : Les caractéristiques phonétiques des consonnes en coda du vietnamien (consonnes non relâchées) ont des conséquences sur les stratégies de simplification des groupes consonantiques du français.
- **Hypothèse 5** : Les combinaisons de phonèmes consonantiques plus marquées selon les universaux du langage sont plus difficiles à produire que les combinaisons moins marquées. Et dans ce cadre, selon que les clusters respectent

ou violent le Principe de Sonorité, leur niveau de difficulté de production doit être différent.

5.2.1. Méthodologie

5.2.1.1. Corpus

Pour cette quatrième expérience, le même corpus que celui de l'expérience 2B sur la perception des séquences et clusters biconsonantiques du français a été utilisé (cf. table 5.7). Rappelons que ce corpus contient tous les groupes de consonnes présents en français et en vietnamien. Le matériel expérimental se compose donc de mots du français de structure #CCV, VC.CV et VCC#. En raison d'ambiguïté en graphie et phonie, certains mots ont été retirés de la liste. Au total, 32 mots ont été retenus. Ils sont insérés dans la phrase porteuse « *Tu prononces ... trois fois* ». Le corpus est constitué de 3 répétitions de 32 phrases mises dans un ordre aléatoire pour tous les sujets. L'ensemble du corpus se trouve annexe 10.

Table 5.8. Unités lexicales du français sélectionnées pour l'expérience et classées selon leur position dans le cluster ou la séquence consonantique ainsi que dans le mot et dans la syllabe.

C ₁	#CC	C.C	CC#
p	pneu, psychique, ptose	chapka, hypnose, capsule, capture	ellipse, adepte
t		atlas, wattman, ethnique	rythme
k	cloche, knout	snack-bar, bec-fin, check-list, acné, taxi, dictée	cycle, phénix, insecte
m		rumba, gymnase	hymne
n		sprinter	weekend, jeans, sprint

5.2.1.2. Participants

Quarante sujets étudiants du Département de Langue et de Civilisation Françaises de l'École Supérieure des Langues Étrangères, de l'Université Nationale de Hanoi ont participé à l'expérience. Ces sujets n'ont pas participé aux expériences précédentes.

Comme dans les expériences précédentes, ces étudiants ont été classés en deux groupes selon leur nombre d'années d'apprentissage du français : le niveau avancé contient 20 étudiants qui apprennent le français depuis plus de 5 ans alors que les 20 étudiants du niveau intermédiaire ont moins de 5 ans d'apprentissage. Ils sont tous de langue maternelle du dialecte du Nord et âgés de 18 à 20 ans (moyenne d'âge 19 ans pour les hommes comme pour les femmes). Leur entourage parle le même dialecte. Ils connaissent tous l'anglais

comme deuxième langue étrangère. Aucun étudiant n'a effectué un séjour dans les pays francophones.

5.2.1.3. Déroulement du test

L'enregistrement s'est déroulé dans la salle multimédia de l'École Supérieure des Langues Etrangères de l'Université Nationale de Hanoi avec un enregistreur Marantz PMD 670 et micro Behringer B-5 à directivité cardioïde. Chaque participant était confortablement assis devant la liste de phrases. Ils ont été invités à lire à voix haute le corpus avec un débit normal. Deux pauses ont été faites au cours de la lecture, après chaque tiers du corpus. Un entraînement leur était préalablement proposé avec trois phrases n'appartenant pas au corpus du test proprement dit. La séance durait environ 10 minutes pour chaque participant. L'enregistrement a été numérisé à 44,1 kHz sous format .wav.

Après l'enregistrement, il a été demandé aux participants de cocher pour chaque mot si oui ou non ils avaient connaissance de ce mot afin de savoir si les stimuli du test sont pour ces étudiants des mots ou des pseudo-mots. La fiche « Connaissez-vous ce mot ? » se trouve annexe XII.

5.2.1.4. Traitement des données

Les 96 mots (32 mots x 3 répétitions) (soit 3,5 heures de parole lue) contenant des clusters et des séquences de consonnes produits par chacun des 40 étudiants ont été segmentés et analysés semi-automatique à l'aide du logiciel Praat, puis transcrit finement avec l'alphabet phonétique international (API version révisée de 2005) à partir de l'observation du signal acoustique et du spectrogramme. Une fois cette étape terminée, les données obtenues pour tous les items et tous les participants ont été analysées statistiquement avec le logiciel SPSS®.

5.2.2. Résultats

Sur l'ensemble des résultats, il est constaté que les apprenants vietnamiens, même après plusieurs années de pratique, présentent une réelle difficulté à prononcer les groupes consonantiques du français. Les résultats montrent que 66,7 % (sur un total de 3 840 clusters), n'ont pas été correctement prononcés par les sujets.

Un test t sur deux échantillons (mots et pseudo-mots) a été mené pour observer la corrélation entre le score de production correcte et la connaissance des stimuli en fonction du niveau des apprenants. Pour les apprenants de niveau intermédiaire, le score de production correcte est plus élevé si les stimuli sont reconnus comme mots (29,6 % vs. 23,3 %) ($p = 0,044$), alors qu'aucun effet significatif entre le score de production et la reconnaissance de stimuli n'a été observé chez les apprenants de niveau avancé, que ces stimuli soient des mots ou

pseudo-mots (39,04 % vs. 39,18 %) ($p = 0,97$). La connaissance des stimuli a permis donc seulement aux apprenants de niveau intermédiaire d'augmenter le score de production correcte de 23,3 % à 29,6 %.

Trois aspects des résultats sont ci-dessous observés. Dans un premier temps, sont examinées les difficultés de production des groupes consonantiques selon leur position dans la syllabe, puis la fréquence des erreurs selon les types de combinaisons de consonnes, et enfin, dans la dernière section, est proposée une analyse détaillée des différentes stratégies de simplification utilisées par les sujets.

5.2.2.1. Effet de la position des clusters

Des analyses ANOVA à mesures répétées ont été effectuées en considérant comme facteur intra-sujet la position des séquences et comme facteur inter-sujet le niveau des étudiants, la variable dépendante étant le taux d'erreur pour tous les sujets et tous les items.

Les résultats statistiques présentent une différence significative du taux d'erreur en production des clusters entre les positions inter-syllabique et finale [$F(1,38) = 84,9$; $p = 0$]. En effet, 73,7 % des clusters en coda sont produits modifiés par rapport à la cible alors qu'ils sont 53,9 % à l'être à la frontière syllabique intra-mot (cf. figure 5.18). En d'autres termes, il est plus difficile pour les apprenants vietnamiens, quel que soit leur niveau de français, de prononcer par exemple le mot *acte* [akt] que *acteur* [aktœʁ]. À remarquer aussi que la production des clusters impliquant une nasale /m/ ou /n/ en C₁ sont particulièrement moins réussies en finale qu'en inter-syllabique (taux d'erreur de 63 % contre 13 %).

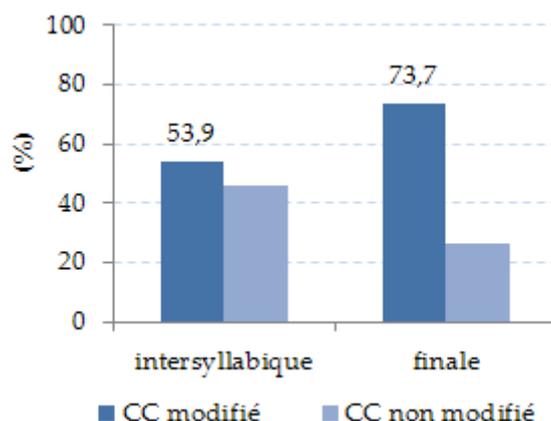


Figure 5.18. Proportions de clusters produits modifiés par rapport à la cible selon leur position (frontière syllabique intra-mot vs. coda finale de mot) tout niveau d'apprentissage confondu.

Un effet significatif de la différence du taux d'erreur pour les trois positions a été observé avec les groupes contenant /p/ ou /k/ en C₁ [$F(2,76) = 133,8$; $p = 0$]. Ces groupes sont mieux prononcés en initiale de mot (attaque de syllabe) qu'en inter-syllabique ou finale. La

différence est significative entre la position initiale et l'ensemble des autres positions [$F(1,38) = 158 ; p = 0$].

La figure 5.19 montre que le taux d'erreur des groupes /p/+C et /k/+C est deux fois plus présent en inter-syllabique et coda qu'en initiale de syllabe (respectivement 80,5 % et 79,7 % contre 39 %). Une interaction entre le niveau des étudiants et la production des groupes /p/+C et /k/+C est trouvée [$F(2,76) = 4,629 ; p = 0,013$]. En général, les étudiants avancés ont produit ces groupes avec moins d'erreur que ceux de niveau intermédiaire, notamment lorsque ces groupes se trouvent en attaque de mot.

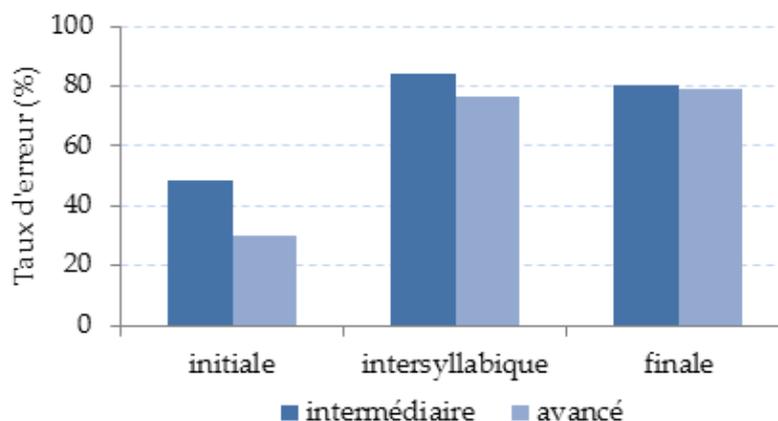


Figure 5.19. Proportions de modifications des clusters /p/+C et /k/+C par rapport à la cible selon leur position dans le mot et selon le niveau des apprenants.

Les résultats montrent aussi que dans chaque position, le mode et le lieu des premiers segments (C_1) des groupes C_1C_2 influencent la performance des apprenants en production, quel que soit leur niveau. En effet, un groupe commençant par une plosive pose beaucoup plus de difficulté (taux d'erreur 79,2 %) qu'un groupe commençant par une nasale (42,7 %), que ce soit en position inter-syllabique ou finale de mot [$F(1,38) = 257,05 ; p = 0$]. Les groupes impliquant une coronale en C_1 sont significativement plus difficiles à prononcer que les autres séquences de consonnes (taux d'erreur de 75,1 % contre 70,3 % des groupes avec C_1 labiale et 70 % des groupes avec C_1 vélaire) [$F(1,38) = 7,092 ; p = 0,011$]. Ainsi, la prononciation de /tn/ dans *ethnique* [ɛt.nik] est moins souvent réussie que /pn/ dans *hypnose* [ip.nɔz] ou /kn/ dans *acné* [ak.ne].

5.2.2.2. Effet du type de consonne dans les combinaisons

5.2.2.2.1. Les combinaisons en attaque

La difficulté dans la production des clusters varie également de manière significative selon la nature des phonèmes impliqués. La figure 5.20 montre qu'en début de mot, la suite Plosive+Fricative présente significativement un taux d'erreur plus élevé quel que soit le

niveau intermédiaire ou avancé (respectivement 90 % et 52 % d'erreurs) [$F(1,38) = 56,03$; $p = 0$]. Au contraire, la suite Plosive+Latérale pose beaucoup moins de problèmes aux apprenants (par niveau, respectivement 10 % et 3 %).

Les étudiants avancés sont plus performants que les intermédiaires dans la prononciation des attaques complexes. Un peu plus d'une fois sur trois, ils n'ont pas produit la séquence demandée (taux d'erreur de 31,6 %), alors que les intermédiaires n'ont réussi qu'une fois sur deux environ (taux d'erreur de 52,9 %). Une interaction est trouvée entre le taux d'erreur et le niveau des apprenants [$F(3,114) = 3,224$; $p = 0,025$].

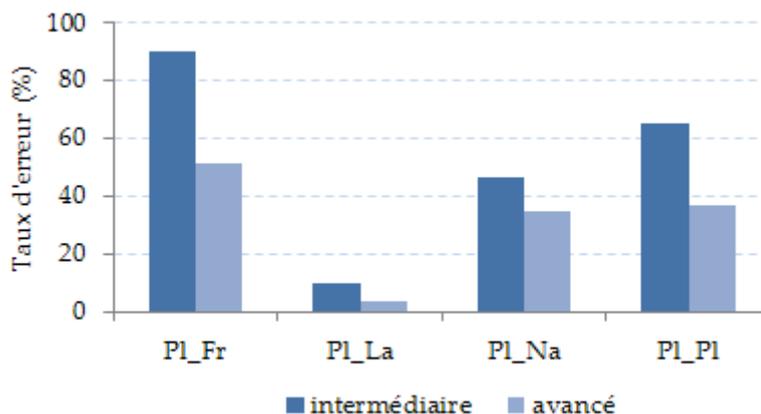


Figure 5.20. Taux d'erreurs de production des attaques complexes selon le type de combinaison des consonnes dans les séquences et le niveau des étudiants.

5.2.2.2.2. Les combinaisons en coda

La figure 5.21 indique que les clusters impliquant une fricative ou une plosive en C₂ présentent un taux d'erreur très élevé, peu importe que le cluster commence par une plosive ou une nasale. Les étudiants, quel que soit leur niveau, rencontrent particulièrement des difficultés à réaliser les codas complexes contenant deux plosives (taux d'erreur de 97,5 %). Dans la plupart des cas, la première plosive est supprimée dans la production des étudiants (cf. section 5.2.2.3.2). Par contre, ils ont significativement moins de difficultés à prononcer les codas Plosive+Latérale (taux d'erreur de 13 % parmi les groupes à C₁ plosive) ($p < 0,05$ dans tous les cas considérés). La combinaison Nasale+Nasale contient significativement le moins d'erreurs par rapport à l'ensemble des autres clusters impliquant une nasale en C₁ (45,8 % vs. 80 % d'erreurs pour les clusters de type Nasale+Fricative et 81,25 % d'erreurs pour les clusters de type Nasale+Plosive) [$F(2,78) = 13,93$; $p = 0$].

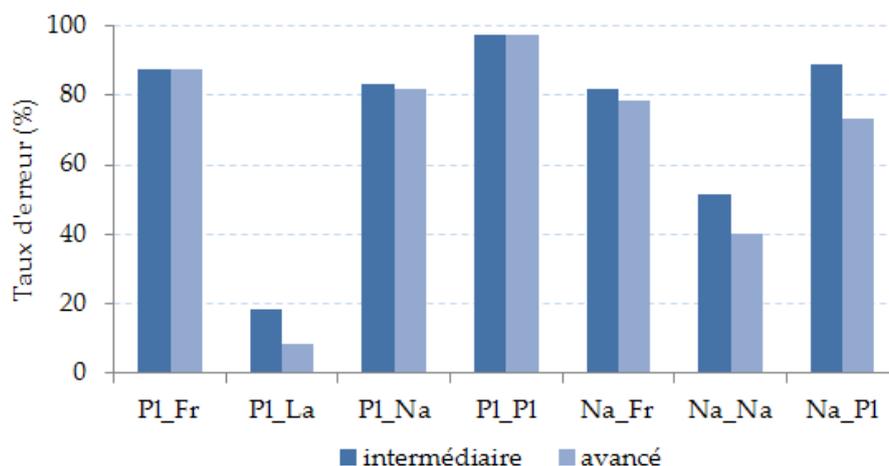


Figure 5.21. Taux d'erreurs dans la production des codas complexes du français selon le type de combinaison consonantique et le niveau des étudiants.

5.2.2.2.3. Les combinaisons inter-syllabe

L'analyse des erreurs dans la réalisation des groupes biconsonantiques répartis de part et d'autre d'une frontière syllabique montre que les clusters impliquant une nasale en C₁ sont très bien produits, les étudiants avancés ayant même réussi à produire sans aucune erreurs les clusters contenant deux nasales (cf. figure 5.22).

La séquence Plosive+Latérale présente le taux d'erreur le moins élevé parmi les combinaisons à première consonne plosive (46,6 %) [F(1,38) = 109,1 ; p = 0]. Comme les résultats obtenus pour les autres positions, les combinaisons Plosive+Fricative, Plosive+Plosive sont toujours les séquences les moins bien produites par les étudiants des deux niveaux (taux d'erreur de 88 % et 92 % respectivement).

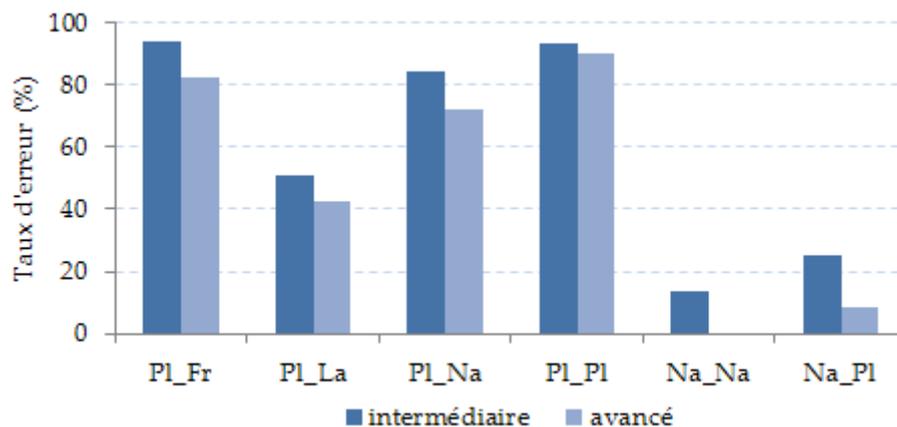


Figure 5.22. Taux d'erreur de production de séquences de consonnes inter-syllabe selon le type de combinaison des consonnes impliquées et le niveau des étudiants.

5.2.2.3. Les stratégies de simplification

Lorsque les apprenants sont incapables de produire les groupes consonantiques demandés, ils ont utilisé différentes stratégies de simplification. En effet, sur les 2 562 clusters modifiés par les apprenants des deux niveaux, ont pu être observés cinq types de stratégie : la modification par le non relâchement d'une des deux consonnes du groupe, la suppression d'une ou des deux consonnes du groupe, l'insertion d'un son épenthétique, le remplacement d'une consonne par une autre (substitution) ou le déplacement des deux consonnes (cf. figure 5.23).

Le nombre d'erreurs total estimé sur 2 562 clusters et séquences inter-syllabe mal produits est de 3 115. Ce chiffre s'explique par le fait que 553 clusters contiennent deux erreurs. Par exemple, lorsqu'un apprenant a prononcé *psychique* attendu [psɪʃik] réalisé [bɪʃikə], un élément du groupe consonantique a été supprimé ([s]) et un autre [p] remplacé par [b]. Ce type d'erreur entre donc dans les catégories « suppression » et « substitution ». Le schwa inséré en dehors du cluster cible n'a pas été pris en compte.

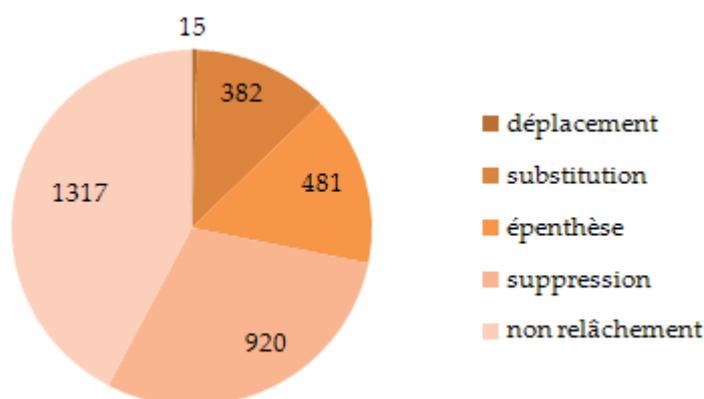


Figure 5.23. Nombre de modifications observées dans les groupes de consonnes (clusters et séquences inter-syllabe) en fonction des stratégies utilisées par les apprenants des deux niveaux.

Les résultats sur l'ensemble des erreurs montrent que le déplacement des consonnes de cluster est la stratégie la moins utilisée (15 % des erreurs). Elle est relevée seulement dans la production des trois clusters du corpus (un en attaque : *psychique* > [psɪʃik] et deux en coda : *rythme* > [rimt], *phénix* > [fenisk]). Cette stratégie touche seulement 9 apprenants sur 40 (6 de niveau intermédiaire et 3 de niveau avancé). Le nombre de ces cas de déplacement étant très limité, il n'apparaîtra plus dans les analyses qui suivent.

5.2.2.3.1. Modification par non relâchement

Cette stratégie rassemble des cas où les plosives sont produites sans phase de détente, l'occlusion n'étant pas suivie d'un bruit caractéristique d'explosion. La modification par non relâchement est la stratégie la plus utilisée par les apprenants, que le niveau soit intermédiaire ou avancé (51,4 % des modifications des clusters). Cette stratégie ne concerne

que les groupes C₁C₂ impliquant une plosive /p/, /t/ ou /k/ en C₁ et uniquement en coda de syllabe, c'est-à-dire lorsque les groupes de consonnes se situent dans les positions inter-syllabique et finale de mot. Comme le montre l'exemple du mot *hot-dog* (cf. figure 5.24), le bruit d'explosion de [t] est absent avant la mise en place du voisement de [d].

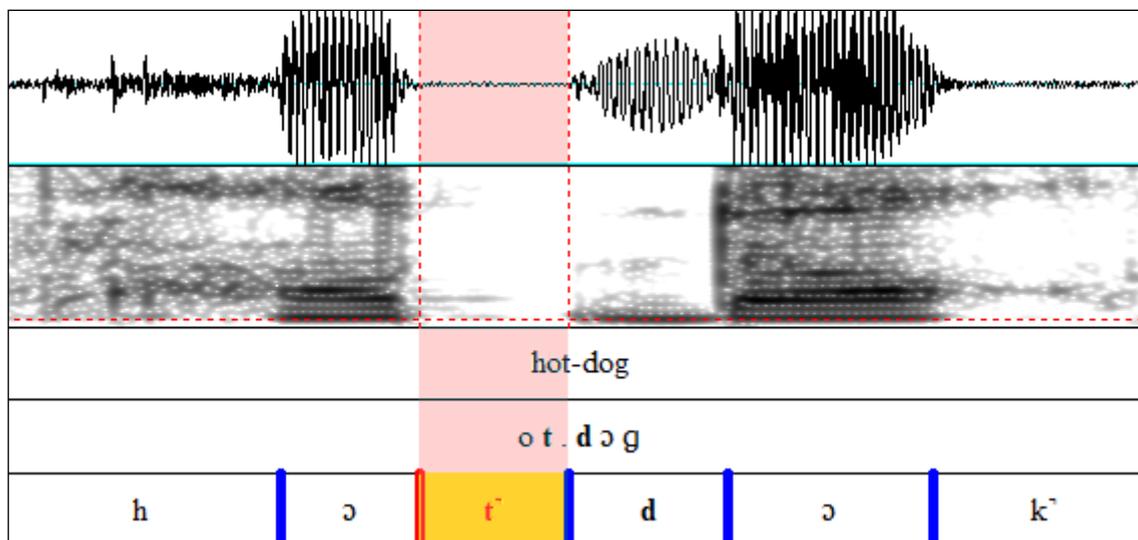


Figure 5.24. Production du mot « hot-dog » par un apprenant de niveau intermédiaire. Cas du non relâchement de la première consonne du groupe [td].

Les codas complexes impliquant une plosive en C₁ ont subi la même tendance. La figure 5.25 montre que l'apprenant simplifie le cluster [pt] en ne relâchant pas l'articulation⁵⁷ de C₁. On voit aussi qu'un schwa d'appui a été ajouté à la fin du mot. La structure syllabique du mot [adept] prononcé [adɛp^htə] a été donc modifiée.

⁵⁷ L'analyse est basée sur les indices acoustiques (descentes de F2, F3 à la fin du segment vocalique [ɛ]) pour associer [p] à la catégorie « non relâchement ». La frontière entre la plosive non relâchée [p^h] et la plosive sourde [t] nous a été indiquée grâce au changement de l'intensité.

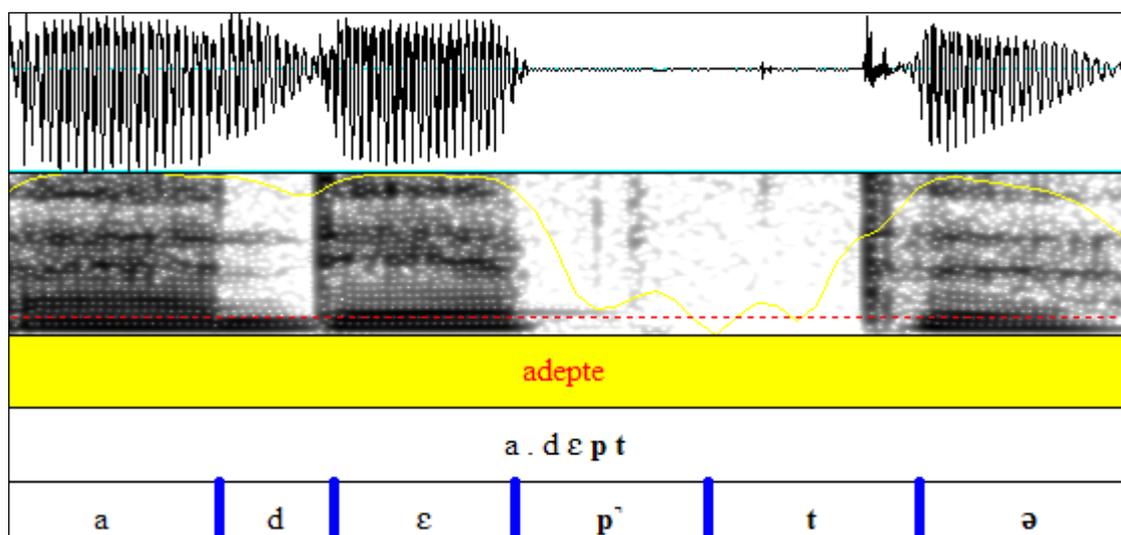


Figure 5.25. Production du mot « adepte » par un apprenant de niveau intermédiaire. Cas du non relâchement de la première consonne du groupe [pt].

Globalement, les consonnes C_1 qui se trouvent à la frontière syllabique ($C_1.C_2$) perdent leur burst davantage que celles en coda de mot ($C_1C_2\#$). Cette différence est significative quel que soit le niveau des apprenants [$F(1,38) = 36,5$; $p = 0$]. Si on regarde en détail chaque consonne, la tendance globale est significativement observée pour /t/ et /k/ [$F(1,39) = 84,3$; $p = 0$], alors que /p/ perd le burst davantage en C_1 d'un cluster en finale de mot qu'en inter-syllabique (52,8 % vs. 44,4 %) [$F(1,39) = 11,54$; $p = 0,002$].

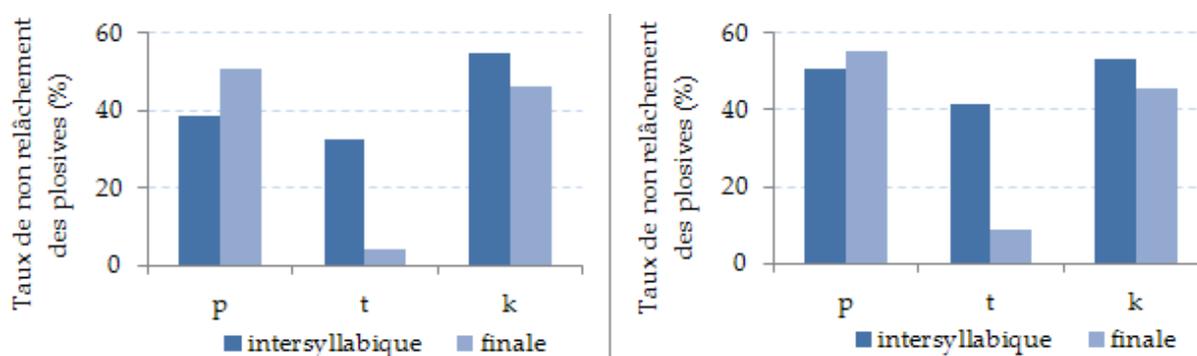


Figure 5.26. Taux de non relâchement des plosives en C_1 des groupes inter-syllabe et en coda, selon le niveau intermédiaire (à gauche) et avancé (à droite) des apprenants.

Le taux de non relâchement de /t/ en C_1 des groupes consonantiques est significativement moins élevé que /p/ et /k/ quelle que soit la position des groupes [$F(1,39) = 187,7$; $p = 0$] ([t¹] observé dans 254 cas vs. 554 cas de [k¹] et 512 cas de [p¹]). Il a été observé que lorsque cette consonne /t/ initie un groupe en inter-syllabique ($C_1.C_2$), elle perd son burst six fois plus que lorsqu'elle est en C_1 d'un cluster en finale de mot ($C_1C_2\#$) (37 % vs. 6,4 %).

La différence du taux de non relâchement est significative selon les types de combinaisons des consonnes en finale de mot [$F(3,114) = 108,2 ; p = 0$]. Les combinaisons qui subissent davantage des modifications par non relâchement sont celles qui impliquent les deux plosives (cf. table 5.9).

Table 5.9. Taux de non relâchement des C_1 initiant des clusters en coda en fonction du type de consonnes impliquées en C_2 .

$C_1 \backslash C_2$	Plosive	Fricative	Nasale	Latérale
Plosive	90 %	57,7 %	25,8 %	0,83 %

La même tendance est observée en inter-syllabique pour les séquences de consonnes à C_1 plosive [$F(3,114) = 84 ; p = 0$]. Huit fois sur dix, une plosive est produite sans bruit d'explosion quand elle est suivie par une autre plosive (cf. table 5.10).

Table 5.10. Taux de non relâchement des C_1 initiant des séquences inter-syllabe en fonction du type de consonnes impliquées en C_2 .

$C_1 \backslash C_2$	Plosive	Fricative	Nasale	Latérale
Plosive	80,3 %	48,6 %	50,4 %	12,9 %

5.2.2.3.2. Suppression

Cette stratégie regroupe les cas d'omission où il n'y a aucune trace de la consonne sur le spectrogramme. C'est la deuxième stratégie utilisée par les apprenants. Quel que soit leur niveau, les apprenants ont une nette tendance à cette stratégie lorsque le cluster se trouve en coda de mot plutôt qu'une autre position ($p = 0$).

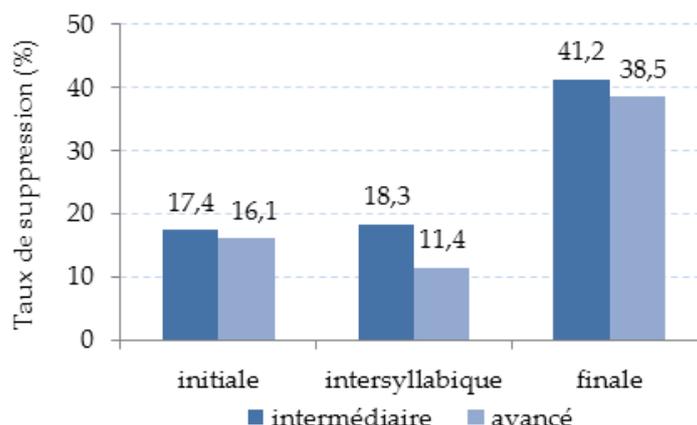


Figure 5.27. Taux de suppression d'un ou des deux éléments du groupe en fonction de la position des groupes consonantiques dans le mot, la syllabe et le niveau des apprenants.

La différence du taux de suppression est significative selon les types de combinaisons des consonnes en coda de mot [$F(7,266) = 71,82 ; p = 0$].

Les combinaisons qui ont subi le plus de suppression en coda sont celles qui impliquent une fricative en C₂, ou une plosive en C₂ lorsque C₁ est une nasale (cf. table 5.11) ou lorsque les deux consonnes du groupe partagent le même lieu d'articulation (coronal) (cf. table. 5.12), et ceci quel que soit le niveau de français des apprenants.

Table 5.11. Taux de suppression dans les clusters en coda en fonction du mode des consonnes impliquées.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	11,3 %	85,4 %	29,2 %	4,2 %
Nasale	78,3 %	78,3 %	37,5 %	

Table 5.12. Taux de suppression dans les clusters en coda en fonction du lieu des consonnes impliquées.

C ₁ \ C ₂	Labial	Coronal	Vélaire
Labial		24,2 %	
Coronal	29,2 %	78,3 %	
Vélaire		29,4 %	

La même tendance est observée à la frontière inter-syllabe pour les séquences de consonnes à C₁ plosive. La stratégie varie en effet en fonction de la combinaison des consonnes [$F(3,114) = 17,4$; $p = 0$]. Les apprenants ont beaucoup plus recours à la suppression pour la combinaison Plosive+Fricative (cf. table 5.13). Par contre, les clusters impliquant une nasale en C₁ sont nettement mieux prononcés dans cette position, quel que soit le niveau des apprenants.

Table 5.13. Taux de suppression dans les séquences inter-syllabe en fonction du mode des consonnes impliquées.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	6,3 %	85 %	11,3 %	26,3 %
Nasale	3,8 %		1,7 %	

Si on regroupe les consonnes selon leur lieu d'articulation, il est également observé que le taux de suppression à la frontière inter-syllabe varie en fonction des combinaisons des consonnes [$F(6,228) = 10,883$; $p = 0,002$] (cf. table 5.14).

Table 5.14. Taux de suppression dans les séquences inter-syllabe en fonction du lieu des consonnes impliquées.

C ₁ \ C ₂	Labial	Coronal	Vélaire
Labial	5,83 %	9,4 %	4,2 %
Coronal	3,3 %	23,5 %	
Vélaire	2,5 %	20 %	

L'effet est également significatif si on considère le taux de suppression selon les combinaisons en attaque [F(3,114) = 27,71 ; p = 0]. La production des clusters de type Plosive+Fricative pose nettement plus de problèmes aux apprenants que les autres combinaisons (cf. table 5.15). Quel que soit le niveau des apprenants, en position d'attaque, les clusters Labial+Coronal sont supprimés 4 fois plus souvent que les clusters Vélaire+Coronal (27,2 % vs. 6,3 %) [F(1,38) = 33,98 ; p = 0].

Table 5.15. Taux de suppression des clusters en attaque en fonction du mode des consonnes impliquées.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	19,2 %	44,2 %	15 %	0,83 %

Dans les 920 cas de suppression, toutes positions confondues, les C₂ sont les plus supprimées (58,15 %) (cf. table 5.16). Les étudiants suppriment moins les C₁ (39,57 %) et encore beaucoup plus rarement les deux consonnes du groupe (2,28 %).

Table 5.16. Fréquence de suppression des consonnes des groupes biconsonantiques.

C ₁ C ₂	C ₁ supprimée	C ₂ supprimée	C ₁ C ₂ absentes	Total
/p/+C	150	38	0	188
/t/+C	20	130	0	150
/k/+C	159	85	0	244
/m/+C	18	36	0	54
/n/+C	17	246	21	284
Total	364	535	21	920

Il a été remarqué que l'élément moins sonore dans le groupe est effacé dans 71,76 % des cas de suppression. C'est le cas des plosives et fricatives en C₂ des groupes initiés par une nasale (87,04 %). La même tendance est observée pour les groupes à C₁ plosive, les apprenants préfèrent conserver le segment plus sonore (plosive voisée, fricative, nasale, latérale) en C₂ et supprimer le segment moins sonore dans le groupe dans 63,76 % des cas (exemple cf. figure 5.28).

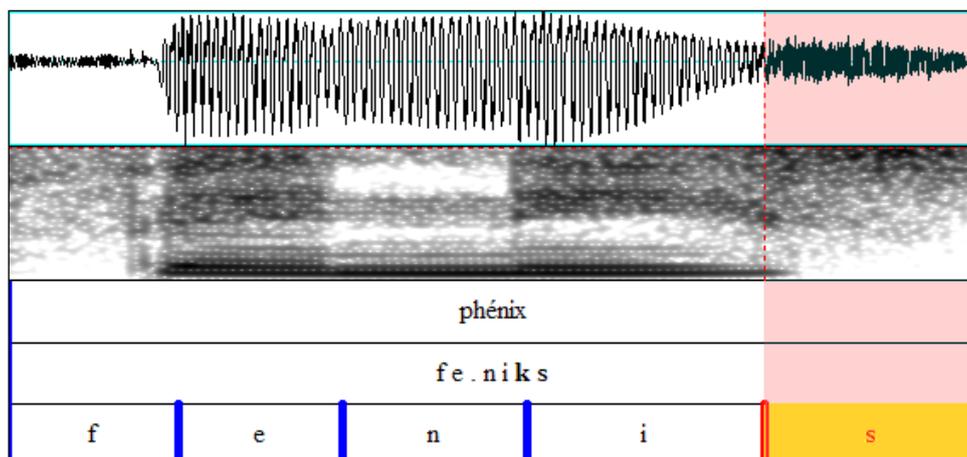


Figure 5.28. Exemple de prononciation du mot « phénix » par un apprenant de niveau avancé où l'élément moins sonore [k] du groupe est supprimé.

Parmi l'ensemble des groupes consonantiques observés, ceux du type /n/+C subissent le plus la suppression : les apprenants des deux niveaux suppriment les C₂ dans 59,5 % des productions des clusters impliquant /n/ (ex. [wikend] > [wikɛn], [ʒins] > [ʒin] comme le montre la figure 5.29). Il est à remarquer que 21 cas pour lesquels les traces des deux consonnes du groupe sont absentes sur le spectrogramme, concernent trois mots « jeans », « sprint » et « weekend ». Les apprenants (8 de niveau intermédiaire, 2 de niveau avancé) ayant la difficulté à retrouver le lien graphie-phonie, les ont prononcés avec les voyelles nasales à la place (respectivement [ʒã], [spɛẽ] et [wikã]).

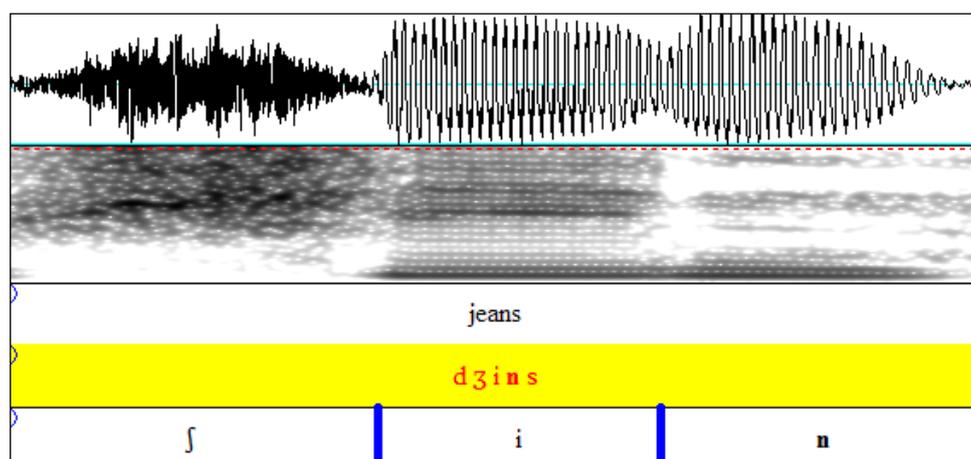


Figure 5.29. Exemple de production du mot « jeans » par un apprenant de niveau intermédiaire.

5.2.2.3.3. Epenthèse

Le taux d'épenthèse selon la position des clusters est présenté figure 5.30 ci-dessous. Les sons épenthétiques sont davantage insérés dans les clusters en coda qu'en attaque. En position inter-syllabique, l'épenthèse est plutôt rare et l'élément inséré est très varié (à la fois

vocalique ou consonnantique). Ce sont des cas très isolés. Ils représentent à peine 2 % des erreurs.

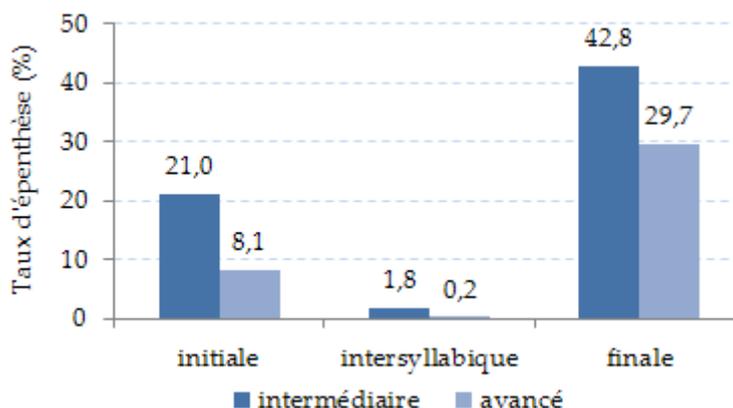


Figure 5.30. Taux d'épenthèses en fonction de la position des clusters dans le mot et le niveau des étudiants.

La différence du taux d'épenthèse est significative selon les types de combinaisons des consonnes en coda de mot [$F(5,190) = 48,379$; $p = 0$]. Aucune interaction entre le niveau des apprenants et les types de combinaison n'a été trouvée [$F(5,190) = 2,25$; $p = 0,81$]. La combinaison Plosive+Latérale en coda de mot présente le taux d'épenthèse le plus élevé (cf. table 5.17).

Table 5.17. Taux d'épenthèse dans les clusters en coda en fonction du mode d'articulation des consonnes successives.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	45,4%	11,7%	53,5%	87,5%
Nasale	4,6%		45%	

Tous les éléments épenthétiques extra-coda sont des vocoïdes centraux soit de type schwa (valeurs moyennes de F1, F2, F3 respectivement 580 Hz, 1 500 Hz, 2 600 Hz) soit de type [ɪ] plus fermé (400 Hz, 1 700 Hz, 2 800 Hz en moyenne) (cf. figure 5.31). L'épenthèse est particulièrement utilisée par les apprenants des deux niveaux lors de la prononciation du mot *cycle* [sikl] (cf. figure 5.32). Ce groupe Plosive+Latérale connaît l'ajout vocalique en final dans 115 des 120 productions.

En position d'attaque, un vocoïde de type schwa est inséré entre les deux consonnes du cluster dans la plupart des cas (62,5 %). Une assimilation des timbres vocaliques (harmonie vocalique) dans lequel les apprenants ont inséré une voyelle « compatible » de même classe, voire la même voyelle entre les deux consonnes, a également été observée (par exemple, *knout* > [kunut], *psychique* [pisiiʃik], *ptose* > [potoz] (cf. figure 5.33)). Le cas de l'ajout de /o/ représente 20,5 % des épenthèses en attaque.

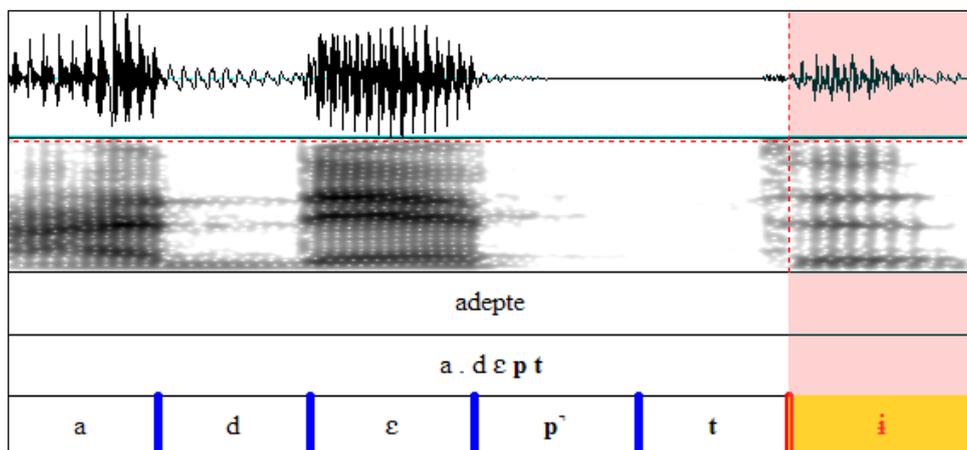


Figure 5.31. Exemple de production du mot « adepte » par un apprenant de niveau intermédiaire.

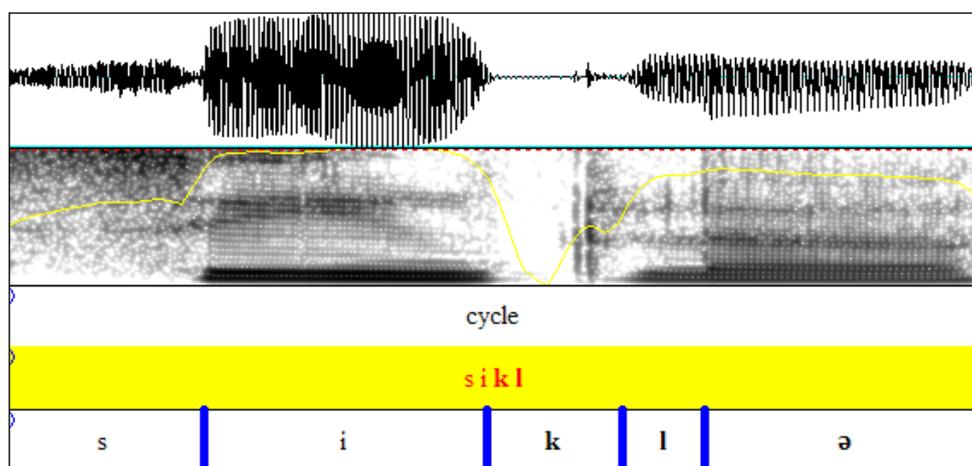


Figure 5.32. Exemple de production du mot « cycle » par un apprenant de niveau avancé.

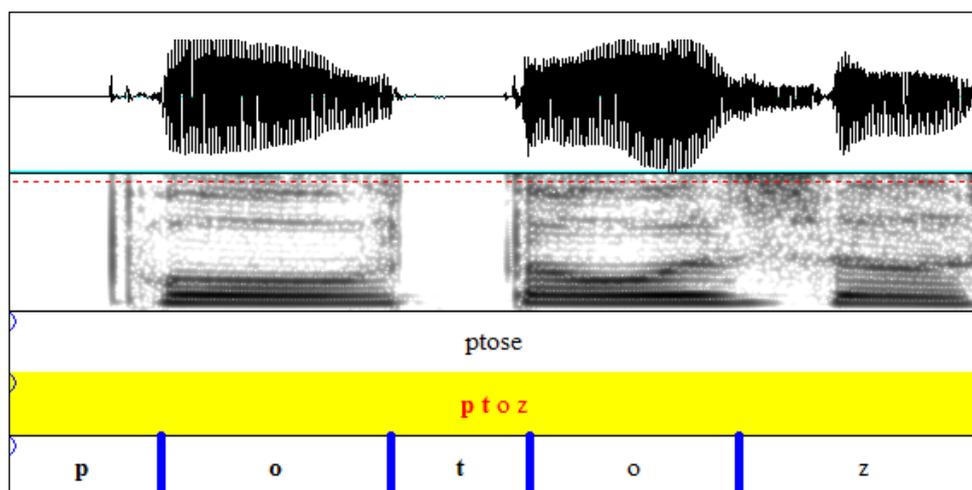


Figure 5.33. Exemple de production du mot « ptose » par un apprenant de niveau intermédiaire.

Il y a quelques autres cas isolés : *knout* [knut] est prononcé [kanut]; *psychique* [psiʃik] est prononcé [pəsiʃik] ou [pskiʃik] par les étudiants de niveau intermédiaire.

5.2.2.3.4. Substitution

Le taux de substitution en fonction de la position des groupes biconsonantiques et le niveau des apprenants est présenté dans la table 5.18.

Table 5.18. Taux de substitution selon la position des groupes consonantiques et le niveau des étudiants.

Position des groupes	Intermédiaire	Avancé
Initiale	13,6%	6,4%
Inter-syllabique	16,2%	6,1%
Finale	12,8%	6,3%

Aucune différence significative du taux de substitution n'a été trouvée en fonction des positions des groupes dans le mot [$F(1,38) = 0,854$; $p = 0,361$]. Par contre, un effet significatif selon le niveau des apprenants a été trouvé [$F(1,38) = 15,1$; $p = 0$]. Les étudiants intermédiaires ont un taux d'erreurs par substitution significativement plus élevé que les avancés (en moyenne 14,2 % contre 6,3 %).

En coda, la différence du taux de substitution est significative selon le type des consonnes impliquées dans le cluster [$F(7,266) = 11,05$; $p = 0$]. Aucune interaction entre le niveau des étudiants et les types de combinaisons [$F(7,266) = 0,96$; $p = 0,46$]. La combinaison Plosive+Nasale en coda de mot (seul /tm/ rentre dans cette catégorie) connaît le taux de substitution le plus élevé (cf. table 5.19).

Table 5.19. Taux de substitution des clusters en coda de mot en fonction de la combinaison des consonnes.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	7,9%	1,7%	27,5%	0,8%
Nasale	7,1%	2,5%	6,7%	

Le cluster [tm] dans *rythme* connaît deux types de changements : soit du mode d'articulation où [ritm] attendu est réalisé [rism] ; soit du lieu d'articulation où [ritm] > [ripm] comme l'illustre la figure 5.34. On voit sur le spectrogramme les transitions de F2, F3 qui indiquent une fermeture labiale après la tenue vocalique [i]. La partie d'occlusion de [p] est suivie par un petit burst, très faible de 52dB.

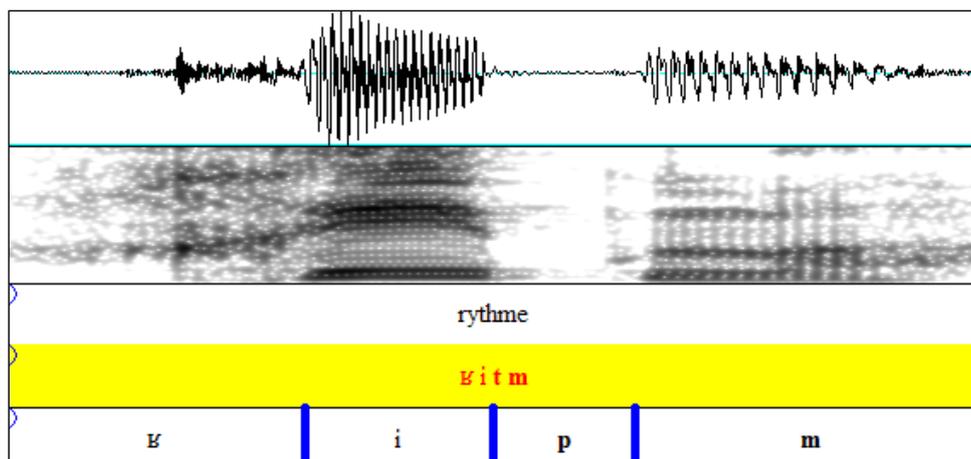


Figure 5.34. Exemple de production du mot « rythme » par un apprenant de niveau avancé.

À la frontière inter-syllabe, la même tendance est observée pour les clusters à C₁ plosive. Cette stratégie varie en fonction de la combinaison des consonnes [F(3,114) = 7,629 ; p = 0]. Les étudiants modifient plus les traits articulatoires de la consonne quand il s'agit des combinaisons Plosive+Nasale (cf. table 5.20).

Table 5.20. Taux de suppression des séquences inter-syllabe en fonction du type de consonnes impliquées.

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	12,5%	3,3%	17,1%	8,8%
Nasale	14,2%		5%	

Les cas les plus fréquents de substitution à la frontière inter-syllabique sont :

- a. Changement de mode :
 - nasale > latérale (C₂ concernée) : *acné* [akne] > [akle], *hypnose* [ipnoz] > [iplɔz], *ethnique* [ɛtnik] > [ɛtlik], *gymnase* [ʒimnaz] > [ʒimlaz]
 - plosive > fricative (C₁ concernée) : *ethnique* [ɛtnik] > [ɛsnik], *wattman* [watman] > [wasman], *taxi* [taksi] > [taxsi]
- b. Changement de voisement :
 - sonore > sourd (C₂ concernée) : *snack-bar* [snakbaʁ] > [snakpaʁ], *rumba* [rumba] > [rumpa] (cf. figure 5.35 : le spectrogramme indique l'absence de voisement pour [b]). Cette consonne est ainsi réalisée sourde [p]).
- c. Changement de lieu :
 - alvéolaire > vélaire : *ethnique* [ɛtnik] > [ɛknik]

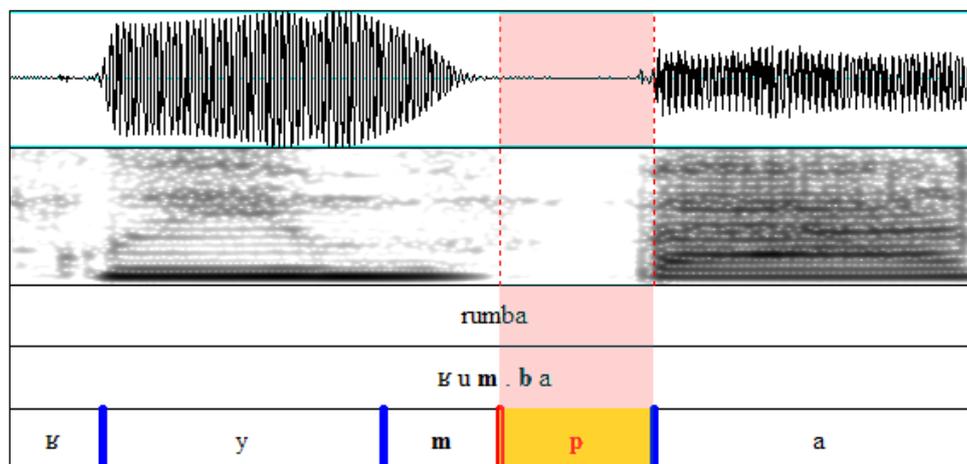


Figure 5.35. Exemple de production du mot « rumba » par un apprenant de niveau avancé.

En attaque, l'effet est également significatif si on considère le taux de substitution selon les types de consonnes qui entrent dans la composition des clusters [F(3,114) = 16,15 ; p = 0]. La séquence Plosive+Fricative est réalisée avec nettement plus de substitutions que les autres combinaisons (cf. table 5.21).

Table 5.21. Taux de substitution dans les clusters en attaque de mot en fonction de la combinaison des consonnes (C₁ = plosives).

C ₁ \ C ₂	Plosive	Fricative	Nasale	Latérale
Plosive	5%	33,3%	8,3%	0,83%

Ainsi, le mot *psychique* a été produit par les apprenants des deux niveaux en changeant le mode et le lieu d'articulation des consonnes : [psiʃik] > [fisik], [p^hisik], [ʃiʃik], [biʃik], [pziʃik] ... Un exemple de substitution en attaque pour ce mot est présenté figure 5.36, laquelle montre que la fricative antérieure [s] a été modifiée en une fricative beaucoup plus postérieure [χ].

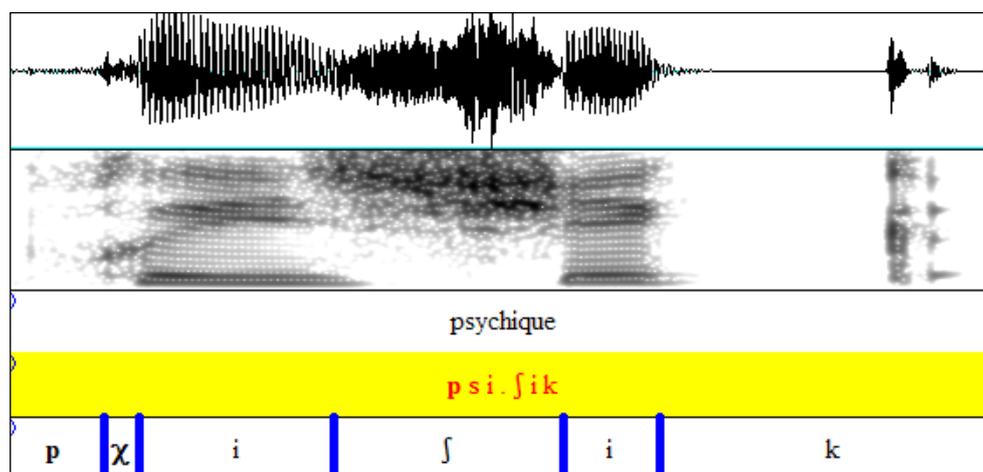


Figure 5.36. Illustration d'une des productions du mot « psychique » par un apprenant de niveau intermédiaire.

5.2.3. Discussion

Cette étude de la production des groupes de consonnes du français par des apprenants vietnamiens de deux niveaux (intermédiaire et avancé) a pour objectif de vérifier si la structure syllabique complexe, absente de la L1, a un effet sur leur acquisition. Les résultats montrent que les apprenants vietnamiens, même après plusieurs années de pratique, ont une réelle difficulté à prononcer les groupes consonantiques du français. Deux fois sur trois, les sujets n'ont pas réussi à prononcer exactement la structure syllabique attendue. Selon *l'Hypothèse de la Différence de Marquage* (Eckman, 1977), ces structures syllabiques complexes testées dans le cadre de cette étude (à savoir CCV, VCC) sont difficiles à acquérir pour les apprenants vietnamiens non seulement parce qu'elles sont inexistantes dans leur langue maternelle, mais aussi parce qu'il s'agit de structures marquées du point de vue des universaux typologiques, par rapport aux attaques et codas simples (Rousset, 2004). Les erreurs de production sont résistantes quel que soit le niveau d'apprentissage. De manière générale, les étudiants avancés font des erreurs aussi fréquentes et identiques à celles produites par les étudiants de niveau intermédiaire. Seul dans quelques cas, une progression est décelable au niveau des attaques complexes mais les véritables raisons de ces erreurs semblent perdurer.

Les résultats obtenus à l'issue des expériences en perception et production des groupes de consonnes du français permettent de discuter et de vérifier les hypothèses posées au début de ce chapitre.

Hypothèse 1: *La structure syllabique CVC du vietnamien impacte l'apprentissage des structures syllabiques fermées du français.*

Plusieurs études ont souligné que les difficultés dans la perception et production des langues étrangères en général (entre autres Polivanov, 1931 ; Troubetzkoy, 1939 ; Lado, 1957 ; Best, 1995 ; Flege, 1995) et des clusters en langue seconde en particulier (entre autres Tarone, 1980 ; Sato, 1984 ; Broselow, 1993 ; Osburne, 1996 ; Carlisle, 2001 ; Detey, Durand & Nespoulous, 2005 ; Sperbeck & Strange, 2008, 2010) résultent de l'influence de la structure phonologique de la langue première. Troubetzkoy écrit ainsi que « l'homme s'approprie le système de sa langue maternelle » et qu'il filtre tout ce qui arrive à son oreille à travers le crible phonologique de sa langue déjà constitué depuis sa naissance. D'autres études montrent aussi que la stratégie syllabique appropriée à la langue première est principalement utilisée lors de la segmentation lexicale d'une langue seconde par les auditeurs de L1, la syllabe prenant ainsi une place essentielle en tant qu'unité de segmentation à un niveau intermédiaire entre phonème et unité lexicale (Cutler et al., 1986 ; Mehler et al., 1992 ; Otake et al., 1993 ; Dupoux et al., 1997).

En accord avec ces études antérieures, le présent travail met en évidence un rôle de la structure syllabique de la langue native dans l'acquisition des groupes de consonnes d'une langue seconde. Rappelons que l'observation de la fréquence des types syllabiques à partir de la base G-ULSID a permis de confirmer la structure CVC comme la plus fréquente en vietnamien (69,47 % des syllabes) (cf. section 3.2.4). Cette structure CVC a effectivement été attestée comme la structure principalement produite par les apprenants lors de la production des groupes biconsonantiques du français. Il a été en effet observé que l'une des deux consonnes dans les structures *CVCC#*, *#CCVC* a été plus souvent supprimée, laissant ainsi la place à une structure CVC, la plus favorisée de la langue vietnamienne. Dans plus de 70 % des cas, c'est l'élément plus sonore dans le groupe qui est conservé et l'élément moins sonore qui est supprimé, confortant la tendance à préférer un élément plus sonore dans la séquence dans les résultats des études antérieures (Davidson, 2001, 2003, 2006 ; Davidson et al, 2004 ; Hancin-Bhatt et Bhatt, 1997 ; Hansen, 2001 ; Kim, 2000 ; Tropic, 1987). Les travaux antérieurs sur l'acquisition des clusters de l'anglais par des apprenants vietnamiens ont attesté le même constat sur la préférence pour la structure syllabique CVC (Sato, 1984 ; Benson, 1988 ; Osburne, 1996 ; Riney, 1999).

Selon l'hypothèse du « *Principe de récupérabilité* » formulée par Weinberger (1987), dans une tâche formelle sans contexte linguistique comme la nôtre (lecture de phrases porteuses), l'épenthèse devrait se produire plus que la suppression car l'apprenant prête plus d'attention à la forme qu'au contenu, la tendance étant à l'inverse dans une tâche moins formelle comme une conversation spontanée. Cependant, les résultats de l'expérience 3 montrent que la suppression est la stratégie de modification la plus utilisée par les apprenants en attaque comme en coda de syllabe. Comme l'étude de Sato (1984) sur le vietnamien, il a été relevé que la proportion de suppressions est plus importante que la proportion d'épenthèses en position initiale (40,36 % vs. 31,43 %) ou finale de syllabe (37,61 % vs. 27,36 %). Ce résultat met en évidence l'influence de la structure CVC de la langue native lors de la production des groupes de consonnes en français : si l'épenthèse est rencontrée dans un groupe *#CCVC* ou *CVCC#*, la structure syllabique sera modifiée en faveur de la structure CV qui est 3 fois moins fréquente que CVC en vietnamien (24,09 % vs. 69,47 %) (cf. section 3.2.4).

Hypothèse 2 : *Les clusters du français seront plus difficiles à récupérer et à produire que les séquences inter-syllabiques.*

L'apprentissage de la phonologie d'une langue étrangère a été associé depuis Troubetzkoy au concept de *crible phonologique*, sur lequel repose en partie l'*Hypothèse de l'Analyse Contrastive* de Lado (1957). Mais cette hypothèse ne peut pas prédire la difficulté relative des éléments de la langue à apprendre, notamment entre deux éléments différents de la langue native, ce qui conduit Eckman à poser l'*Hypothèse de la Différence de Marquage* (1977) : les formes du français différentes et plus marquées que celles du vietnamien

constitueront les zones de difficultés plus problématiques. Rappelons que la présente étude vise à confronter les apprenants à des groupes sonores identiques aux deux langues au niveau des consonnes qui les composent, mais différentes au niveau de leur position dans le mot et dans la syllabe : il s'agit des séquences biconsonantiques à la frontière de syllabe en français, aussi en vietnamien, mais des clusters intra-mot en français uniquement. L'ensemble des résultats de nos expériences, tant sur le plan perceptif des mots et des pseudo-mots correspondant à des schémas phonotactiques du français, que sur le plan productif des items lexicaux dans cette langue seconde, confirme en partie l'hypothèse formulée et montre que seuls les clusters en coda du français, absents en vietnamien, donc différents et plus marqués, sont moins bien récupérés et moins bien produits que les séquences inter-syllabes, et ce quel que soit le niveau des apprenants. Il est ainsi peu difficile pour un apprenant vietnamien d'identifier et réaliser un cluster comme par exemple dans *acte* [akt] que *acteur* [aktœʁ]. Les clusters en initiale de mot connaissent dans la plupart des cas un meilleur score d'identification et de production que les séquences inter-syllabe (la différence de score de bonnes réponses n'est pas attestée entre ces deux positions lors du test d'identification des clusters du français), mais entre les scores de production correcte une différence est relevée.

Plusieurs travaux, notamment les tenants du *modèle de la Phonologie Articulatoire*, ont montré des arguments en faveur d'un encodage des gestes articulatoires pouvant fournir les explications aux erreurs de production (Browman & Goldstein, 1992 ; Saltzman & Munhall, 1989 ; Byrd, 1992, 1996 ; Chitoran, Goldstein & Byrd, 2002 ; Fougeron, 2005 ; Kürhert & Hoole, 2006a ; Kürhert, Hoole & Mooshammer, 2006b). La notion centrale dans tous ces travaux est la coproduction gestuelle : les gestes se chevaucheraient, dans le temps et dans l'espace, pour produire l'enchaînement de trajectoires qui caractérise la langue parlée. La raison pour laquelle les groupes en initiale de mot ont moins d'erreurs que d'autres positions dans le mot et dans la syllabe semble résider dans la coordination temporelle entre les gestes. Il a été montré en effet que les groupes consonantiques en initiale de mot (ou de syllabe) tendent à connaître un moindre chevauchement inter-gestes que les mêmes groupes en position finale de mot ou à travers une frontière syllabique (Byrd, 1996). Un autre argument a été évoqué pour expliquer des différences de synchronisation entre des gestes est la notion de « récupérabilité » perceptive des gestes dans la séquence (Byrd, 1996 ; Silverman et Jun, 1994 ; Silverman, 1995 ; Wright, 1996 ; Chitoran et al., 2002). Ainsi, « dans les cas où le chevauchement pourrait masquer des corrélats perceptifs importants de l'une des consonnes de la suite, ce chevauchement sera moindre » (Kühnert et al., 2006a). Ces arguments basés sur la coordination gestuelle peuvent aider d'une part à expliquer l'asymétrie positionnelle dans l'acquisition des clusters intra-syllabiques du français (mieux perçus lorsqu'ils sont en attaque de syllabe plutôt qu'en coda car moins de chevauchement des gestes) et d'autre part à valider l'hypothèse 3.

Cependant les clusters combinant en attaque une consonne labiale suivie d'une consonne coronale ont un fort taux d'erreurs représentant un quart des modifications des clusters par stratégie de suppression. Cette configuration Labiale suivie de Coronale a pourtant été étudiée dans le cadre des structures syllabiques CVC et dans le cadre des attaques de deux syllabes consécutives. Les travaux de MacNeilage & Davis (2000), Rousset (2004), Valée, Rossato & Rousset (2009), Cissé (2009), Carrissimo-Bertola (2011) montrent que ces séquences Labiale suivies de Coronale sont plus de deux fois plus fréquentes en moyenne dans les langues que le patron inverse consonne Coronale suivie d'une consonne Labiale. Une explication à cette tendance universelle, appelée par MacNeilage & Davis (2000) « *effet Labial-Coronal* » ou « effet LC » a été proposée par Rochet-Capellan & Schwartz (2005). Selon ces auteurs, la tendance est le résultat d'une plus grande cohésion articulatoire entre consonnes si une labiale est produite avant une coronale, plutôt que l'inverse : une labiale permet l'anticipation de la coronale, donc une meilleure synchronisation des gestes pour les deux consonnes, sans gêner la perception des deux segments consonantiques, alors que l'anticipation d'un geste de fermeture labiale dans la production d'une coronale ne peut se faire sans perturber la réalisation d'une coronale. Cette explication, contestée par Carrissimo-Bertola (2011) car incapable de s'appliquer à un effet de consonne Labial-Vélaire qui n'est pas observé dans les langues du monde par rapport à Vélaire-Labial, trouve ici avec le résultat concernant les stratégies de modification par suppression, des clusters Labial + Coronal un autre argument de remise en question.

Hypothèse 3 : *Les apprenants vietnamiens modifient plus fréquemment les codas complexes du français que les attaques complexes, ces dernières étant moins marquées.*

Les analyses effectuées sur la production des groupes de consonnes initiés par /p/ et /k/ mettent en évidence les difficultés des apprenants vietnamiens à réaliser ces groupes plus en coda qu'en attaque de mot. Ce résultat est en conformité avec plusieurs études antérieures (entre autres Sato, 1984 ; Broselow, Chen & Wang, 1998 ; Hancin-Bhatt & Bhatt, 1997 ; Major, 1999) qui confirment l'*Hypothèse de la Différence de Marquage* selon laquelle les apprenants d'une langue seconde ont une plus grande difficulté à produire les codas complexes, ces dernières étant plus marquées. Rousset (2004) montre, à partir des lexiques syllabés de 16 langues, que les groupes de consonnes, présents seulement dans 7,86 % des quelques 160 500 syllabes de la base G-ULSID, occupent plus fréquemment la position d'attaque (67 %) que coda de syllabe (33 %). La tendance des clusters à apparaître majoritairement en attaque est peut être liée à des contraintes sur le nombre et la position des consonnes dans le cycle mandibulaire (Redford, 1999), la phase d'abaissement étant plus longue que la remontée, elle laisserait plus de place à l'articulation consonantique en position initiale de syllabe.

Il a été montré dans des études articulatoires sur les données de l'anglais (Byrd, 1996), du géorgien (Chitoran et al., 2002) et du français (Kühnert et al., 2006) que dans les séquences

de deux plosives, le degré de chevauchement gestuel est lié au lieu d'articulation. Un chevauchement moindre est observé dans des groupes de consonnes postérieure + antérieure (ex. [kt]) comparées à des suites antérieure + postérieure (ex. [pt]). Les résultats de notre étude s'inscrivent dans les résultats de ces études en montrant que quel que soit le mode de C₂, la suppression d'un des deux segments des clusters en attaque /p/+coronal (antérieure + postérieure) est 4 fois plus fréquente que pour des clusters de type /k/+coronal (postérieure + antérieure) (27,2 % vs. 6,3 %). C₁ est supprimée davantage (ex. /p/ dans *ptose* > [tɔz]). La raison semble résider dans l'organisation temporelle entre les gestes : plus de chevauchement dans une suite de consonnes [antérieure + postérieure] que [postérieure + antérieure]. Kühnert et collègues (2006), dans une étude sur la cohésion temporelle des groupes C+/l/ initiaux, ajoutent que ce pattern relève non seulement de facteurs de récupérabilité perceptuelle, mais aussi de contraintes simples du système moteur d'exécution : « le chevauchement limité dans le contexte [k] relève de contraintes sur la configuration globale de la langue pendant la production d'une vélaire. Dans le contexte d'un [p], non-lingual, la langue est libre pour se déplacer et anticiper l'articulation du [l] dès la phase de fermeture labiale. Par conséquent, le plateau de constriction peut être atteint plus tôt. Les articulations vélares, en revanche, exigent un mouvement holistique de la langue qui implique également le système bout/lame de la langue. Les caractéristiques spécifiques des plosives vélares, en particulier, sont connues pour contraindre fortement les mouvements de la totalité de la langue pendant l'intervalle de fermeture, ainsi que l'ont démontré maintes études physiologiques (Perkell, 1969 ; Löfqvist, Gracco, 2002 cité par Kühnert). Ainsi, dans le contexte [kl], la lame de la langue ne peut pas exécuter l'articulation du geste de [l] aussi tôt que pour [pl] » (Kühnert et al., 2006).

Hypothèse 4 : *Les caractéristiques phonétiques des consonnes en coda du vietnamien (consonnes non relâchées) ont des conséquences sur les stratégies de simplification des groupes consonantiques du français.*

Quoi que les séquences de consonnes existent dans les deux langues et qu'elles puissent être constituées de segments proches voire identiques, elles ne sont pas similaires sur le plan acoustique : les consonnes post-vocaliques en vietnamien, premier constituant des séquences de consonnes, présentent des particularités de réalisation phonétique. Rappelons que les plosives finales de syllabe sont plus souvent non relâchées en vietnamien, l'étude acoustique (cf. section 4.1.2.2) a montré l'absence du burst dans 73 % des réalisations des plosives en coda. Cette caractéristique typique de la langue maternelle a nettement influencé la production des consonnes en groupes du français. Il a été montré que plus de 53 % des erreurs sont dues à la modification par non relâchement des plosives en coda de syllabe ou de mot, parmi plusieurs autres stratégies utilisées par les apprenants (la suppression d'une ou des consonnes, la modification des caractéristiques des consonnes, l'ajout d'un élément

épenthétique ou le déplacement). Les habitudes productives modelées par le système sonore de la langue native semblent se projeter sur la langue cible. Les plosives, qui se trouvent en coda de syllabe du français, représentées par le même symbole API que les plosives du vietnamien bien que des analyses acoustiques révèlent des différences entre les deux, semblent être classées comme *similaires* aux plosives du vietnamien, selon le *Speech Learning Model* de Flege (1995). En effet, notre étude montre que les apprenants vietnamiens, quel que soit leur niveau de français, ont beaucoup de difficultés à réaliser les plosives en coda de syllabe du français (position post-vocalique) qui sont *semblable* (mais pas *identique*) à des plosives dans cette même position de la langue maternelle : plus la moitié (55,3 %) des plosives en C₁ sont réalisées non relâchées devant la frontière syllabique par les apprenants (cf. section 5.2.2.3.1). Les groupes initiés par les plosives /p/, /t/ et /k/ sont cependant plutôt bien récupérés qu'ils soient présents dans des pseudo-mots (respectivement 94,9 %, 89,76 %, 88,94 %) ou des mots du français (respectivement 88,29 %, 68,99 %, 80,6 %). Rappelons que les corrélats acoustico-perceptifs des plosives /p/, /t/ et /k/ en coda du vietnamien, identifiés dans les transitions VC#, sont plutôt bien récupérés par les sujets vietnamiens (taux d'identification correcte de 74 %, 84 % et 83 %) (cf. section 4.2.2). Or, lors de la production de ces plosives initiales des groupes de consonnes du français, les apprenants vietnamiens peinent à discerner la différence *phonétique* entre une plosive relâchée en coda de L2 et une plosive non relâchée en coda de L1. Il semble que pour les apprenants, ces deux types de plosives (non relâchées du vietnamien et relâchées du français) soient *similaires* (car bien récupérées à partir du signal audio dans les deux langues, voir section 4.2.2 et 5.1 (expériences 2A et 2B)) et doivent comporter les mêmes caractéristiques phonétiques lors de la production. Ainsi, il a été observé qu'en position inter-syllabique, la séquence Plosive+Plosive (ex. *capture* [kaptɥʁ], *chapka* [ʃapka], *snack-bar* [snakbæʁ]) est la plus réussie en perception, que cette combinaison soit insérée dans des pseudo-mots ou mots du français. Cette séquence semble être perçue comme un *bon exemplaire* en fonction de sa similarité par rapport à des segments présents dans l'espace phonologique de la langue native (Best, 1995). Au contraire, cette séquence totalise le plus de modifications en production par les étudiants des deux niveaux en raison du non relâchement très fréquent de C₁ dans ce groupe. Le taux de production correcte des séquences initiées par /p/, /t/ et /k/ à l'inter-syllabique est de 14,6 %, 16,04 % et 24,4 %. Ce qui suggère une disparité pouvant exister entre perception et production et que cette relation entre perception et production semble fortement influencée par les caractéristiques phonético-phonologiques du système sonore de la langue première. Flege (1995) note que les sons de la L2 pourraient être d'abord identifiés comme des *allophones* de la L1. Au fur et à mesure, lors d'une étape plus avancée de l'apprentissage, les apprenants discerneraient la différence phonétique entre des sons de la L2 avec ceux les plus proches de la L1. A ce stade, de nouvelles catégories *phonétiques* peuvent être établies pour ces nouveaux sons de la L2 qui sont indépendants des représentations déjà établies pour ceux de la L1. L'introduction d'une nouvelle valeur d'exemplaire (*allophone*) qui n'est

pas présente dans les systèmes phonologiques du français peut générer des difficultés de perception chez un locuteur natif du français.

Hypothèse 5 : *Les combinaisons de phonèmes consonantiques plus marquées selon les universaux du langage seront plus difficiles à produire que les combinaisons moins marquées. Et dans ce cadre, selon que les clusters respectent ou violent le Principe de Sonorité, leur niveau de difficulté d'apprentissage devrait être différent.*

Rappelons que les groupes de consonnes observés commencent soit par une plosive, soit par une nasale. L'objectif est de confronter les sujets vietnamiens apprenant du FLE à des séquences sonores identiques dans les deux langues au niveau des consonnes qui les composent, mais différentes au niveau de leur position intra-syllabe ou inter-syllabe. Les résultats montrent que le mode et le lieu d'articulation des C_1 initiales des clusters C_1C_2 influencent la performance des apprenants en production, quel que soit leur niveau. Un cluster commençant par une plosive pose deux fois plus de difficulté qu'un cluster commençant par une nasale, que ce soit en position inter-syllabique ou finale de mot.

Il a été observé dans notre étude que les difficultés de production varient selon le type de consonnes C_1+C_2 dans les combinaisons. Quelle que soit la position des clusters, les combinaisons de même mode d'articulation Plosive+Plosive ainsi que les clusters Plosive+Fricative et Nasale+Fricative présentent un taux d'erreurs plus élevé que les autres combinaisons. Au contraire, la combinaison Plosive+Latérale, quelle que soit sa position dans le mot, occasionne le moins de difficultés aux apprenants. L'Hypothèse du « *Paramètre Minimal de la Distance de Sonorité* », formulé par Broselow et Finer (1991), selon laquelle le marquage d'une attaque complexe dépend de la distance de sonorité entre ses composants, semble pouvoir expliquer ce résultat, non seulement pour les clusters en attaque, mais aussi en coda. Les résultats de l'expérience 3 montrent que les clusters Plosive + Plosive, Plosive + Fricative, Plosive + Nasale dont la distance de sonorité des composants est soit nulle soit très proche, posent plus de problème lors de la production, quelle que soit leur position, attaque ou coda. À l'inverse, les clusters contenant une plosive et une latérale, donc des consonnes dont la distance sur l'échelle de sonorité est plus grande, sont mieux récupérés et réalisés par les apprenants dans toutes les positions. Il a été aussi démontré que les clusters de même mode d'articulation ont une fréquence faible dans les langues du monde, idem pour Nasale+Liquide (Kawasaki, 1982 ; Rousset, 2004), puisque selon Kawasaki (1982), Krakow (1999), en opposition aux groupes Obstruente+Liquide, ces groupes manquent de modulation acoustique.

L'hypothèse concernant la distance de sonorité ne peut cependant pas expliquer tous les résultats observés. La combinaison Plosive+Plosive plus marquée est donc prédite plus difficile à apprendre que Plosive+Fricative. Ce constat est valable pour les combinaisons en

coda où celles impliquant deux plosives obtiennent jusqu'à 97,5 % d'erreurs, les clusters contenant une plosive et une fricative ont 87,5 % d'erreurs en production par rapport à la cible [$F(1,39) = 8,3$; $p = 0,006$]. Alors que le résultat est à l'inverse si on observe les combinaisons en attaque : la combinaison Plosive+Fricative présente significativement un taux d'erreurs plus élevé que Plosive+Plosive (70,8 % vs. 50,8 %) [$F(1,39) = 11,32$; $p = 0,002$].

Ce chapitre consacré à la présentation d'une série d'études en perception et production des consonnes simples et en séquences du français par des apprenants vietnamiens de deux niveaux, intermédiaire et avancé, a pour l'objectif d'estimer l'impact de la syllabe et de ses frontières dans la L1 dans l'acquisition des groupes de consonnes d'une langue seconde. Plus généralement, cette partie de la thèse cherche à montrer la présence ou l'absence d'un effet de la syllabe de la L1 et de ses frontières dans le système phonologique interlangue. Les résultats suggèrent une forte influence des structures syllabiques de la L1 sur l'apprentissage de la L2.

Toutes les expériences ont mis en évidence les effets de la position sur la perception et la production des consonnes simples et en séquences du français. Les analyses attestent que les étudiants montrent de grandes difficultés à percevoir et particulièrement à récupérer les consonnes en position finale par rapport aux positions initiale ou inter-syllabique. Ces résultats, en conformité avec les études antérieures (entre autres Anderson, 1987 ; Benson, 1986 ; Major, 1996, 1999 ; Sato, 1984 ; Weinberger, 1987 ; Hansen, 2001 ; Broselow, Chen & Wang, 1998 ; Hancin-Bhatt & Bhatt, 1997) qui démontrent que l'acquisition des clusters comme des consonnes simples en coda reste relativement problématique par rapport à d'autres positions.

Il a été également montré dans la présente étude que non seulement la structure syllabique CVC dominante de la langue native, mais aussi les caractéristiques phonétiques de consonnes en coda du vietnamien (consonnes non relâchées) ont de fortes conséquences sur les stratégies de réalisation des groupes consonantiques du français : le taux de non relâchement des C_1 en coda de syllabe et la suppression de l'élément le moins sonore dans le groupe recouvrent la plupart des erreurs de production chez les apprenants.

Les combinaisons les mieux peçues et les mieux produites par les étudiants de deux niveaux ont été analysées. D'une manière générale, les combinaisons plus marquées selon les tendances universelles ont été plus difficiles à récupérer et à réaliser.

Les résultats obtenus dans le cadre de cette étude ont permis de discuter les aspects théoriques. Par rapport aux modèles en perception et production des langues secondes (Best, 1995 ; Flege, 1995), les sujets sont tous soit dans des catégories de la L1, soit dans des catégories de L1 et L2. Les erreurs sont des interprétations des éléments sonores de la L2 par rapport à la L1, au niveau des segments comme au niveau des structures possibles de la syllabe.

Discussion générale et perspectives

L'objet central de cette présente étude porte sur la difficulté à réaliser les groupes de consonnes du français par des apprenants vietnamiens. Afin de répondre à la question centrale de la thèse « *Quelles sont les véritables raisons de la difficulté persistante à réaliser des clusters chez des apprenants vietnamiens ?* », une série d'expériences acoustiques et perceptives des séquences de consonnes communes aux deux langues ont été menées. Les éléments principaux suivants ont été mis en évidence ou confirmés :

Concernant la perception et la production des consonnes en séquences du vietnamien, langue source (chapitre IV) :

1. Les plosives sont plus souvent non relâchées en coda de syllabe à la frontière intra-mot plutôt qu'aux frontières inter-mot. Dans les cas où le burst apparaît, il est plus bref et peu intense par rapport à leur réalisation en attaque (section 4.1).
2. Les plosives en coda, même non relâchées, peuvent être bien identifiées grâce à des informations sur leur lieu d'articulation récupérées dans la transition avec la voyelle précédente ($\Delta F1$, $\Delta F2$ et Intensité) (section 4.2). La présence d'un burst améliore l'identification.
3. Le type de frontière syllabique influence la réalisation des consonnes en coda de même contexte vocalique : consonne plus longue en coda de mot que de syllabe 1 de composé lexical ; différences de $\Delta F0$, $\Delta F1$, ΔI dans la partie de transition voyelle-plosive en fonction de la frontière syllabique (section 4.1).
4. Les éléments sub-syllabiques comportent des indices acoustiques porteurs d'information sur les frontières de la syllabe. Les durées de la consonne finale et de la voyelle pré-consonantique sont les corrélats acoustiques pouvant permettre l'identification du type de frontière (section 4.3).

Concernant la perception et la production des consonnes simples, clusters et séquences consonantiques inter-syllabe du français, langue cible (chapitre V) :

5. La position des consonnes dans la syllabe, dans le mot, affecte le taux de réussite et donc l'apprentissage : il est plus difficile de percevoir et produire correctement une consonne ou un groupe de consonnes en coda qu'en inter-syllabe ou en attaque.
6. Le fait que le groupe consonantique soit tauto-syllabique (clusters) gêne la production et dans une moindre mesure, la perception, mais les effets sont plus remarquables en coda qu'en attaque.
7. Le non relâchement des codas en C1 et la suppression des segments les moins sonores en C2 sont les stratégies les plus utilisées pour simplifier la production des groupes de consonnes du français.
8. Les combinaisons plus marquées sont plus difficiles à percevoir et à produire.

Les résultats obtenus à l'issue des expériences des chapitres IV et V amènent des éléments de discussion sur les hypothèses générales suivantes, posées en entrée de cette recherche.

1. Etudier la perception de la parole dans la langue source et dans la langue cible peut rendre compte des difficultés que les apprenants rencontrent.
2. Analyser les différences de réalisation des séquences de consonnes de la langue source et de la langue cible peut permettre également de dégager des éléments du crible phonologique.

Une explication souvent avancée aux difficultés rencontrées par des apprenants d'une langue seconde avec la réalisation des phonèmes de la L2 qui sont différents de ceux de la L1 est que ces catégories de sons de la L2 en question pourraient ne pas être exactement perçues. Le lien unissant perception et production a été d'ailleurs mis en exergue en général par plusieurs recherches pour le langage humain (Schwartz, Abry, Boë, Cathiard, 2002 ; Lindblom, 1990 ; Liberman et Mattingly, 1985 ; Stevens, 1996 entre autres), aussi en neuro- et psycho-linguistique (Pulvermüller et Fadiga, 2010 ; Graski, Lamalle et Sato, 2011). En apparence, cette explication est plutôt raisonnable, au moins en acquisition de la L1 : chez les enfants, la perception des contrastes phonémiques est toujours développée avant leur production (Smith, 1973 ; Ingram, 1976 ; Menyuk, 1977 ; Vihman, 1996). Selon cette approche, si les apprenants ne peuvent pas percevoir les catégories de sons de la L2 correctement, ils ne pourront pas alors les produire avec succès. La notion de *crible phonologique* à l'initiative de Troubetzkoy (1939) et largement étudiée par la suite dans de nombreux travaux sur l'acquisition des langues étrangères (entre autres Lado, 1957 ; Best, 1995 ; Flege, 1995) soutient largement cette idée en soulignant que « pour bien produire un son, il faut d'abord l'avoir bien perçu » (Borrell, 1990). Le rapport étroit entre perception et production a été utilisé comme base de la méthode verbo-tonale de correction phonétique en didactique des langues secondes et/ou langues étrangères. Ainsi, Renard (1989) affirme que :

« Si nous voulons donc reproduire un message donné, il nous faudra d'abord le percevoir correctement. Compte tenu de ce qui a été dit du caractère sélectif de la perception, cela impose au pédagogue l'impérieuse nécessité d'apprendre à percevoir » (Renard, 1989 : 27)

Selon les tenants de cette méthode de correction phonétique, pour bien faire (re)produire un son d'une autre langue, il convient d'abord de le faire bien percevoir malgré la résistance du « crible phonologique » de la langue maternelle. Un son correctement perçu ne devrait alors poser aucun problème de réalisation. Pourtant, les études peinent à montrer une telle réalité. Par exemple, Borrell (1996) remarque qu'une bonne perception d'un élément sonore du langage ne garantit pas automatiquement une reproduction convenable de celui-ci. Il est possible de percevoir un son d'une langue étrangère sans être capable de le reproduire. Il cite le cas, par exemple, de la vibrante alvéolaire /r/ de l'espagnol que beaucoup d'apprenants

étrangers n'arrivent jamais à réaliser correctement alors qu'ils la perçoivent sans aucune difficulté. C'est aussi le cas de /ç/ en allemand que certains apprenants français reconnaissent parfaitement mais qu'ils confondent toujours, sur le plan articulatoire, avec /ʃ/ français. La proximité articulatoire des deux sons (respectivement médio et pré-palatal, mais dorsale pour la première et apicale pour la seconde), expliquerait cette difficulté : les sons proches perceptivement sont plus difficiles à acquérir (Best, 1995 ; Flege, 1995).

L'ensemble des expériences en perception et production des clusters du français menées dans le cadre du chapitre V montre que les tests de perception sont largement mieux réussis que les tests de production. Dans la tâche de répétition des consonnes simples du français, les apprenants intermédiaires et avancés ont réussi à 75 % moyenne à reproduire les consonnes cibles, et respectivement à 86,7 % et 84 % à produire correctement les consonnes /p t k m n ŋ/. Ces dernières existent dans les deux langues mais en tant que coda en vietnamien, donc en tant qu'élément C₁ des séquences biconsonantiques. Lorsque ces dernières sont insérées dans des séquences pseudo-mots, elles sont identifiées correctement (en moyenne à 79,8 % et 80,7 %, respectivement selon le niveau intermédiaire ou avancé des apprenants). Les scores restent toujours bien supérieurs au seuil du hasard lors de la perception des groupes de consonnes dans les mots (respectivement 78,8 % et 76,7 %). Pourtant, la production de ces consonnes en groupe est à l'encontre de ce qui était attendu : malgré les performances en perception, les étudiants présentent des difficultés à réaliser correctement les clusters et les séquences de consonnes (taux de réussite de 33,3 %), alors que la tâche consistait à lire une liste de phrases porteuses, donc une tâche sans contexte linguistique et considérée comme formelle, à même de fournir les meilleures réalisations selon certains auteurs (Labov, 1969 ; Schmidt, 1977 ; Dickerson & Dickerson, 1977 ; Tarone 1979, 1982, 1983, 1985).

Il est à remarquer que certaines combinaisons consonantiques Plosive+Plosive et Plosive+Fricative les moins bien récupérées dans l'expérience 5.1.2B sur la perception des groupes de consonnes dans les mots sont effectivement moins bien produites dans l'expérience 5.2. Par exemple, /kt/ dans *insecte* [ɛ̃sɛkt] perçu correctement dans 53 % des cas, n'est réalisé correctement que dans 3 % des cas par les apprenants de niveau intermédiaire, les scores ne s'améliorent pas pour autant chez les apprenants de niveau avancé (68 % en perception et 2 % en production). /ks/ dans [feniks] *phénix* a été correctement perçu à 54 % et réalisé correctement dans 13 % des cas par des apprenants intermédiaires. La même tendance est observée chez d'autres groupes /td/ *hot-dog*, /kt/ *dictée*, /ps/ *psychique*. Les combinaisons Plosive+Latérale /kl/ dans *check-list* [ʃɛk.list], *cycle* [sikl], *cloche* [klɔʃ] mieux récupérées sont également observées mieux produites par les apprenants des deux niveaux (score moyen supérieur de 80 % dans tous les cas considérés). Cependant, la figure 6.1 montre globalement qu'aucune corrélation entre perception et production de chacun des 32 groupes de consonnes du français testés dans l'étude n'est observée chez les étudiants de

niveau intermédiaire (ddl = 30 ; $r = 0,3$; $p > 0,05$), ni chez les avancés (ddl = 30 ; $r = 0,1$; $p > 0,05$).

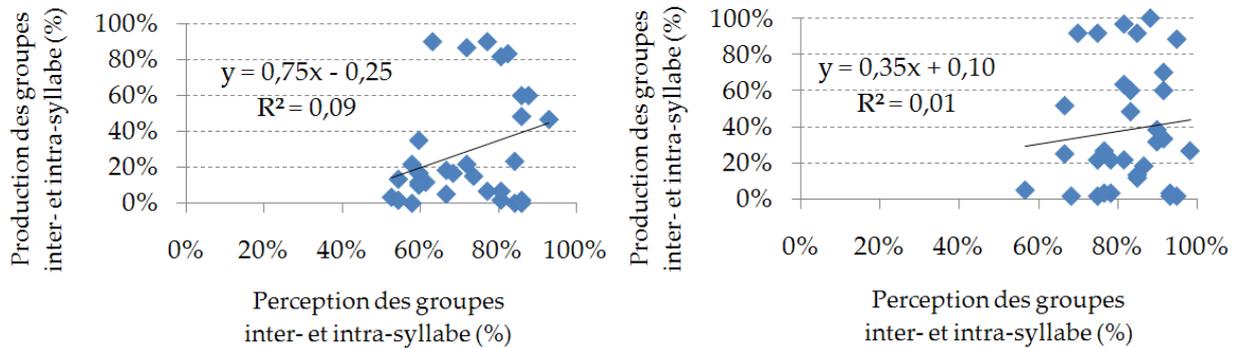


Figure 6.1. Corrélation entre scores de perception correcte (en abscisse) et scores de production correcte (en ordonnée) de chacun des 32 groupes de consonnes inter- et intra-syllabe chez des apprenants de niveau intermédiaire (à gauche) et avancé (à droite).

Ces résultats mettent en évidence la dissociation qui peut exister entre perception et production et rejettent l'hypothèse qu'en général, un cluster bien perçu est forcément bien produit. Une bonne perception auditive est une condition nécessaire pour l'acquisition et la production du langage (cf. MacNeilage, 1998) mais ce n'est apparemment pas suffisant pour aider l'apprenant à avoir une prononciation authentique.

Comment expliquer alors le taux de réussite assez élevé lors du test d'écoute et de répétition de consonnes simples du français par les apprenants vietnamiens ? La tâche de répétition avait pour objectif de fixer une forme sonore et d'éveiller l'aptitude à la reconnaissance et à la production de la consonne à l'intérieur des pseudo-mots. La répétition se fait « dans l'empan de la mémoire immédiate ... l'auditeur passe à l'encodage (du signal acoustique à des unités linguistiques et/ou phonétiques) et devient ensuite le locuteur en transformant l'image sonore en mouvements articulatoires par des commandes neuromusculaires » (Lhote, 1988). Le son dans ce cas n'a pas à être interprété phonologiquement, mais seulement reconnu phonétiquement et imité. Il suffit d'avoir une bonne oreille (Borrell, 1996).

Si les apprenants montrent qu'ils disposent d'une capacité de traitement de l'information auditive des matériaux acoustiques que ce soit de la langue source (cf. section 4.2 sur la perception des consonnes en séquences du vietnamien) ou de la langue cible (cf. section 5.2 sur la perception des consonnes en séquences du français), la problématique se pose en ces termes : pourquoi la réactivation de ces informations dans une tâche de production en langue seconde (cf. section 5.3 sur la production des groupes de consonnes du français) se fait-elle aussi difficilement ?

La production en langue seconde semble être influencée par des facteurs linguistiques. Harmegnies, Delvaux, Huet & Piccaluga (2005) attestent qu'« en didactique des langues,

l'idée est communément admise que l'équipement psycholinguistique correspondant à la langue maternelle constitue un frein à la performance en langue seconde, l'apprenant ayant naturellement tendance à traiter la langue constituant la cible de l'apprentissage de la même manière qu'il traite sa langue maternelle ; d'où divers inconvénients liés au fait que les langues différentes ayant des propriétés différentes, le système de traitement qui convient à l'une ne peut au mieux qu'imparfaitement convenir à l'autre ». Plusieurs résultats des expériences menées dans la présente étude montrent que les habitudes productives modelées par le système sonore de la langue source se projettent sur la langue cible. La plupart des erreurs dans la production des groupes de consonnes du français par des apprenants vietnamiens retrouvent la source dans la langue de départ (suppression des consonnes permettant de retrouver la structure CVC ; C₁ souvent non relâchées en coda de syllabe...). Même si les Vietnamiens apprenants du FLE sont capables d'identifier correctement une consonne en coda dans leur langue native, d'ailleurs souvent produites sans relâchement audible (cf. section 4.3) ou dans la langue seconde (cf. section 5.2), lorsqu'ils sont confrontés à la production de cette consonne à l'intérieur d'un cluster ou d'une séquence de consonnes du français, ils appliquent le système de traitement de leur langue maternelle inadapté au cas du français.

Bien que des facteurs phonético-phonologiques présents dans la L1 jouent un rôle dans les erreurs de la production de groupes de consonnes en L2, les résultats du chapitre V montrent que l'influence du système phonologique et phonotactique de la L1 ne peut pas expliquer toutes les erreurs. D'autres sources sont nécessaires pour expliquer les difficultés des apprenants qui ne peuvent pas être directement liées aux différences entre la L1 et la L2 (par exemple, la difficulté particulière pour la combinaison Plosive+Plosive et Plosive+Fricative en production par rapport à d'autres combinaisons). L'Hypothèse de la Différence de Marquage (Eckman, 1977) et l'hypothèse du « *Paramètre Minimal de la Distance de Sonorité* », formulé par Broselow et Finer (1991), basées sur les préférences typologiques universelles et son rôle dans l'apprentissage des langues, trouvent dans les résultats de cette étude, une nouvelle illustration. Comme le montrent les études antérieures (Beebe, 1987 ; Broselow, 1987 ; Eckman, 1977 ; Greenberg, 1965 ; Sato, 1984, 1985 ; Tarone, 1987 ; Weinberger, 1987), la marque influence la phonologie de l'interlangue, comme elle influence celle des langues natives. Les résultats de cette présente étude montrent clairement que même si les clusters en attaque et en coda sont absents dans la langue source, les apprenants acquièrent plus facilement les clusters les moins marqués (ceux situés en position d'attaque). Sur ce point, la perception comme la production de ces clusters est meilleure.

Selon Tarone, Frauenfelder & Selinker (1978), Frauenfelder & Porquier (1979), Perdue & Porquier (1980), la production pourrait être aussi influencée par des facteurs extralinguistiques. Ces auteurs considèrent que l'apprentissage est un processus global à caractère commun à tous les apprenants comportant une série de variables fixes (âge,

motivations, besoins, aptitudes) ou aléatoires (stratégies personnelles, attitudes, durée d'exposition à la langue, apprentissage antérieur). Ces facteurs sociaux, personnels, affectifs ou situationnels forment un « crible psychologique » qui entraîne rapidement une fossilisation de fautes. L'acquisition dépend alors des variables dont on ignore le taux de variabilité dans l'apprentissage, ce qui va poser problèmes en didactique où le phénomène temps est souvent la première variable à considérer. Un paradoxe apparaît : la pédagogie est souvent pensée en termes de groupe alors que l'acquisition se fait en termes d'individu.

Les participants à ces tests avaient des connaissances acquises préalables de la langue cible, certes. Mais l'exposition à cette langue n'est que d'une dizaine d'heures par semaine. En dehors des cours, les étudiants replongent dans le bain linguistique de leur langue maternelle. Leur cours contient des éléments théoriques de base de la phonétique seulement en troisième année à l'Université. Des groupes de consonnes sont particulièrement négligés dans le programme de l'enseignement du français. Les corrections phonétiques dans les heures de pratique de la langue concernent souvent le segment vocalique ou consonantique dans sa forme isolée. Or, l'acquisition de la phonologie d'une langue seconde ne se limite pas à la capacité à produire des segments isolément. L'apprenant doit aussi savoir combiner ces segments à d'autres pour produire les séquences acceptables dans la langue cible. Si par exemple, la suppression se produit aussi fréquemment hors contexte linguistique (par exemple, [k] dans [taks] est supprimé, *taxe* deviendra *tasse* [tas]), cela peut entraîner une grande possibilité d'ambiguïté et gêner ainsi la compréhension de l'interlocuteur. Un manque de conscience de la difficulté dans la production des groupes de consonnes entraînerait par conséquent une acquisition incertaine.

Andersen (1983) compare les productions des apprenants à une pidginisation de la langue cible, du fait même que, souvent, elles semblent se figer surtout en matière phonéto-phonologique (structures phonémiques et prosodiques). Cette pidginisation est définie comme étant une acquisition linguistique dans des conditions où d'une part, l'exposition à la langue est restreinte et d'autre part, la saisie des données linguistiques est fortement réduite à cause des facteurs socio-affectifs, de contraintes de temps ou pour des raisons de complexité linguistique. Cela s'applique donc à la langue étrangère et pose un problème pour l'enseignement.

On est donc ici dans une double perspective théorique, didactique et phonéto-phonologique. Faire le lien entre les problèmes d'acquisition des clusters du français par les apprenants vietnamiens, les différences existant entre la réalisation des séquences consonantiques dans les deux langues est une étape indispensable et préalable à toute proposition d'une méthode d'apprentissage. Cette recherche constitue ainsi une étape nécessaire et préliminaire à la conception d'une méthode qui devrait être capable de donner aux apprenants les moyens d'acquérir de nouvelles habitudes articulatoires leur permettant de produire des clusters de manière acceptable. Il faudra, dans ce cadre, réfléchir à des

techniques diverses destinées à déconditionner l'apprenant de ses habitudes auditives et articulatoires acquises lors de l'apprentissage du système de la langue maternelle et à la reconditionner au système de la langue étrangère.

La présente étude est focalisée sur les séquences de consonnes communes aux deux langues. Il est important de l'élargir avec d'autres types de clusters comme les études antérieures auprès d'apprenants vietnamiens (Benson, 1988 ; Osburne, 1996 ; Sato, 1984) dans le but d'observer les hiérarchies de marquage des clusters selon les généralisations universelles de Greenberg (1965) et de comparer les difficultés rencontrées entre groupes consonantiques présents *vs.* absents de la langue source. Il sera aussi important de mener une étude longitudinale en perception et production des clusters auprès des mêmes groupes d'apprenants, ce que nous n'avons pas eu la possibilité de faire pendant la durée impartie de cette recherche.

La recherche en acquisition d'une langue seconde a montré que la variation de tâche a une influence sur la phonologie de l'interlangue (Labov, 1971 ; Schmidt, 1977 ; Tarone 1979, 1982, 1983 ; Lin, 2001, 2003 ; Major, 1987 ; Weinberger, 1987). La tâche utilisée dans notre étude était la lecture de phrases porteuses, ce qui peut conduire à des résultats différents d'une tâche de parole spontanée. Il sera intéressant de comparer l'acquisition des séquences communes aux deux langues dans d'autres types de tâches moins formelles.

Enfin, l'aspect novateur de la partie phonétique de la langue vietnamienne est d'avoir montré l'existence d'une influence de la frontière syllabique sur la réalisation des consonnes en coda, qui suggère des degrés de coarticulation différents en fonction de la frontière syllabique. Il est déjà prévu dans la suite de cette étude de vérifier le degré de coarticulation des consonnes en séquences dans les deux langues avec des données de production par des Vietnamiens natifs, obtenues avec un articulographe électromagnétique EMA® (Carstens). L'analyse de ces données dans le cadre théorique de la phonologie articulatoire permettra d'analyser plus finement les productions des sujets vietnamiens dans une tâche de réalisation des clusters du français et de vérifier si des éléments dans la superposition des gestes entraînent des effets de masquage qui pourraient être des facteurs gênant la perception.

Au-delà de la problématique de l'acquisition des systèmes sonores des langues du monde et de ses aspects didactiques, les études dans ce champ de recherche concernent le fonctionnement du langage en général. Nos résultats, en confirmant l'importance de la marque, dans l'acquisition d'une langue cible, et plus particulièrement pour des structures n'appartenant pas à la langue source (langue native), montrent la nécessité de comprendre l'origine des tendances universelles dans les systèmes sonores des langues du monde, notamment des liens entre les systèmes phonologiques et les capacités sensorimotrices. Dans ce cadre, l'étude de la phonologie interlangue, de l'émergence de ses unités et de ses

structures, son évolution constitue un observatoire sans doute capable de refléter des éléments fondamentaux du fonctionnement du langage humain en général et de ses unités.

Bibliographie

- Abramson, A. S. & Tingsabadh, K. (1999). Thai Final Stops : Cross-Language Perception. *Phonetica* 56(3-4): 111-122.
- Agrawal, A. & Wen, C. L. (1975). Aspects of voiced speech parameters on the intelligibility of PB words. *Journal of the Acoustical Society of America* 57(1): 217-222.
- Akahane-Yamada, R., Tohkura, Y., Bradlow, A. R. & Pisoni, D. B. (1996). Does training in speech perception modify speech production? In *4th International Conference on Spoken Language Processing* Philadelphia, PA, USA.
- Albareda-Castellot, B., Pons, F. & Sebastián-Gallés, N. (2010, in press). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm *Developmental Science*
- Altenberg, E. P. (2005). The judgment, perception and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching* 43: 53–80.
- Anan, F. (1981). A study of intonation interference (translation from Japanese). *The Bulletin of The Phonetic Society of Japan* 167: 9-13.
- Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. In *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 279-291 (Eds G. Ioup and S. Weinberger). New York: Newbury House/Harper and Row.
- Aoyama, K., Flege, J. E., Guion, S., Akahane-Yamada, R. & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–250.
- Archibald, J. (1998). *Second language phonology*. John Benjamins Publishing Company.
- Aubergé, V., Boel, J. & Lefèvre, J. P. (1988). Lexiques et groupes consonantiques. In *XVIIèmes Journées d'Etude sur la Parole*, 55-60 Nancy.
- Barlow, J. A. (2005). Phonological change and the representation of consonant clusters in Spanish: A case study. *Clinical Linguistics & Phonetics* 19(8): 659–679.
- Bayley, R. (1996). Competing constraints on variation in the speech of Chinese learners of English. In *Second language acquisition and linguistic variation*, 97-120 (Eds R. Bayley and D. R. Preston). Amsterdam: Benjamins.
- Beebe, L. (1987). Myths about interlanguage phonology. In *Interlanguage phonology*, 165-175 (Eds G. Ioup and S. Weinberger). Cambridge, MA.: Newbury House.
- Bell-Berti, F., Raphael, L. J., Pisoni, D. B. & Sawusch, J. R. (1979). Some relationship between Speech Production and Perception. *Phonetica* 36: 373-383.
- Benson, B. (1988). Universal preference for the open syllable as an independent process in

- interlanguage phonology. *Language Learning* 38: 221-242.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research*, 171-204 (Ed W. Strange). Timonium: MD: York Press.
- Best, C. T., McRoberts, G. W. & Sithole, N. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14: 345-360.
- Best, C. T. & Strange, W. (1992). Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants. *Haskins Laboratories Status Report on Speech Research* 109/110: 89-108.
- Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In *Second language speech learning: The role of language experience in speech perception and production*, 13-34 (Eds M. J. Munro and O.-S. Bohn). Amsterdam: John Benjamins.
- Billières, M. (1995). Didactique des langues et phonétique. La place du verbo-tonal. *revue de Phonétique Appliquée* 114: 43-63.
- Blevins, J. (1995). The syllable in phonological theory. In *A Handbook of Phonological Theory*, 206-244 (Ed J. Goldsmith). Oxford: Blackwell.
- Bloomfield, L. (1933/1984). *Language*. New York: Holt.
- Bohn, O.-S. & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14: 131-158.
- Borrell, A. (1990). Perception et (re)production dans l'apprentissage des langues étrangères. Quelques réflexions sur les aspects phonético-phonologiques. *revue de Phonétique Appliquée* 95-96-97: 107-114.
- Borrell, A. (1991). Importance de la phonétique dans l'enseignement / apprentissage des langues secondes et étrangères. *revue de Phonétique Appliquée* 99-100-101.
- Borrell, A. (1992). Les rapports entre perception et (re)production dans l'acquisition des langues secondes et/ou étrangères. *Cahiers du Centre Interdiscipline des Sciences du Langage* 8(Université de Toulouse II): 29-41.
- Borrell, A. (1993). Relation entre les aspects articulatoires et les aspects acoustiques en phonétique. *revue de Phonétique Appliquée* 7: 97-111.
- Borrell, A. (1996). Parallèle entre perception et production ? Complexité du lien entre reconnaissance et production des unités phonético-phonologiques. *La Linguistique* 32:

- 105-116.
- Borrell, A. (1996). Systématisation des erreurs de production et donc de perception ... *revue de Phonétique Appliquée* 118-119: 1-15.
- Borrell, A. (1996). Systématisation des erreurs de production et donc de perception chez les apprenants étrangers ? *revue de Phonétique Appliquée* 118-119: 1-16.
- Boubnova, G. (2006). Correction phonétique : enseignement du français/du russe à des apprenants russophones/francophones. *Revue française de linguistique appliquée* Vol. XI / 1: p.7 - p.19.
- Boutin, B. A., Lyche, C. & Prignitz, G. (2007). Les enquêtes PFC en Afrique. In *Enseignement du français & Travaux en cours, Phonologie du Français Contemporain*, Vol. Bulletin n°7, 297-330 (Eds S. Detey and D. Nouveau).
- Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants*. Paris: Editions Odile Jacob.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B. & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61 (5): 977-985.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299-2310.
- Brière, E. J. (1966). An Investigation of Phonological Interference. *Language* 42(4): 768-796.
- Broselow, E. (1983). Non-obvious transfer: On predicting epenthesis errors. In *Language Transfer and Language Learning*, 269-280 (Eds S. Gass and L. Selinker). Rowley: MA: Newbury House.
- Broselow, E. & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research* 7: 35-59.
- Browman, C. & Goldstein, L. (1992). Articulatory Phonology: an overview. *Phonetica* 49: 155-180.
- Brunelle, M. (2003). Tone coarticulation in Northern Vietnamese. In *Proceedings of the 15th international congress of phonetic sciences*, 2673-2676 Barcelona.
- Brunelle, M. (2009). Northern et Southern Vietnamese Tone Coarticulation: A Comparative Case Study. *Journal of Southeast Asian Linguistics* 1: 49-62.
- Brunelle, M. (2009). Tone perception in Northern et Southern Vietnamese. *Journal of Phonetics* 37: 79-96.
- Brunelle, M., Nguyễn, D. D. & Nguyễn, K. H. (2010). A Laryngographic and Laryngoscopic Study of Northern Vietnamese Tones. *Phonetica* 67: 147-169.

- Burfin, S., Savariaux, C., Granjon, L., Sanchez, C., Tran, T. T. H., Faraco, S. S. & Kandel, S. (2011). Overcoming phonological deafness in L2 conversations by perceiving the facial the movements of the speaker. In *Workshop on Bilingualism : Neurolinguistic and Psycholinguistic Perspectives*, 41 Aix en Provence, France.
- Butt, M. (1992). Sonority and the Explanation of Syllable Structure. *Linguistische Berichte* 137: 45-67.
- Byrd, D. (1992). Perception of Assimilation in Consonant Clusters:A Gestual Model. *Phonetica* 49: 1-24.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24: 209 – 244.
- Cairns, C. E. & Feinstein, M. H. (1982). Markedness and the Theory of Syllable Structure. *Linguistic Inquiry* 13(2): 193-215.
- Calabrino, A. (2006). Effets acoustiques du débit sur la production de la parole chez des locuteurs enfants et adultes. In *X^e Colloque des étudiants en sciences du langage*, 61-83.
- Cao Xuân-Hạo (1975). Le problème du phonème vietnamien *Etudes Vietnamiennes* 40: 99-127.
- Cao Xuân-Hạo (1985). *Phonologie et linéarité. Réflexions critiques sur les postulats de la phonologie contemporaine*. Paris: SELAF.
- Cao Xuân-Hạo (2001). *Tiếng Việt : mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa* (Tr. Le vietnamien : quelques problèmes phonétiques, grammaticales et sémantiques). Hochiminh City: Nhà xuất bản Giáo dục (Maison d'édition de l'Education).
- Carlisle, R. (1988). The effect of markedness on epenthesis in Spanish/ English interlanguage phonology. *Issues and Developments in English and Applied Linguistics* 3: 15-23.
- Carlisle, R. (1994). Markedness and environment as internal constraints on the variability of interlanguage phonology. In *First and Second Language Phonology*, 223-249 (Ed M. Yavas). San Diego, CA: Singular.
- Carlisle, R. (1998). The acquisition of onsets in a markedness relationship. A longitudinal study. *Studies in Second Language Acquisition* 20: 245-260.
- Carlisle, R. (2001). Syllable Structure Universals and Second Language Acquisition. *International Journal of English Studies* 1(1): 1-19.
- Carrissimo-Bertola, M. (2011). Tendances et universaux des structures syllabiques : l'effet Labial-Coronal en italien. Université Stendhal, Grenoble (Mémoire de Master 2 Recherche en Sciences du Langage).
- Carton, F. (1974). *Introduction à la phonétique du français*. Paris: Bordas.
- Castelli, E. & Hierholz, A. (2006). « Locus equation » pour les consonnes /b/, /d/ et /g/ du

- vietnamien. In *XXVIèmes Journées d'Étude sur la Parole* Rennes.
- Chang, Y.-C., Hong, J. & Hallé, P. (2007). English Cluster Perception by Taiwanese Mandarin Speakers. In *ICPhS XVI*, 797-800 Saarbrücken, Germany.
- Chin, S. B. (1996). The role of sonority hierarchy in delayed phonological systems. In *Pathologies of speech and language: Contributions of clinical phonetics and linguistics*, 109-117 (Ed T. W. Powell). New Orleans, LA: International Clinical Phonetics and Linguistics Association.
- Chitoran, I., Goldstein, L. & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In *Papers in Laboratory Phonology VII*, 419-448 (Eds C. Gussenhoven, T. Rietfield and N. Warner). Mouton de Gruyter.
- Cho, M.-H. (2008). Asymmetries in the perception and production of the English incomplete off-gliding diphthongs by Korean speakers. *Studies in Phonetics, Phonology and Morphology* 14(3): 483-499.
- Cissé, I. A. H. (2009). Comparaison de deux langues en contact, le fulfulde et le bambara, dans une perspective typologique : Structures phonémiques, syllabiques et lexicales. Université Stendhal (Mémoire de Master 2 Recherche).
- Clements, N. (1990). The Role of the Sonority Cycle in Core Syllabification. In *Papers in Laboratory Phonology I*, 283-333 (Eds J. Kingston and M. Beckman). Cambridge: Cambridge University Press.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385-400.
- Davidson, L. (2001). Hidden rankings in the final state of the English grammar. In *Ruling Papers II*, 21-48 (Eds G. Horwood and S. K. Kim). New Brunswick: Rutgers University.
- Davidson, L. (2002). The effects of hidden rankings on the acquisition of consonant clusters. In *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second Language Speech*, 87-96 (Eds A. James and J. Leather). Klagenfurt: University of Klagenfurt.
- Davidson, L. (2003). Articulatory and perceptual influences on the production of non-native consonant clusters. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Eds M. J. Solé, D. Recasens and J. Romero). Barcelona, Spain.
- Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34: 104-137.
- Davidson, L. & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In *Proceedings of WCCFL22* (Eds G. Garding and M. Tsujimura). Somerville, MA: Cascadilla Press.

- Davis, J. M., Elfenbein, J., Schum, R. & Bentler, R. A. (1986). Effects of mild and moderate hearing impairments on language, educational, and psychosocial behavior of children. *Journal of Speech and Hearing Disorders* 51: 53–62.
- Delattre, P. (1963). Le jeu des transitions de formants et la perception des consonnes In *ICPhS IV*, 407-418 La Haye, Mouton.
- Delattre, P., Liberman, A. M. & Cooper, F. S. (1955). Acoustic Loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27: 769-774.
- Demuth, K. & McCullough, E. (2009). The longitudinal development of clusters in French. *Journal of Child Language* 36: 425-448.
- Detey, S., Durant, J. & Nespoulous, J.-L. (2005). Apprentissage de la Phonologie du Français par des Japonais et Multimodalité : une didactique nécessairement cognitive In *1er Colloque International de Didactique Cognitive* Toulouse.
- Diaz, B., Baus, C., Escera, C., Costa, A. & Sebastien-Galles, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America* 105: 16083-16088.
- Dickerson, L. & Dickerson, W. (1977). Interlanguage phonology: Current research and future directions. In *The notions of simplification, interlanguages and pidgins: Actes du 5ème Colloque de Linguistique Appliquée de Neufchatel*, 18-30 (Eds S. P. Corder and E. Roulet).
- Diệp Quang-Ban (2004). *Ngữ pháp tiếng Việt (Tr. La grammaire du vietnamien)*. Hanoi: Nhà xuất bản Giáo dục (Maison d'édition de l'Education).
- Đỗ Thế Dũng & Lê Thanh Thủy (1994). *Le vietnamien sans peine (Méthode d'apprentissage de la langue vietnamienne)*. Chennevières sur Marne: Méthode Assimil.
- Đỗ Thế Dũng, Trần Thiên-Hương & Boulakia, G. (1998). Intonation in Vietnamese. In *Intonation systems : a survey of twenty languages*, 395-416 (Eds D. Hirst and A. Di Cristo). Cambridge: Cambridge University Press.
- Đỗ Trọng Tú & Takara, T. (2004). Precise tone generation for Vietnamese Text-to-Speech system. *Acoustical science and technology* 25(5): 347-353.
- Đoàn Thiện Thuật (1999). *Ngữ âm tiếng Việt (Tr. La phonétique du vietnamien)*. Hanoi: Nhà xuất bản Đại học Quốc gia Hà Nội (Maison d'édition de l'Université Nationale de Hanoi).
- Docherty, G., Hay, J. & Walker, A. (2006). Sociophonetic patterning of phrase-final /t/ in New Zealand English In *Proceedings of the 11th Australian International Conference on Speech Science & Technology*(Ed P. W. C. I. Watson). University of Auckland, New Zealand.
- Dorman, M. F., Studdert-Kennedy, M. & Raphael, L. J. (1977). Stop-consonant recognition:

- Release bursts and formant transition as functionally equivalent, context-dependent cues. *Perception & Psychophysics* 22(2): 109-122.
- Dubesset, V. A. (2005). La Langue française Parlée Complétée (LPC) : Production et Perception Vol. Docteur de l'Institut National Polytechnique de Grenoble, Spécialité : « Sciences Cognitives » GRENOBLE: Institut de la Communication Parlée, UMR CNRS 5009.
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B. & Mével, J.-P. (2001). *Dictionnaire de linguistique*. Paris: Larousse
- Dupoux, E., Pallier, C., Sebastián-Gallés, N. & Mehler, J. (1997). A distressing “deafness” in French? *Journal of Memory and Language* 36: 406-421.
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315-330.
- Eckman, F. R. (2004). Research on Second Language Phonology. *Studies in Second Language Acquisition* 26(4): 513-545.
- Eckman, F. R. (2008). Typological markedness and second language phonology. In *Phonology and second language acquisition*(Eds J. G. H. Edwards and M. L. Zampini). John Benjamins.
- Eckman, F. R. & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9: 234-252.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science* 171(968): 303-306.
- Elfenbein, J. L., Hardin-Jones, M. A. & Davis, J. M. (1994). Oral communication skills of children who are hard of hearing. *Journal of Speech and Hearing Research* 37: 216-226.
- Ellis, R. (1985). *Understanding Second Language Acquisition*. Oxford University Press.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ferlus, M. (1975). Vietnamien et Proto Vietmuong. *Asie du Sud-Est et Monde Insulindien* 6(4): 21-55.
- Ferlus, M. (1977). L'infixé instrumental -rn- en khamou et sa trace en vietnamien. *Cahiers de linguistique - Asie orientale* 2(1): 51-55.
- Ferlus, M. (1982). Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien. *Cahiers de Linguistique Asie Orientale* XI (n°1): 83-106.
- Ferlus, M. (1989). Sur l'origine géographique des langues Viet-muong. *Mon-Khmer Studies* 18-19: 52-59.

-
- Ferlus, M. (1992). Histoire abrégée de l'évolution des consonnes initiales du Vietnamien et du Sino-Vietnamien. *Mon-Khmer Studies* 20: 111-125.
- Ferlus, M. (2001). The origin of tones in Viet-Muong. In *The 11th Annual Meeting of the Southeast Asian Linguistics Society* Bangkok, Institute of Languages and Culture, Mahidol University.
- Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47-65.
- Flege, J. E. (1995). Second Language Speech Learning. Theory, Findings and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-language research.*, 233-277 (Ed W. Strange). Timonium: MD : Yord Press.
- Flege, J. E. (1997a). English vowel production by Dutch talkers: more evidence for the "similar" vs "new" distinction. In *Second-Language Speech*, 11-52 (Eds A. James and J. Leather). Berlin: Mouton de Gruyter.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception In *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, 319-355 (Eds A. Meyer and N. Schiller). Berlin: Mouton de Gruyter.
- Flege, J. E., Bohn, O.-S. & Jang, S. (1997b). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437-470.
- Flege, J. E. & Hillenbrand, J. (1987). Limits on phonetic accuracy in foreign language speech production. In *Interlanguage Phonology: Acquisition of a Second Language Sound System*, 176-203 (Eds G. Loup and S. Weinberger). Cambridge: MA: Newbury House.
- Flege, J. E., Mackay, I. R. A. & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106(5): 2973-2987.
- Flege, J. E., Munro, M. J. & Mackay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16(1): 1-26.
- Flege, J. E., Munro, M. J. & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125-3134.
- Fougeron, C. (2005). Introduction à la Phonologie Articulatoire. In *Phonologie et phonétique (Traité IC2, série Cognition et Traitement de l'Information)*, 265-290 (Eds N. Nguyen, S. Wauquier-Gravelines and J. Durand). Paris: Hermès.
- Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728-3740.

- Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A. & Hallé, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics* 36: 649-663.
- Gadet, F. (1994). La g n se du concept de marque (1926-1931). *Cahiers de l'Institut de linguistique et des Sciences du Langage* 5: 81-92.
- Galantucci, B., Fowler, C. A. & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13(3): 361-377.
- Galazzi, E. (1995). Phon tique/Universit /Enseignement des langues   la fin du XIX  si cle. *Histoire Epist mologie Langage* 17/I: 95-114.
- Gerlach, S. (2004). Another look at effects of environment on L2 epenthesis : Evidence for transfer of ranked constraints. In *Laboratory Approaches To Spanish Phonology*, 143-170 (Ed L. F. Timothy). Mouton de Gruyter.
- Gibbon, F., Hardcastle, W. & Nicolaidis, K. (1993). Temporal and spatial aspects of lingual co-ordination in /kl/ sequences: a cross-linguistic investigation. *Speech and Language* 36: 261-277.
- Gierut, J. A. (1999). Syllable onsets: Clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research* 42: 708-726.
- Giulivi, S., Whalen, D. H., Goldstein, L. M., Nam, H. & Levitt, A. G. (2011). An Articulatory Phonology Account of Preferred Consonant-Vowel Combinations. *Language Learning and Development* 7: 202-225.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R" *Neuropsychologia* 9(3): 317-323.
- Greenberg, J. (1966). *Language universals*. The Hague, The Netherlands: Mouton.
- Greenberg, J. (1978/1984). Some generalizations concerning initial and final consonant clusters. In *Universals of human language*, Vol. 2, 243-279 (Eds J. Greenberg, C. Ferguson and E. Moravcsik). Stanford, CA: Stanford University Press Hawkins.
- Guberina, P. (1974). La parole dans la m thode structo-globale audio-visuelle. *Le Franais dans le monde, Hommage   G. Gouguenheim* 103: 49-54.
- Guberina, P. (1984). Bases th oriques de la m thode SGAV, une linguistique de la parole. In *Aspects d'une politique de diffusion du franais langue  trang re depuis 1945* (Ed D. Coste). Paris: CREDIF/Hatier.
- Gundel, J., Houlihan, K. & Sanders, G. (1986). Markedness distribution in phonology and syntax. In *Markedness*, 107-138 (Eds F. Eckman, E. Moravcsik and J. Wirth). New York: Plenum Press.

- Hallé, P. A., De Boysson-Bardies, B. & Vihman, M. M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34: 299-318.
- Han, M. S. & Kim, K. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2: 223-232.
- Hancin-Bhatt, B. & Bhatt, R. M. (1997). Optimal L2 Syllables: Interactions of Transfer and Developmental Effects. *Studies in Second Language Acquisition* 19: 331-378.
- Hansen, J. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22(3): 338-365.
- Hansen, J. G. E. (2008). Social factors and variation in production in L2 phonology. 251-282 (Eds J. Hansen and M. L. Zampini). Amsterdam, The Netherlands: John Benjamins.
- Harmegnies, B. (1987). SGAV, informatisation et recherche expérimentale. *revue de Phonétique Appliquée* 82, 83, 84: 149-160.
- Harmegnies, B., Delvaux, V., Huet, K. & Piccaluga, M. (2005). Oralité et cognition : pour une approche raisonnée de la pédagogie du traitement de la matière phonique *Revue Parole* 34-35-36: 277-348.
- Henrich, N. (2001). Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception. Université Pierre et Marie Curie - Paris VI.
- Hiligsmann, P., Beheydt, L., Degand, L., Godin, P. & Vanderlinden, S. (2005). *Les études néerlandaises en France et en Belgique francophone* Bruylant.
- Hoàng Giao (1975). Notes sur quelques caractéristiques typologiques de la langue vietnamienne. *Etudes Vietnamiennes* 4(Hanoi): 49-66.
- Hume, E., Johnson, K., Seo, M. & Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, 2069-2072 San Francisco.
- Ingram, D. (1976). *Phonological disability in children*. New York: Elsevier.
- Ingram, J. & Nguyễn Thị-Anh-Thư (2006). Stress, tone and word prosody in Vietnamese compounds In *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 193-198 University of Auckland, New Zealand.
- Ioup, G. (1984). Is there a structural foreign accent ? *Language Learning* 34(2): 1-17.
- Ioup, G. & Weinberger, S. H. (1987). *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge, MA: Newbury House.
- Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals* The Hague:

- Mouton (Translation by R. Keiler of original German version).
- Johansson, F. A. (1973). *Immigrant Swedish phonology*. Lund, Sweden: Gleerup.
- Jusczyk, P. W. (1993). How Word Recognition may Evolve from Infant Speech Perception Capacities. In *Cognitive Models Of Speech Processing : The Second Sperlonga Meeting*, 27-56 (Eds G. Altmann and R. Shillcock). Hove, East Sussex, UK: Erlbaum.
- Kabak, B. (2003). *The Perceptual Processing of Second Language Consonant Clusters*. University of Delaware, USA.
- Kamiyama, T. (2009). Apprentissage phonétique des voyelles du français langue étrangère chez des apprenants japonophones. Université Sorbonne Nouvelle, Paris 3.
- Keating, P. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics* 11 410-406.
- Keating, P., Wright, R. & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* 97: 157-173.
- Kent, R. (1992). The biology of phonological development. In *Phonological development: Models, Research and implications*, 65-90 (Eds C. Ferguson, L. Menn and C. Stoel-Gammon). Timonium: MD: York Press
- Kim, C. K. (2000). *Phonological Processes in English and Korean: Direct Optimality Theory Approach*. University of South Carolina.
- Kirchner, R. (1999). Preliminary thoughts on 'phonologization' within an exemplar-based speech processing system. *UCLA Working Papers in Linguistics - Papers in Phonology* 2 1: 207-231.
- Kishon-Rabin, L., Dayan, M. & Michaeli, O. (2003). Effect of second-formant transitions on the perception of Hebrew voiced stop consonants. *Journal of Basic and Clinical Physiology and Pharmacology* 14(2): 151-164.
- Klein, W. (1989). *L'acquisition de langue étrangère (traduction de l'allemand par Colette Noyau)*. Paris: Armand Colin.
- Kluge, D. C., Rauber, A. S., Reis, M. S. & Bion, R. A. H. (2007). The Relationship between the Perception and Production of English Nasal Coda by Brazilian Learners of English In *Interspeech*, 2297-2300 Antwerp, Belgium.
- Krakow, R. (1999). Physiological organization of syllables: a review. *Journal of Phonetics* 27: 23-54.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50(2): 93-107.

- Kuhl, P. K. & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science* 218(4577): 1138-1141.
- Kühnert, B. & Hoole, P. (2006a). Cohésion temporelle dans les groupes C1/l/ initiaux en français. In *Actes des XXVIe Journées d'Etude sur la Parole*, 545-548 Dinard.
- Kühnert, B., Hoole, P. & Mooshammer, C. (2006b). Gestural overlap and C-center in selected French consonant clusters. In *Proc. 7th International Seminar on Speech Production*, 327-334 (Eds H. C. Yehia, D. Demolin and R. Laboissière). UFMG Belo Horizonte.
- Labov, W. (1969). The study of language in its social context. *Studium Generale* 23: 30-87.
- Labrune, L. (2006). La phonologie du japonais. Paris: Peeters / Société de Linguistique de Paris.
- Ladefoged, P. (2001). *Vowels and consonants: an introduction to the sounds of languages*. Oxford: Blackwell.
- Ladefoged, P. & Maddieson, I. (1996). *The Sounds of World's Languages*. Oxford: Blackwell Publishers Ltd.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan: Press: Ann Arbor.
- Lane, H. (1971). Production et perception de la parole. *Phonetica* 23(2): 94-125.
- Larsen-Freeman, D. & Long, M. (1991). *An Introduction to Second Language Acquisition Research*. New York: Longman.
- Lauret, B. (2007). *Enseigner la prononciation du français : questions et outils*. Paris: Hachette Français Langue Etrangère.
- Lê Thị-Xuyến (1989). Etude contrastive de l'intonation expressive en français et en vietnamien. Paris 3: Université de la Sorbonne Nouvelle
- Lê Việt Bắc (2006). Reconnaissance automatique de la parole pour des langues peu dotées. Grenoble: Université Joseph Fourier.
- Lee, B. (2005). A comparative study of English /oe/ and /e/ produced by early and late Korean and Japanese bilinguals. *Studies in Phonetics, Phonology and Morphology* 11(3): 459-476.
- Lee, M. (2008). Korean consonant cluster reduction: focus on markedness-oriented PREC. *Studies in Phonetics, Phonology and Morphology* 14(3): 427-444.
- Lee, S. (2006). Post Obstruent Tensification in Korean. In *Proceedings of the 2006 annual conference of the Canadian Linguistic Association*.
- Lee, S. (2006). The relationship between perception and production in L2 English stress

- acquisition. *Studies in Phonetics, Phonology and Morphology* 12(3): 661-685.
- Lee, S. (2007). The effect of an experimental method on speech perception. *Studies in Phonetics, Phonology and Morphology* 13(3): 559-583.
- Lee, S. & Cho, M.-H. (2002). Sound replacement in the acquisition of English consonant clusters: a constraint-based approach *Studies in Phonetics, Phonology and Morphology* 8(2): 255-271.
- Léon, P. (1992). *Phonétisme et prononciation du français*. Paris: Nathan.
- Lepetit, D. (1989). Cross-Linguistic Influence in Intonation: French/Japanese and French/English*. *Language Learning* 39(3): 397-413.
- Lepetit, D. (1989). Cross-Linguistic Influence in Intonation: French/Japanese and French/English*. *Language Learning* 39(3): 397-413.
- Levitt, A. G. & Wang, Q. (1991). Evidence for Language-Specific Rhythmic Influences in the Reduplicative Babbling of French-and English-Learning Infants. *Language and Speech* 34 (3): 235-249
- Levy, E. S. & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America* 128(3): 1290-1305.
- Lhote, E. (1988). L'approche paysagiste de l'oral en didactique des langues. In *Sociolinguistique : "Evaluation" et Analyse de discours (Cahiers du français des années quatre-vingt)*, 151-164 Paris: Ecole Normale Supérieure de Fontenay - CREDIF.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74(6): 431-461.
- Liberman, A. M., Delattre, P. C., Cooper, F. S. & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied* 68(8): 1-13.
- Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21: 1-36.
- Lin, Y.-H. (2001). Syllable simplification strategies: A stylistic perspective. *Language Learning* 51: 681-718.
- Lin, Y.-H. (2003). Interphonology variability—Sociolinguistic factors affecting L2 simplification strategies. *Applied Linguistics* 24(4): 439-464.
- Lindblom, B. (1983). Economy of speech gestures. In *The Production of Speech*, 217-246 (Ed P. F. MacNeilage). Berlin: Springer-Verlag.
- Lindblom, B. (1990). Explaining phonetic variation : a sketch of the H& H theory In *Speech*

- production and speech modelling*, 403-439 (Eds W. J. Hardcastle and A. Marchal). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Lisker, L. (1999). Perceiving Final Voiceless Stops without Release: Effects of Preceding Monophthongs versus Nonmonophthongs. *Phonetica* 56(1-2): 44-55.
- Llisterri, J. (1995). Relationships between Speech Production and Speech Perception in a Second Language. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 4, 92-99 Stockholm, Sweden: Stockholm: KTH / Stockholm University.
- MacNeilage, P. F. & Davis, B. (1990). Acquisition of speech production: Frame then content. In *Attention and Performance XIII: Motor Representation and Control* (Ed M. Jeannerod). Hillsdale, NJ: Lawrence Erlbaum.
- MacNeilage, P. F. & Davis, B. (2000). Deriving speech from nonspeech: A view from Ontogeny. *Phonetica* 57(2-4): 284-296.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Maddieson, I. (1986). The Size and Structure of Phonological Inventories: Analysis of UPSID. In *Experimental Phonology*, 105-123 (Ed J. J. Ohala). New York: Academic Press.
- Maddieson, I. (1993). The structure of segment sequences. *UCLA Working Papers in Phonetics* 83: 1-8.
- Magen, C., Billières, M. & Gaillard, P. (2005). Surdit  phonologique et cat gorisation. Perception des voyelles franaises par les hispanophones. *Revue Parole* 33: 9-34.
- Mai Ngọc-Ch , Vĩ Đức-Nghi u & Ho ng Tr ng-Phi n (1997). *Cơ sở ngôn ngữ học và tiếng Việt (Tr. Les bases linguistiques et la langue vietnamienne)*. Hanoi: Nhà xuất bản Đại học và Giáo dục chuyên nghiệp (Edition de l'enseignement sup rieur et de l' ducation professionnelle).
- Major, R. C. (1987). Phonological Similarity, Markedness, and Rate of L2 Acquisition. *Studies in Second Language Acquisition* 9 63-82.
- Major, R. C. & Faudree, M. C. (1996). Markedness Universals and the Acquisition of Voicing Contrasts by Korean Speakers of English. *Studies in Second Language Acquisition* 18(1): 69-90.
- Mal cot, A. (1974). Frequency of Occurrence of French Phonemes and Consonant Clusters. *Phonetica* 29(3): 158-170.
- Mal cot, A. (1975). The Glottal Stop in French. *Phonetica* 31(1): 51-63.
- Marin, S. & Pouplier, M. (2008). Organization of Complex Onsets and Codas in American English: Evidence for a Competitive Coupling Model In *8th International Seminar on Speech Production*, 437-440.

- Marin, S., Pouplier, M. & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research. *The Journal of the Acoustical Society of America* 127(1): 445-461.
- Martinet, A. (1960). *Eléments de linguistique générale*. Paris.
- Maupeu, M. (2005). Traitement de données lexicales pour l'analyse de structures syllabiques des langues du monde. Grenoble: Département Statistiques et Traitement Informatique des Données, IUT 2 et Institut de la Communication Parlée (Rapport de stage).
- McLeod, S., Doorn, J. V. & Reed, V. A. (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology* 10(2): 99-110.
- Mehler, J. & Dupoux, E. (1992). *Naître humain*. Paris: Odile Jacob.
- Ménard, L., Chrétien, J., Lachapelle, R. & Marleau, I. (2010). Corrélat acoustiques de la perception des voyelles produites par des locuteurs sourds. *Spectrum (Revue de la langue française portant sur la communication humaine et ses troubles* 2: 19-31.
- Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S. & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America* 111: 1892-1905.
- Menyuk, P. (1977). *Language and maturation*. Cambridge: MIT Press
- Michaud, A. (2004). Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica* 61(2-3): 119-146.
- Michaud, A. (2005). Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés. Paris: Université Paris 3 - Sorbonne Nouvelle.
- Michaud, A. (2010). Les systèmes de tons en Asie orientale : typologie, schémas évolutifs et modélisation. *Faits de Langues* n°35-36: (consultation sur <http://halshs.archives-ouvertes.fr/halshs-00436458/en/>).
- Michaud, A., Vũ Ngọc-Tuấn, Amelot, A. & Roubeau, B. (2006). Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. *Mon-Khmer Studies* 36 121-137.
- Michaud, A., Vũ Ngọc-Tuấn, Amelot, A. & Roubeau, B. (2006). Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. *Mon-Khmer Studies* 36 121-137.
- Moffit, A. R. (1971). Consonant Cue Perception by Twenty- to Twenty-Four-Week-Old Infants. *Child Development* 42(3): 717-731.
- Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology* 14(3): 477-492

- Moulton, W. G. (1962). *The Sounds of English and German*. Chicago: University of Chicago Press.
- Munro, M. J. & Bohn, O.-S. (2007). The study of second language speech. A brief review. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, 3-12 (Eds O.-S. Bohn and M. J. Munro). Amsterdam: NLD: John Benjamins Publishing Company.
- Nearey, T. M. & Rochet, B. L. (1994). *Effects of place of articulation and vowel context on VOT production and perception for French and English stops*. Cambridge, ROYAUME-UNI: Cambridge University Press.
- Nguyễn Khắc-Khâm (1960). Originalité de la culture vietnamienne. Archive de l'Organisation des Nations Unies pour l'Education, la Science et la Culture.
- Nguyen, N. (2005). La perception de la parole. In *Phonologie et phonétique : Forme et substance*, 425-447 (Eds N. Nguyen, S. Wauquier-Gravelines and J. Durand). Paris: Hermès.
- Nguyen, N., Wauquier, S., Lancia, L. & Tuller, B. (2007). Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In *Segmental and Prosodic Issues in Romance Phonology*, 3-23 (Ed J. M. a. M.-J. S. Pilar Prieto). John Benjamins, Amsterdam.
- Nguyễn Ngân (2008). Interlanguage Phonology and the Pronunciation of English Final Consonant Clusters by Native Speakers of Vietnamese. *Ohio University Working Papers in Linguistics and Language Teaching* 16.
- Nguyễn Phú-Phong (1989). Le vietnamien, un cas de romanisation inachevée *Cahiers d'études vietnamiennes* 10(Université de Paris VII): 25-32.
- Nguyễn Phú-Phong (2005). *Việt Nam - Chữ viết, Ngôn ngữ và Xã hội (Vietnam - Ecriture, Langage et Société)*. Hochiminh ville: Đại học sư phạm TP Hồ Chí Minh (Université de Pédagogie, HCM ville).
- Nguyễn Quốc-Cường (2002). Reconnaissance de la parole en langue vietnamienne. Institut National Polytechnique de Grenoble.
- Nguyễn Tài-Cần (1995). *Giáo trình lịch sử ngữ âm tiếng Việt (sơ thảo) (Tr. Cours de phonétique historique de la langue vietnamienne (version esquisse))*. Hanoi: Nhà xuất bản Giáo dục (Maison d'édition de l'Education).
- Nguyễn Thị-Anh-Thư & Ingram, J. (2007). Acoustic and perceptual cues for compound-phrasal contrasts in Vietnamese. *Journal of the Acoustical Society of America* 122(3): 1746-1757.
- Nguyễn Thị-Bình-Minh (2000). Regards sur l'enseignement de la phonétique dans la formation des étudiants en F.L.E. à l'Université Pédagogique de Ho Chi Minh ville.

- Université de Rouen.
- Nguyễn Thiện-Giáp (1998). *Từ vựng học tiếng Việt (Tr. Lexicologie du vietnamien)*. Hanoi: Nhà xuất bản Giáo dục (Maison d'édition de l'Education).
- Nguyễn Thị-Ngân-Hà (2001). Interférences phonologiques de la langue maternelle dans l'apprentissage du français, Etude du cas des étudiants de l'école des sciences de Hue et proposition pédagogiques. Université de Rouen.
- Nguyễn Thị-Thanh-Hoa (2004). Contribution à l'étude de la prosodie du vietnamien (Variations de l'intonation dans les modalités assertives, interrogatives et impératives). Paris 7: Université Denis Diderot.
- Nguyễn Tiến-Dũng, Mixdorff, H., Lương Chi-Mai, Ngô Hoàng-Huy & Vũ Kim-Bằng (2004). Fujisaki Model based F0 contours in Vietnamese TTS. In *8th International Conference on Spoken Language Processing*, 1429-1432 Korea.
- Nguyễn Văn-Lợi & Edmondson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: instrumental case studies. *The Mon-Khmer Studies Journal* 28: 1-18.
- Nguyễn Việt-Son (2009). Etude de caractéristiques de la langue vietnamienne en vue de sa synthèse et de sa reconnaissance automatique. Aspects statiques et dynamiques.: Telecom-ParisTech et Institut Polytechnique de Hanoi.
- Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115: 39-57.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, MA: Cambridge University Press.
- Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders* 32: 397-422.
- Osburne, A. G. (1996). Final cluster reduction in English L2 speech: A case study of a Vietnamese speaker. *Applied Linguistics* 17 (2): 164-181.
- Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32: 258-278.
- Pennington, M. C. & Ku, P. (1992). Realization of English Final Stops by Chinese Speakers *Hong Kong Journals Online* 4.
- Peperkamp, S. (2002). La perception de l'accent tonique chez des bilingues français-espagnol. *Annales de la Fondation Fyssen* 16: 91-102.
- Petinou, K. C., Schwartz, R. G., Gravel, J. S. & Raphael, L. J. (2001). A preliminary account of phonological and morphophonological perception in young children with and without otitis media. *International Journal of Language and Communication Disorders* 36: 21-42.

- Phạm, A. H. (2003). *Vietnamese tone: a new analysis*. New York: Routledge.
- Pickett, J. M., Bunnell, H. T. & Revoile, S. G. (1995). Phonetics of Intervocalic Consonant Perception: Retrospect and Prospect. *Phonetica* 52(1): 1-40.
- Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle linguistique de Prague* 4.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52: 37-52.
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S. & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America* 106: 10598-10602.
- Pulvermüller, F. & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience* 11 (5): 351-360.
- Redard, F. (1973). *Contrastive phonemics: Notes and exercices*. Neuchatel, Switzerland: Centre de Linguistique.
- Redford, M. A. (1999). *An Articulatory Basis for the Syllable*. Austin: The University of Texas.
- Renard, R. (1989). *La méthode verbo-tonale de correction phonétique*. Paris: Didier-Erudition.
- Riney, T. J. & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20(2): 213-243.
- Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In *Speech perception and linguistic experience: Issues in cross-language research*, 379-410 (Ed W. Strange). Baltimore: York Press.
- Rochet-Capellan, A. & Schwartz, J.-L. (2005). The Labial-Coronal effect and CVCV stability during reiterant speech production: an articulatory analysis. In *Proceedings of the 9th International Conference on Speech Communication and Technology*, 1013-1016 Lisbonne.
- Rosch, E., Mervis, C., Gray, W., Johnson, D. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* 8: 382-349.
- Rossato, S. (2000). *Du son au geste, inversion de la parole, le cas des voyelles nasales*. Grenoble: Institut National Polytechnique de Grenoble (INPG).
- Rousset, I. (2004). *Structures syllabiques et lexicales des langues du monde. Données, typologiques, tendances universelles et contraintes substantielles*. Grenoble: Université Stendhal.
- Rvachew, S., Rafaat, S. & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology* 8:

33-43.

- Ryalls, J., Simon, M. & Thomason, J. (2004). Voice Onset Time production in older Caucasian- and African-Americans. *Journal of Multilingual Communication Disorders* 2(1): 61 - 67.
- Saltzman, E. L. & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Haskins Laboratories Status Report on Speech Research SR-99 /100*: 38-68.
- Sato, C. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning* 34: 43-57.
- Sato, C. (1985). Task variation in interlanguage phonology. In *Input in Second Language Acquisition* (Eds S. M. Gass and C. G. Madden). Cambridge, Mass: Newbury House.
- Schmidt, R. (1977). Sociolinguistic variation and language transfer in phonology. *Working Papers in Bilingualism* 12: 79-95.
- Schwab, S. & Grosjean, F. (2004). La perception du débit en langue seconde. *Phonetica* 61(2-3): 84-94.
- Schwartz, J.-L., Abry, C., Boë, L.-J. & Cathiard, M. (2002). Phonology in a Theory of Perception-for-Action-Control. In *Phonology, Phonetics and Cognition*, 254-280 (Eds J. Durand and B. Laks). Oxford: Oxford University Press.
- Schwartz, J.-L., Ménard, L., Basirat, A. & Sato, M. (2010). The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*: 1-19.
- Schwartz, J.-L., Sato, M. & Fadiga, L. (2011). Le langage commun de la perception et de l'action dans la communication parlée: une perspective neurocognitive. *Faits de Langues* 37: 117-136.
- Sebastián-Gallés, N. (2010 (in press)). Bilingual language acquisition: Where does the difference lie? *Human Developmental*.
- Sebastián-Gallés, N. & Baus, C. (2005). On the Relationship between Perception and Production in L2 Categories. In *Twenty-First Century Psycholinguistics: Four Cornerstones*, 279-292 (Ed A. Cutler). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Sebastián-Gallés, N., Rodríguez-Fornells, A., Diego-Balaguer, R. d. & Díaz, B. (2006). First and second language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience* 18: 1277-1291.
- Selkirk, E. (1984). On the major class features and syllable theory. In *Language Sound Structure*, 107-136 (Eds M. Aronoff and R. T. Oehrle). Cambridge, MA: MIT Press.

- Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3(3): 243-261.
- Sherwin, S. (1999). The Sonority Sequencing Principle in Interlanguage Phonology. In *Working Papers in Linguistics* Vol. 6, 55-74 Fairfax, Virginia: George Mason University.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L. & Wilson, D. L. (1997). The speech disorders classification system (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research* 40: 723-740.
- Silverman, D. (1995). Phasing and Recoverability. UCLA Ph.D dissertation.
- Silverman, D. & Jun, J. (1994). Aerodynamic evidence for articulatory overlap in Korean. *Phonetica* 51: 210-220.
- Smith, N. V. (1973). *The acquisition of phonology : a case study*. London: C.U.P.
- Snoeren, N. D., Halle, P. A. & Segui, J. (2006). A voice for the voiceless : Production and perception of assimilated stops in French. *Journal of Phonetics* 34: 241-268.
- Sperbeck, M. & Strange, W. (2008). An epenthetic vowel between consonantal sequences in perception and production by Japanese. *The Journal of the Acoustical Society of America* 124(4): 2595-2595.
- Sperbeck, M. & Strange, W. (2010). The perception of complex onsets in English: universal markedness? In *Proceedings of the 33rd Annual Penn Linguistics Colloquium*, Vol. Vol. 16.
- Sperbeck, M. & Strange, W. (2010). The production and perception of English consonant sequences by Japanese-speaking learners of English. *The Journal of the Acoustical Society of America* 127(3): 1955-1955.
- Sperbeck, M., Strange, W. & Ito, K. (2005). Training Japanese L2 learners to perceive difficult American vowel contrasts. *The Journal of the Acoustical Society of America* 117(4): 2400-2400.
- Stelmachowicz, P. G. (2001). The Importance of High-Frequency Amplification for Young Children. In *Proceedings from the 2nd International Conference: "A Sound Foundation Through Early Amplification"*, Vol. Chapter 13, 167-175
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.
- Stevens, M. & Hajek, J. (2006). Blocking of word-boundary consonant lengthening in Siense Italian: some auditory and acoustic evidence In *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 176-181 University of Auckland, New Zealand.
- Stockman, I. J. & Pluut, E. (1992). Segment Composition as a Factor in the Syllabification

- Errors of Second-Language Speakers. *Language Learning* 42: 21–45.
- Stockwell, Bowen & Martin (1965). *The Grammatical Structures of English and Spanish*. University of Chicago.
- Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: a longitudinal study. *Journal of Speech and Hearing Research* 28: 505-512.
- Strange, W. (1995). Phonetics of second language acquisition: past, present and future. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 4, 84-91 Stockholm, Sweden.
- Sundara, M. & Polka, L. (2008). Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction. *Cognition* 106(1): 234-258.
- Sundara, M., Polka, L. & Molnar, M. (2008). Development of coronal stops perception: Bilingual infants keep pace with their monolingual peers. *Cognition* 108: 232-242.
- Sung, E.-k. (2006). L2 sound perception and production by Korean adults and children. *Studies in Phonetics, Phonology and Morphology* 12(3): 577-596.
- Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning* 26: 233–253.
- Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics in Language Teaching* 18: 139-152.
- Tarone, E. E. (1979). Interlanguage as chameleon. *Language Learning* 29: 181-191.
- Tarone, E. E. (1982). Systematicity and attention in interlanguage. *Language Learning* 12: 69-84.
- Tarone, E. E. (1983). On the variability of interlanguage system. *Applied Linguistics* 4: 142-163.
- Tarone, E. E. (1985). Variability in Interlanguage Use : A Study of Style-Shifting in Morphology and Syntax. *Language Learning* 35(3): 373-403.
- Tarone, E. E. (1989). On Chameleons and Monitors. In *The Dynamic Interlanguage. Empirical Studies in Second Language Variation*(Ed M. R. Eisenstein). New York: Plenum Press.
- Teodor, P. (2001). Acquisition of marked consonant clusters in German as a foreign language. *Poznań Studies in Contemporary Linguistics* 37: 157-186.
- Terrell, T. D. (1989). Teaching Spanish Pronunciation in a Communicative Approach. In *American Spanish Pronunciation: Theoretical and Applied Perspectives*, 196-214 (Eds B. P. C. and R. M. Hammond). Washington DC: Georgetown University Press.
- Trần Đỗ-Đạt, Castelli, E., Serignat, J.-F., Lê Xuân-Hùng & Trịnh Văn-Loan (2005). Influence of F0 on Vietnamese syllable perception. In *Interspeech*, 1697-1700 Lisbon.

- Trần Thị-Thúy-Hiền & Vallée, N. (2009). An acoustic study of interword consonant sequences in vietnamese. *Journal of Southeast Asian Linguistics* 1: 231–249.
- Trehub, S. E. (1973). Infant's sensitivity to vowel and tonal contrasts. *Developmental Psychology* 9: 91-96.
- Tropf, H. (1987). Sonority as a variable factor in second language phonology. In *Sound Patterns in Second Language Acquisition*, 173-191 (Eds A. James and J. Leather). Dordrecht: Foris.
- Troubetzkoy, N. S. (1939/2005). *Principes de phonologie (Grundzüge der Phonologie)*. Klincksieck.
- Trương Văn-Chình (1970). *Structure de la langue vietnamienne*. Paris: Librairie Orientaliste Paul Geuthner.
- Tsao, F.-M., Liu, H.-M. & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Development* 75(4): 1067 – 1084.
- Tsukada, K. (2004). Cross-language perception of final stops in Thai and English: A comparison of native and non-native listeners. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 563-568 Macquarie University, Sydney.
- Tsukada, K., Bialystok, E., Mack, M., Sung, H., Birdsong, D. & Flege, J. (2002). Phonetic Realization of Word-Final Stops in English by Korean Adults and Children: A Developmental Study. In *LabPhon 8* New Haven, CT, USA.
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H. & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33: 263–290.
- Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E. & Flege, J. (2004). Release Bursts in English Word-Final Voiceless Stops Produced by Native English and Korean Adults and Children. *Phonetica* 61(2-3): 67-83.
- Vaissière, J. (2007). *La phonétique (Collection "Que sais-je ?")*. Paris: P.U.F.
- Vallée, N., Boë, L.-J., Maddieson, I. & Rousset, I. (2000). Des lexiques aux syllabes des langues du monde. Typologies et structures In *XXIIIèmes Journées d'Étude sur la Parole* Aussois, France.
- Vallée, N., Rossato, S. & Rousset, I. (2009). Favoured syllabic patterns in the world's languages and sensorimotor constraints. In *Approaches to Phonological Complexity* (Eds F. Pellegrino, E. Marsico, I. Chitoran and C. Coupé). Berlin: Mouton de Gruyter.
- Van Dommelen, W. A. (1983). Parameter Interaction in the Perception of French Plosives.

- Phonetica* 40(1): 32-62.
- Van Dommelen, W. A. (1985). Further Experiments in the Perception of French Plosives. *Phonetica* 42(2-3): 100-115.
- Velleman, S. L. & Vihman, M. m. (2006). Phonological development in infancy and early childhood: Implications for theories of language learning. In *Phonology in Context*, 25-50 (Ed M. C. Pennington). Luton: Macmillan.
- Vihman, M. (1996). *Phonological Development : The Origins of Language in the Child*. Cambridge: Wiley-Blackwell.
- Vũ Minh-Quang (2007). Exploitation de la prosodie pour la segmentation et l'analyse automatique de signaux de parole : Institut National Polytechnique de Grenoble et Institut Polytechnique de Hanoi.
- Vũ Ngọc-Tuấn, d'Alessandro, C. & Michaud, A. (2005). Using open quotient for the characterisation of Vietnamese glottalised tones. In *Eurospeech-Interspeech : 9th European Conference on Speech Communication and Technology* Lisbonne.
- Vũ Ngọc-Tuấn, D'Alessandro, C. & Rosset, S. (2002). A phonetic study of vietnamese tones : acoustic and electroglottographic measurements. In *7th International Conference on Spoken Language Processing* Denver, Colorado, USA.
- Vương Lộc & et al. (2006). *Tiếng Việt (Tr. Le vietnamien)* Hanoi: (Document consulté en interne du Département de Sciences du Langage, Université Nationale de Hanoi)
- Weinberger, S. (1994). Functional and Phonetic Constraints on Second Language Phonology. In *First and Second Language Phonology*(Ed M. S. Yavas). San Diego, CA: Singular Publishing Group Inc.
- Weinberger, S. H. (1987). The influence of linguistic context on syllable structure simplification. In *Interlanguage phonology: The acquisition of a second language sound system*, 401-417 (Eds G. Ioup and S. H. Weinberger). Rowley, MA: Newbury House.
- Weinreich, U. (1953 / 1968). *Languages in contact*. The Hague, Mouton.
- Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49-63.
- Wiltshire, C. R. (2006). Word-final Consonant and Cluster Acquisition in Indian English(es). In *30th Boston University Conference on Language Development*.
- Wright, R. (1996). Consonant clusters and cue preservation in Tsou. *Dissertations in Linguistics (UCLA Ph.D dissertation)* 20.
- Yao, Y. (2009). Understanding VOT Variation in Spontaneous Speech. In *Proceedings of the 18th International Congress of Linguists (CIL XVIII)* Seoul, South Korea: Korea University.

- Yeon, S.-H. (2004). Teaching English Word-Final Alveolopalatals to native speakers of Korean. The Graduate School of the University of Florida, USA.
- Yoo, H. (2004). A longitudinal study of consonant cluster acquisition. *Studies in Phonetics, Phonology and Morphology* 10(3): 481-503.
- Zydorowicz, P. (2010). Consonant Clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. *Poznań Studies in Contemporary Linguistics* 46(4): 565–588

Annexes

Date 23 octobre 2007, Corpus lu du Vietnamien, Variété du Nord, Locuteur (prénom, nom, sexe masculin, âge), Lieu de l'enregistrement (Chambre Sourde ICP Campus), Matériel (Enregistreur numérique Marantz PMD 671, Micro AKG C 1000 S).

Entraînement

- a) Bạn sẽ gặp từ tia **tiếp tục** xuất hiện trong bài khoá
- b) Bạn sẽ gặp từ khoan **tiếp tục** xuất hiện trong bài khoá
- c) Bạn sẽ gặp từ xin **liên tiếp** xuất hiện trong bài khoá
- d) Bạn sẽ gặp từ tha **liên tiếp** xuất hiện trong bài khoá

Corpus lu

1. Bạn sẽ gặp từ cá tiếp tục xuất hiện trong bài khoá
2. Bạn sẽ gặp từ tán tiếp tục xuất hiện trong bài khoá
3. Bạn sẽ gặp từ má tiếp tục xuất hiện trong bài khoá
4. Bạn sẽ gặp từ tám tiếp tục xuất hiện trong bài khoá
5. Bạn sẽ gặp từ ná tiếp tục xuất hiện trong bài khoá
6. Bạn sẽ gặp từ cán tiếp tục xuất hiện trong bài khoá
7. Bạn sẽ gặp từ tá tiếp tục xuất hiện trong bài khoá
8. Bạn sẽ gặp từ ngang tiếp tục xuất hiện trong bài khoá
9. Bạn sẽ gặp từ cá tiếp tục xuất hiện trong bài khoá
10. Bạn sẽ gặp từ tám tiếp tục xuất hiện trong bài khoá

11. Bạn sẽ gặp từ má tiếp tục xuất hiện trong bài khoá
12. Bạn sẽ gặp từ ná tiếp tục xuất hiện trong bài khoá
13. Bạn sẽ gặp từ cán tiếp tục xuất hiện trong bài khoá
14. Bạn sẽ gặp từ tán tiếp tục xuất hiện trong bài khoá
15. Bạn sẽ gặp từ tám tiếp tục xuất hiện trong bài khoá
16. Bạn sẽ gặp từ tá tiếp tục xuất hiện trong bài khoá
17. Bạn sẽ gặp từ má tiếp tục xuất hiện trong bài khoá
18. Bạn sẽ gặp từ ná tiếp tục xuất hiện trong bài khoá
19. Bạn sẽ gặp từ tán tiếp tục xuất hiện trong bài khoá
20. Bạn sẽ gặp từ ngang tiếp tục xuất hiện trong bài khoá
21. Bạn sẽ gặp từ cá tiếp tục xuất hiện trong bài khoá
22. Bạn sẽ gặp từ cán tiếp tục xuất hiện trong bài khoá
23. Bạn sẽ gặp từ tá tiếp tục xuất hiện trong bài khoá
24. Bạn sẽ gặp từ tám tiếp tục xuất hiện trong bài khoá
25. Bạn sẽ gặp từ má tiếp tục xuất hiện trong bài khoá
26. Bạn sẽ gặp từ ná tiếp tục xuất hiện trong bài khoá
27. Bạn sẽ gặp từ tán tiếp tục xuất hiện trong bài khoá
28. Bạn sẽ gặp từ ngang tiếp tục xuất hiện trong bài khoá
29. Bạn sẽ gặp từ tá tiếp tục xuất hiện trong bài khoá
30. Bạn sẽ gặp từ cán tiếp tục xuất hiện trong bài khoá
31. Bạn sẽ gặp từ cá tiếp tục xuất hiện trong bài khoá

32. Bạn sẽ gặp từ ngang tiếp tục xuất hiện trong bài khoá
33. Bạn sẽ gặp từ bát ngát liên tiếp xuất hiện trong bài khoá
34. Bạn sẽ gặp từ ác liệt liên tiếp xuất hiện trong bài khoá
35. Bạn sẽ gặp từ sáng tác liên tiếp xuất hiện trong bài khoá
36. Bạn sẽ gặp từ cát liên tiếp xuất hiện trong bài khoá
37. Bạn sẽ gặp từ kháng cự liên tiếp xuất hiện trong bài khoá
38. Bạn sẽ gặp từ áp lực liên tiếp xuất hiện trong bài khoá
39. Bạn sẽ gặp từ pháp lệnh liên tiếp xuất hiện trong bài khoá
40. Bạn sẽ gặp từ đáng sợ liên tiếp xuất hiện trong bài khoá
41. Bạn sẽ gặp từ tác dụng liên tiếp xuất hiện trong bài khoá
42. Bạn sẽ gặp từ khát nước liên tiếp xuất hiện trong bài khoá
43. Bạn sẽ gặp từ đáng tiếc liên tiếp xuất hiện trong bài khoá
44. Bạn sẽ gặp từ sáng tạo liên tiếp xuất hiện trong bài khoá
45. Bạn sẽ gặp từ bán kết liên tiếp xuất hiện trong bài khoá
46. Bạn sẽ gặp từ khoác lác liên tiếp xuất hiện trong bài khoá
47. Bạn sẽ gặp từ nát liên tiếp xuất hiện trong bài khoá
48. Bạn sẽ gặp từ sáng kiến liên tiếp xuất hiện trong bài khoá
49. Bạn sẽ gặp từ lác đác liên tiếp xuất hiện trong bài khoá
50. Bạn sẽ gặp từ các liên tiếp xuất hiện trong bài khoá
51. Bạn sẽ gặp từ tát liên tiếp xuất hiện trong bài khoá
52. Bạn sẽ gặp từ đám cưới liên tiếp xuất hiện trong bài khoá

53. Bạn sẽ gặp từ mát liên tiếp xuất hiện trong bài khoá
54. Bạn sẽ gặp từ tháng chạp liên tiếp xuất hiện trong bài khoá
55. Bạn sẽ gặp từ sáng tác liên tiếp xuất hiện trong bài khoá
56. Bạn sẽ gặp từ đáng sợ liên tiếp xuất hiện trong bài khoá
57. Bạn sẽ gặp từ bát ngát liên tiếp xuất hiện trong bài khoá
58. Bạn sẽ gặp từ pháp luật liên tiếp xuất hiện trong bài khoá
59. Bạn sẽ gặp từ nát liên tiếp xuất hiện trong bài khoá
60. Bạn sẽ gặp từ thán phục liên tiếp xuất hiện trong bài khoá
61. Bạn sẽ gặp từ phát giác liên tiếp xuất hiện trong bài khoá
62. Bạn sẽ gặp từ áp lực liên tiếp xuất hiện trong bài khoá
63. Bạn sẽ gặp từ pháp lệnh liên tiếp xuất hiện trong bài khoá
64. Bạn sẽ gặp từ đáng kính liên tiếp xuất hiện trong bài khoá
65. Bạn sẽ gặp từ tát liên tiếp xuất hiện trong bài khoá
66. Bạn sẽ gặp từ sáng kiến liên tiếp xuất hiện trong bài khoá
67. Bạn sẽ gặp từ các liên tiếp xuất hiện trong bài khoá
68. Bạn sẽ gặp từ khát nước liên tiếp xuất hiện trong bài khoá
69. Bạn sẽ gặp từ kháng cự liên tiếp xuất hiện trong bài khoá
70. Bạn sẽ gặp từ lác đác liên tiếp xuất hiện trong bài khoá
71. Bạn sẽ gặp từ áp dụng liên tiếp xuất hiện trong bài khoá
72. Bạn sẽ gặp từ đám cưới liên tiếp xuất hiện trong bài khoá
73. Bạn sẽ gặp từ cát liên tiếp xuất hiện trong bài khoá

74. Bạn sẽ gặp từ gián tiếp liên tiếp xuất hiện trong bài khoá
75. Bạn sẽ gặp từ bán kết liên tiếp xuất hiện trong bài khoá
76. Bạn sẽ gặp từ bát ngát liên tiếp xuất hiện trong bài khoá
77. Bạn sẽ gặp từ thán phục liên tiếp xuất hiện trong bài khoá
78. Bạn sẽ gặp từ khát vọng liên tiếp xuất hiện trong bài khoá
79. Bạn sẽ gặp từ thảng chập liên tiếp xuất hiện trong bài khoá
80. Bạn sẽ gặp từ khoác lác liên tiếp xuất hiện trong bài khoá
81. Bạn sẽ gặp từ pháp luật liên tiếp xuất hiện trong bài khoá
82. Bạn sẽ gặp từ phát giác liên tiếp xuất hiện trong bài khoá
83. Bạn sẽ gặp từ khát vọng liên tiếp xuất hiện trong bài khoá
84. Bạn sẽ gặp từ ác liệt liên tiếp xuất hiện trong bài khoá
85. Bạn sẽ gặp từ tác liên tiếp xuất hiện trong bài khoá
86. Bạn sẽ gặp từ tác dụng liên tiếp xuất hiện trong bài khoá
87. Bạn sẽ gặp từ phát giác liên tiếp xuất hiện trong bài khoá
88. Bạn sẽ gặp từ đáng sợ liên tiếp xuất hiện trong bài khoá
89. Bạn sẽ gặp từ bán kết liên tiếp xuất hiện trong bài khoá
90. Bạn sẽ gặp từ tác liên tiếp xuất hiện trong bài khoá
91. Bạn sẽ gặp từ khoác lác liên tiếp xuất hiện trong bài khoá
92. Bạn sẽ gặp từ sáng tác liên tiếp xuất hiện trong bài khoá
93. Bạn sẽ gặp từ mát liên tiếp xuất hiện trong bài khoá
94. Bạn sẽ gặp từ mác liên tiếp xuất hiện trong bài khoá

95. Bạn sẽ gặp từ sáng tạo liên tiếp xuất hiện trong bài khoá
96. Bạn sẽ gặp từ thán phục liên tiếp xuất hiện trong bài khoá
97. Bạn sẽ gặp từ khát nước liên tiếp xuất hiện trong bài khoá
98. Bạn sẽ gặp từ nát liên tiếp xuất hiện trong bài khoá
99. Bạn sẽ gặp từ ác liệt liên tiếp xuất hiện trong bài khoá
100. Bạn sẽ gặp từ đáng tiếc liên tiếp xuất hiện trong bài khoá
101. Bạn sẽ gặp từ áp dụng liên tiếp xuất hiện trong bài khoá
102. Bạn sẽ gặp từ mác liên tiếp xuất hiện trong bài khoá
103. Bạn sẽ gặp từ đáng kính liên tiếp xuất hiện trong bài khoá
104. Bạn sẽ gặp từ các liên tiếp xuất hiện trong bài khoá
105. Bạn sẽ gặp từ đám cưới liên tiếp xuất hiện trong bài khoá
106. Bạn sẽ gặp từ phát giác liên tiếp xuất hiện trong bài khoá
107. Bạn sẽ gặp từ tác dụng liên tiếp xuất hiện trong bài khoá
108. Bạn sẽ gặp từ pháp lệnh liên tiếp xuất hiện trong bài khoá
109. Bạn sẽ gặp từ áp lực liên tiếp xuất hiện trong bài khoá
110. Bạn sẽ gặp từ lác đác liên tiếp xuất hiện trong bài khoá
111. Bạn sẽ gặp từ pháp luật liên tiếp xuất hiện trong bài khoá
112. Bạn sẽ gặp từ tháng tám liên tiếp xuất hiện trong bài khoá
113. Bạn sẽ gặp từ tác liên tiếp xuất hiện trong bài khoá
114. Bạn sẽ gặp từ mát liên tiếp xuất hiện trong bài khoá
115. Bạn sẽ gặp từ áp dụng liên tiếp xuất hiện trong bài khoá

116. Bạn sẽ gặp từ sáng kiến liên tiếp xuất hiện trong bài khoá
117. Bạn sẽ gặp từ tháng chạp liên tiếp xuất hiện trong bài khoá
118. Bạn sẽ gặp từ tát liên tiếp xuất hiện trong bài khoá
119. Bạn sẽ gặp từ đáng kính liên tiếp xuất hiện trong bài khoá
120. Bạn sẽ gặp từ sáng tạo liên tiếp xuất hiện trong bài khoá
121. Bạn sẽ gặp từ thán phục liên tiếp xuất hiện trong bài khoá
122. Bạn sẽ gặp từ gián tiếp liên tiếp xuất hiện trong bài khoá
123. Bạn sẽ gặp từ bán kết liên tiếp xuất hiện trong bài khoá
124. Bạn sẽ gặp từ mác liên tiếp xuất hiện trong bài khoá
125. Bạn sẽ gặp từ mát liên tiếp xuất hiện trong bài khoá
126. Bạn sẽ gặp từ cát liên tiếp xuất hiện trong bài khoá
127. Bạn sẽ gặp từ đáng tiếc liên tiếp xuất hiện trong bài khoá
128. Bạn sẽ gặp từ khát vọng liên tiếp xuất hiện trong bài khoá
129. Bạn sẽ gặp từ sáng tác liên tiếp xuất hiện trong bài khoá
130. Bạn sẽ gặp từ bát ngát liên tiếp xuất hiện trong bài khoá
131. Bạn sẽ gặp từ tát liên tiếp xuất hiện trong bài khoá
132. Bạn sẽ gặp từ kháng cự liên tiếp xuất hiện trong bài khoá
133. Bạn sẽ gặp từ gián tiếp liên tiếp xuất hiện trong bài khoá
134. Bạn sẽ gặp từ tác dụng liên tiếp xuất hiện trong bài khoá
135. Bạn sẽ gặp từ tác liên tiếp xuất hiện trong bài khoá
136. Bạn sẽ gặp từ pháp lệnh liên tiếp xuất hiện trong bài khoá

137. Bạn sẽ gặp từ thánng chạp liên tiếp xuất hiện trong bài khoá
138. Bạn sẽ gặp từ đám cưới liên tiếp xuất hiện trong bài khoá
139. Bạn sẽ gặp từ ác liệt liên tiếp xuất hiện trong bài khoá
140. Bạn sẽ gặp từ pháp luật liên tiếp xuất hiện trong bài khoá
141. Bạn sẽ gặp từ sáng kiến liên tiếp xuất hiện trong bài khoá
142. Bạn sẽ gặp từ áp lực liên tiếp xuất hiện trong bài khoá
143. Bạn sẽ gặp từ thánng tám liên tiếp xuất hiện trong bài khoá
144. Bạn sẽ gặp từ đánng kính liên tiếp xuất hiện trong bài khoá
145. Bạn sẽ gặp từ nát liên tiếp xuất hiện trong bài khoá
146. Bạn sẽ gặp từ đánng sợ liên tiếp xuất hiện trong bài khoá
147. Bạn sẽ gặp từ thánng tám liên tiếp xuất hiện trong bài khoá
148. Bạn sẽ gặp từ lác đác liên tiếp xuất hiện trong bài khoá
149. Bạn sẽ gặp từ sáng tạo liên tiếp xuất hiện trong bài khoá
150. Bạn sẽ gặp từ khoác lác liên tiếp xuất hiện trong bài khoá
151. Bạn sẽ gặp từ thánng tám liên tiếp xuất hiện trong bài khoá
152. Bạn sẽ gặp từ mác liên tiếp xuất hiện trong bài khoá
153. Bạn sẽ gặp từ khát nước liên tiếp xuất hiện trong bài khoá
154. Bạn sẽ gặp từ gián tiếp liên tiếp xuất hiện trong bài khoá
155. Bạn sẽ gặp từ cát liên tiếp xuất hiện trong bài khoá
156. Bạn sẽ gặp từ áp dụng liên tiếp xuất hiện trong bài khoá
157. Bạn sẽ gặp từ đánng tiếc liên tiếp xuất hiện trong bài khoá

158. Bạn sẽ gặp từ các liên tiếp xuất hiện trong bài khoá
159. Bạn sẽ gặp từ khát vọng liên tiếp xuất hiện trong bài khoá
160. Bạn sẽ gặp từ kháng cự liên tiếp xuất hiện trong bài khoá

Date 17 décembre 2007, Corpus lu du Vietnamien, Variété du Nord, Locuteur (prénom, nom, sexe masculin, âge), Lieu de l'enregistrement (Chambre Sourde ICP Campus), Matériel (Enregistreur numérique Marantz PMD 671, Micro AKG C 1000 S).

Entraînement

- a) Bạn sẽ gặp từ tia tiếp tục xuất hiện trong bài khoá
- b) Bạn sẽ gặp từ khoan tiếp tục xuất hiện trong bài khoá
- c) Bạn sẽ gặp từ xin liên tiếp xuất hiện trong bài khoá
- d) Bạn sẽ gặp từ tha liên tiếp xuất hiện trong bài khoá

Corpus lu

1. Bạn sẽ gặp từ máng tiếp tục xuất hiện trong bài khoá
2. Bạn sẽ gặp từ mán tiếp tục xuất hiện trong bài khoá
3. Bạn sẽ gặp từ cáng tiếp tục xuất hiện trong bài khoá
4. Bạn sẽ gặp từ ngán tiếp tục xuất hiện trong bài khoá
5. Bạn sẽ gặp từ nám tiếp tục xuất hiện trong bài khoá
6. Bạn sẽ gặp từ cảm tiếp tục xuất hiện trong bài khoá
7. Bạn sẽ gặp từ táng tiếp tục xuất hiện trong bài khoá
8. Bạn sẽ gặp từ cáng tiếp tục xuất hiện trong bài khoá
9. Bạn sẽ gặp từ mán tiếp tục xuất hiện trong bài khoá

10. Bạn sẽ gặp từ cảm tiếp tục xuất hiện trong bài khoá
11. Bạn sẽ gặp từ máng tiếp tục xuất hiện trong bài khoá
12. Bạn sẽ gặp từ táng tiếp tục xuất hiện trong bài khoá
13. Bạn sẽ gặp từ nám tiếp tục xuất hiện trong bài khoá
14. Bạn sẽ gặp từ ngán tiếp tục xuất hiện trong bài khoá
15. Bạn sẽ gặp từ cáng tiếp tục xuất hiện trong bài khoá
16. Bạn sẽ gặp từ máng tiếp tục xuất hiện trong bài khoá
17. Bạn sẽ gặp từ mán tiếp tục xuất hiện trong bài khoá
18. Bạn sẽ gặp từ cảm tiếp tục xuất hiện trong bài khoá
19. Bạn sẽ gặp từ nám tiếp tục xuất hiện trong bài khoá
20. Bạn sẽ gặp từ táng tiếp tục xuất hiện trong bài khoá
21. Bạn sẽ gặp từ ngán tiếp tục xuất hiện trong bài khoá
22. Bạn sẽ gặp từ cáng tiếp tục xuất hiện trong bài khoá
23. Bạn sẽ gặp từ cảm tiếp tục xuất hiện trong bài khoá
24. Bạn sẽ gặp từ mán tiếp tục xuất hiện trong bài khoá
25. Bạn sẽ gặp từ máng tiếp tục xuất hiện trong bài khoá
26. Bạn sẽ gặp từ ngán tiếp tục xuất hiện trong bài khoá
27. Bạn sẽ gặp từ táng tiếp tục xuất hiện trong bài khoá
28. Bạn sẽ gặp từ nám tiếp tục xuất hiện trong bài khoá
29. Bạn sẽ gặp từ xác đáng liên tiếp xuất hiện trong bài khoá
30. Bạn sẽ gặp từ ngáp liên tiếp xuất hiện trong bài khoá

31. Bạn sẽ gặp từ mát dạ liên tiếp xuất hiện trong bài khoá
32. Bạn sẽ gặp từ khám phá liên tiếp xuất hiện trong bài khoá
33. Bạn sẽ gặp từ áp bức liên tiếp xuất hiện trong bài khoá
34. Bạn sẽ gặp từ ngát liên tiếp xuất hiện trong bài khoá
35. Bạn sẽ gặp từ ngáp liên tiếp xuất hiện trong bài khoá
36. Bạn sẽ gặp từ phát động liên tiếp xuất hiện trong bài khoá
37. Bạn sẽ gặp từ xám xịt liên tiếp xuất hiện trong bài khoá
38. Bạn sẽ gặp từ hát ví liên tiếp xuất hiện trong bài khoá
39. Bạn sẽ gặp từ ngác liên tiếp xuất hiện trong bài khoá
40. Bạn sẽ gặp từ tấp liên tiếp xuất hiện trong bài khoá
41. Bạn sẽ gặp từ khám phá liên tiếp xuất hiện trong bài khoá
42. Bạn sẽ gặp từ pháp lý liên tiếp xuất hiện trong bài khoá
43. Bạn sẽ gặp từ gác bút liên tiếp xuất hiện trong bài khoá
44. Bạn sẽ gặp từ xác đáng liên tiếp xuất hiện trong bài khoá
45. Bạn sẽ gặp từ hát ví liên tiếp xuất hiện trong bài khoá
46. Bạn sẽ gặp từ mát dạ liên tiếp xuất hiện trong bài khoá
47. Bạn sẽ gặp từ xác đáng liên tiếp xuất hiện trong bài khoá
48. Bạn sẽ gặp từ bám trụ liên tiếp xuất hiện trong bài khoá
49. Bạn sẽ gặp từ gác bút liên tiếp xuất hiện trong bài khoá
50. Bạn sẽ gặp từ cấp liên tiếp xuất hiện trong bài khoá
51. Bạn sẽ gặp từ tấp liên tiếp xuất hiện trong bài khoá

52. Bạn sẽ gặp từ phát động liên tiếp xuất hiện trong bài khoá
53. Bạn sẽ gặp từ xác đáng liên tiếp xuất hiện trong bài khoá
54. Bạn sẽ gặp từ ngác liên tiếp xuất hiện trong bài khoá
55. Bạn sẽ gặp từ cán sự liên tiếp xuất hiện trong bài khoá
56. Bạn sẽ gặp từ bám trụ liên tiếp xuất hiện trong bài khoá
57. Bạn sẽ gặp từ ngát liên tiếp xuất hiện trong bài khoá
58. Bạn sẽ gặp từ ngát liên tiếp xuất hiện trong bài khoá
59. Bạn sẽ gặp từ gác bút liên tiếp xuất hiện trong bài khoá
60. Bạn sẽ gặp từ tấp liên tiếp xuất hiện trong bài khoá
61. Bạn sẽ gặp từ xám xịt liên tiếp xuất hiện trong bài khoá
62. Bạn sẽ gặp từ khám phá liên tiếp xuất hiện trong bài khoá
63. Bạn sẽ gặp từ ngáp liên tiếp xuất hiện trong bài khoá
64. Bạn sẽ gặp từ áp bức liên tiếp xuất hiện trong bài khoá
65. Bạn sẽ gặp từ hát ví liên tiếp xuất hiện trong bài khoá
66. Bạn sẽ gặp từ pháp lý liên tiếp xuất hiện trong bài khoá
67. Bạn sẽ gặp từ bám trụ liên tiếp xuất hiện trong bài khoá
68. Bạn sẽ gặp từ ngáp liên tiếp xuất hiện trong bài khoá
69. Bạn sẽ gặp từ ngác liên tiếp xuất hiện trong bài khoá
70. Bạn sẽ gặp từ phát động liên tiếp xuất hiện trong bài khoá
71. Bạn sẽ gặp từ cán sự liên tiếp xuất hiện trong bài khoá
72. Bạn sẽ gặp từ cấp liên tiếp xuất hiện trong bài khoá

73. Bạn sẽ gặp từ pháp lý liên tiếp xuất hiện trong bài khoá
74. Bạn sẽ gặp từ cán sự liên tiếp xuất hiện trong bài khoá
75. Bạn sẽ gặp từ ngát liên tiếp xuất hiện trong bài khoá
76. Bạn sẽ gặp từ ngác liên tiếp xuất hiện trong bài khoá
77. Bạn sẽ gặp từ mát dạ liên tiếp xuất hiện trong bài khoá
78. Bạn sẽ gặp từ tấp liên tiếp xuất hiện trong bài khoá
79. Bạn sẽ gặp từ áp bức liên tiếp xuất hiện trong bài khoá
80. Bạn sẽ gặp từ xám xịt liên tiếp xuất hiện trong bài khoá
81. Bạn sẽ gặp từ gác bút liên tiếp xuất hiện trong bài khoá
82. Bạn sẽ gặp từ khám phá liên tiếp xuất hiện trong bài khoá
83. Bạn sẽ gặp từ cấp liên tiếp xuất hiện trong bài khoá
84. Bạn sẽ gặp từ bám trụ liên tiếp xuất hiện trong bài khoá
85. Bạn sẽ gặp từ pháp lý liên tiếp xuất hiện trong bài khoá
86. Bạn sẽ gặp từ hát ví liên tiếp xuất hiện trong bài khoá
87. Bạn sẽ gặp từ cán sự liên tiếp xuất hiện trong bài khoá
88. Bạn sẽ gặp từ xám xịt liên tiếp xuất hiện trong bài khoá
89. Bạn sẽ gặp từ cấp liên tiếp xuất hiện trong bài khoá
90. Bạn sẽ gặp từ áp bức liên tiếp xuất hiện trong bài khoá
91. Bạn sẽ gặp từ mát dạ liên tiếp xuất hiện trong bài khoá
92. Bạn sẽ gặp từ phát động liên tiếp xuất hiện trong bài khoá

Annexe 2. Valeurs moyennes et écarts-types des pentes de F1, F2, F3 (Hz) en fonction de la consonne en coda /p/, /t/ et /k/ et du ton de la syllabe qui suit la consonne cible (S = A1, M = D1, D = D2) et valeurs relevées de 50 % à 90 % de la durée de la voyelle avec un pas de 10

%	Δ		p_S	t_S	k_S	p_M	t_M	k_M	p_D	t_D	k_D
50 %	F1	Moy.	10,30	4,24	25,02	-5,07	-3,10	14,06	8,89	-1,19	13,49
		Ecart.	18,07	11,01	32,22	38,70	8,42	22,07	9,24	35,16	29,32
	F2	Moy.	3,87	25,44	6,23	16,50	21,18	-14,71	10,34	13,70	-4,81
		Ecart.	30,63	29,25	18,31	14,35	21,69	34,87	17,44	15,12	15,74
	F3	Moy.	7,26	19,76	-32,93	-28,33	10,27	-27,05	13,03	4,40	11,88
		Ecart.	108,14	62,82	89,24	30,77	54,59	38,84	16,58	32,92	41,32
60 %	F1	Moy.	27,06	16,85	26,52	-16,1	-15,5	21,15	-8,29	-10,22	20,44
		Ecart.	39,34	36,54	47,96	36,50	22,83	24,28	26,79	34,69	44,13
	F2	Moy.	-7,43	3,32	-4,18	20,7	28,56	-20,7	6,35	34,45	-24,09
		Ecart.	38,33	46,52	37,42	21,26	16,05	26,13	23,78	18,93	39,85
	F3	Moy.	8	3,31	-22,8	22,27	15,62	-35,4	22,1	4,57	8,23
		Ecart.	114,74	77,92	103,49	78,22	116,79	74,13	23,78	101,59	52,32
70 %	F1	Moy.	15,77	30,53	36,92	-37,3	-34,3	-6,5	-15,5	-19,21	-7,83
		Ecart.	51,71	29,72	59,04	38,30	39,09	42,53	37,74	40,89	45,89
	F2	Moy.	-31,22	18,41	6,18	-4,15	50,53	-48	-2,41	55,94	-30,63
		Ecart.	81,47	37,47	22,38	24,16	23,33	29,48	25,00	25,22	38,84
	F3	Moy.	6,23	15,78	-30,3	19,59	69,55	-53,8	16	37,44	51,99
		Ecart.	122,14	100,08	41,49	81,46	78,59	71,52	42,28	94,66	80,74
80 %	F1	Moy.	-9,78	4,74	9,42	-79,8	-67,9	-46,3	-47,1	-47,26	-22
		Ecart.	55,14	53,62	42,16	30,86	54,11	36,00	71,87	61,39	52,43
	F2	Moy.	-35,85	18,4	-22,4	-16,1	58,69	-81,7	-21	63,91	-34,21
		Ecart.	37,29	38,74	43,13	23,69	18,34	65,40	36,23	40,31	59,68
	F3	Moy.	-35,75	6,62	3,38	11,9	63,01	-11,9	-0,17	27,73	64,56
		Ecart.	133,57	130,33	79,29	85,48	152,71	100,87	60,18	129,88	95,55
90 %	F1	Moy.	-46,32	-58,16	-57,7	-120	-112,9	-76,4	-104	-97,57	-85,48
		Ecart.	55,12	125,73	60,59	44,80	59,70	70,34	73,30	36,43	61,84
	F2	Moy.	-82,35	17,41	-43,2	-67,5	49,97	-97,2	-62,9	85,49	-49,9
		Ecart.	72,01	52,20	73,65	40,05	30,85	81,01	57,91	64,74	65,56
	F3	Moy.	-3,5	54,61	23,19	5,5	72,93	20,3	5,38	46,05	88,95
		Ecart.	111,67	124,82	77,34	92,46	169,09	125,35	65,44	144,31	114,73

Annexe 3. Valeurs moyennes et écarts-types des pentes de l'intensité (dB) en fonction de la consonne en coda (p, t, k) et du ton de la syllabe qui suit (S = A1, M = D1, D = D2) de 50 % à 90 % de la durée de la voyelle

%	ΔI	p_S	t_S	k_S	p_M	t_M	k_M	p_D	t_D	k_D
50 %	Moy.	-0,288	-0,103	-0,239	-0,135	-0,14	-0,369	-0,07	-0,094	-0,333
	Ecart.	0,33	0,31	0,35	0,38	0,29	0,24	0,32	0,22	0,28
60 %	Moy.	-0,798	-0,36	-0,651	-0,339	-0,392	-0,992	-0,276	-0,337	-0,87
	Ecart.	0,53	0,60	0,76	0,69	0,61	0,47	0,57	0,45	0,38
70 %	Moy.	-1,324	-0,74	-1,195	-0,588	-0,679	-1,804	-0,481	-0,68	-1,623
	Ecart.	0,76	0,86	1,16	0,93	0,79	0,63	0,86	0,75	0,57
80 %	Moy.	-1,752	-1,359	-2,037	-1,061	-1,232	-3,107	-1,016	-1,222	-2,89
	Ecart.	0,99	1,07	1,59	1,09	1,04	0,74	1,11	0,91	0,52
90 %	Moy.	-2,751	-2,432	-3,918	-3,319	-3,154	-5,889	-3,067	-3,148	-5,522
	Ecart.	0,99	1,13	2,39	1,47	1,28	1,12	1,33	1,16	0,77

Annexe 4. Valeurs moyennes et écarts-types des pentes de F0 (dB) en fonction de la consonne en coda (p, t, k) et du ton de la syllabe qui suit (S = A1, M = D1, D = D2) de 50 % à 90 % de la durée de la voyelle

%	$\Delta F0$	p_S	t_S	k_S	p_M	t_M	k_M	p_D	t_D	k_D
50 %	Moy.	5,77	5,71	6,2	3,71	3,24	4,74	3,83	3,3	6,94
	Ecart.	1,74	1,63	1,21	5,32	1,52	1,82	2,09	1,86	6,38
60 %	Moy.	225,91	226,21	229,56	219,60	227,27	221,41	231,12	230,89	229,14
	Ecart.	53,90	52,60	54,43	52,27	53,25	53,72	57,29	56,00	55,66
70 %	Moy.	20,25	21,59	21,64	14,38	11,53	16,75	12,16	13,38	17,46
	Ecart.	5,85	4,13	4,02	4,59	4,14	5,97	4,53	3,73	8,26
80 %	Moy.	28,46	30,44	30,32	20,00	16,23	22,75	17,87	18,04	22,50
	Ecart.	8,26	6,48	6,76	7,47	5,72	9,07	7,31	5,93	9,63
90 %	Moy.	36,70	39,72	38,39	23,82	19,40	27,94	22,39	21,96	26,51
	Ecart.	12,29	9,99	9,52	9,81	7,06	12,18	8,95	8,11	11,61

Date 17 mars 2008, Corpus lu, Locuteur (prénom, nom, sexe, âge, langue maternelle), Lieu de l'enregistrement (Chambre Sourde du Département Parole et Cognition de GIPSA-lab Campus), Matériel (Enregistreur numérique Marantz PMD 671, Micro AKG C 1000 S)

Entraînement

- a) Tu prononces [k a] trois fois
- b) Tu prononces [a b a] trois fois
- c) Tu prononces [a t] trois fois

Corpus lu

- 1 - Tu prononces [a g a] trois fois
- 2 - Tu prononces [a p] trois fois
- 3 - Tu prononces [a ŋ] trois fois
- 4 - Tu prononces [a z] trois fois
- 5 - Tu prononces [a n] trois fois
- 6 - Tu prononces [ɲ a] trois fois
- 7 - Tu prononces [a s a] trois fois
- 8 - Tu prononces [m a] trois fois
- 9 - Tu prononces [a s] trois fois

10 - Tu prononces [a z a] trois fois

11 - Tu prononces [ʃ a] trois fois

12 - Tu prononces [m a] trois fois

13 - Tu prononces [b a] trois fois

14 - Tu prononces [a n] trois fois

15 - Tu prononces [a ŋ] trois fois

16 - Tu prononces [a v] trois fois

17 - Tu prononces [g a] trois fois

18 - Tu prononces [a s a] trois fois

19 - Tu prononces [ʃ a] trois fois

20 - Tu prononces [z a] trois fois

21 - Tu prononces [a f a] trois fois

22 - Tu prononces [a b a] trois fois

23 - Tu prononces [a n a] trois fois

24 - Tu prononces [a l a] trois fois

25 - Tu prononces [f a] trois fois

26 - Tu prononces [a b] trois fois

27 - Tu prononces [a t a] trois fois

28 - Tu prononces [a g a] trois fois

29 - Tu prononces [a k] trois fois

30 - Tu prononces [a p a] trois fois

31 - Tu prononces [a f a] trois fois

32 - Tu prononces [a z] trois fois

33 - Tu prononces [ʒ a] trois fois

34 - Tu prononces [a ʒ] trois fois

35 - Tu prononces [a ʃ a] trois fois

36 - Tu prononces [a l] trois fois

37 - Tu prononces [ʃ a] trois fois

38 - Tu prononces [a ʃ] trois fois

39 - Tu prononces [a m] trois fois

40 - Tu prononces [a g a] trois fois

41 - Tu prononces [s a] trois fois

42 - Tu prononces [g a] trois fois

43 - Tu prononces [a s] trois fois

44 - Tu prononces [p a] trois fois

45 - Tu prononces [a v] trois fois

46 - Tu prononces [a d a] trois fois

47 - Tu prononces [a b] trois fois

48 - Tu prononces [p a] trois fois

49 - Tu prononces [b a] trois fois

50 - Tu prononces [s a] trois fois

51 - Tu prononces [a ɲ] trois fois

52 - Tu prononces [t a] trois fois

53 - Tu prononces [a f] trois fois

54 - Tu prononces [a ɲ a] trois fois

55 - Tu prononces [ʒ a] trois fois

56 - Tu prononces [a v a] trois fois

57 - Tu prononces [a v] trois fois

58 - Tu prononces [a l] trois fois

59 - Tu prononces [a ɲ a] trois fois

60 - Tu prononces [a k] trois fois

61 - Tu prononces [a f a] trois fois

62 - Tu prononces [a ɲ] trois fois

63 - Tu prononces [a d a] trois fois

- 64 - Tu prononces [ʁ a] trois fois
- 65 - Tu prononces [k a] trois fois
- 66 - Tu prononces [n a] trois fois
- 67 - Tu prononces [d a] trois fois
- 68 - Tu prononces [a ʃ a] trois fois
- 69 - Tu prononces [a m] trois fois
- 70 - Tu prononces [a ʁ] trois fois
- 71 - Tu prononces [a z a] trois fois
- 72 - Tu prononces [a d] trois fois
- 73 - Tu prononces [ʁ a] trois fois
- 74 - Tu prononces [a g] trois fois
- 75 - Tu prononces [l a] trois fois
- 76 - Tu prononces [a f] trois fois
- 77 - Tu prononces [a k a] trois fois
- 78 - Tu prononces [a v a] trois fois
- 79 - Tu prononces [a ʃ] trois fois
- 80 - Tu prononces [v a] trois fois
- 81 - Tu prononces [a t] trois fois

82 - Tu prononces [a p] trois fois

83 - Tu prononces [a l a] trois fois

84 - Tu prononces [d a] trois fois

85 - Tu prononces [a s a] trois fois

86 - Tu prononces [a s] trois fois

87 - Tu prononces [f a] trois fois

88 - Tu prononces [a b a] trois fois

89 - Tu prononces [f a] trois fois

90 - Tu prononces [a d a] trois fois

91 - Tu prononces [a t a] trois fois

92 - Tu prononces [a ʁ] trois fois

93 - Tu prononces [d a] trois fois

94 - Tu prononces [a g] trois fois

95 - Tu prononces [a d] trois fois

96 - Tu prononces [a k a] trois fois

97 - Tu prononces [a p] trois fois

98 - Tu prononces [k a] trois fois

99 - Tu prononces [t a] trois fois

- 100 - Tu prononces [a ʁ a] trois fois
- 101 - Tu prononces [ʒ a] trois fois
- 102 - Tu prononces [ʁ a] trois fois
- 103 - Tu prononces [v a] trois fois
- 104 - Tu prononces [g a] trois fois
- 105 - Tu prononces [a n a] trois fois
- 106 - Tu prononces [n a] trois fois
- 107 - Tu prononces [a ʒ a] trois fois
- 108 - Tu prononces [m a] trois fois
- 109 - Tu prononces [ɲ a] trois fois
- 110 - Tu prononces [a ʃ a] trois fois
- 111 - Tu prononces [a t] trois fois
- 112 - Tu prononces [a ŋ] trois fois
- 113 - Tu prononces [a m a] trois fois
- 114 - Tu prononces [a ʁ a] trois fois
- 115 - Tu prononces [a z a] trois fois
- 116 - Tu prononces [a n a] trois fois
- 117 - Tu prononces [a ʒ] trois fois

118 - Tu prononces [a b a] trois fois

119 - Tu prononces [z a] trois fois

120 - Tu prononces [a t] trois fois

121 - Tu prononces [a ʒ] trois fois

122 - Tu prononces [s a] trois fois

123 - Tu prononces [a k] trois fois

124 - Tu prononces [a k a] trois fois

125 - Tu prononces [p a] trois fois

126 - Tu prononces [a d] trois fois

127 - Tu prononces [n a] trois fois

128 - Tu prononces [a p a] trois fois

129 - Tu prononces [a ɲ] trois fois

130 - Tu prononces [v a] trois fois

131 - Tu prononces [a p a] trois fois

132 - Tu prononces [a ʁ a] trois fois

133 - Tu prononces [a l] trois fois

134 - Tu prononces [z a] trois fois

135 - Tu prononces [a ʁ] trois fois

136 - Tu prononces [a m a] trois fois

137 - Tu prononces [a l a] trois fois

138 - Tu prononces [a g] trois fois

139 - Tu prononces [a n] trois fois

140 - Tu prononces [a ʒ a] trois fois

141 - Tu prononces [b a] trois fois

142 - Tu prononces [a b] trois fois

143 - Tu prononces [a f] trois fois

144 - Tu prononces [a z] trois fois

145 - Tu prononces [a t a] trois fois

146 - Tu prononces [l a] trois fois

147 - Tu prononces [a ɲ a] trois fois

148 - Tu prononces [a m] trois fois

149 - Tu prononces [k a] trois fois

150 - Tu prononces [a ʒ a] trois fois

151 - Tu prononces [ɲ a] trois fois

152 - Tu prononces [l a] trois fois

153 - Tu prononces [t a] trois fois

154 - Tu prononces [a m a] trois fois

155 - Tu prononces [a v a] trois fois

156 - Tu prononces [a ʃ] trois fois

NGHIÊN CỨU VỀ NGỮ ÂM THỰC NGHIỆM

PHIẾU THÔNG TIN

Số thứ tự :

Xin bạn vui lòng cho biết một số thông tin cá nhân trước khi làm bài khảo sát. Tôi xin khẳng định chỉ sử dụng các thông tin này cho nghiên cứu của tôi và không sử dụng trong bất kỳ trường hợp nào khác. Rất cảm ơn bạn đã tham gia!

Tuổi: Giới tính: Nam Nữ
 Giọng nói: Hà Nội Vùng khác :
 Giọng nói của người thân:
 Bố : Hà Nội miền Bắc (khác) miền Trung miền Nam
 Mẹ : Hà Nội miền Bắc (khác) miền Trung miền Nam
 Chồng (vợ) : Hà Nội miền Bắc (khác) miền Trung miền Nam
 Các nơi ở Việt Nam mà bạn đã ở hơn 6 tháng:.....
 Trình độ học vấn : Năm 1 Năm 3 Email.....
 Bạn đã học tiếng Pháp bao lâu ?..... Ngoại ngữ khác (nếu có).....
 Bạn có từng ở nước ngoài không? Không Có Nước nào..... Bao lâu

Trình tự các bài khảo sát về phụ âm tiếng Pháp

Bài 1 Nghe phát âm và nhắc lại (Phòng thu âm): Các bạn hãy lắng nghe các âm được đọc và nhắc lại
Bài 2 Nghe và chọn âm (Phòng máy): Các bạn hãy lắng nghe các tiếng / các từ được phát âm và xác định phụ âm hoặc nhóm phụ âm nào xuất hiện trong tiếng / từ đó bằng cách nhấn chuột vào ô chữ tương ứng (phần khoanh tròn trong hình minh họa bên dưới đây).

Lưu ý : các bạn nhập tên “Fichier de Sortie” (.xls) bằng cách đánh số tên fichier d’entrée (đã có sẵn ở phần trên) + số thứ tự của bạn. Ví dụ :

- ⇒ Fichier d’Entrée : 1.txt
- ⇒ Fichier de Sortie : 1.1 nếu bạn có số thứ tự là số 1 (1.5 nếu số thứ tự là 5, ...)

(1) tên fichier d’entrée (1.txt)

(2) Nhập số 3 +

(3) Nhập số 1 +

(4) Click vào

Click chuột

Après avoir rempli les cases, taper Modifications

Annexe 7 : liste des séquences communes aux deux langues avec leur nombre d'occurrences dans le lexique de 15 000 mots du français et 5 000 mots du vietnamien

	Séquences communes aux deux langues	Occurrences		Séquences communes aux deux langues	Occurrences
Fr.	p . k	2	Fr.	k . s	178
Vn.	p . k	3	Vn.	k . s	2
Fr.	p . n	5	Fr.	k . t	276
Vn.	p . n	1	Vn.	k . t	9
Fr.	p . s	29	Fr.	m . k w	1
Vn.	p . s	2	Vn.	m . k w	5
Fr.	p . t	95	Fr.	m . b	3
Vn.	p . t	5	Vn.	m . b	7
Fr.	t . b	2	Fr.	m . d	2
Vn.	t . b	8	Vn.	m . d	9
Fr.	t . d	5	Fr.	m . k	1
Vn.	t . d	19	Vn.	m . k	12
Fr.	t . l	11	Fr.	m . l	1
Vn.	t . l	15	Vn.	m . l	6
Fr.	t . m	10	Fr.	m . n	21
Vn.	t . m	8	Vn.	m . n	5
Fr.	t . n	2	Fr.	m . s	3
Vn.	t . n	4	Vn.	m . s	5
Fr.	t . s	3	Fr.	m . t	2
Vn.	t . s	2	Vn.	m . t	10
Fr.	k . b	3	Fr.	m . v	5
Vn.	k . b	5	Vn.	m . v	5
Fr.	k . d	7	Fr.	n . d	7
Vn.	k . d	18	Vn.	n . d	29
Fr.	k . f	3	Fr.	n . f	2
Vn.	k . f	5	Vn.	n . f	9
Fr.	k . k	1	Fr.	n . k	2
Vn.	k . k	3	Vn.	n . k	22
Fr.	k . l	1	Fr.	n . k w	1
Vn.	k . l	13	Vn.	n . k w	4
Fr.	k . m	3	Fr.	n . l	3
Vn.	k . m	14	Vn.	n . l	24
Fr.	k . n	13	Fr.	n . m	1
Vn.	k . n	6	Vn.	n . m	15

Annexe VII – Liste des séquences de consonnes en inter-syllabique communes aux deux langues.

Fr.	n . s	4	Fr.	n . v	2
Vn.	n . s	4	Vn.	n . v	8
Fr.	n . t	9	Fr.	ŋ . b	1
Vn.	n . t	40	Vn.	ŋ . b	19
Fr.	ŋ . s	1	Vn.	ŋ . s	2

Date 17 mars 2008, Corpus lu, Locuteur (prénom, nom, sexe, âge, langue maternelle), Lieu de l'enregistrement (Chambre Sourde du Département Parole et Cognition de GIPSA-lab Campus), Matériel (Enregistreur numérique Marantz PMD 671, Micro AKG C 1000 S)

Entraînement

- a) Tu prononces [i b l a] trois fois
- b) Tu prononces [k f u] trois fois
- c) Tu prononces [o t m] trois fois

Corpus lu

- 1 - Tu prononces [a t . b a] trois fois
- 2 - Tu prononces [a n . s a] trois fois
- 3 - Tu prononces [a n . f a] trois fois
- 4 - Tu prononces [a p . t a] trois fois
- 5 - Tu prononces [a k . k a] trois fois
- 6 - Tu prononces [a n . d a] trois fois
- 7 - Tu prononces [k l a] trois fois
- 8 - Tu prononces [a t . l a] trois fois
- 9 - Tu prononces [a m . n a] trois fois

- 10 - Tu prononces [a k . n a] trois fois
- 11 - Tu prononces [a t m] trois fois
- 12 - Tu prononces [k l a] trois fois
- 13 - Tu prononces [a k . t a] trois fois
- 14 - Tu prononces [a k l] trois fois
- 15 - Tu prononces [a p . t a] trois fois
- 16 - Tu prononces [a m . d a] trois fois
- 17 - Tu prononces [a k t] trois fois
- 18 - Tu prononces [a m . l a] trois fois
- 19 - Tu prononces [a t s] trois fois
- 20 - Tu prononces [p t a] trois fois
- 21 - Tu prononces [a k . k a] trois fois
- 22 - Tu prononces [a k . m a] trois fois
- 23 - Tu prononces [a t s] trois fois
- 24 - Tu prononces [a m . n a] trois fois
- 25 - Tu prononces [a p . k a] trois fois
- 26 - Tu prononces [a m . d a] trois fois
- 27 - Tu prononces [a ŋ s] trois fois

- 28 - Tu prononces [a k s] trois fois
- 29 - Tu prononces [a n . s a] trois fois
- 30 - Tu prononces [a k . l a] trois fois
- 31 - Tu prononces [a m n] trois fois
- 32 - Tu prononces [a t . l a] trois fois
- 33 - Tu prononces [a m . b a] trois fois
- 34 - Tu prononces [p s a] trois fois
- 35 - Tu prononces [a m n] trois fois
- 36 - Tu prononces [a p s] trois fois
- 37 - Tu prononces [a n . t a] trois fois
- 38 - Tu prononces [a k s] trois fois
- 39 - Tu prononces [a ŋ . s a] trois fois
- 40 - Tu prononces [a n t] trois fois
- 41 - Tu prononces [k s a] trois fois
- 42 - Tu prononces [a ŋ . s a] trois fois
- 43 - Tu prononces [a m . b a] trois fois
- 44 - Tu prononces [a k l] trois fois
- 45 - Tu prononces [a k . m a] trois fois

- 46 - Tu prononces [p s a] trois fois
- 47 - Tu prononces [t s a] trois fois
- 48 - Tu prononces [a m . l a] trois fois
- 49 - Tu prononces [a t . b a] trois fois
- 50 - Tu prononces [a p s] trois fois
- 51 - Tu prononces [a k . t a] trois fois
- 52 - Tu prononces [a n s] trois fois
- 53 - Tu prononces [a n . t a] trois fois
- 54 - Tu prononces [a p . k a] trois fois
- 55 - Tu prononces [a t . s a] trois fois
- 56 - Tu prononces [p n a] trois fois
- 57 - Tu prononces [a k . n a] trois fois
- 58 - Tu prononces [a k s] trois fois
- 59 - Tu prononces [a n . k a] trois fois
- 60 - Tu prononces [a k . f a] trois fois
- 61 - Tu prononces [a n . t a] trois fois
- 62 - Tu prononces [a t . m a] trois fois
- 63 - Tu prononces [a n . d a] trois fois

- 64 - Tu prononces [a k . k a] trois fois
- 65 - Tu prononces [a m . v a] trois fois
- 66 - Tu prononces [a n . m a] trois fois
- 67 - Tu prononces [a t . s a] trois fois
- 68 - Tu prononces [a m . v a] trois fois
- 69 - Tu prononces [a p s] trois fois
- 70 - Tu prononces [p n a] trois fois
- 71 - Tu prononces [a n . s a] trois fois
- 72 - Tu prononces [p s a] trois fois
- 73 - Tu prononces [k t a] trois fois
- 74 - Tu prononces [a p . s a] trois fois
- 75 - Tu prononces [p n a] trois fois
- 76 - Tu prononces [a t . b a] trois fois
- 77 - Tu prononces [a n s] trois fois
- 78 - Tu prononces [a p . n a] trois fois
- 79 - Tu prononces [a t m] trois fois
- 80 - Tu prononces [p t a] trois fois
- 81 - Tu prononces [a m . s a] trois fois

- 82 - Tu prononces [p t a] trois fois
- 83 - Tu prononces [a k l] trois fois
- 84 - Tu prononces [a n . l a] trois fois
- 85 - Tu prononces [a p . n a] trois fois
- 86 - Tu prononces [a t m] trois fois
- 87 - Tu prononces [a p . t a] trois fois
- 88 - Tu prononces [a n . d a] trois fois
- 89 - Tu prononces [k s a] trois fois
- 90 - Tu prononces [a t . l a] trois fois
- 91 - Tu prononces [a n . d a] trois fois
- 92 - Tu prononces [a n . k a] trois fois
- 93 - Tu prononces [a p s] trois fois
- 94 - Tu prononces [a m . v a] trois fois
- 95 - Tu prononces [a n s] trois fois
- 96 - Tu prononces [a n . l a] trois fois
- 97 - Tu prononces [a p t] trois fois
- 98 - Tu prononces [a n . v a] trois fois
- 99 - Tu prononces [a m . d a] trois fois

- 100 - Tu prononces [a n s] trois fois
- 101 - Tu prononces [a k . t a] trois fois
- 102 - Tu prononces [a k m] trois fois
- 103 - Tu prononces [a n . k a] trois fois
- 104 - Tu prononces [k t a] trois fois
- 105 - Tu prononces [a p t] trois fois
- 106 - Tu prononces [a m . l a] trois fois
- 107 - Tu prononces [k l a] trois fois
- 108 - Tu prononces [k n a] trois fois
- 109 - Tu prononces [a k . f a] trois fois
- 110 - Tu prononces [p k a] trois fois
- 111 - Tu prononces [a m . s a] trois fois
- 112 - Tu prononces [k t a] trois fois
- 113 - Tu prononces [k s a] trois fois
- 114 - Tu prononces [a ŋ . b a] trois fois
- 115 - Tu prononces [a t . m a] trois fois
- 116 - Tu prononces [a n t] trois fois
- 117 - Tu prononces [a k s] trois fois

118 - Tu prononces [a ŋ . b a] trois fois

119 - Tu prononces [t s a] trois fois

120 - Tu prononces [a n t] trois fois

121 - Tu prononces [a m . t a] trois fois

122 - Tu prononces [p k a] trois fois

123 - Tu prononces [m n a] trois fois

124 - Tu prononces [a k . f a] trois fois

125 - Tu prononces [a m . s a] trois fois

126 - Tu prononces [k n a] trois fois

127 - Tu prononces [a m . v a] trois fois

128 - Tu prononces [a p . s a] trois fois

129 - Tu prononces [a t . n a] trois fois

130 - Tu prononces [a k t] trois fois

131 - Tu prononces [a ŋ s] trois fois

132 - Tu prononces [a p . n a] trois fois

133 - Tu prononces [a k . l a] trois fois

134 - Tu prononces [a k m] trois fois

135 - Tu prononces [a m . k a] trois fois

136 - Tu prononces [a m . b a] trois fois

137 - Tu prononces [a t s] trois fois

138 - Tu prononces [a n . t a] trois fois

139 - Tu prononces [a n . m a] trois fois

140 - Tu prononces [a n d] trois fois

141 - Tu prononces [a k . t a] trois fois

142 - Tu prononces [t s a] trois fois

143 - Tu prononces [p s a] trois fois

144 - Tu prononces [a n . l a] trois fois

145 - Tu prononces [a k . f a] trois fois

146 - Tu prononces [a m . k a] trois fois

147 - Tu prononces [a t . m a] trois fois

148 - Tu prononces [a k . l a] trois fois

149 - Tu prononces [a m . t a] trois fois

150 - Tu prononces [p k a] trois fois

151 - Tu prononces [a t . b a] trois fois

152 - Tu prononces [a k . k a] trois fois

153 - Tu prononces [a n d] trois fois

- 154 - Tu prononces [a m . l a] trois fois
- 155 - Tu prononces [a n . f a] trois fois
- 156 - Tu prononces [a ŋ . b a] trois fois
- 157 - Tu prononces [m n a] trois fois
- 158 - Tu prononces [a k . s a] trois fois
- 159 - Tu prononces [a p . k a] trois fois
- 160 - Tu prononces [k t a] trois fois
- 161 - Tu prononces [a k . s a] trois fois
- 162 - Tu prononces [a k . m a] trois fois
- 163 - Tu prononces [a n . m a] trois fois
- 164 - Tu prononces [a k t] trois fois
- 165 - Tu prononces [k l a] trois fois
- 166 - Tu prononces [a t . n a] trois fois
- 167 - Tu prononces [k n a] trois fois
- 168 - Tu prononces [a t . s a] trois fois
- 169 - Tu prononces [a ŋ . s a] trois fois
- 170 - Tu prononces [a n t] trois fois
- 171 - Tu prononces [a t m] trois fois

- 172 - Tu prononces [a m n] trois fois
- 173 - Tu prononces [m n a] trois fois
- 174 - Tu prononces [p n a] trois fois
- 175 - Tu prononces [a p . n a] trois fois
- 176 - Tu prononces [a p t] trois fois
- 177 - Tu prononces [a t . s a] trois fois
- 178 - Tu prononces [a m . t a] trois fois
- 179 - Tu prononces [k n a] trois fois
- 180 - Tu prononces [a m . k a] trois fois
- 181 - Tu prononces [a n . v a] trois fois
- 182 - Tu prononces [t s a] trois fois
- 183 - Tu prononces [m n a] trois fois
- 184 - Tu prononces [a k m] trois fois
- 185 - Tu prononces [a m . k a] trois fois
- 186 - Tu prononces [a k . n a] trois fois
- 187 - Tu prononces [a t s] trois fois
- 188 - Tu prononces [a p . t a] trois fois
- 189 - Tu prononces [a m n] trois fois

- 190 - Tu prononces [a m . d a] trois fois
- 191 - Tu prononces [a ŋ s] trois fois
- 192 - Tu prononces [a p . s a] trois fois
- 193 - Tu prononces [p t a] trois fois
- 194 - Tu prononces [a p t] trois fois
- 195 - Tu prononces [a n . f a] trois fois
- 196 - Tu prononces [a k . s a] trois fois
- 197 - Tu prononces [a n d] trois fois
- 198 - Tu prononces [a n . v a] trois fois
- 199 - Tu prononces [a n d] trois fois
- 200 - Tu prononces [a t . l a] trois fois
- 201 - Tu prononces [a k l] trois fois
- 202 - Tu prononces [a n . l a] trois fois
- 203 - Tu prononces [a n . s a] trois fois
- 204 - Tu prononces [a k . m a] trois fois
- 205 - Tu prononces [a k t] trois fois
- 206 - Tu prononces [a p . s a] trois fois
- 207 - Tu prononces [a m . n a] trois fois

- 208 - Tu prononces [a t . n a] trois fois
- 209 - Tu prononces [a k m] trois fois
- 210 - Tu prononces [a ŋ s] trois fois
- 211 - Tu prononces [k s a] trois fois
- 212 - Tu prononces [a m . n a] trois fois
- 213 - Tu prononces [a k . s a] trois fois
- 214 - Tu prononces [a n . f a] trois fois
- 215 - Tu prononces [a m . b a] trois fois
- 216 - Tu prononces [a ŋ . s a] trois fois
- 217 - Tu prononces [a m . s a] trois fois
- 218 - Tu prononces [a p . k a] trois fois
- 219 - Tu prononces [a k . l a] trois fois
- 220 - Tu prononces [a m . t a] trois fois
- 221 - Tu prononces [a t . m a] trois fois
- 222 - Tu prononces [p k a] trois fois
- 223 - Tu prononces [a ŋ . b a] trois fois
- 224 - Tu prononces [a n . m a] trois fois
- 225 - Tu prononces [a k . n a] trois fois

226 - Tu prononces [a n . v a] trois fois

227 - Tu prononces [a n . k a] trois fois

228 - Tu prononces [a t . n a] trois fois

Date 17 mars 2008, Corpus lu du français, Locuteur (prénom, nom, sexe, âge, langue maternelle), Lieu de l'enregistrement (Chambre Sourde du Département Parole et Cognition de GIPSA-lab Campus), Matériel (Enregistreur numérique Marantz PMD 671, Micro AKG C 1000 S)

Entraînement

- a) Tu prononces « classe » trois fois
- b) Tu prononces « option » trois fois
- c) Tu prononces « miracle » trois fois

Corpus lu

- 1 - Tu prononces « insecte » trois fois
- 2 - Tu prononces « acteur » trois fois
- 3 - Tu prononces « weekend » trois fois
- 4 - Tu prononces « snack-bar » trois fois
- 5 - Tu prononces « ethnie » trois fois
- 6 - Tu prononces « hypnose » trois fois
- 7 - Tu prononces « cloche » trois fois
- 8 - Tu prononces « come-back » trois fois
- 9 - Tu prononces « adepte » trois fois

- 10 - Tu prononces « cycle » trois fois
- 11 - Tu prononces « hymne » trois fois
- 12 - Tu prononces « apside » trois fois
- 13 - Tu prononces « milk-bar » trois fois
- 14 - Tu prononces « pneu » trois fois
- 15 - Tu prononces « snack-bar » trois fois
- 16 - Tu prononces « ethnie » trois fois
- 17 - Tu prononces « vexer » trois fois
- 18 - Tu prononces « ptose » trois fois
- 19 - Tu prononces « boucle » trois fois
- 20 - Tu prononces « chapka » trois fois
- 21 - Tu prononces « acteur » trois fois
- 22 - Tu prononces « bec-fin » trois fois
- 23 - Tu prononces « dictée » trois fois
- 24 - Tu prononces « snack-bar » trois fois
- 25 - Tu prononces « cloche » trois fois
- 26 - Tu prononces « apside » trois fois
- 27 - Tu prononces « gymnase » trois fois
- 28 - Tu prononces « stockfisch » trois fois

- 29 - Tu prononces « jeans » trois fois
- 30 - Tu prononces « acte » trois fois
- 31 - Tu prononces « bloc-notes » trois fois
- 32 - Tu prononces « capsule » trois fois
- 33 - Tu prononces « atteler » trois fois
- 34 - Tu prononces « check-list » trois fois
- 35 - Tu prononces « taxe » trois fois
- 36 - Tu prononces « drachme » trois fois
- 37 - Tu prononces « rumba » trois fois
- 38 - Tu prononces « stockfisch » trois fois
- 39 - Tu prononces « psychose » trois fois
- 40 - Tu prononces « phénix » trois fois
- 41 - Tu prononces « check-list » trois fois
- 42 - Tu prononces « atteler » trois fois
- 43 - Tu prononces « boucle » trois fois
- 44 - Tu prononces « drachme » trois fois
- 45 - Tu prononces « ellipse » trois fois
- 46 - Tu prononces « boucle » trois fois
- 47 - Tu prononces « taxi » trois fois

- 48 - Tu prononces « weekend » trois fois
- 49 - Tu prononces « hymne » trois fois
- 50 - Tu prononces « sprinter » trois fois
- 51 - Tu prononces « capture » trois fois
- 52 - Tu prononces « psychose » trois fois
- 53 - Tu prononces « cycle » trois fois
- 54 - Tu prononces « come-back » trois fois
- 55 - Tu prononces « hypnose » trois fois
- 56 - Tu prononces « ethnique » trois fois
- 57 - Tu prononces « taxi » trois fois
- 58 - Tu prononces « stockfisch » trois fois
- 59 - Tu prononces « cycle » trois fois
- 60 - Tu prononces « ellipse » trois fois
- 61 - Tu prononces « check-list » trois fois
- 62 - Tu prononces « dictée » trois fois
- 63 - Tu prononces « acte » trois fois
- 64 - Tu prononces « knout » trois fois
- 65 - Tu prononces « sprint » trois fois
- 66 - Tu prononces « hymne » trois fois

- 67 - Tu prononces « drachme » trois fois
- 68 - Tu prononces « chapka » trois fois
- 69 - Tu prononces « acné » trois fois
- 70 - Tu prononces « capsule » trois fois
- 71 - Tu prononces « insecte » trois fois
- 72 - Tu prononces « sprinter » trois fois
- 73 - Tu prononces « knout » trois fois
- 74 - Tu prononces « optique » trois fois
- 75 - Tu prononces « rythme » trois fois
- 76 - Tu prononces « ellipse » trois fois
- 77 - Tu prononces « bloc-notes » trois fois
- 78 - Tu prononces « apte » trois fois
- 79 - Tu prononces « hot-dog » trois fois
- 80 - Tu prononces « psychose » trois fois
- 81 - Tu prononces « jeans » trois fois
- 82 - Tu prononces « cloche » trois fois
- 83 - Tu prononces « adepte » trois fois
- 84 - Tu prononces « insecte » trois fois
- 85 - Tu prononces « psychique » trois fois

- 86 - Tu prononces « atlas » trois fois
- 87 - Tu prononces « bec-fin » trois fois
- 88 - Tu prononces « sprinter » trois fois
- 89 - Tu prononces « taxe » trois fois
- 90 - Tu prononces « ethnique » trois fois
- 91 - Tu prononces « wattman » trois fois
- 92 - Tu prononces « rumba » trois fois
- 93 - Tu prononces « come-back » trois fois
- 94 - Tu prononces « hot-dog » trois fois
- 95 - Tu prononces « apte » trois fois
- 96 - Tu prononces « jeans » trois fois
- 97 - Tu prononces « bec-fin » trois fois
- 98 - Tu prononces « acteur » trois fois
- 99 - Tu prononces « climat » trois fois
- 100 - Tu prononces « acné » trois fois
- 101 - Tu prononces « acte » trois fois
- 102 - Tu prononces « knout » trois fois
- 103 - Tu prononces « optique » trois fois
- 104 - Tu prononces « atlas » trois fois

- 105 - Tu prononces « ptose » trois fois
- 106 - Tu prononces « apside » trois fois
- 107 - Tu prononces « milk-bar » trois fois
- 108 - Tu prononces « ptose » trois fois
- 109 - Tu prononces « laps » trois fois
- 110 - Tu prononces « weekend » trois fois
- 111 - Tu prononces « rythme » trois fois
- 112 - Tu prononces « optique » trois fois
- 113 - Tu prononces « climat » trois fois
- 114 - Tu prononces « psychique » trois fois
- 115 - Tu prononces « gymnase » trois fois
- 116 - Tu prononces « laps » trois fois
- 117 - Tu prononces « rumba » trois fois
- 118 - Tu prononces « gymnase » trois fois
- 119 - Tu prononces « ethnie » trois fois
- 120 - Tu prononces « atteler » trois fois
- 121 - Tu prononces « bloc-notes » trois fois
- 122 - Tu prononces « pneu » trois fois
- 123 - Tu prononces « chapka » trois fois

- 124 - Tu prononces « wattman » trois fois
- 125 - Tu prononces « rythme » trois fois
- 126 - Tu prononces « laps » trois fois
- 127 - Tu prononces « phénix » trois fois
- 128 - Tu prononces « capture » trois fois
- 129 - Tu prononces « acné » trois fois
- 130 - Tu prononces « psychique » trois fois
- 131 - Tu prononces « wattman » trois fois
- 132 - Tu prononces « ethnique » trois fois
- 133 - Tu prononces « sprint » trois fois
- 134 - Tu prononces « phénix » trois fois
- 135 - Tu prononces « vexer » trois fois
- 136 - Tu prononces « adepte » trois fois
- 137 - Tu prononces « climat » trois fois
- 138 - Tu prononces « capsule » trois fois
- 139 - Tu prononces « apte » trois fois
- 140 - Tu prononces « pneu » trois fois
- 141 - Tu prononces « hypnose » trois fois
- 142 - Tu prononces « sprint » trois fois

143 - Tu prononces « capture » trois fois

144 - Tu prononces « dictée » trois fois

145 - Tu prononces « taxi » trois fois

146 - Tu prononces « atlas » trois fois

147 - Tu prononces « vexer » trois fois

148 - Tu prononces « hot-dog » trois fois

149 - Tu prononces « milk-bar » trois fois

150 - Tu prononces « taxe » trois fois

Date : 28 avril 2009

Corpus lu du français

Locuteur : nom et prénom, sexe, âge, langue maternelle, numéro

Lieu de l'enregistrement : Département de Langue et de Civilisation Françaises, École Supérieure des Langues Étrangères, Université Nationale du Vietnam à Hanoi

Matériel : Enregistreur numérique Marantz PMD 670

Micro : Behringer B-5

Entraînement

- a) Tu prononces « classe » trois fois
- b) Tu prononces « option » trois fois
- c) Tu prononces « miracle » trois fois

Corpus lu

- 1 - Tu prononces « taxi » trois fois
- 2 - Tu prononces « adepte » trois fois
- 3 - Tu prononces « weekend » trois fois
- 4 - Tu prononces « knout » trois fois
- 5 - Tu prononces « rumba » trois fois
- 6 - Tu prononces « insecte » trois fois
- 7 - Tu prononces « sprinter » trois fois
- 8 - Tu prononces « bec-fin » trois fois
- 9 - Tu prononces « ptose » trois fois

- 10 - Tu prononces « acné » trois fois
- 11 - Tu prononces « capture » trois fois
- 12 - Tu prononces « ellipse » trois fois
- 13 - Tu prononces « wattman » trois fois
- 14 - Tu prononces « rythme » trois fois
- 15 - Tu prononces « adepte » trois fois
- 16 - Tu prononces « cloche » trois fois
- 17 - Tu prononces « sprint » trois fois
- 18 - Tu prononces « dictée » trois fois
- 19 - Tu prononces « ethnique » trois fois
- 20 - Tu prononces « knout » trois fois
- 21 - Tu prononces « capsule » trois fois
- 22 - Tu prononces « rythme » trois fois
- 23 - Tu prononces « sprint » trois fois
- 24 - Tu prononces « atlas » trois fois
- 25 - Tu prononces « ellipse » trois fois
- 26 - Tu prononces « insecte » trois fois
- 27 - Tu prononces « weekend » trois fois
- 28 - Tu prononces « hot-dog » trois fois
- 29 - Tu prononces « capsule » trois fois
- 30 - Tu prononces « hypnose » trois fois
- 31 - Tu prononces « gymnase » trois fois

- 32 - Tu prononces « phénix » trois fois
- 33 - Tu prononces « check-list » trois fois
- 34 - Tu prononces « rythme » trois fois
- 35 - Tu prononces « sprinter » trois fois
- 36 - Tu prononces « capture » trois fois
- 37 - Tu prononces « phénix » trois fois
- 38 - Tu prononces « rumba » trois fois
- 39 - Tu prononces « cloche » trois fois
- 40 - Tu prononces « hymne » trois fois
- 41 - Tu prononces « dictée » trois fois
- 42 - Tu prononces « capsule » trois fois
- 43 - Tu prononces « adepte » trois fois
- 44 - Tu prononces « atlas » trois fois
- 45 - Tu prononces « bec-fin » trois fois
- 46 - Tu prononces « check-list » trois fois
- 47 - Tu prononces « taxi » trois fois
- 48 - Tu prononces « snack-bar » trois fois
- 49 - Tu prononces « jeans » trois fois
- 50 - Tu prononces « hymne » trois fois
- 51 - Tu prononces « insecte » trois fois
- 52 - Tu prononces « cycle » trois fois
- 53 - Tu prononces « pneu » trois fois

- 54 - Tu prononces « hot-dog » trois fois
- 55 - Tu prononces « sprinter » trois fois
- 56 - Tu prononces « chapka » trois fois
- 57 - Tu prononces « ptose » trois fois
- 58 - Tu prononces « sprint » trois fois
- 59 - Tu prononces « gymnase » trois fois
- 60 - Tu prononces « atlas » trois fois
- 61 - Tu prononces « check-list » trois fois
- 62 - Tu prononces « psychique » trois fois
- 63 - Tu prononces « snack-bar » trois fois
- 64 - Tu prononces « hymne » trois fois
- 65 - Tu prononces « gymnase » trois fois
- 66 - Tu prononces « capture » trois fois
- 67 - Tu prononces « bec-fin » trois fois
- 68 - Tu prononces « psychique » trois fois
- 69 - Tu prononces « jeans » trois fois
- 70 - Tu prononces « taxi » trois fois
- 71 - Tu prononces « weekend » trois fois
- 72 - Tu prononces « acné » trois fois
- 73 - Tu prononces « cycle » trois fois
- 74 - Tu prononces « snack-bar » trois fois
- 75 - Tu prononces « hypnose » trois fois

- 76 - Tu prononces « pneu » trois fois
- 77 - Tu prononces « chapka » trois fois
- 78 - Tu prononces « jeans » trois fois
- 79 - Tu prononces « ethnique » trois fois
- 80 - Tu prononces « phénix » trois fois
- 81 - Tu prononces « ptose » trois fois
- 82 - Tu prononces « knout » trois fois
- 83 - Tu prononces « ellipse » trois fois
- 84 - Tu prononces « wattman » trois fois
- 85 - Tu prononces « cycle » trois fois
- 86 - Tu prononces « rumba » trois fois
- 87 - Tu prononces « psychique » trois fois
- 88 - Tu prononces « dictée » trois fois
- 89 - Tu prononces « hypnose » trois fois
- 90 - Tu prononces « cloche » trois fois
- 91 - Tu prononces « pneu » trois fois
- 92 - Tu prononces « wattman » trois fois
- 93 - Tu prononces « acné » trois fois
- 94 - Tu prononces « chapka » trois fois
- 95 - Tu prononces « hot-dog » trois fois
- 96 - Tu prononces « ethnique » trois fois

NGHIÊN CỨU VỀ NGỮ ÂM THỰC NGHIỆM

PHIẾU THÔNG TIN

Số thứ tự :

Xin bạn vui lòng cho biết một số thông tin cá nhân trước khi thực hiện khảo sát. Tôi xin khẳng định chỉ sử dụng các thông tin này cho nghiên cứu của tôi và không sử dụng trong bất kỳ trường hợp nào khác. Rất cảm ơn bạn đã tham gia!

Tên : Tuổi : Giới tính : Nam Nữ

Giọng nói: Hà Nội Vùng khác :

Giọng nói của người thân:

Bố : Hà Nội miền Bắc (khác) miền Trung miền Nam

Mẹ : Hà Nội miền Bắc (khác) miền Trung miền Nam

Chồng (vợ) : Hà Nội miền Bắc (khác) miền Trung miền Nam

Các nơi ở Việt Nam mà bạn đã ở hơn 6 tháng :

Trình độ học vấn : Năm 1 Năm 2 Năm 3 Năm 4 Cao đẳng

Bạn đã học tiếng Pháp bao lâu?..... Ngoại ngữ khác (nếu có).....

Bạn có từng ở nước ngoài không? Không Có Nước nào..... Bao lâu.....

Email :

Xin bạn hãy đọc một số lưu ý sau đây trước khi đọc corpus :

- Tư thế ngồi ngay ngắn, đối diện với micro và cách micro 30 cm
- Đọc to, rõ ràng, tốc độ bình thường (không nhanh, không quá chậm)
- Đọc số thứ tự (bằng tiếng Việt) trước mỗi câu
- Nếu bạn đọc sai, không vấn đề gì ! Bạn hãy đọc lại câu ngay sau đó (có kèm theo số thứ tự)
- Khi thu âm, xin bạn hãy cố gắng không cử động mạnh gây ra tiếng ồn, không giở trang giấy trong lúc đọc
- Đeo kính nếu bạn bị cận hay viễn thị.

Connaissez-vous ce mot ?

Ce questionnaire n'est pas une évaluation ! Rassurez-vous que c'est juste une analyse !
Cochez la case « Oui » si vous pensez connaître le sens de ces mots (même si le sens n'est pas bon) et « Non » si vous ne le connaissez pas. Merci de votre participation.

N°	Mots	Oui	Non
1	pneu		
2	psychique		
3	ptose		
4	chapka		
5	hypnose		
6	capsule		
7	capture		
8	ellipse		
9	adepte		
10	rythme		
11	hot-dog		
12	atlas		
13	wattman		
14	ethnique		
15	cloche		
16	knout		
17	snack-bar		
18	bec-fin		
19	check-list		
20	acné		
21	taxi		
22	acteur		
23	cycle		
24	phénix		
25	insecte		
26	rumba		
27	gymnase		
28	hymne		
29	sprinter		
30	weekend		
31	jeans		
32	sprint		

- N° :
- Nom et prénom :

