THÈSE

présentée par

Ahmed Réda BERRAH

pour obtenir le grade de DOCTEUR
de l'Institut National Polytechnique de Grenoble
(Arrêté ministériel du 30 mars 1992)

Spécialité : **Sciences Cognitives**

Évolution d'une Société Artificielle d'Agents de Parole : Un Modèle pour l'Émergence des Structures Phonétiques

Soutenue le 12 juin 1998 devant le jury composé de Messieurs :

Jean-Luc SCHWARTZ	Président
Ian MADDIESON	Rapporteur
Luc STEELS	Rapporteur
Jean-Marie HOMBERT	Examinateur
Louis-Jean BOË	Directeur de thèse

Thèse préparée au sein de l'Institut de la Communication Parlée (UPRESA CNRS 5009) et du Laboratoire Leibniz-IMAG (UMR CNRS 5522)

THÈSE

présentée par

Ahmed Réda BERRAH

pour obtenir le grade de DOCTEUR
de l'Institut National Polytechnique de Grenoble
(Arrêté ministériel du 30 mars 1992)

Spécialité : **Sciences Cognitives**

Évolution d'une Société Artificielle d'Agents de Parole : Un Modèle pour l'Émergence des Structures Phonétiques

---

Soutenue le 12 juin 1998 devant le jury composé de Messieurs :

Jean-Luc SCHWARTZ	Président
Ian MADDIESON	Rapporteur
Luc STEELS	Rapporteur
Jean-Marie HOMBERT	Examinateur
Louis-Jean BOË	Directeur de thèse

Thèse préparée au sein de l'Institut de la Communication Parlée (UPRESA CNRS 5009) et du Laboratoire Leibniz-IMAG (UMR CNRS 5522)

Remerciements

Je remercie très vivement Monsieur Jean-Luc Schwartz, Directeur de Recherche à l'Institut de la Communication Parlée (ICP), pour son aide et ses précieux conseils durant ces années de thèse. Je lui sais gré également d'avoir minutieusement examiné ce mémoire et d'avoir présidé le jury.

Je tiens à remercier Monsieur Ian Maddieson, Professeur à University of California - Los Angeles, et Monsieur Luc Steels, Professeur au Vrije Universiteit Brussels et Directeur du Artificial Intelligence Laboratory à Bruxelles, de m'avoir fait l'honneur de rapporter sur ce travail. Je les remercie d'autant plus pour leurs remarques pertinentes et pour leur commentaires constructifs.

Mes remerciements vont également à Monsieur Jean-Marie Hombert, Professeur à l'Université Louis Lumière - Lyon II et Directeur du Laboratoire Dynamique Du Langage, pour l'intérêt qu'il aura accordé à mes travaux et pour l'honneur qu'il m'a fait en acceptant de participer à ce jury.

J'exprime ma gratitude à Monsieur Louis-Jean Boë, responsable de recherche à l'ICP pour son encadrement et pour sa rigueur scientifique. Ses conseils et la confiance qu'il m'a accordée le long de ces années m'ont permis de mener à bien ce travail.

Je voudrais remercier également Monsieur Pierre Bessière, Chargé de Recherche au Laboratoire Leibniz, pour m'avoir accueilli au sein de son équipe et pour son soutien. Ces années de thèse m'ont permis d'apprécier ses qualités nombreuses, tant humaines que scientifiques.

C'est avec grand plaisir que j'exprime ma vive reconnaissance à Monsieur Rafael Laboissière, Chargé de Recherche à l'ICP, pour avoir suivi une partie importante de ce travail. J'ai beaucoup apprecié durant ces années ses multiples compétences. Je tiens à le remercier également pour sa disponibilité et pour les encouragements qu'il m'a continuellement prodigués. Enfin, j'ai été très sensible à son sens profond des relations humaines. Travailler avec Rafael, c'est super !

J'adresse également mes remerciements à Monsieur Philippe Jorrand, Directeur du

Laboratoire Leibniz, et à Monsieur Pierre Escudier, Directeur de l'ICP, pour m'avoir accueilli au sein de leur laboratoire.

Et un grand merci aussi à tous les membres de l'équipe Laplace du Laboratoire Leibniz pour avoir su créer une excellente ambiance. Je ne manquerai pas de remercier tous les membres de l'ICP - gare ou campus - pour les bons souvenirs que je vais garder de mon séjour au laboratoire. Un petit clin d'œil à Nathalie Vallée pour les longues et fructueuses discussions que nous avons eues.

Mes plus chaleureux remerciements vont à mes amis inconditionnels d'Alger, ainsi qu'à ceux que j'ai eu plaisir à connaître et à apprécier durant mon séjour à Grenoble et avec lesquels j'ai partagé bon nombre de moments inoubliables. Je ne voudrais surtout pas oublier de remercier mes complices dans la messagerie électronique ; je n'ose imaginer ce qu'aurait été la rédaction de ce mémoire si elle n'avait pas été ponctuée par des instants de détente et d'évasion que les «mails» me procuraient. Je ne me hasarderai pas à citer des noms de peur d'en oublier beaucoup.

Enfin, je suis conscient que cette thèse n'aurait jamais pu aboutir sans le soutien affectif et les encouragements de ma famille. Que mon père, ma mère, mon frère et chacune de mes deux sœurs trouvent ici l'expression de ma plus profonde gratitude. Merci d'avoir toujours cru en moi, et c'est tout naturel que je vous dédie à vous cette thèse !

Ahmed Réda BERRAH

Juin 1998

Table des matières

Table des figures

Liste des tableaux

Introduction

La question de l'origine des langues suscite bon nombre de recherches à l'heure actuelle. Cet intérêt relativement nouveau est en passe de faire changer de statut scientifique à cette problématique qui a été, il est bon de le rappeler, carrément exclue par la Société Linguistique de Paris de ses premiers statuts en 1866. Certes, ce thème reste spéculatif. Néanmoins, les scénarios proposés sont guidés par un grand nombre de données, collectées dans divers domaines tels l'archéologie et la paléontologie. Certaines études génétiques estiment que la période à laquelle le premier mot a été prononcé par l'homme, remonte à 100 000 ans avant notre ère. Le linguiste Ruhlen pense que ce premier mot fut le monosyllabe *tik*, qui aurait désigné le doigt. Suite à des déformations progressives, ce mot aurait évolué en *digit* en latin, puis en *toe* en anglais.

Actuellement, la communauté linguistique classe les langues en quelque 200 familles. Quel que soit le système linguistique considéré, la parole peut être considérée, d'un certain point de vue, comme une succession de gestes de fermeture et d'ouverture du conduit vocal : les syllabes composées de consonnes *C* et de voyelles *V*. Majoritairement, les syllabes sont de type CV bien qu'il existe des structures beaucoup plus complexes (CCCVCC) (Maddieson et Precoda 1990). À partir des années 1930, les linguistes ont établi les inventaires des ensembles de consonnes et de voyelles dans lesquels les langues puisent leurs phonèmes. Les descriptions phonologiques d'échantillons représentatifs ont été dressées et l'inventaire obtenu est constitué de 920 phonèmes possibles dont 654 consonnes, 177 voyelles et 89 diphtongues (Maddieson et Precoda 1990). La plupart des langues utilisent 5 voyelles et une vingtaine de consonnes. Or, la combinatoire C_{177}^{5} pour V et C_{654}^{20} pour C est très élevée, et pourtant, les typologies menées depuis Troubetskoy (1939) jusqu'à Sedlak (1969), Crothers (1978), Maddieson (1986) et Vallée (1994) montrent que les systèmes phonologiques ne sont pas constitués de façon arbitraire, et qu'ils relèvent, en revanche, de principes d'optimisation dont les paramètres ne sont pas tous connus. Il est possible, de ce fait, de raisonner en termes de tendances universelles des structures sonores des langues du monde.

La justification de ces aspects universels des systèmes phonologiques a d'abord été recherchée dans l'étude des changements phonétiques. Les linguistes de la première partie du XIX^e siècle expliquaient l'existence et la régularité des changements phonétiques par une tendance générale de l'organisme vers un moindre effort. À ce principe, Roudet (1910) ajoute *«la tendance à maintenir les distinctions phonétiques nécessaires à l'intelligence du langage»*. Dans ce même courant d'idées, Martinet (1970) énonce ainsi le mécanisme général d'évolution des systèmes de sons : *«les unités distinctives, les phonèmes qui coexistent, tendront naturellement à utiliser au mieux les latitudes que leur offrent les organes dits de parole ; ils tendront à être aussi aussi distants de leurs voisins qu'il est loisible pour eux de l'être tout en restant faciles à articuler et faciles à percevoir»*.

Plusieurs modèles de prédictions ont été proposés à partir de ces hypothèses (Stevens 1972, Liljencrants et Lindblom 1972, Lindblom, MacNeilage et Studdert-Kennedy 1984, Lindblom 1986, Schwartz, Boë, Vallée et Abry 1997*b*). Ces modèles permettent d'une part de prévoir la structure phonologique des systèmes, et d'autre part d'expliquer, d'une manière générale, pourquoi certains sons se sont imposés dans les langues du monde. Par contre, ils n'abordent ni les processus évolutifs ni les aspects liés à l'interaction entre locuteurs qui sont à l'origine de ces tendances universelles.

Ces réflexions constituent le point de départ des travaux présentés dans ce mémoire. Plus précisément, nous nous intéressons aux modèles de prédictions des systèmes sonores du langage, et nous désirons élaborer une modèle qui détermine, cette fois, l'émergence des inventaires phonologiques à partir de principes d'interaction entre agents communicants. Cette approche est novatrice dans le domaine et s'inscrit dans le champ ouvert récemment par des travaux tels ceux de Steels (1997*b*) dans le cadre de la Vie Artificielle.

Le **chapitre I** sera dédié à une petite revue préliminaire de questions sur l'émergence des structures sonores. Nous présenterons les deux approches visant à expliquer les inventaires phonologiques afin de positionner notre étude dans cette voie de recherche.

Nous expliciterons dans le **chapitre II**, l'architecture et le mode de fonctionnement du modèle d'interactions que nous proposons.

Le **chapitre III** sera consacré à la présentation des résultats du modèle proposé au chapitre II, appliqué à la prédiction des systèmes vocaliques. Il contiendra une large discussion sur ces prédictions, ainsi que sur celles obtenues avec les différentes variantes testées du modèle.

Enfin, nous détaillerons et discuterons dans le **chapitre IV** les applications de notre modèle visant à expliquer l'apparition de traits supplémentaires dans les larges systèmes vocaliques, en intégrant le principe du MUAF (Ohala 1980).

Chapitre I

L'émergence des structures sonores

Résumé

Avant d'aborder la description du modèle de prédiction (objet de cette étude), il nous semble opportun de nous attarder quelque peu sur la problématique des tendances universelles des langues du monde.

Dans ce chapitre, nous énoncerons d'abord les deux courants antagonistes en linguistique, visant à justifier les régularités des systèmes sonores des langues. Nous présenterons brièvement ensuite la base de données phonologiques UPSID, ainsi qu'une typologie des systèmes vocaliques de cette base. UPSID nous servira de référence pour nos modélisations.

Par ailleurs, nous décrirons quelques travaux effectués conformément à chacune des deux approches (globale et locale) expliquant la structure des inventaires phonologiques. Puis nous trancherons, à partir de ces descriptions, sur l'approche que nous retiendrons dans le cadre de ce travail de thèse.

Nous délimiterons, en guise de conclusion, le cadre de cette étude.

I.1 Problématique

Seule, parmi les espèces vivantes, l'espèce humaine possède cette faculté linguistique qui permet à tous les peuples ou groupes ethniques de mettre en œuvre un outil de communication à usage social et culturel : le langage. Le caractère universel des langues en tant qu'outil de communication caractérise cette activité humaine fondamentale.

En effet, un non spécialiste ne peut être que saisi par le nombre de langues parlées dans le monde. Aujourd'hui, les quelques six milliards d'habitants de la Terre parlent entre 4 000 et 6 000 langues distinctes. Le nombre de langues est approximatif et varie selon les études et les critères considérés. Pour sa part, Ruhlen (1987) le chiffre à environ 5 000.

Le système sonore d'une langue peut être caractérisé par ses phonèmes, leurs réalisations ainsi que la manière de les combiner entre eux. Toutes les langues fonctionnent à partir d'un système phonologique qui peut être subdivisé en deux types de sous-systèmes : les systèmes vocaliques et les systèmes consonantiques. Mais toutes les langues ne disposent pas du même nombre et des mêmes phonèmes.

Grâce à la constitution et à l'observation d'inventaires des systèmes de sons des langues du monde, il a été constaté que, sous cette diversité apparente, se cachait une certaine régularité. Il a été noté de plus que la variété des langues naturelles n'impliquait pas pour autant la diversité des systèmes de sons existants. En effet, les premiers classements proposés pour les systèmes de sons ont permis de mettre en évidence des tendances de ces systèmes : une sorte de restriction limite les choix parmi un ensemble de sons pour former les systèmes phonologiques des langues du monde. L'existence de tendances universelles est aujourd'hui admise, c'est-à-dire la généralisation de phonèmes ou de propriétés des langues du monde. Avec la mise en évidence de ces régularités, un intérêt s'est développé pour la recherche des tendances universelles du langage, ainsi que pour la définition des capacités cognitives du cerveau humain en ce qui concerne le langage. Actuellement, ce n'est plus la mise en évidence des tendances universelles qui suscite des questionnements dans ce domaine : la théorie qui expliquerait ces tendances universelles est maintenant devenue une véritable préoccupation de recherche (Vallée 1994).

Pour justifier ces régularités, les linguistes, du courant générativiste estiment que les tendances universelles sont axiomatiquement imposées sans aucun critère de naturalité (Chomsky et Halle 1968). Chomsky (1986) pense que la syntaxe[1] reflète une aptitude innée très complexe : n'importe quel enfant peut apprendre n'importe laquelle des langues humaines, indépendamment de la langue de ses géniteurs, sans faire des fautes de grammaire auxquelles on pourrait s'attendre si cette aptitude n'était pas innée. Selon

[1]cf. annexe A pour la définition de la «syntaxe».

Chomsky et son école, le langage humain serait apparu soudainement, dès lors qu'une modification génétique aurait coïncidé avec l'apparition indépendante d'autres caractères, comme le contrôle conscient de l'émission des sons. Ceci suppose que le langage est basé sur une sorte de grammaire universelle. Selon cette théorie, les êtres humains naissent tous munis du même ensemble de traits distinctifs[2]. Puis cet ensemble est filtré de façon à ne garder que les traits distinctifs du lexique maternel. Néanmoins, il n'est pas nécessaire d'épouser la thèse chomskyenne d'une grammaire universelle innée pour reconnaître que l'apprentissage d'une première langue réclame une prédisposition génétique identique pour tous les hommes (Victorri 1997).

Pour notre part, nous adopterons une position différente de celle défendue par Chomsky : du fait que tous les hommes possèdent la faculté du langage et disposent tous des mêmes potentiels de production et de perception des sons, les tendances universelles du langage humain sont les conséquences de facteurs naturels et de contraintes biologiques, et de ce fait sont prévisibles. La recherche d'explications sur la généralisation des tendances universelles dans les langues est orientée vers les limites du système biologique humain ainsi que vers l'optimisation des processus de production et de perception des sons. Selon cette hypothèse, les facteurs qui entrent en jeu dans la formation des systèmes phonologiques ont leur origine dans les processus d'interaction entre locuteurs et auditeurs. En d'autres termes, le système d'une langue naturelle serait le résultat d'adaptations, à caractère fonctionnel, liées aux contraintes que s'imposent mutuellement locuteurs et auditeurs, utilisateurs de ce langage.

Les contraintes peuvent être mieux précisées en reprenant la classification de Maddieson (1991) qui propose trois types de limites du système biologique humain en ce qui concerne la parole :

1. les limites de l'amplitude et de la vitesse du mouvement des articulateurs du conduit vocal,
2. les limites de la capacité du système auditif à percevoir deux sons comme qualitativement distincts (seuils de perception et de catégorisation),
3. les limites de capacité de stockage des connaissances linguistiques dans le cerveau.

La constitution d'un échantillon représentatif de l'inventaire des langues du monde est indispensable si l'on souhaite examiner le contenu et la structure des systèmes phonologiques des langues parlées sur la surface du globe.

Ces langues ont été regroupées selon des degrés de similarité au niveau lexical, phonologique, sémantique ou syntaxique. Un groupe rassemble des langues génétiquement plus proches les unes des autres. Ce procédé est appelé *regroupement génétique* des langues du monde (Greenberg 1966, Ruhlen 1987).

Nous allons présenter succinctement, maintenant, une base de données phonolo-

[2]cf. annexe A pour la définition d'un «trait distinctif».

giques particulière regroupant bon nombre des langues du monde : la base UPSID élaborée par Maddieson (1986).

I.2 Une base de données phonologiques : UPSID

La base UPSID (*UCLA Phonological Segment Inventory Database*) élaborée par Maddieson (1986) est un échantillon de systèmes phonologiques actuels des langues du monde. UPSID se veut être représentative d'un point de vue génétique et géographique d'au moins 5% de l'ensemble des langues du monde. Dix-sept familles plus trois langues isolées, réparties sur les cinq continents y sont représentées. Au moins un dialecte par groupe de langues (sous-famille) est inclus dans l'inventaire, sur la base d'une distance génétique de séparation des langues d'au moins 1 500 ans. Il s'agit d'une marge d'assurance pour l'indépendance de l'évolution des langues issues d'une même souche. Une telle mesure permet d'écarter des similarités liées à la parenté génétique qui biaiserait la représentativité de l'échantillonnage.

La base UPSID constitue un précieux matériau pour l'étude des systèmes phonologiques des langues du monde, en permettant, d'une part, de dresser un inventaire des types de systèmes, et d'exploiter, d'autre part, des tendances universelles sur la fréquence d'occurrences de phonèmes particuliers. Toutefois, la base UPSID peut présenter certaines faiblesses, notamment :

- la description des sons repose sur une notation auditive : elle dépend donc du transcripteur,
- problèmes au niveau du choix des allophones[3] : l'allophone le plus commun, le plus utilisé en prononciation isolée est généralement choisi lors des transcriptions.

Malgré ces quelques critiques, UPSID continue à être actuellement la seule source d'étude de systèmes phonologiques des langues du monde permettant de dresser un inventaire des types de systèmes.

Des taxinomies des systèmes vocaliques ont été élaborées à partir d'une première version de la base UPSID, contenant 317 systèmes. Puis, des tendances fortes ont émergé de ces taxinomies grâce à une étude fine des statistiques d'UPSID_{317}. Ceci a conduit Vallée (1994) à proposer, à partir d'UPSID_{317}, une typologie des systèmes vocaliques, ainsi qu'un certain nombre de propositions sur leur organisation que nous décrivons ci-après.

[3] cf. annexe A pour la définition du terme «allophone».

I.3 Typolologie des systèmes vocaliques

Les systèmes vocaliques des langues du monde[4] recrutent de 3 à 24 voyelles (Maddieson 1986, Vallée 1994). Les deux tiers de ces systèmes comportent entre 5 et 7 voyelles.

La figure I.1 présente les 37 symboles vocaliques de la base $UPSID_{317}$, alors que la figure I.2 représente un classement, par ordre décroissant de fréquence, des systèmes vocaliques selon le nombre de voyelles d'après la typologie élaborée par Vallée (1997).

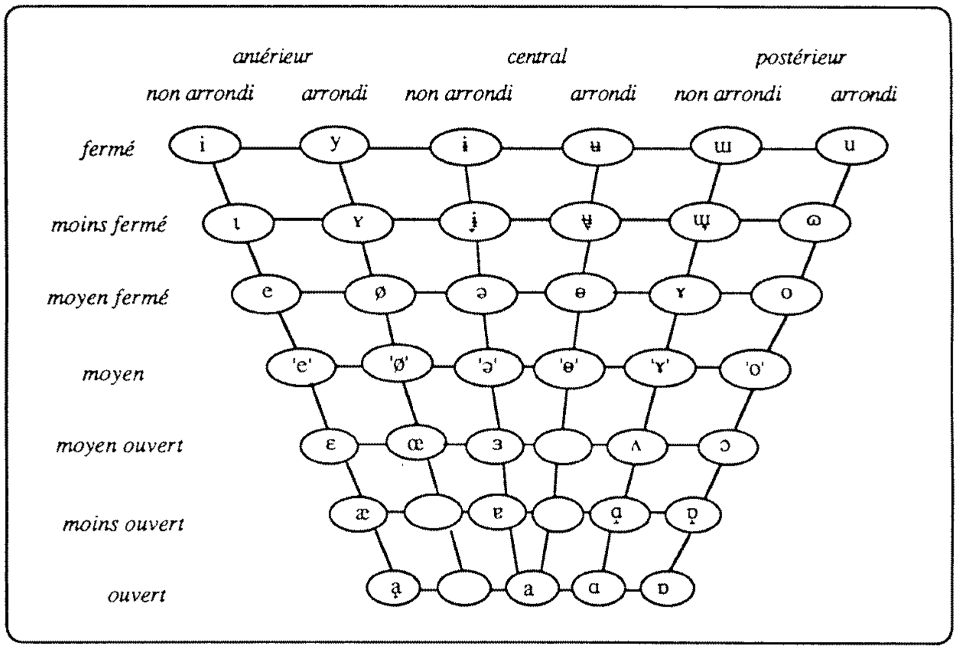


FIG. I.1: Description des 37 symboles vocaliques d'$UPSID_{317}$

Notons que les systèmes vocaliques représentés dans la figure I.2 ont été obtenus suite à une opération de *regroupement* de certains systèmes selon des critères bien établis (Vallée 1994). Le regroupement a pour but de fournir les principales tendances des systèmes vocaliques des langues du monde. La connaissance de ces tendances est une condition sine qua non à la validation des modèles de prédiction que nous aborderons ultérieurement.

La figure I.3 schématise quelques équivalences admises par Vallée (1994) afin de regrouper les systèmes vocaliques.

L'observation des systèmes les plus fréquents (cf. figure I.2) met en évidence un ordre d'apparition des voyelles dans les systèmes. On retrouves les 3 voyelles vedettes

[4]Désormais, nous entendons par *langues du monde* les langues représentés dans $UPSID_{317}$.

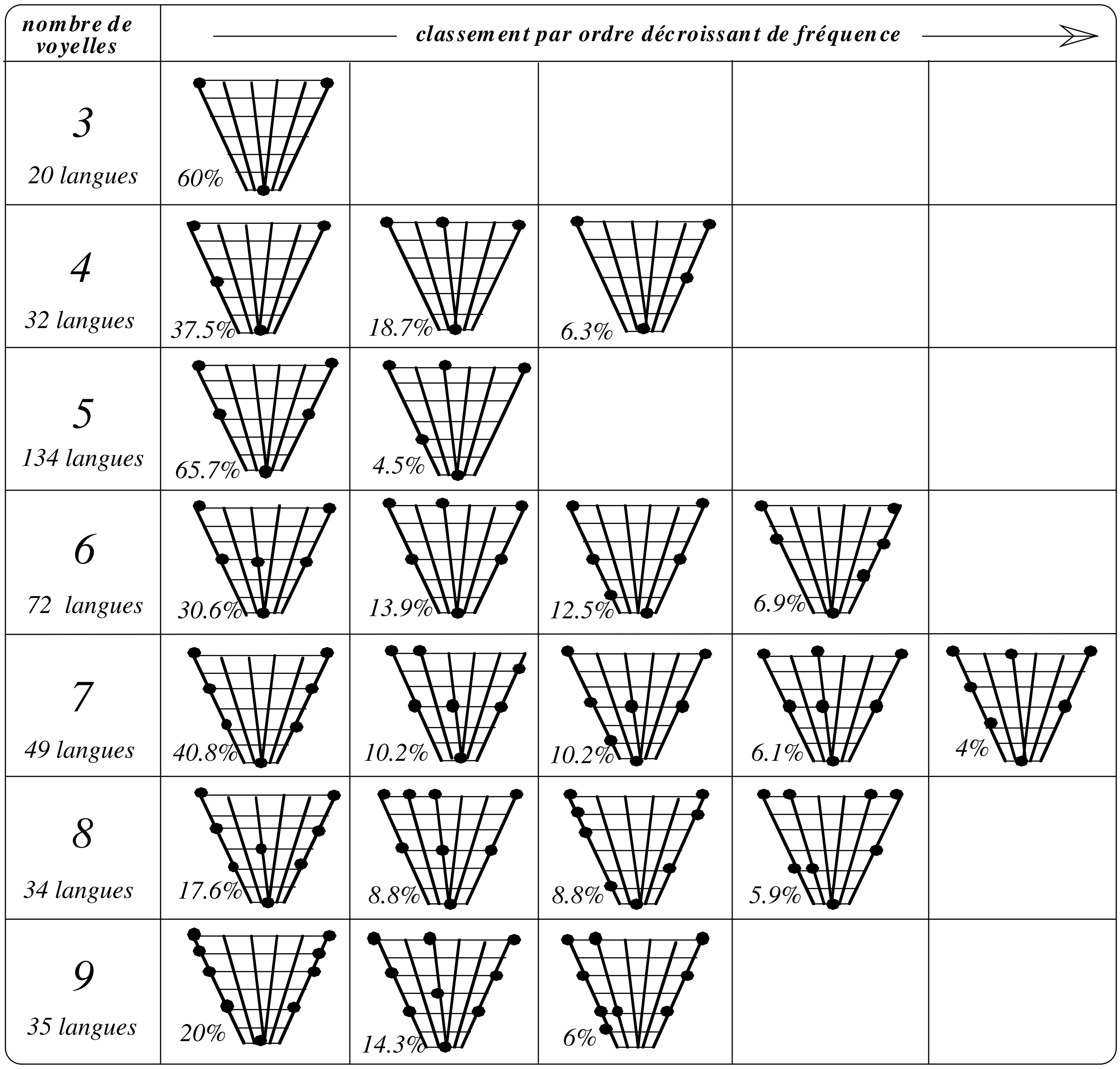


FIG. I.2: Hiérarchie universelle des systèmes vocaliques (d'après Vallée et al. (1997)).

/i a u/ dans pas moins de 97 % des langues. Une voyelle antérieure (/ˈeˈ/) s'y ajoute pour le système à 4 le plus représenté. Le système /i ˈeˈ a ˈoˈ u/ est le plus commun dans les systèmes à 5 voyelles. Les systèmes périphériques sont les plus fréquents pour les systèmes à 7 et à 9 voyelles, alors que les systèmes à 6 et à 8 voyelles les plus fréquents s'assimilent respectivement à des systèmes à 5 et à 7 voyelles ajoutés d'une voyelle centrale notée /ə/.

Des études typologiques ont montré que 9 était généralement le nombre maximum de timbres vocaliques de base que peut contenir un système (Vallée 1994). Autrement

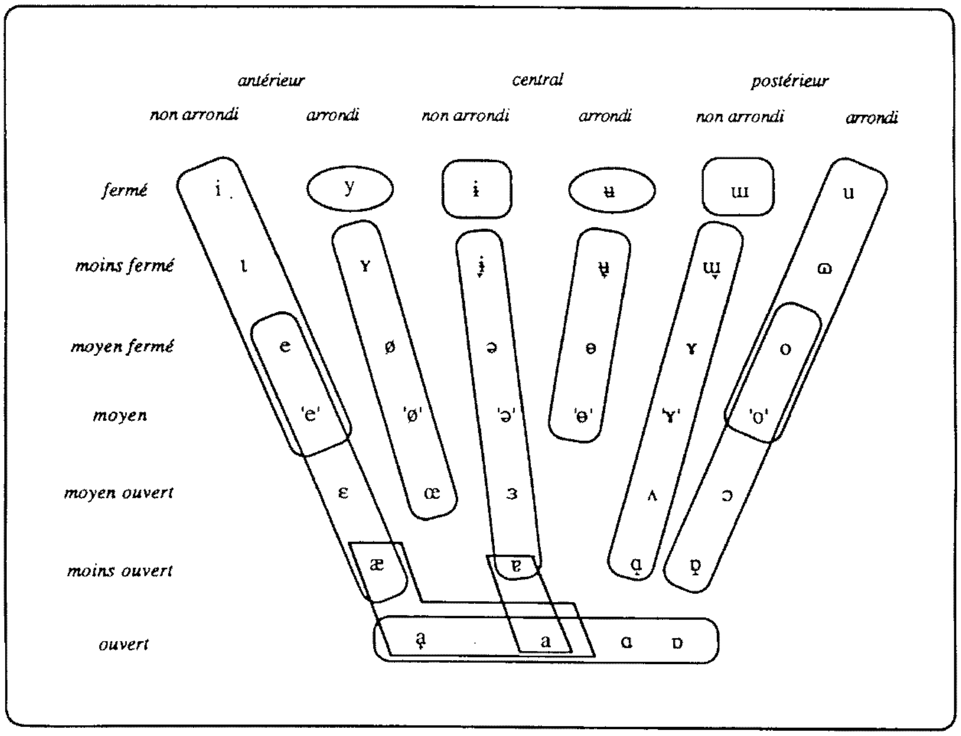


FIG. I.3: Équivalence admises entre phonèmes pour regrouper les systèmes vocaliques (d'après Vallée (1994)).

dit, les systèmes de plus de 9 voyelles ne développent pas de nouveaux timbres afin d'augmenter leur inventaire mais utilisent une autre dimension telle que la nasalité ou la quantité. Nous reviendrons avec plus de détails sur les particularités des larges systèmes vocaliques, ainsi que sur le concept de timbre de base et de dimension supplémentaire au cours du chapitre IV.

Le problème de l'explication des inventaires phonologiques des langues du monde peut être abordé selon deux approches différentes :

- une approche globale (classique en Sciences Phonétiques) fondée sur un compromis entre des critères d'efficacité perceptive et des critères d'économie articulatoire,

- une approche locale, peu usuelle, qui repose sur l'évolution et les interactions au sein d'une société d'agents communicants.

Nous allons étudier, dans la partie restante de ce chapitre, les deux approches afin de retenir celle que nous allons adopter.

I.4 Approche globale

L'interaction locuteur/auditeur est supposée être basée sur deux principes majeurs qui sont :

- la simplicité articulatoire (appelée aussi critère du moindre effort ou principe de l'économie articulatoire),
- la distinctivité perceptive (principe du contraste) : les unités fonctionnelles, consonnes ou voyelles, sont primairement caractérisées par leur fonction distinctive dans toutes les langues.

Dans le processus de communication parlée, des moyens sont a priori mis en œuvre par le locuteur pour, d'une part, minimiser l'effort articulatoire et donc le coût de production, et pour, d'autre part, maintenir un contraste perceptif suffisant entre les phonèmes. Ce contraste est nécessaire si l'on souhaite éviter toute confusion dans l'interprétation, augmentant ainsi les chances de bonne perception du message par l'auditeur, et ce, dans toutes les situations de communication. Les aspects universels des systèmes de sons sont, donc selon ces hypothèses, les conséquences des propriétés anthropomorphiques du système de production et du système de perception de la parole.

Deux théories complémentaires ont alimenté le débat autour de ces hypothèses dans les années soixante dix. Il s'agit de la *Théorie Quantique* proposée par Stevens (1972, 1989) et de la *Théorie de la Dispersion* énoncée par Liljencrants & Lindblom (1972). Ce sont deux courants de recherche orientés vers la notion d'efficacité fonctionnelle, notion à la base de la fonction distinctive qui forme le rôle caractéristique des sons du langage.

Dans un premier temps, passons en revue les principes de ces deux théories avant de nous attarder sur la *Théorie de la Dispersion-Focalisation* (Schwartz et al. 1997*b*) née d'une amélioration de la théorie de la Dispersion.

I.4.1 La théorie Quantique

Le fondement de la théorie Quantique est que la formation des systèmes de sons a des origines fonctionnelles qui se retrouveraient dans les unités (voyelles ou consonnes) et leurs traits distinctifs. En ne considérant pas l'existence des phonèmes comme axiomatique, mais comme étant tout a fait démontrable, et en leur cherchant donc un fondement, Stevens (1972, 1989) se propose d'expliquer, par la théorie Quantique, la raison d'être des phonèmes en tant qu'unités discrètes et hiérarchiquement organisées dans les langues. L'hypothèse de base est que le langage est adapté à des systèmes biologiques humains spécialisés pour la communication orale. Il s'agit du système articulatoire (pour la production des sons) et du système perceptif (perception humaine des sons). Il est donc supposé que la sélection des sons dans une langue est gouvernée par des contraintes physiques et physiologiques. Les sons du langage ne sont qu'une partie du potentiel humain de production des sons. Ces sons sont sélectionnés par des contraintes et forment ce qui est appelé *l'espace des sons possibles* qui est parfaitement délimitable. Les langues puisent leur inventaire de sons à l'intérieur de cet espace. La notion quantique du langage traduit la notion de discrétisation ou de catégorisation des sons dans cet espace.

Le postulat de base de cette théorie est la non linéarité du passage de l'articulatoire à l'acoustique (cf. figure I.4). D'où l'existence de régions de l'espace articulatoire pour lesquelles :

- le résultat acoustique est *stable* (régions I et III sur la figure),
- une petite variation du paramètre articulatoire implique une variation abrupte du paramètre acoustique (région II).

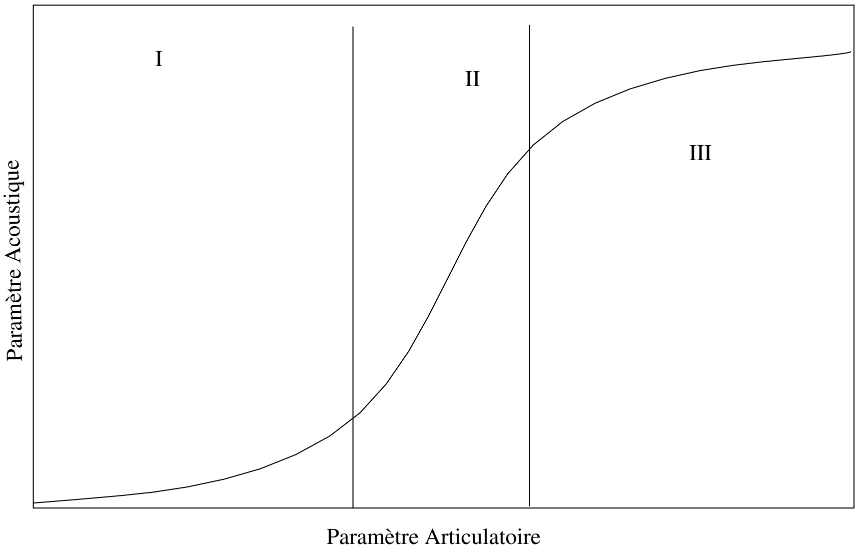


FIG. I.4: Non linéarité du passage d'un paramètre de commande articulatoire à un paramètre de sortie acoustique dans le postulat de base de la Théorie Quantique (d'après Stevens (1972)).

Dans les régions I et III, les phonèmes ne requièrent qu'une faible précision articulatoire. Ce serait un critère de sélection des voyelles et des consonnes pour les systèmes. La région II affecte la perception : les sons ne peuvent y être produits qu'avec une grande précision articulatoire. Cette région est une région de discontinuité dans laquelle se situe le seuil acoustique perceptif de deux réalisations sonores. Selon Stevens, la région II est ainsi la région de discrimination des sons : le changement acoustique rapide qui s'y produit provoque, chez les auditeurs, un changement net de réponse aux stimuli puisqu'ils sont différemment perçus. Cette région expliquerait la distribution entre les sons d'une même langue, et par suite, la structure phonologique des langues par le biais des traits distinctifs.

Stevens introduit la notion de stabilité dont la fonction serait accrue dans le processus auditif. Un phonème est d'autant plus fréquent dans les langues du monde qu'il est stable. Ce phonème sera qualifié de *quantique*. Stevens considère qu'un phonème acoustiquement stable possède une forme perceptive plus riche de par sa stabilité auditive et est, par conséquent, plus aisément discriminable.

Analyse critique

Les arguments les plus convaincants fournis par la théorie Quantique portent sur l'articulation des structures sonores. À ce propos, Lindblom & Engstrand (1989) énoncent : «*the essence of an intuition that underlies the QT is that continuous articulatory motion produces clear discrete acoustic segments*». Cependant, les arguments avancés par Stevens concernant le lieu d'articulation sont plus controversés[5] (Abry, Boë et Schwartz 1989).

D'autre part, il apparaît que la règle selon laquelle les voyelles quantiques sont les plus rencontrées dans les langues du monde n'est pas vérifiée. Vallée (1994) remarque que la voyelle /y/ devrait être, au même titre que la voyelle /i/, très fréquente dans les systèmes, alors qu'en réalité, /y/ est rencontrée approximativement dix fois moins que /i/.

I.4.2 La théorie de la Dispersion

À l'instar de la théorie Quantique, la théorie de la Dispersion a pour objectif d'expliquer la structure phonologique des systèmes vocaliques et consonantiques. Elle a été initiée par Liljencrants & Lindblom (1972).

Étant donné que les systèmes vocaliques ne présentent qu'un ensemble limité de combinaisons possibles de voyelles, Liljencrants & Lindblom (1972) ont étudié les systèmes vocaliques d'un certain nombre de langues afin de prédire leur structure phonologique, et d'analyser la manière dont les voyelles se distribuent dans les différents systèmes. Les typologies sur les systèmes de sons des langues naturelles montrent qu'il existe des voyelles universellement favorisées. Le postulat de base de la théorie est que les sons favorisés sont ceux qui maintiennent entre eux, à l'intérieur des systèmes, une certaine distance reliable à la notion de discrimination : les systèmes sont organisés sur la base de contrastes sonores si bien que les voyelles qui les composent sont fortement discriminables entre elles. La prédiction de l'existence de certaines voyelles ne peut donc se faire que par rapport à d'autres voyelles.

[5]On trouvera un éventail de critiques dans le numéro spécial du volume **17** du *Journal of Phonetics* de 1989 *"On the Quantal nature of speech"*.

À la différence de la théorie Quantique, les propriétés du contraste, dans la théorie de la Dispersion sont déterminées par les relations que les voyelles entretiennent entre elles à l'intérieur du système, et non pas par les caractéristiques acoustiques et articulatoires intrinsèques à chacune d'elles.

Le critère de dispersion visant à maximiser les distances perceptives entre les voyelles constitue le principe du *contraste maximal*. Le principe du contraste maximal permet de générer, par simulation, dans un espace de réalisation, un système donné où les voyelles s'organisent pour présenter un maximum de distinctivité globale.

Le modèle de prédiction de Liljencrants & Lindblom (1972) repose sur un espace vocalique bidimensionnel issu du modèle articulatoire réalisé par Lindblom & Sundberg (1971). Dans ce modèle de prédiction, le système vocalique optimal minimise l'expression suivante :

$$G = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\frac{1}{d_{ij}})^2 \tag{I.1}$$

n étant le nombre total de voyelles du système.

La distance perceptive d_{ij} entre deux voyelles i et j est définie par :

$$d_{ij} = ((M_1^j - M_1^i)^2 + (M_2^{'j} - M_2^{'i})^2)^{1/2} \tag{I.2}$$

M_1 et M_2' représentent les paramètres perceptifs d'une voyelle obtenus en convertissant les formants F_1 et F'_2 exprimés en Hertz en l'unité auditive *Mels*. Le second formant effectif F'_2 est lune pondération de F_2 et F_3.

Le principe de dispersion permet de trouver les valeurs phonétiques des systèmes vocaliques en fonction du nombre de voyelles des systèmes.

Les résultats du modèle de dispersion (cf. figure I.5) sont proches des données typologiques et sont corrects pour un nombre total de voyelles inférieur ou égal à 6. En effet, les prédictions de systèmes de plus de 6 voyelles présentent un trop grand nombre de hautes /ɨʉɯ/ entre /i/ et /u/. Ce phénomène n'est pas observé dans les tendances universelles des langues du monde.

Depuis cette première version, le modèle n'a cessé d'être complété et amélioré, en appliquant au principe de base qu'est la dispersion maximale, les résultats issus du développement des connaissances sur le système de perception humain.

En 1975, Lindblom (1975) a ajouté une pondération qui fait augmenter la contribution du premier formant F_1. Ces travaux montrent l'existence, dans les systèmes vocaliques, d'une suprématie des voyelles aussi bien entre /i/ et /a/ qu'entre /a/ et /u/

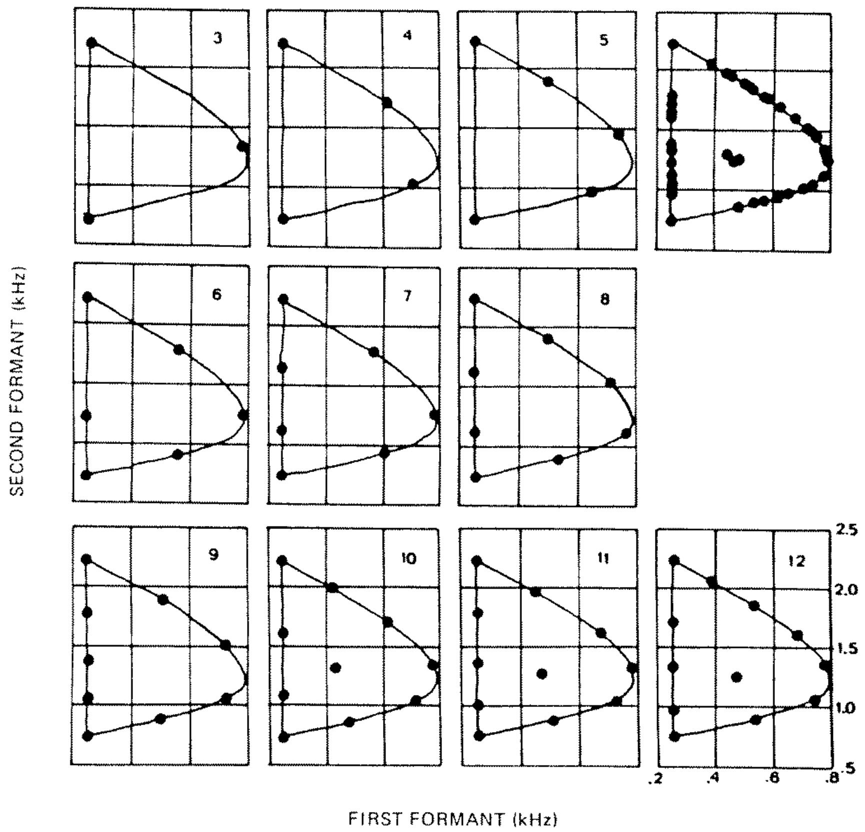


FIG. I.5: Résultats des prédictions de systèmes de 3 à 12 voyelles représentés sur un plan F_1/F_2 (d'après Liljencrants & Lindblom (1972)).

tandis que les voyelles entre /i/ et /u/ sont rarement rencontrées. Lindblom argumente que les voyelles exploiteraient plus le premier formant F_1 que les autres formants, du moment que, F_1 est plus intense et donc statistiquement plus résistant aux bruits d'après les préceptes de la théorie acoustique élémentaire. Cette primauté sera confirmée en 1985 par les travaux expérimentaux de Lindblom & Lubker. Ces travaux ont en effet abouti à ce que la sensibilité proprioceptive semble être plus développée pour les mouvements haut et bas de la mâchoire que pour les mouvements avant et arrière de la langue. Les auteurs déduisent que F_1 serait plus exploité dans la différenciation perceptive que les autres formants (Lindblom et Lubker 1985).

Crothers (1978), Bladon & Lindblom (1981), Lindblom (1986) ont proposé plusieurs révisions du modèle permettant ainsi d'obtenir des prédictions plus conformes aux typologies des langues du monde. Lindblom (1986) introduit la notion de *contraste perceptif suffisant* qui, selon l'auteur, est mieux appropriée pour expliquer le structure phonologique des systèmes vocaliques : la fréquence des voyelles /i/, /a/ et /u/ augmente avec le nombre de voyelles ce qui laisse supposer qu'une distance maximale entre les éléments du système n'est plus indispensable à la fonction distinctive des sons. Il est plutôt question de contraste perceptif suffisamment nécessaire pour la distinctivité.

Jusque là, les facteurs articulatoires ont été exclus de la théorie de la Dispersion vocalique. Lindblom (1986) propose, en perspective de ses travaux, de rajouter, à la composante perceptive, une composante basée sur la mémoire et une autre sensori-motrice (articulatoire). À ce propos, Lindblom (1986, p. 38) énonce : « *it make sense, in fact, to talk about sufficient contrast also in the contexts of memory retrieval operations and the feedback control of speech gestures. If we were to look at the optimization of vowel systems from this broader perspective, an optimal system might be one that meets not only perceptual but also memory-based and sensori-motor conditions of distinctiveness. For instance, how, in that case, could the detailed geometry of the articulatory space influence the design of sound inventories.*»

Lindblom (1990) approfondit le travail sur l'explication des tendances universelles en cherchant à augmenter la participation des intérêts du locuteur dans le modèle de prédiction. Le critère d'optimisation est révisé afin de prendre en compte la simplicité articulatoire. Cette adaptation simultanée entre les intérêts du locuteur et ceux de l'auditeur est à la base d'une nouvelle théorie : celle de la *dispersion adaptative* (Lindblom 1990).

Pour minimiser la somme des coûts articulatoires tout en maximisant les distances perceptives, l'auteur propose l'équation suivante :

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\frac{A_{ij}}{d_{ij}})^2 \tag{I.3}$$

avec :

d_{ij} : distance perceptive entre les phonèmes i et j,

A_{ij} : coût articulatoire entre les phonèmes i et j.[6]

Les voyelles préférentielles dans les systèmes sont celles qui, au niveau articulatoire, sont produites avec une déviation minimale par rapport à une position supposée neutre, tout en conservant une distinction perceptive maximale.

Analyse critique

La théorie de la Dispersion est toujours en quête de nouveaux facteurs susceptibles d'améliorer les prédictions. Le premier modèle reprenant les fondements de cette théorie (Liljencrants et Lindblom 1972) permet de générer un système donné par simulation, dans un espace de réalisation 2D : les voyelles s'organisent pour présenter un maximum

[6] L'auteur n'a pas précisé la manière de calculer ce coût articulatoire.

de distinctivité acoustico-perceptive globale, c'est-à-dire des distances maximales entre chaque paire de voyelles à l'intérieur du système. Une sérieuse lacune des premières simulations de ce modèle reside dans le fait qu'il génère un important nombre de voyelles fermées pour les systèmes de plus de 7 voyelles. Ce problème a amené Lindblom (1975) à proposer de retoucher l'espace vocalique acoustique en augmentant le poids de F_1 au dépens de F_2 ou de F'_2.

Le second problème se rapporte aux voyelles hautes non périphériques. L'analyse de la base $UPSID_{317}$ révèle que, dans les systèmes à configuration non périphérique, les voyelles antérieures arrondies sont préférées aux voyelles postérieures non arrondies. Autrement dit, après la série /i ɨ u/, la série /i y u/ constitue un meilleur candidat que la série /i ɯ u/ dans les langues du monde. Or cette série est défavorisée par le critère de dispersion acoustique.

Pour pallier ces imperfections, Schwartz et al. (1997*b*) ont élaboré un modèle de prédiction basé sur des principes perceptifs extra-linguistiques et proposent :

- d'augmenter, grâce à un paramètre pondérateur, le poids de F_1 ; ainsi, les voyelles simulées viendront se positionner à la périphérie de l'espace vocalique plutôt qu'entre /i/ et /u/,
- d'associer à la théorie du contraste maximal un nouveau principe qui permette la prédiction des voyelles antérieures arrondies (telles le /y/) ; ce principe est la *focalisation*.

Ces principes forgent les bases de la Théorie de *la Dispersion-Focalisation* dont les fondements sont retracés ci-dessous.

I.4.3 La théorie de la Dispersion-Focalisation

La théorie la Dispersion-Focalisation (TDF) reprend les principes de la théorie de la Dispersion, mais lui apporte toutefois des modifications. Schwartz et al. (1997*b*) introduisent un paramètre, dans le terme de dispersion, susceptible de contrôler les poids qui s'appliquent d'un côté à F_1 et d'un autre côté aux formants hauts. Ce paramètre doit ainsi permettre de résoudre le problème des voyelles périphériques insuffisantes dans les simulations. Un coût perceptif local (la focalisation) est ensuite intégré par addition. Ce coût est contrôlé par un second paramètre qui doit améliorer le cas des voyelles antérieures arrondies mal prédites par la théorie de la Dispersion.

Les voyelles sont décrites dans le modèle à quatre formants (F_1, F_2, F_3, F_4), avec F_4 fixé à 3560 Hz[7]. Toutes les valeurs sont exprimées en Bark qui est une échelle perceptive

[7] F_4 varie peu pour les voyelles. Il a donc été considéré constant pour simplifier les calculs.

calculée suivant la formule proposée par Schroeder, Atal et Hall (1979) :

$$F_{Bark} = 7 \text{ argsh}(F_{Hz}/650) \tag{I.4}$$

Puis, le second formant effectif F_2', connu pour représenter un raisonnable équivalent de F_2 et des hauts formants dans la détermination de la qualité vocalique (Schwartz et al. 1997*b*), est calculé à partir de F_2, F_3 et F_4, conformément au modèle basé sur le concept du centre de gravité et sur une distance critique de 3.5 Barks (Mantakas, Schwartz et Escudier 1986).

L'algorithme de calcul de F_2' est décrit sur la figure I.6.

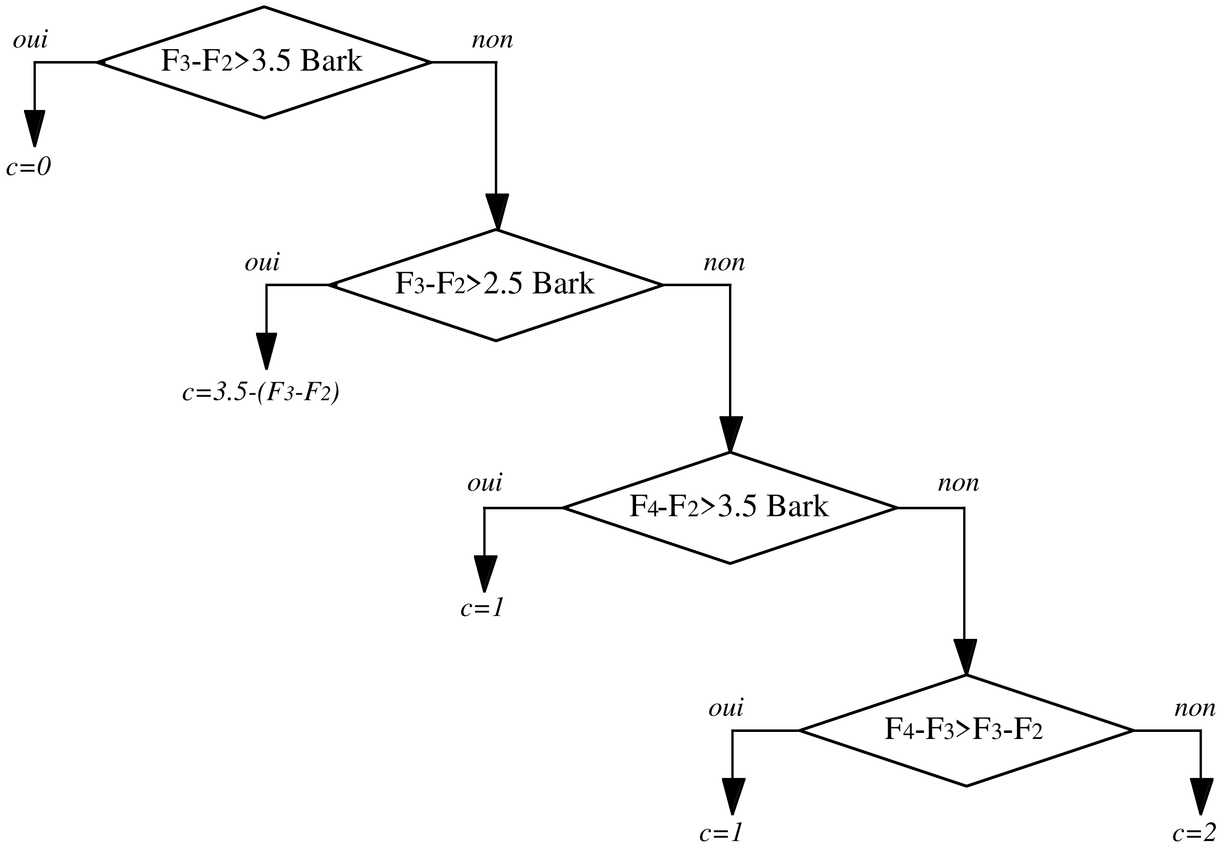


FIG. I.6: Algorithme de calcul de F'$_2$. $F_2' = (c_2F_2 + c_3F_3 + c_4F_4)/(c_2 + c_3 + c_4)$ avec c_2, c_3 et c_4 définis par : si $c \leq 1$ alors $c_2 = 1,\ c_3 = 0.5c,\ c_4 = 0$ et si $c = 2$ alors $c_2 = 0,\ c_3 = 1,\ c_4 = 0.5$.

Le système optimal selon la théorie TDF est celui qui minimise la fonction d'énergie suivante :

$$E_{DF} = E_D + E_F \tag{I.5}$$

avec :

E_D : terme de dispersion

E_F : terme de focalisation.

Le terme de dispersion est décrit par :

$$E_D = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\frac{1}{d_{ij}})^2 \tag{I.6}$$

où :

$$d_{ij} = ((F_1^j - F_1^i)^2 + \lambda^2 (F_2^{'j} - F_2^{'i})^2)^{1/2} \tag{I.7}$$

Schwartz et al. (1997*b*) ont intégré la proposition de Lindblom (1975) qui consiste à dilater la dimension F_1 de l'espace acoustique en pondérant d_{ij}[8] dans le plan F_1/F'_2 : le paramètre λ peut prendre toute valeur inférieure à 1, de manière à ce que les formants hauts soient pondérés par ce facteur. Ils ont, de ce fait, un rôle moins important que les formants bas dans la détermination des qualités vocaliques.

En outre, la TDF se distingue de la théorie de la Dispersion par l'introduction de la *focalisation*. Ce second terme de la fonction d'énergie diminue ainsi la valeur de l'énergie des configurations qui possèdent des voyelles avec F_1 proche de F_2, ou F_2 proche de F_3 ou F_3 proche de F_4, de manière à rendre ces configurations plus stables perceptivement. La proximité entre deux résonances a pour effet de renforcer l'énergie donnant ainsi aux voyelles une qualité *focale* comparable à celle des couleurs qualifiées ainsi (Brown et Lenneberg 1954). Le terme de focalisation est défini par :

$$E_F = \alpha(E_{12} + E_{23} + E_{34}) \tag{I.8}$$

avec :

$$E_{12} = -\sum_i \frac{1}{(F_2^i - F_1^i)^2} \tag{I.9}$$

$$E_{23} = -\sum_i \frac{1}{(F_3^i - F_2^i)^2} \tag{I.10}$$

$$E_{34} = -\sum_i \frac{1}{(F_4^i - F_3^i)^2} \tag{I.11}$$

[8]Nous avons retenu dans ce manuscrit la même notation pour exprimer une distance perceptive, quel que soit le mode de calcul de celle-ci.

α est le second paramètre de la TDF. Il détermine le poids entre le terme de dispersion (coût énergétique de la structure du système) et le terme de focalisation (coûts énergétiques locaux).

Les simulations de ce modèle consistent à répartir, pour chaque système de cardinal n, les n prototypes vocaliques dans l'espace selon les principes de la TDF, à savoir minimiser l'énergie totale donnée par l'équation (I.5). Les deux paramètres (λ, α) sont à manipuler pour tester les capacités du modèle à prédire les configurations vocaliques des systèmes universellement favorisés.

Les résultats des prédictions ont été comparés aux systèmes d'UPSID_{317}. Des plages de valeurs du couple (λ, α) ont été déterminées pour les différentes tailles de systèmes vocaliques en effectuant une exploration extensive de l'espace des phases. Il s'est agi de trouver une zone λ et α permettant de prédire les systèmes préférentiels de 3 à 9 voyelles. Les résultats[9] montrent que cette zone existe, et qu'elle est délimitée par :

$$0.2 \leq \lambda \leq 0.3 \text{ et } 0 \leq \alpha \leq 0.4$$

De plus, pour assurer la stabilité des systèmes comportant la voyelle /y/, une contrainte supplémentaire sur α est considérée : $0.3 \leq \alpha \leq 0.4$.

Analyse critique

La TDF constitue une théorie qui clarifie certains aspects au niveau des les composantes structurelles des systèmes vocaliques, et permet de prédire les grandes tendances phonologiques. Toutefois, elle repose sur les paramètres λ et α fixés empiriquement. La TDF, à l'instar des théories de l'approche globale, ne permet pas de rendre compte de la limite à 9 du cardinal des systèmes vocaliques.

Les modèles de l'approche globale expliquent, d'une manière générale, pourquoi les tendances universelles se sont imposées dans les langues du monde. En revanche, ils n'abordent ni les processus évolutifs ni les aspects de l'interaction qui sont à l'origine de ces tendances universelles. Ces derniers points constituent le propre de l'approche locale décrite ci-après.

[9] On trouvera dans l'article de Schwartz et al. (1997*b*) de plus amples informations sur ces expériences ainsi qu'une discussion étoffée autour des résultats obtenus.

I.5 Approche locale

L'objectif d'une approche locale est de déterminer la manière dont les inventaires phonologiques et lexicaux émergent de principes d'interaction entre agents communicants. Dans ce contexte, Steels (1996, 1997*b*) perçoit le langage comme étant un *phénomène d'émergence*, et ce, de deux manières différentes. Steels (1996) ajoute : «*Language is a 'mass phenomenon' actualised by the different agents interacting with each other. In this sense, language is like a cloud of birds which attains and keeps its coherence based on individual rules enacted by each bird. No single individual has a complete view of language nor does anyone control a language [...] Second, language might be emergent in the sense that it could spontaneously form itself once the appropriate physiological, psychological and social conditions are satisfied.*».

À ce jour, à notre connaissance, il existe deux études portant sur l'émergence du code phonétique qui utilisent des principes d'interactions. L'une a été menée à l'*Institut de la Communication Parlée* (ICP à Grenoble) par Glotin (1995) et a abouti au modèle AGORA. L'autre a été conduite par De Boer à *Artificial Intelligence Laboratory du Vrije Universiteit Brussel* (1997*a*). Passons en revue ces deux modèles. Nous fixerons, ensuite, le cadre de notre étude.

I.5.1 Le modèle AGORA

Dans ses travaux menés à l'ICP, Glotin (1995) a contribué à la réalisation d'un cadre de modélisation d'échanges de voyelles dans une population de robots parlants appelé «Algorithme Génétique Orienté en Recherche Articulatori-acoustique (AGORA)». Ces robots interagissent entre eux dans le but de faire converger leurs lexiques vers des symboles connus dans les inventaires phonologiques. Les robots parlants utilisés sont en fait des instances du modèle articulatoire de Maeda (1988, 1989).

Le modèle articulatoire de Maeda

Ce modèle articulatoire fut développé à partir d'une analyse statistique des clichés cinéradiographiques d'un sujet féminin ayant prononcé dix phrases en français (Botherel, Simon, Wioland et Zerling 1986). A partir de sept paramètres (deux pour les lèvres, un pour la mâchoire, un pour le larynx et trois pour la langue), le modèle articulatoire de Maeda (1988, 1989) génère la coupe sagittale du conduit vocal (cf. figure I.7). Les paramètres articulatoires de commande du modèle sont normalisés par rapport à une valeur moyenne qui correspond au contour moyen du conduit vocal.

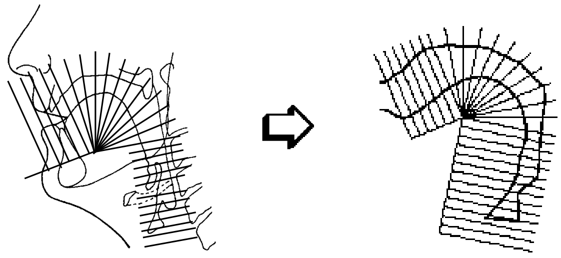


FIG. I.7: Modèle de Maeda : de l'analyse à la génération de la coupe sagittale

La fonction d'aire (Perrier, Boë et Sock 1992, Beautemps, Badin et Laboissière 1995) et la fonction de transfert (Badin et Fant 1984, Jospa 1994) sont calculées à partir de la coupe sagittale. La fonction de transfert fournit les formants (maxima du spectre des fréquences acoustiques de résonance correspondantes). Une fois les formants calculés, il est possible de synthétiser et d'entendre le son vocalique par la méthode des pôles et zéros proposée par Feng (1983).

Les interaction entre les robots

Il est demandé aux robots parlants d'établir des communications orales deux par deux en utilisant comme symboles des voyelles. La seule spécification au départ porte sur le nombre de symboles du lexique, chaque robot ayant son inventaire de symboles initialisé de façon aléatoire. L'inventaire de symboles propre à chaque robot est décrit en termes de résultats à la fois acoustiques et articulatoires. Lors d'un échange entre deux robots, le robot jouant le rôle de locuteur choisit un symbole de son lexique au hasard et, au moyen de techniques d'inversion acquises, produit les mouvements articulatoires correspondants. Le robot jouant le rôle d'auditeur «décode» le symbole du locuteur selon son propre inventaire. Puisqu'en général, les inventaires ne coïncident pas, l'auditeur retouche des items de manière à s'adapter au nouveau symbole perçu. L'adaptation consiste à retoucher l'inventaire de l'auditeur de manière à, d'une part, rapprocher un de ses symboles du symbole perçu, et d'autre part, faire éloigner les autres afin d'éviter tout risque de confusion et de maximiser les distinctivités perceptives. Un coût d'adaptation, croissant avec la quantité d'ajustement effectué sur les symboles du lexique, est alors calculé.

La spécificité du modèle AGORA est que les robots ont une longévité bien fixe et la population est renouvelée après un nombre d'échanges bien déterminé, ce qui constitue une nouvelle *génération*. L'auteur utilise un algorithme pseudo-génétique dans le sens qu'après une génération, les robots sont évalués et classés selon un critère de *fitness* donné.

Les agents se dupliquent et donnent naissance dans la nouvelle génération à un certain nombre de descendants. La taille de la population restant constante, le nombre de descendants d'un agent est fonction de sa *fitness*. Puis, une moitié des agents, tirée au hasard dans cette nouvelle population, croisent leur lexique alors que l'autre moitié reste intacte. Ce dernier croisement aurait pour but d'accélérer la convergence. Le même processus d'échanges se répète pour chaque nouvelle génération et la simulation est arrêtée lorsque la convergence est jugée satisfaisante.

Expériences

Ce modèle a été validé au départ sans faire appel au modèle articulatoire, les items du lexique étant représentés uniquement par le vecteur des formants. Les agents évoluent dans le seul espace des formants et leur *fitness* n'est évaluée que par les variations acoustiques. L'espace acoustique est délimité par un triangle ayant pour sommet /i/, /a/ et /u/. Une simulation a été effectuée avec une population de 30 robots possédant chacun 4 items. Au bout de 60 générations, les lexiques des robots ont convergé vers le système préférentiel [i e a u] dans les langues du monde. La figure I.8 montre l'évolution de la valeur de la *fitness* d'une simulation réalisée sur une population de 12 agents.

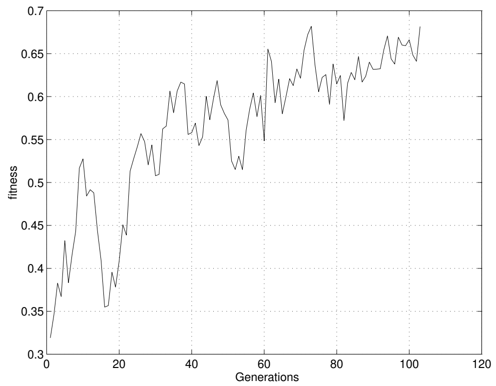


FIG. I.8: Évolution de la fitness pour une population de 12 robots (d'après Glotin (1995)).

Par la suite, AGORA a été testé avec le modèle articulatoire tel qu'il a été décrit ci-dessus. Cette combinaison permet d'ôter tout a priori sur les frontières de l'espace acoustique vu que les contraintes découlent naturellement de la physiologie de l'appareil articulatoire. La *fitness*, à son tour, est fonction des ajustements articulatoires. Les lexiques des différents robots ont convergé vers un système semblable à celui illustré par la figure I.9.

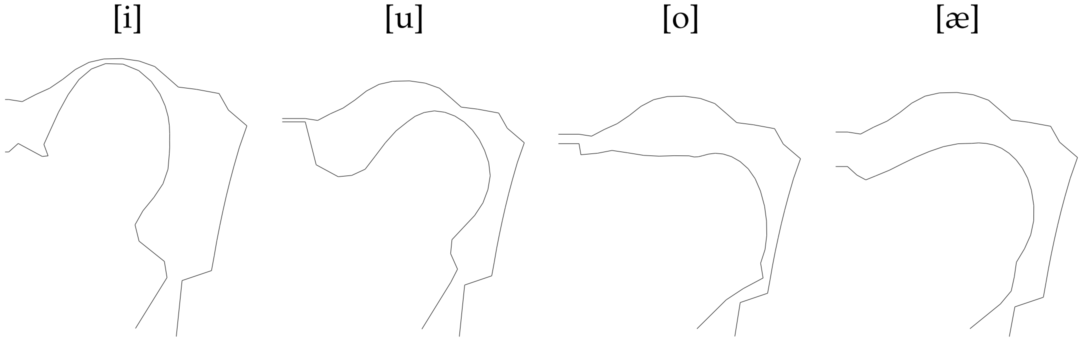


FIG. I.9: Coupes sagittales du conduit vocal du lexique d'un robot au cours d'une simulation de genèse de système vocalique à 4 voyelles. La simulation comporte 4 robots. Les lexiques des 4 robots sont similaires. Les quatre voyelles obtenues sont des réalisations assez voisines des types phonémiques [i], [o], [u] et [æ] (d'après Glotin (1995)).

Analyse critique

AGORA constitue la première étude menée à l'ICP autour de l'émergence des inventaires phonologiques où des techniques de *Vie Artificielle* ont été mises en œuvre. Quand bien même, tel qu'il a été initialement décrit et implémenté, AGORA demeure critiquable et peu exploitable. En effet, l'auteur, dans le souci de concevoir un modèle aussi réaliste que possible, a introduit plusieurs notions non toujours justifiables. Notons par exemple que l'utilisation d'un algorithme pseudo-génétique pour le renouvellement de la population a plus tendance à complexifier le processus d'échanges et à augmenter le temps d'exécution qu'à accélérer la convergence. En effet, le croisement des lexiques d'une partie de la population n'engendre en aucun cas une population plus stable. Par ailleurs, AGORA est fortement pénalisé par les temps d'exécution des simulations qui nécessitent, le plus souvent, plusieurs jours de calculs[10], ce qui se ressent inévitablement sur la qualité des résultats[11].

I.5.2 Le modèle de De Boer

Le second modèle basé sur les interactions inter-agents est le modèle de De Boer (1997*a*). Ce modèle vise à présenter une explication fonctionnelle de certains tendances universelles des systèmes vocaliques. Il est basé sur des interactions entre des individus d'une population : les agents s'échangent des sons (voyelles) en suivant un protocole

[10]Sur une station de travail de type HP9000, deux jours de calculs son nécessaires pour une simulation en acoustique seule pour une population de 20 individus ayant chacun 4 voyelles. En revanche, il a fallu cinq jours pour une simulation dans les deux espaces avec seulement une société de 4 agents possédant chacun 4 voyelles.

[11]Les diverses simulations rapportées n'ont porté que sur des agents possédant 4 voyelles.

bien particulier appelé *jeu d'imitation (imitation game)*. Avant de définir ce qu'est un jeu d'imitation, nous allons commencer par présenter l'architecture d'un agent de ce modèle.

L'architecture d'un agent

Chaque agent du système a sa propre liste de voyelles. Cette liste, initialement vide, se remplit au fur et à mesure des interactions avec les autres agents. Les voyelles sont représentées par des paramètres articulatoires.

Les voyelles des agents sont produites par un synthétiseur et sont reconnues par une unité perceptive. Une unité spéciale de contrôle régule les actions des agents et l'évaluation des voyelles. La figure I.10 illustre l'architecture interne d'un agent.

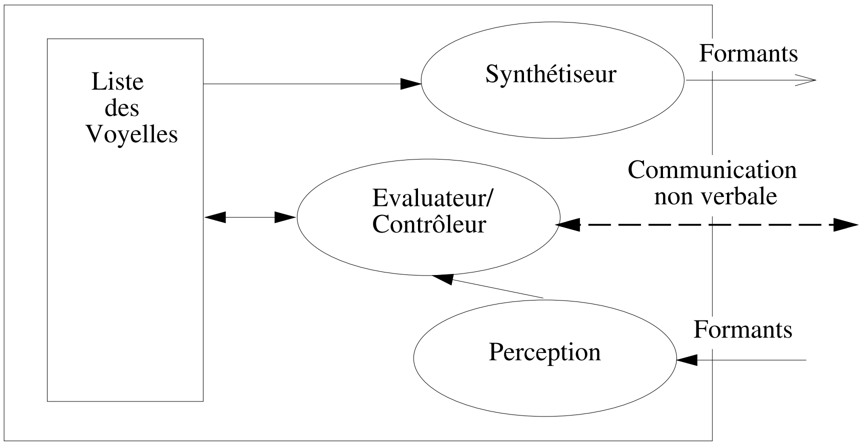


FIG. I.10: Architecture d'un agent (d'après De Boer (1997*a*)).

Une caractéristique des expériences menées par De Boer est la prise en compte du bruit en plus des fréquences formantiques produites par les agents. L'ajout du bruit rend les échanges plus naturels.

Jeu d'imitation

Les échanges entre agents sont appelés jeux d'imitation (Steels et Vogt 1997). La structure des jeux d'imitation est fondée sur les idées de Steels sur l'origine du langage (Steels 1996, Steels 1997*b*). En effet, Steels considère la langage comme un phénomène culturel dont la cohérence est maintenue grâce à son auto-organisation. Selon l'auteur toujours, le langage est appris en effectuant activement des hypothèses sur la forme de celui-ci et en les testant sur des interactions linguistiques qu'il appelle *jeux de langages (language games)*.

Les règles de base d'un jeu d'imitation sont très simples :

- deux agents sont choisis aléatoirement parmi la population, l'un jouera le rôle de l'*initiateur* et l'autre celui de l'*imitateur*,
- l'initiateur sélectionne un de ces items et le produit,
- l'imitateur interprète cet item en fonction de son lexique, et produit l'item qu'il pense avoir reconnu (l'item le plus proche du perçu),
- l'initiateur perçoit cette *imitation* et la compare à son tour aux items de son propre lexique,
- le succès du jeu d'imitation dépendra de ce que l'item reconnu par l'initiateur s'identifie à celui qu'il a produit ou non.

On remarquera ici une ressemblance avec le principe d'interaction dans AGORA (§ I.5.1).

Deux informations seront stockées pour chaque item des listes des agents : le nombre de fois qu'il a été utilisé et le score. Le score est incrémenté à chaque fois qu'un jeu d'imitation utilisant cet item, est couronné de succès. La qualité d'un item est donnée par le rapport entre son score et son nombre d'utilisations.

Les inventaires initialement vides des agents évoluent au cours des échanges. Les agents créent d'abord un item correspondant à une configuration articulatoire aléatoire. Les processus de création et de suppression des items sont décrits dans (De Boer 1997*a*). Notons juste que lorsque l'imitation réussit, l'imitateur rapproche l'item reconnu[12] de l'item imité. Ainsi, l'imitation est améliorée. Notons également que les items de basse qualité sont supprimés à la longue. Par ailleurs, les items voisins au sein d'un même inventaire sont fusionnés pour éviter tout risque de confusion. Comme nous venons de le voir, ce modèle utilise une sorte de principe de dispersion entre l'attraction de l'item imité sur l'item reconnu d'une part et, la fusion des items les plus proches d'autre part. En fait, cette dispersion est une conséquence du processus de génération aléatoire des items qui couvrent l'espace de production.

Expériences

De nombreuses simulations de ces jeux d'imitation ont été conduites (De Boer 1997*a*, De Boer 1997*b*, Steels 1997*c*). Les expériences consistaient à réaliser un nombre fixe d'itérations des jeux d'imitation sur une population homogène d'agents. Le but était d'ob-

[12] Si celui-ci n'a pas un score élevé car ceci signifierait qu'il est une bonne imitation d'autres items, auquel cas, un nouvel item, identique à l'imité, serait créé.

server les types de systèmes sonores obtenus en faisant varier la taille de la population sous diverses conditions de bruit.

Les résultats rapportés montrent qu'en général, lorsque le bruit est faible, les différents agents convergent vers les mêmes phonèmes. À titre d'exemple, lorsque les réalisations acoustiques sont soumises à un bruit de 10%, le système vocalique obtenu pour une population de 5 agents après 1 000 jeux est [i,ɛ,a,ɑ,ɤ] (De Boer 1997*a*). Dans les mêmes conditions de bruit, une expérience a porté sur 2 000 jeux d'imitation avec 20 agents et les agents avaient tous pour système une variante de [i,ə,ɛ,ɐ,ɔ,ɯ,u] (De Boer 1997*b*). Avec un bruit important, la population a tendance à converger vers de petits systèmes (de l'ordre de 2). Ceci étant, les systèmes obtenus n'ont pas la prétention de révéler les systèmes les plus fréquents dans les langues du monde comme le précise De Boer (1997*a*) : «*the goal of the experiments was to investigate whether it was possible to develop a successful sound system in a population of initially empty agents, and what form this sound system will take under different conditions of noise, and for different population size*».

Analyse critique

Le modèle proposé par De Boer a montré qu'il était tout à fait possible de générer des systèmes vocaliques dans une population d'agents distribués qui tentent de s'imiter mutuellement sous certaines conditions.

L'originalité de ce modèle réside dans le fait que la taille des systèmes engendrés n'est pas constante. Bien au contraire, les inventaires des agents, initialement vides, évoluent au fur et à mesure des simulations, soit en créant de nouveaux phonèmes, soit en supprimant d'autres dont la qualité est médiocre (en-deça d'un seuil préalable). Toutefois, le revers de la médaille des systèmes ouverts, dans l'état actuel des choses, est que les résultats restent peu exploitables dans le domaine de la prédiction des tendances universelles des langues du monde. En effet, il est important de noter que les populations évoluent éternellement avec le temps, dans la mesure où il n'existe pas de critère d'arrêt pour les simulations. Ceci incite, a fortiori, l'expérimentateur à suivre de près la progression des inventaires. Une conséquence directe de ce que nous venons de souligner est qu'il est impossible de connaître le nombre optimal de voyelles d'une simulation. Enfin, et ce qui n'est pas le plus dépourvu d'intérêt pour nous, il serait difficile d'affirmer que les systèmes générés à partir de ce modèle et décrits dans les expériences reflètent des systèmes vocaliques rencontrés fréquemment dans les langues du monde.

I.6 Conclusion

La recherche sur les tendances universelles des langues du monde a drainé une multitude de travaux. En particulier, ce thème de recherche est au centre des intérêts de l'ICP. Vallée (1994) a affiné une typologie des voyelles des langues du monde et a proposé des prototypes articulatoires de 37 voyelles de la base UPSID_{317}. En outre, cette base, pour être exploitée d'un point de vue statistique et sélectif, a été implémentée par L.-J. Boë et N. Vallée sous forme de banque de données facilement consultable et exploitable. Pour la prédiction des systèmes vocaliques, Schwartz et al. (1997*b*) ont adopté une approche globale et ont proposé la Théorie de la Dispersion-Focalisation (cf. § I.4.3). Cette théorie a permis de réaliser de bonnes prédictions de systèmes vocaliques.

En ce qui nous concerne, nous nous plaçons toujours dans cette même lignée de recherche. Toutefois, nous nous inspirerons, pour notre part, des techniques de la *Vie Artificielle* (Heudin 1994, Langton 1995) afin de réaliser des simulations rappelant quelque peu ce qu'effectuent les êtres humains. La Vie Artificielle offre des perspectives fascinantes dans ce créneau de recherche et a suscité de nombreuses recherches : MacLennan (1991), MacLennan & Burghardt (1994) et Werner & Dyer (1991) ont conduit des expériences, basées sur des évolutions génétiques, sur l'origine de la communication. Ces expériences ont montré l'importance de la coopération entre les différents agents lors du processus de communication. Toujours dans le même ordre d'idées, Cangelosi & Parisi (1996) ont présenté des simulations dans lesquelles, un «langage» émergeait de l'évolution d'une population de réseaux de neurones. Dans cette expérience, le but des communications est de catégoriser les champignons comestibles des vénéneux, tandis que dans l'expérience conduite par Werner & Dyer (1991), les communications avec succès aidaient la reproduction. Par ailleurs, plusieurs études se sont focalisées sur l'étude de la syntaxe du langage (Hashimoto et Ikegami 1996, Kirby 1996, Steels 1997*a*, Batali 1998). Hashimoto & Ikegami (1996) ont montré, à partir de simulations menées, que la complexité d'une grammaire pouvait augmenter suite à des jeux de langages *(langage games)* entre agents. De son côté, Kirby (1996) a effectué des simulations portant sur l'émergence des structures syntaxiques «universelles» du langage. L'auteur a constaté que certaines structures du langage n'émergeaient pas, et a proposé de combiner aux théories dites *functionalist*, des théories qu'il qualifie *of innate, autonomous syntactic knowledge* afin d'expliquer les tendances universelles. Dans un autre registre, Steels (1997*a*) a réalisé des expérimentations sur deux robots munis chacun d'une caméra, et guettant des scènes dans un *écosystème*. Ces expériences ont montré qu'un lexique et des structures syntaxiques pouvaient ressortir du besoin des robots à exprimer des propriétés des objets perçus dans un tel environnement.

À l'aide de techniques pouvant être apparentées à la Vie Artificielle, nous désirons aborder des questions fondamentales portant sur l'origine du langage, la manière de l'acquisition des tendances universelles du langage et les raisons de l'évolution perpétuelle des langues du monde.

Le modèle AGORA a constitué une première tentative d'utilisation d'une approche de ce type pour prédire les systèmes vocaliques. Malheureusement, les simulations d'AGORA sont lentes, recrutent peu d'agents et les résultats obtenus n'ont pas été jugés satisfaisants.

Le présent travail se veut être une application des techniques de la Vie Artificielle en Phonétique. Dans un souci de rigueur, nous reprendrons les choses depuis le début en nous inspirant des principes généraux utilisés dans le cadre d'AGORA. Cependant, il n'existe pas de moyen d'évaluer dans AGORA l'influence des divers paramètres, et d'identifier la source des problèmes. En effet, comme AGORA intègre en plus du protocole d'échanges, un module d'inversion articulatori-acoustique et un algorithme génétique, il est ardu d'identifier à quelle module sont liés les problèmes des simulations. À cet égard, nous allons adopter une approche analytique en procédant par étapes. Nous préférons, au début, simplifier le problème et n'évoluer que dans l'espace acoustique en laissant provisoirement de côté la partie articulatoire. Cette simplification du modèle de communication ne rend pas la validation de ce modèle moins complexe. En effet, nous aurons à manipuler plusieurs paramètres (la taille de la population, le nombre de transactions, etc.). En outre, nous devrons contrôler les simulations testant le modèle sur la prédiction des systèmes vocaliques. De plus, nous ferons en sorte que notre modèle puisse permettre de simuler le principe du MUAF[13] (Ohala 1980), ce qui, à notre connaissance, n'a jamais été proposé auparavant.

[13]Maximum Use of Available distinctive Features

Chapitre II

Un agent d'une Société communicante : L'Agent de Parole

Résumé

Ce chapitre est entièrement dédié à la description du modèle de prédiction. Avant d'entrer dans le vif du sujet, nous présenterons succinctement deux projets de recherche dans lesquels s'inscrit parfaitement cette étude.

Par la suite, nous détaillerons la structure d'un agent de parole, ainsi que les protocoles d'interaction basés sur la coopération entre deux agents de parole. Nous introduirons alors la notion de société d'agents, et nous présenterons les critères adoptés pour apprécier la convergence de la société, et pour présenter les résultats.

Nous terminerons ce chapitre par l'énumération des paramètres de notre modèle et la descritpion de l'algorithme d'une simulation du modèle d'émergence.

II.1 Introduction

Cette étude vise à réaliser une plate-forme rendant compte de l'émergence des structures phonétiques par un principe d'interaction locale au sein d'une communauté d'agents.

Ce travail de thèse s'insère dans le cadre de deux projets de recherche accordés par le *Groupement d'Intérêt Scientifique (GIS) Sciences de la Cognition* et l'*Agence Rhône-Alpes pour les Sciences Sociales et Humaines (ARASSH)*. Ces projets s'intitulent respectivement «Les robots parlent aux robots : Un Générateur des Structures Sonores du Langage (Projet CNA 34)» et «Un Générateur des Structures Sonores du Langage inspiré des contraintes développementales».

Nous présenterons d'abord rapidement le cadre de recherche dans lequel s'insère notre étude. Nous nous attarderons par la suite sur la description proprement dite de notre modèle de simulation.

II.2 Les robots parlent aux robots

La préoccupation commune des projets est dans l'utilisation des avancées récentes dans le domaine de la *robotique de la parole* pour mettre en œuvre une situation inédite : «l'observation» d'une communication entre robots parlants afin de mettre en relation l'évolution de cette société de robots avec les données expérimentales majeures concernant l'apprentissage (ontogenèse) et l'évolution (phylogenèse).

II.2.1 La robotique de la parole

Si l'on considère que la parole est un parcours optimisé (au sens d'une habileté motrice acquise) d'une séquence d'objectifs perceptifs négociés entre un locuteur et un auditeur, deux questions fondamentales se posent :

(i) Comment sont négociés les objectifs ?

(ii) Comment sont optimisés les parcours ?

Les réponses à ces questions dépendent pour une large part de la compréhension en profondeur des mécanismes d'interaction locuteur/auditeur en communication parlée. Les chercheurs de l'ICP se sont inspirés d'une autre discipline, à savoir la robotique, en définissant une *robotique de la parole* (Laboissière 1992) : la parole est construite et régulée par les interactions sensori-motrices, et justement, la robotique est le domaine par excellence des développements des techniques pour l'étude des relations entre un système sensoriel et un système moteur ; le noyau central consiste en la mise en œuvre de processus de passage d'espaces moteurs à espaces sensoriels permettant de résoudre des problèmes complexes d'optimisation ou de satisfaction de contraintes multi-espaces.

II.2.2 Plan de travail des projets

Le plan de travail comporte trois parties relativement autonomes quoique étroitement imbriquées :

- l'élaboration d'un robot parlant,
- l'ontogenèse d'un robot élève et,
- la phylogenèse d'une société de robots parlants.

II.2.3 L'élaboration d'un robot parlant

Par réalisation d'un robot parlant (cf. figure II.1), on entend la simulation d'un système anthropomorphique de production de la parole capable d'inférer les mouvements articulatoires (les causes) à partir des suites de sons à réaliser (les effets). On s'intéresse à l'anthropomorphisme à double titre. Tout d'abord, on désire que le robot parlant ait une flexibilité d'adaptation aux conditions de la communication parlée. En d'autres termes, on souhaite qu'il trouve aux problèmes posés des solutions compatibles avec celles qu'aurait trouvées un être humain. En second lieu, on veut utiliser le robot pour comprendre certains phénomènes typiques de la parole. Or, d'après les rédacteurs des projets, seule une approche de synthèse des systèmes de production et de perception humains peut permettre d'atteindre ces deux objectifs à la fois.

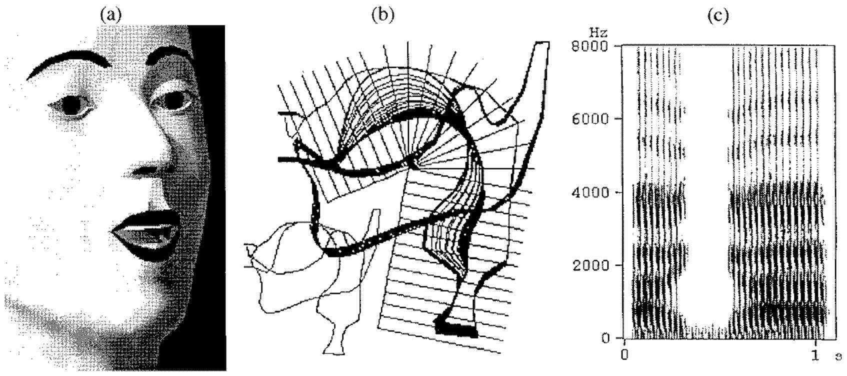


FIG. II.1: Le robot parlant de l'ICP : (a) visage parlant ; (b) mouvements du modèle articulatoire sous-jacent ; (c) synthèse acoustique résultante.

L'ontogenèse d'un robot élève

L'objectif est de proposer des solutions permettant à un robot «élève» de reproduire une séquence acoustico-visuelle produite par son maître, en la recadrant à l'intérieur de son propre espace distal de manière à déterminer les commandes proximales associées (cf. figure II.2).

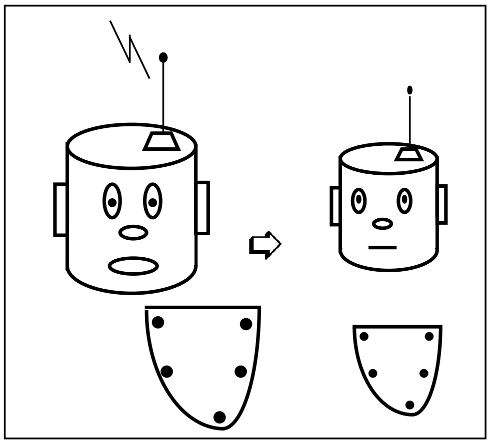


FIG. II.2: L'ontogenèse d'un robot élève : transfert de prototypes avec changement d'espace.

Ceci relève de la problématique classique de la normalisation acoustique dans lequel des stimuli acoustiquement différents sont jugés proches, car produits par des conduits vocaux anatomiquement dissemblables. Il est clairement établi qu'il existe des processus de normalisation «extrinsèques», c'est à dire recadrant un stimulus grâce à une connaissance sur le locuteur (par exemple, de son espace maximal de production, ou d'une estimation de la taille de son conduit vocal) (Mattingly 1966, Nordström 1977, Wakita 1977), et des processus «intrinsèques», permettant un cadrage sans apprentissage à partir des calculs directs sur le signal à normaliser (Lobanov 1971, Matsumoto et Wakita 1986, Zahorian et Jagharghi 1991). Dans ce second cas, soit des algorithmes exploitant la fréquence fondamentale, soit des calculs spectraux utilisant le plus souvent des paramètres formantiques relatifs sont proposés.

Dans le cadre du projet, un robot élève cherchant à reproduire les stimuli émis par un maître doit évidemment résoudre le problème de normalisation. Dans ce contexte, le cadre naturel est ce qui est communément appelé *normalisation articulatoire*, soit l'estimation à partir de stimuli distaux, non seulement des commandes proximales, mais aussi des caractéristiques de base du modèle articulatoire.

La phylogenèse d'une société de robots parlants

L'objectif est cette fois de déterminer la manière dont les inventaires phonologiques et lexicaux émergent. Ce problème peut être abordé par une approche globale (cf. § I.4) fondée sur un compromis entre des critères d'efficacité perceptive et des critères d'économie articulatoire, ou bien par une approche locale (cf. § I.5) dans laquelle l'émergence résulte des principes d'échanges et de négociation entre agents communicants (cf. figure II.3).

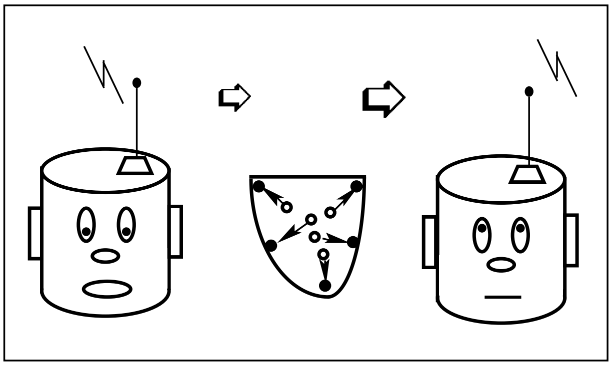


FIG. II.3: La phylogenèse d'une société de robots parlants : émergence d'un système de prototypes négocié entre deux robots partenaires.

Ce point constitue la problématique de notre étude. L'objectif de notre recherche à long terme est d'apporter une pierre à l'édifice d'une théorie solide de la Communication Parlée en fournissant quelques éléments de réponse à des questions portant notamment sur l'origine, l'apprentissage et l'évolution du langage ainsi que l'apparition de dialectes.

Dans le cadre de travail, nous nous focaliserons sur les processus d'interactions entre des agents de parole tels que ceux décrits ci-après.

II.3 Définition d'un Agent de Parole

Un robot parlant (Laboissière 1992) est une réplique en simulation anthropomorphique des systèmes de production et de perception de la parole, ajoutés d'un système de contrôle moteur qui permet le contrôle des systèmes musculaires de production en fonction des buts perceptifs à accomplir. Nous avons conçu une caricature du robot parlant que nous baptisons «Agent de Parole (AP)» afin de permettre de simuler des échanges et des interactions.

II.4 Structure d'un Agent de Parole

La structure interne d'un Agent de Parole (AP) se limite à son lexique. Le lexique d'un AP est matérialisé par les items qui le composent, spécifiés, à leur tour, par les fréquences acoustiques de résonance (formants) (cf. figure II.4). Ces fréquences sont exprimées en Hertz.

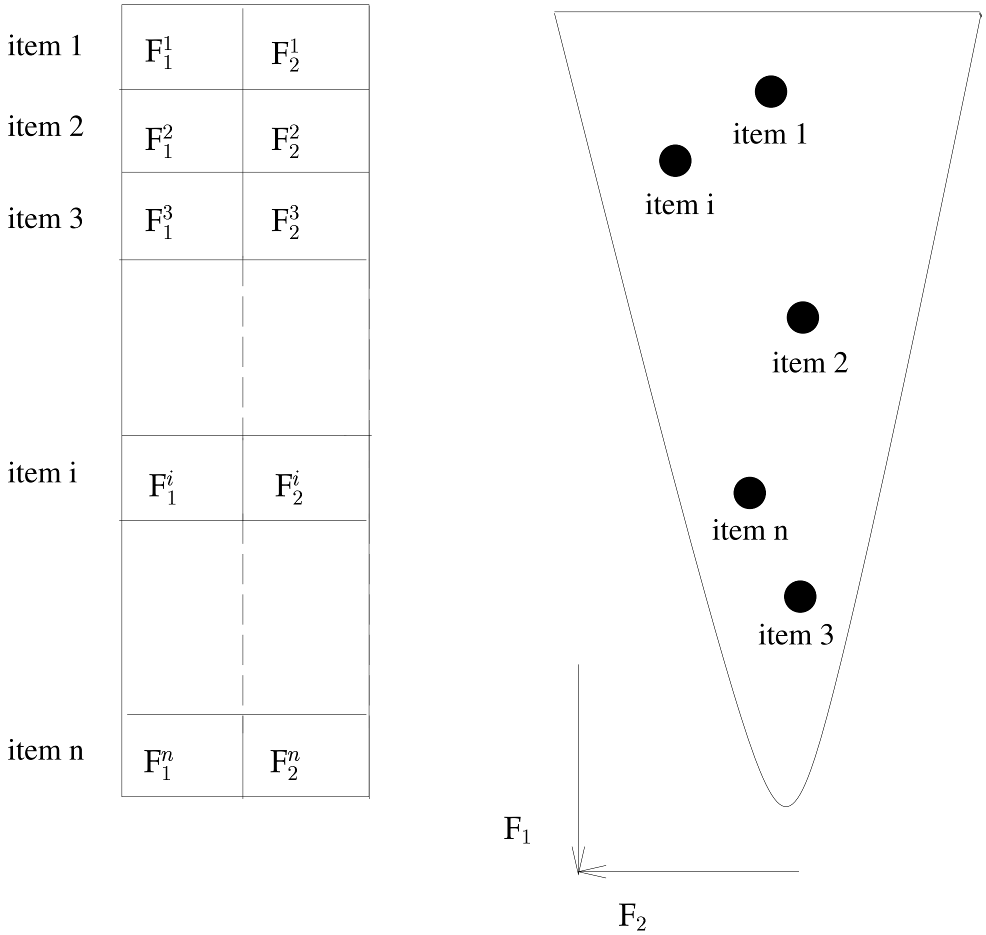


FIG. II.4: Le lexique d'un AP contenant n items : les items sont représentés sur le plan (F_1,F_2) dans un triangle vocalique.

En premier lieu, nous nous fixons la contrainte suivante : toute voyelle doit s'inscrire à l'intérieur de l'*espace vocalique maximal*. Celui-ci a été délimité par Boë et al. (1989) à partir de résultats obtenus à l'issue d'environ 60 000 simulations vocaliques. L'espace vocalique calculé par les auteurs,à partir des projections de ces 60 000 valeurs dans (F_1,F_2) et représenté par la figure II.5, possède les limites déterminées par les équations suivantes :

$$F_1^{max} = -800 + 2.59\ F_2 - (1.04 * 10^{-3})\ F_2^2 \quad pour \quad 510 \leq F_2 \leq 1086\ Hz \tag{II.1}$$
$$F_1^{max} = 248 + 0.93\ F_2 - (4 * 10^{-4})\ F_2^2 \quad pour \quad 1086 \leq F_2 \leq 2322\ Hz \tag{II.2}$$
$$F_1^{min} = 250\ Hz \tag{II.3}$$

F_2^{min} vaut 510 selon une étude menée par Boë et al.(1992). Les paraboles ayant pour équations (II.1) et (II.2) se croisent lorsque F_2 vaut 1086, alors que la droite II.3 rencontre la parabole (II.2) pour une valeur de F_2 égale à 2322.

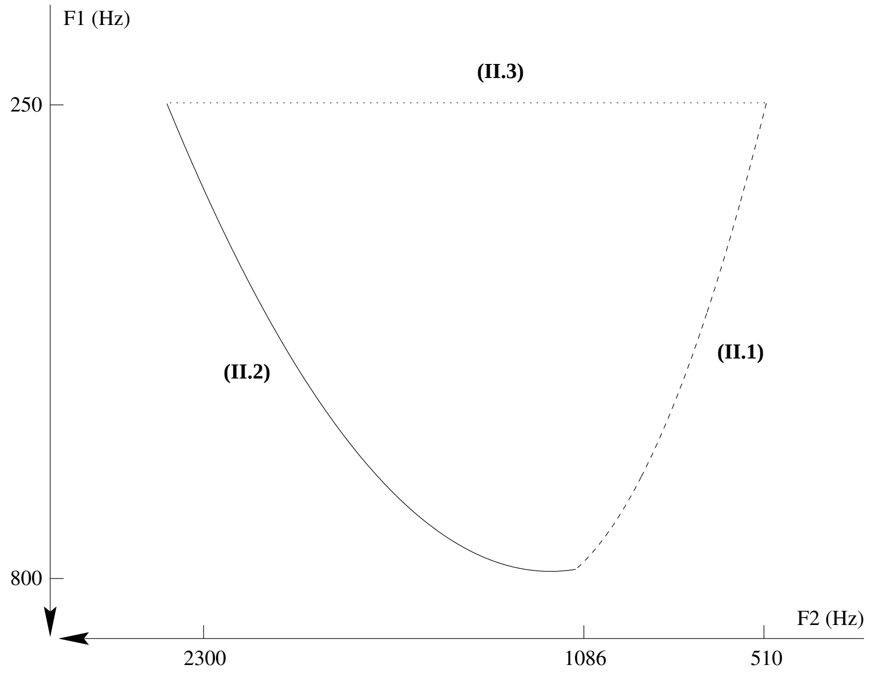


FIG. II.5: Projection de l'espace vocalique maximal sur le plan (F1,F2), les équations du triangle vocalique sont données par les expressions (II.1), (II.2) et (II.3).

Comme nous le verrons ultérieurement, nous serons appelés à faire évoluer les lexiques des APs dans un espace vocalique dont les valeurs sont exprimées en Bark. L'échelle des Barks, rappelons-le, est une échelle de fréquence perceptive, définie sur la base d'un certain nombre d'expériences psycho-acoustiques (Schroeder, Atal et Hall 1979), et calculée suivant la formule exprimée par l'équation (I.4). Par conséquent, il nous est nécessaire de connaître les limites de cet espace en Bark dans le plan (F_1,F_2). En minimisant les moindres carrés, nous obtenons les équations suivantes qui délimitent l'espace vocalique en Bark illustré par la figure II.6 :

$$F_1^{max} = -12.5249 + 4.0308\ F_2 - 0.2061\ F_2^2 \quad pour \quad 5.05 \leq F_2 \leq 12.29\ Bk \tag{II.4}$$
$$F_1^{max} = -72.3223 + 13.7578\ F_2 - 0.6017\ F_2^2 \quad pour \quad 12.29 \leq F_2 \leq 13.90\ Bk \tag{II.5}$$
$$F_1^{min} = 2.63\ Bk \tag{II.6}$$

2.63, 5.05 et 13.90 sont les résultats de la conversion en Bark de respectivement 250, 510 et 2322 Hz.

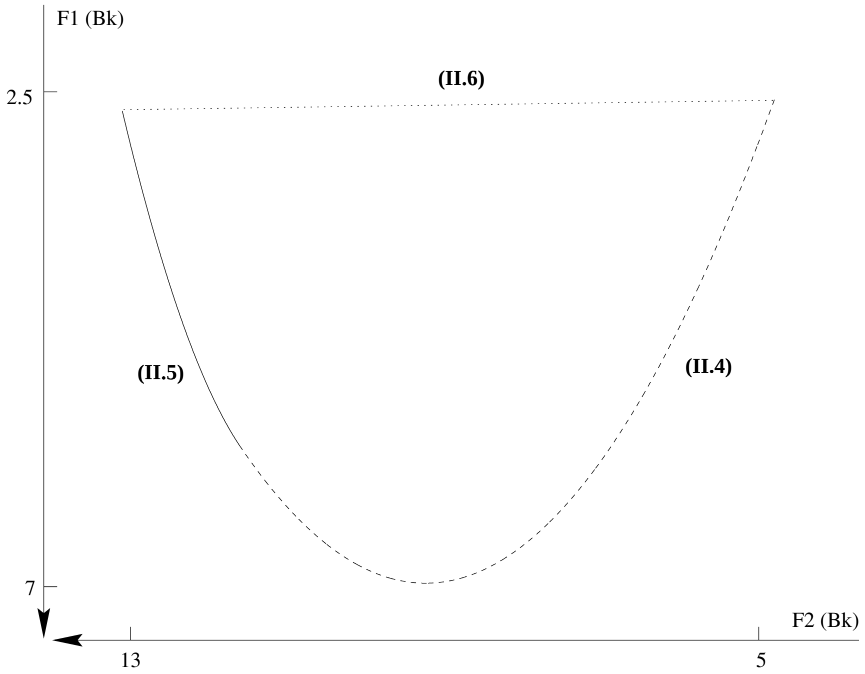


FIG. II.6: Espace vocalique maximal en Bark sur le plan (F1,F2), les équations du triangle vocalique sont données par les expressions (II.4), (II.5) et (II.6)

L'espace vocalique schématisé par la figure II.7 est généré en convertissant les points des bordures de l'espace en Bark (cf. figure II.6) en Hertz.

L'objectif de cette étude est de faire interagir un ensemble d'APs entre eux puis d'observer et d'analyser leurs évolutions. Nous adopterons désormais dans ce mémoire, le terme de *société* d'APs. Avant de décrire les caractéristiques d'une société, attardons-nous sur la présentation en détails d'une interaction entre deux robots. Cette interaction constitue le maillon de base de notre modèle intitulé «**E**mergence dans une **S**ociété d'agents de **P**arole par l'**E**ffet de la **C**oop**E**ration (ESPECE)» . ESPECE se veut être un modèle de simulation basé sur la coopération lors des interactions entre agents.

II.5 Transaction entre robots locuteur et auditeur

Une interaction entre un AP locuteur et un AP auditeur constitue ce que nous appellons une *transaction*.

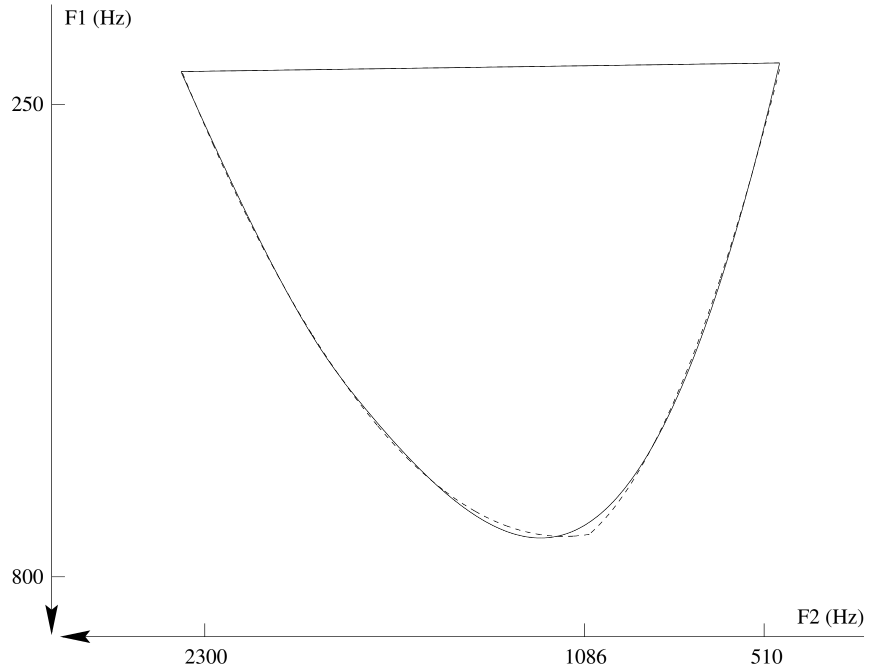


FIG. II.7: Nouvel espace vocalique maximal en Hz qui est une bonne approximation de l'espace de la figure II.5 représenté en traits discontinus.

Lors d'une transaction, l'AP jouant le rôle de locuteur choisit aléatoirement un item de son lexique et le «produit». En fait, cette production revient à transmettre les fréquences formantiques de ce symbole. Cette première étape est illustrée dans la figure II.8 (a). L'AP jouant le rôle d'auditeur «décode» alors le symbole du locuteur conformément à son propre inventaire.

Les inventaires ne coïncidant pas forcément, l'auditeur aura la capacité de retoucher ses items de manière à s'adapter au nouveau symbole perçu. L'adaptation de l'auditeur à un nouveau symbole perçu constitue le point critique des simulations. En effet, c'est d'elle que dépend en grande partie la convergence des lexiques comme nous le verrons ultérieurement. Nous partons du principe qu'une personne qui apprend un nouveau mot (ou un nouveau son) commence par essayer de répéter ce mot (ou son). De même, un robot auditeur qui perçoit un symbole ne figurant pas dans son propre lexique tentera de le «répéter». En termes pratiques, cette répétition consistera à retoucher son inventaire de manière à, d'une part, rapprocher un de ses symboles du symbole perçu, et d'autre part, faire éloigner les autres symboles afin d'éviter tout risque de confusion et de maximiser les distinctivitéss perceptives. Par conséquent, cette adaptation s'appuie essentiellement sur deux concepts que nous appellerons *attraction* et *répulsion*.

Avant de détailler ces deux notions là, il est opportun de présenter préalablement les outils qu'elles utilisent, et en particulier les expressions des distances sur lesquelles

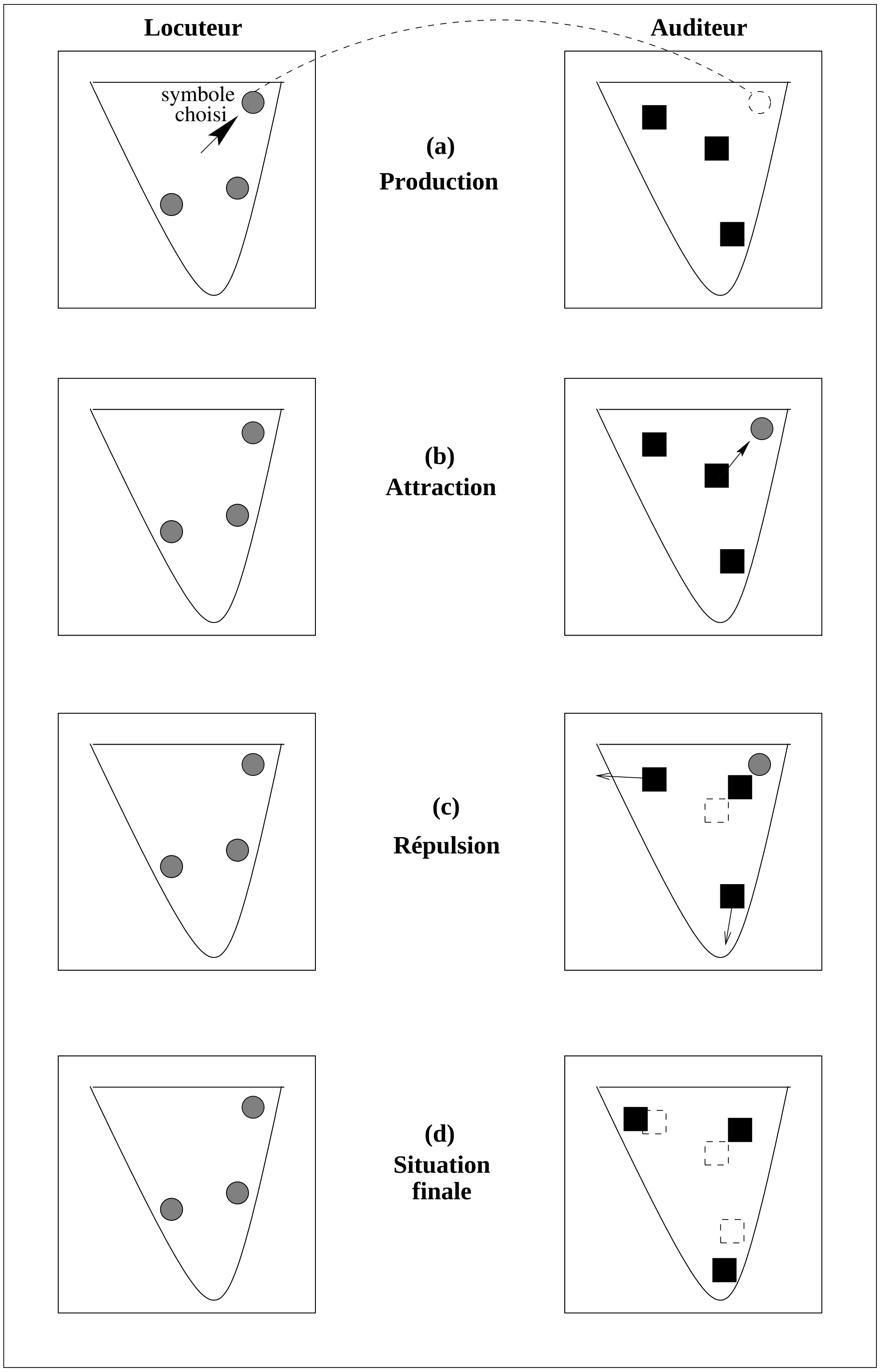


FIG. II.8: Différentes étapes lors d'une transaction elles se fondent.

II.5.1 Distances perceptives

Une fois que l'auditeur aura perçu le symbole produit par le locuteur, il aura à le comparer avec les symboles de son propre inventaire, de manière à identifier son item qui subira une attraction, et ceux qui subiront une répulsion. Ceci revient à calculer les distances perceptives entre le symbole perçu et les symboles du lexique de l'auditeur.

L'expression de la distance perceptive entre des symboles *i* et *j* est donnée par :

$$d_{ij} = ((F_1^j - F_1^i)^2 + \lambda^2(F_2^j - F_2^i)^2)^{1/2} \tag{II.7}$$

A l'instar de Lindblom (1975) et Schwartz et al. (1997*b*), nous avons intégré un paramètre λ qui pondère la dimension F_1 de l'espace acoustique par rapport à F_2 afin de mettre en évidence la suprématie du contraste haut/bas (distinction par l'ouverture) sur le contraste antéro-postérieur (lieu d'articulation). λ peut prendre, par conséquent, toute valeur positive inférieure à 1. Les valeurs des formants sont exprimées en Bark.

II.5.2 Attraction

Le concept général de nos simulations est de mettre en société un ensemble d'APs et de les laisser communiquer entre eux. Au cours de ces échanges, les inventaires des APs évolueront suite à des «coopérations» entre les robots. Nous entendons par coopération le comportement (ou réaction) qu'aura un robot auditeur suite à la perception d'un symbole émis par le locuteur. En effet, ce sont les changements que subissent les lexiques des APs qui font que ces lexiques évoluent et, éventuellement, convergent vers un lexique commun.

L'adaptation du robot auditeur consistera dans un premier temps à rapprocher un de ses items vers l'item perçu. L'item qui subira cette attraction est celui qui est le plus proche de l'item produit par le locuteur (le symbole perçu). Autrement dit, ce symbole est celui dont la distance perceptive, avec le symbole perçu, est la plus petite.

Dans nos simulations, l'attraction revient à déplacer le symbole sélectionné de façon à ce qu'il soit à égale distance de son ancienne position et du symbole perçu. Ce qui signifie en d'autres termes que nous le rapprocherons de la moitié de la distance qui le séparait de l'item attiré.

Cette attraction, donnée par l'équation (II.8) est schématisée sur la figure II.8 (b).

$$item_{attiré} \leftarrow item_{attiré} - (\Delta/2) \tag{II.8}$$

avec :

$$\Delta \;=\; [1\ \lambda] * \begin{pmatrix} F_1^{attiré} - F_1^{percu} \\ F_2^{attiré} - F_2^{percu} \end{pmatrix} \tag{II.9}$$

L'item à attirer sera déplacé dans la direction et le sens de l'item perçu. Ceci est assuré dans l'équation (II.8) par le terme Δ. Comme le montre l'équation (II.9), Δ est un vecteur ayant pour composantes la différence entre les formants de l'item attiré et ceux de l'item perçu. Toujours dans l'objectif d'exploiter la primauté de la dimension F_1 de l'espace acoustique par rapport à F_2 (Lindblom et Lubker 1985), la seconde composante de ce vecteur ($F_2^{attiré} - F_2^{percu}$) est multipliée par le facteur λ. Ainsi, les symboles auront le loisir de mieux exploiter la première dimension.

II.5.3 Répulsion

Le second volet de l'adaptation d'un auditeur suite à la perception d'un symbole est ce que nous appelons *répulsion*. Comme son nom l'indique, la répulsion consiste à éloigner du symbole attiré tous les autres symboles du lexique de l'auditeur. Comme il a été précisé précédemment, cette répulsion vise à augmenter la distinctivité perceptive au sein des inventaires. Autrement dit, elle aspire à atténuer les risques de confusion entre les items d'un même lexique.

Chaque symbole de l'inventaire à éloigner subira un répulsion dont l'intensité est exprimée par :

$$\eta = \xi * (\Delta/||\Delta||) * (1/(1 + d^2_{attiré\ rep})) \tag{II.10}$$

avec :

$$\Delta \;=\; [1\ \lambda] * \begin{pmatrix} F_1^{rep} - F_1^{attiré} \\ F_2^{rep} - F_2^{attiré} \end{pmatrix} \tag{II.11}$$

$$d_{attiré\ rep} \;=\; ((F_1^{attiré} - F_1^{rep})^2 + \lambda^2(F_2^{attiré} - F_2^{rep})^2)^{1/2} \tag{II.12}$$

Δ est un vecteur semblable à celui défini en § II.5.2 mais ayant pour composantes, cette fois, la différence entre les formants de l'item à repousser et ceux de l'item attiré.

L'item à repousser sera déplacé dans la direction de l'item attiré mais dans le sens opposé (comme l'illustre la figure II.8(c)). Ceci est assuré dans l'équation (II.8) par le terme normé ($\Delta/||\Delta||$).

Le paramètre ξ nous permettra uniquement de faire varier l'intensité de la répulsion. Cette intensité est inversement proportionnelle à la distance perceptive entre l'item à repousser et l'item attiré. En effet, nous désirons que la force de répulsion soit d'autant plus importante lorsque les items antagonistes sont proches l'un de l'autre, et inversement, que la force soit moindre si les items sont éloignés. Nous pourrons faire une certaine analogie avec un champ magnétique qui est de plus en plus important dès lors qu'on s'approche de la source : d'où le terme $(1/(1+d^2_{attir\acute{e}\ rep}))$ dans l'équation (II.10). Il a été rajouté la scalaire «1» au dénominateur du terme $1/(1+d^2_{attir\acute{e}\ rep})$ uniquement pour borner cette distance perceptive lorsque la valeur de $d^2_{attir\acute{e}\ rep}$ est très petite ($\ll 1$).

En somme, repousser un item revient à (cf. figure II.8 (c)) :

$$item_{rep} \leftarrow item_{rep} + \eta \tag{II.13}$$

Notons que les ajustements subis par les symboles lors d'une attraction ou d'une répulsion sont effectués dans l'espace acoustique en Bark.

II.5.4 Glissement autour du triangle vocalique

Les frontières de l'espace vocalique maximal découlent directement des contraintes physiologiques de l'appareil articulatoire. Nos simulations évoluant uniquement dans l'espace acoustique, nous devons veiller à ce que les configurations acoustiques des symboles des inventaires soient bornés par le triangle vocalique.

En conséquence, nous testerons, à chaque fois, si tout symbole repoussé n'a pas débordé du triangle vocalique. Si un item a été repoussé en dehors des limites du triangle, nous effectuerons une projection orthogonale de cet item sur les bords du triangle. Ceci est réalisé en recherchant à chaque fois le point du triangle le plus proche de cet item (cf. figure II.9). Cette projection a pour effet de faire glisser le symbole à repousser sur les bords du triangle vocalique.

II.5.5 Synthèse d'une transaction

Une transaction entre un agent locuteur et un agent auditeur peut se résumer en les étapes suivantes :

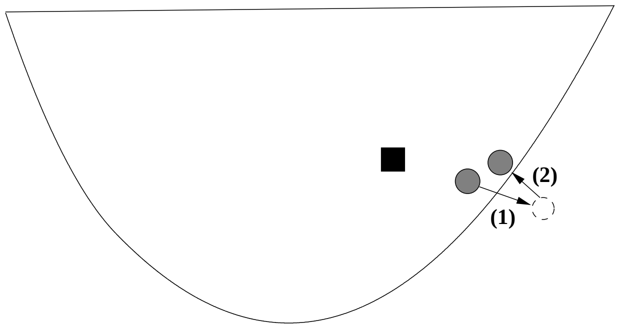


FIG. II.9: Procédure de glissement des symboles sur le triangle vocalique. Ce dernier est représenté cette fois avec la même échelle pour F_1 et F_2 exprimés en Bark.

- une paire d'APs est aléatoirement désignée parmi la population, un agent sera locuteur et l'autre auditeur,
- le locuteur sélectionne un de ces symboles et le produit,
- l'auditeur perçoit ce symbole et adapte son lexique suite à cette perception ; cette adaptation se réalise à partir des deux opérations suivantes :
  (i) l'attraction : l'auditeur rapprochera son symbole le plus proche du symbole perçu vers ce dernier,
  (ii) la répulsion : l'auditeur éloignera ses autres symboles du symbole attiré.

II.6 Société d'agents de parole

Mettre en société des APs signifie, rappelons-le, faire interagir un ensemble d'APs entre eux. Les simulations consistent à mettre en société une communauté d'APs et à les faire coopérer entre eux durant un nombre important de transactions. Nous souhaitons, de cette manière, amener les inventaires de tous les APs à converger sensiblement vers un même lexique. Ainsi, pourrons-nous analyser ce lexique et le comparer à ceux des langues du monde. Si ce lexique permet de rendre compte des tendances universelles des langues du monde, nous pourrons avancer que, la possibilité théorique, selon laquelle les phonèmes des langues sont le fruit, depuis la nuit des temps, d'échanges et de coopérations fondés sur certains principes tels que la distinctivité perceptive, est plausible et cohérente.

Nous affectons, initialement, à tous les items de la société d'APs des petites variations autour d'une voyelle que nous appellerons «neutre». Il est difficile de définir une position neutre de l'espace acoustique sans que cette définition ne soit arbitraire. Nous choisissons cependant comme voyelle neutre la configuration obtenue en mettant tous les paramètres du modèle articulatoire de Maeda (1988, 1989) à zéro, ce qui produit une voyelle centrale proche du /œ/ donnée par la configuration arbitraire $F_1 = 530$ et $F_2 = 1425$.

Ensuite, nous tirons, au hasard, autant de paires d'APs que la taille de la population le permet. les antagonistes d'une transaction sont sélectionnés par un tirage uniforme. Autrement dit, il y a une équiprobabilité pour qu'un AP interagisse avec n'importe lequel de ses partenaires dans la société. Si nous voulions, justement, restreindre les communications d'un AP, ce serait sur cette procédure de sélection de paires qu'il faudrait apporter des restrictions.

Passons maintenant à la description des critères adoptés pour apprécier la convergence des APs, ainsi que ceux adoptés pour l'identification du lexique en commun.

II.6.1 Remplacement

Pour accroître les performances et accélérer la convergence des simulations, nous reprenons le concept de remplacement dans les algorithmes génétiques (Goldberg 1989). Nous proposons de supprimer, après un nombre fixe d'échanges, que nous appellerons *époque*, le «plus mauvais individu», et de le remplacer dans la société par le «meilleur individu». En effet, nous avons constaté dans des premières simulations que souvent, la population tardait à converger (ou ne convergeait pas du tout) en raison d'un individu, tel un minimum local, dont les symboles éprouvaient du mal à se stabiliser et oscillaient d'une position à une autre. Un tel individu possède, par exemple, des doublons (items différents mais très proches sur le plan acoustique). De pareils agents pénalisent la société et perturbent même fortement la convergence de la population.

La «qualité» de chaque individu est évaluée en calculant la somme des ajustements qu'il a effectués au long d'une époque. En effet, plus un robot est mauvais, plus il tarde à opter pour une position stable et plus ses symboles basculent d'une position à une autre ; inversement pour un bon individu. Le plus mauvais individu est naturellement celui dont les symboles ont subi le plus d'ajustements.

Notons que le remplacement n'est pas un principe capital pour notre modèle. En outre, il n'a aucune interprétation cognitive. Il a juste été introduit dans un souci d'accélérer la convergence des populations.

II.6.2 Présentation des résultats

Nous considérerons qu'une population aura convergé lorsque tous les APs auront sensiblement le même lexique. Dans ce cas, les inventaires des robots n'auront plus la possibilité d'évoluer. Un critère de convergence pourrait être : la somme des ajustements des APs au cours des n derniers échanges est inférieur à un seuil donné.

Une fois la convergence établie, afin d'analyser et d'exploiter les résultats de la simulation, nous devons pouvoir extraire d'une population son lexique moyen qui rend le mieux compte des configurations des symboles de ses individus.

Dans ce sens, nous avons élaboré une routine d'*appariement*. Cet algorithme est fondé sur une technique semi-empirique qui tend à minimiser la somme des distances des items à leurs moyennes correspondantes. Partant de chaque item de l'inventaire d'un AP donné, le principe de cette routine est, en d'autres termes, de trouver pour tous les autres APs, leur item le plus proche de cet item et de les faire apparier. Il ne nous restera plus qu'à calculer les moyennes des items appariés pour obtenir le lexique moyen représentatif de l'évolution de la société au terme d'une simulation. Un exemple d'appariement est illustré sur la figure II.10.

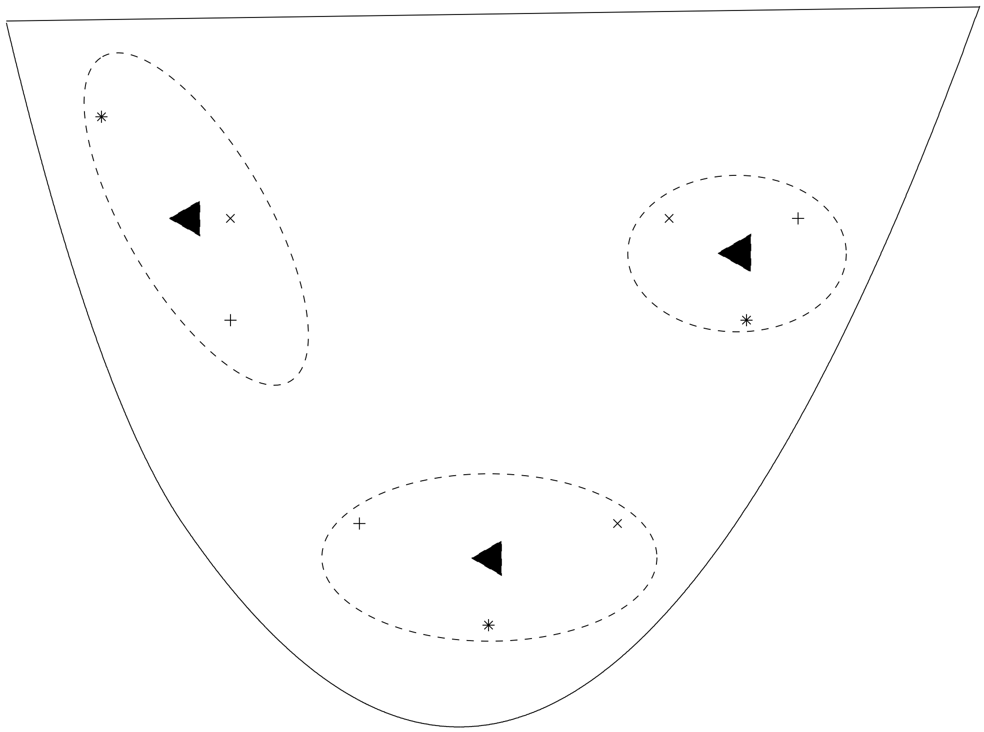


FIG. II.10: Exemple d'appariement : Les symboles (* + x) représentent les items de 3 individus. (1) apparier les items en créant autant de classes qu'un individu a de symboles ; cet appariement est schématisé par les ellipsoïdes. (2) les (▲) illustrent le lexique moyen calculé en effectuant la moyennes des items appariés.

II.7 Les paramètres du modèle ESPECE

Tandis que dans les paragraphes précédents, nous nous sommes attachés a décrire le noyau du modèle de simulation ESPECE, nous allons passer maintenant à la description des sept paramètres sur lesquels ce modèle repose :

1. la taille de la population (nombre d'APs),
2. la taille des lexiques des agents (nombre de symboles de chaque agent),
3. le nombre de transactions constituant une simulation,
4. la force de répulsion : ce paramètre permet de faire varier l'intensité de la force de répulsion dans l'équation (II.10),
5. le nombre de transactions constituant une époque au terme de laquelle il sera procédé à un remplacement (cf. § II.6.1),
6. la semence du générateur de nombres aléatoires : les lexiques sont initialisés aléatoirement et les paires locuteur-auditeur sont également sélectionnées aléatoirement, ce paramètre permet de générer des jeux de simulations différents en prenant des semences différentes ; aussi et ce qui est considérable, il est possible de répliquer des simulations en lui affectant une même valeur,
7. la facteur λ pondèrant la dimension F_1 par rapport à F_2.

II.8 Algorithme d'une simulation du modèle ESPECE

L'algorithme d'une simulation type du modèle ESPECE est donné par l'organigramme de la figure II.11.

La première étape consiste à initialiser les lexiques des agents de la population. Puis, des paires d'agents sur lesquels s'effectueront des transactions sont sélectionnées. Au bout de chaque époque, il sera procédé à un remplacement du plus mauvais individu par le meilleur. Enfin, la simulation prendra fin après quelques milliers de transactions. Le choix de ce critère d'arrêt sera argumenté dans le chapitre suivant, plus précisément en § III.2. Finalement, pour réaliser une étude de cas, il faut pouvoir accomplir un ensemble de simulations de manière à pouvoir établir des statistiques

II.9 Conclusion

Nous avons défini dans ce chapitre la structure d'un agent de parole, ainsi que la notion de société d'agents. Par ailleurs, nous avons décrit une transaction entre deux

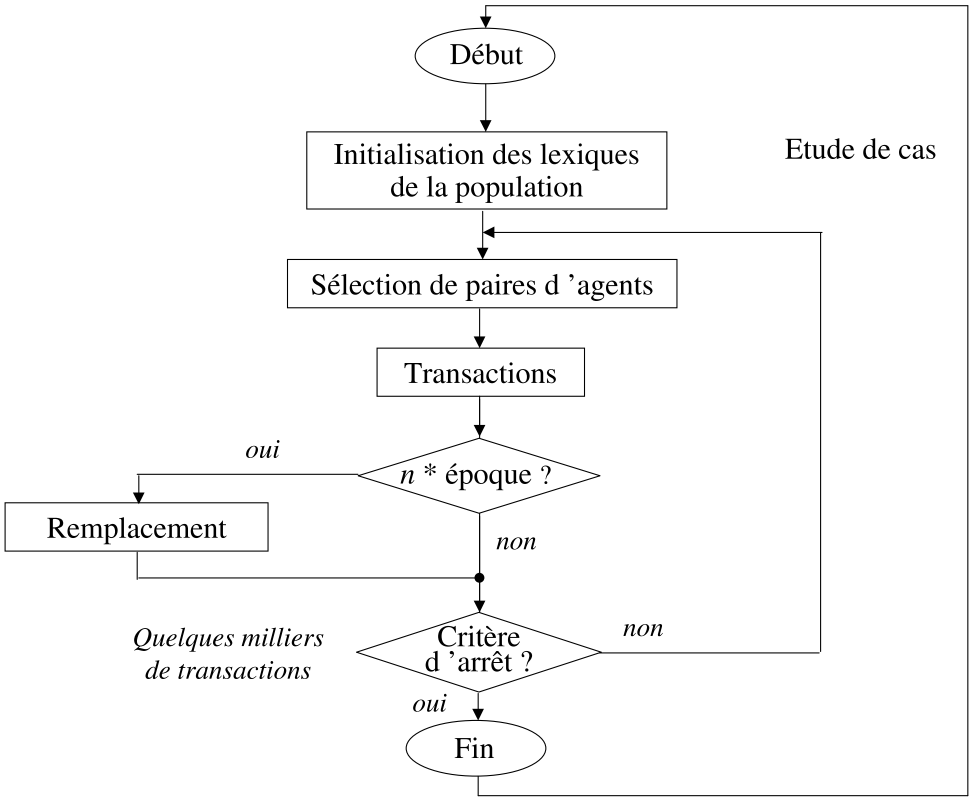


FIG. II.11: Organigramme d'une simulation type du modèle ESPECE.

agents. et nous avons détaillé les processus d'attraction, de répulsion et de calcul de distances perceptives qu'utilise une transaction. Nous avons présenté également les principes généraux, les paramètres du modèle d'interactions entre agents que nous avons baptisé ESPECE, ainsi que l'algorithme d'une simulation.

La partie restante de ce mémoire sera dédiée à la présentation des applications du modèle ESPECE sur l'émergence des systèmes vocaliques des langues du monde. Ces applications auront pour but de valider le modèle proposé.

Chapitre III

Le modèle ESPECE pour la prédiction des systèmes vocaliques

Résumé

Ce chapitre sera consacré au modèle décrit dans le chapitre II, appliqué à la prédiction des systèmes vocaliques. Nous décrirons les caractéristiques, ainsi que les paramètres, de ce modèle. Nous présenterons ensuite les résultats des évolutions de plusieurs sociétés d'agents. Ces prédictions, ainsi que celles obtenues avec différentes variantes du modèle, seront confrontées aux données obtenues à partir des inventaires phonologiques des langues du monde (base UPSID). Il découlera, de cette comparaison, une large discussion. Par ailleurs, nous procéderons à une étude de l'effet des nombreux paramètres du modèle sur les prédictions. Ces analyses permettront non seulement de valider le modèle, mais également de recenser les limites du cadre de simulation.

III.1 Introduction

La première application du modèle ESPECE que nous allons explorer est celle de la prédiction des systèmes vocaliques. Plusieurs travaux dans la lignée de ceux de Liljencrants et Lindblom (1972), portant sur la prédiction des systèmes vocaliques d'une part (Lindblom 1986, Lindblom 1990, Schwartz et al. 1997*b*), et des systèmes syllabiques d'autre part (Lindblom et al. 1984, Berrah, Boë et Schwartz 1995, Boë, Schwartz, Berrah, Vallée et Abry 1996, Berrah, Glotin, Laboissière, Bessière et Boë 1996) ont été réalisés. Ces modèles prédictifs associent à un système de sons donné une fonction d'énergie pouvant être évaluée par des principes classiques d'optimisation, de façon à sélectionner les systèmes optimaux. Cette énergie est fonction des coûts perceptifs et des coûts articulatoires caractérisant chaque son du système.

Ces études ont proposé, pour l'explication de la structure des systèmes sonores, des règles générales permettant d'établir des relations de préférence entre les systèmes phonétiques.

La méthode que nous adoptons se différencie principalement de celle de ces travaux par le fait qu'elle n'est pas basée sur des critères d'optimisation globale, mais sur l'interaction locale entre deux agents.

III.2 Description du modèle et de ses paramètres

Dans le modèle ESPECE appliqué à la prédiction des systèmes vocaliques (Berrah 1997, Berrah et Laboissière 1997), les voyelles composant l'inventaire des APs sont représentées par leur formants (fréquences de résonances acoustiques). Initialement, les symboles des inventaires reçoivent tous la configuration formantique du neutre ayant subi une légère variation. Au cours des transactions, les configurations des symboles, objets des communications inter-agents, évoluent mais tout en s'inscrivant au sein des bornes du triangle vocalique décrit en section II.4. La coopération des agents, lors des échanges, est basée sur les notions d'attraction et de répulsion. Les expressions des forces d'attraction, de répulsion ainsi que des distances perceptives sont identiques à celles décrites en II.5. Aussi, adoptons-nous la même procédure que celle illustrée en II.5.4 pour faire glisser un symbole sur les bords du triangle vocalique lorsqu'il aura été repoussé en dehors du triangle.

L'ensemble des simulations réalisées a été effectué sur une société de 10 individus. Le nombre de transactions est fixe et a pour valeur 3 000. Au cours d'une transaction, 5 paires (soit la moitié de la taille de la population) sont aléatoirement sélectionnées et des échanges entre ces 5 paires sont réalisés en parallèle. Ainsi, au terme d'une simulation, chaque AP aura pris parti dans 3 000 échanges. Soulignons que pour la version parallélisée du modèle ESPECE utilisée, une transaction est constituée de plusieurs échanges effectués en parallèle.

En outre, nous avons relevé pour chaque simulation, toutes les 100 transactions, la somme des ajustements effectués par l'ensemble des agents, suite aux attractions et aux répulsions subies. En effet, comme il a déjà été expliqué dans § II.6.2, la convergence d'une société est appréciée par la stabilité des APs. En d'autres termes, lorsque le facteur «somme des ajustements» se stabilise, nous pouvons considérer sans trop prendre de risques que la population a convergé.

La figure III.1 illustre bien cet effet. Les évolutions de ce facteur pour trois simulations y sont représentées. Nous constatons que c'est lors des premières transactions que la population subit l'essentiel des modifications avant d'opter pour une configuration stable.

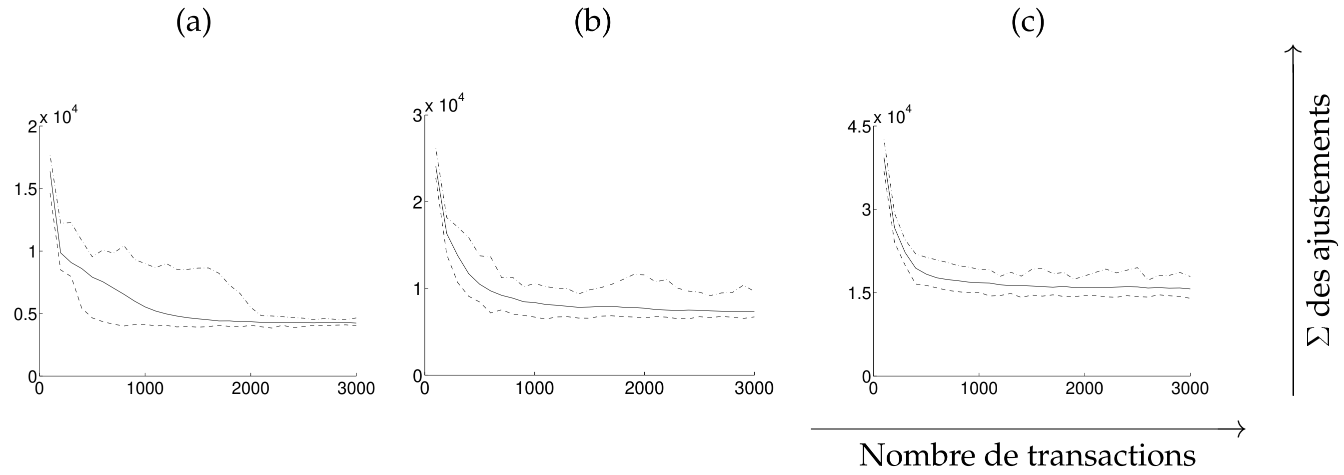


FIG. III.1: Évolution de la somme des ajustements au long des transactions de 3 simulations : (a) avec 4 individus, (b) avec 6 individus et (c) avec 7 individus ; la courbe en trait plein représente la moyenne des valeurs de la somme des ajustements pour l'ensemble des simulations, la courbe en pointillés schématise les valeurs minimales alors que les valeurs maximales sont illustrées par la courbe en traits discontinus.

Pour l'ensemble des simulations décrites ci-après, la force de répulsion a pour valeur 0.5 et une époque est constituée de 100 transactions. Au terme d'une époque, chaque agent aura pris part dans 100 échanges, et aura par conséquent, interagi une dizaine de fois, en moyenne, avec chacun des autres agents. Chaque agent aura de ce fait la possibilité d'échanger, en moyenne, toutes ses voyelles avec chacun des autres agents. Nous estimons alors qu'il est raisonnable d'évaluer la société toutes les 100 transactions et de remplacer le plus mauvais individu par le meilleur (cf. § II.6.1).

La nécessité de pondérer F_1 par rapport à F_2 est justifiée et est indiscutable pour les raisons in-voquées en § I.4.2. Par contre, le débat reste ouvert concernant la valeur à affecter au facteur de pondération λ. La seule propriété faisant l'unanimité est uniquement que λ doit être inférieur à 1. Les tests proprioceptifs de Lindblom & Lubker (1975) montrent que les estimations des sujets concernant les mouvements de la mâchoire couvrent un éventail de valeurs allant de de 0.2 à 2, tandis que celles des mouvements de la langue englobent les valeurs allant de 0.5 à 1.6 (cf. figure III.2). Ceci signifie un accroissement de 10 pour les mouvements de la mâchoire et de 3 pour ceux de la langue. De ce fait, nous serions assez tentés d'affecter à λ le rapport entre ces deux accroissements ($3/10 = 0.3$). Certes, il existe une préservation de la topologie entre le triangle présenté dans la figure III.2 et le triangle acoustique. Cependant, il reste à prouver qu'il est correct de se baser sur des expériences articulatoires pour attribuer une valeur à un paramètre destiné à être utilisé dans une équation perceptive.

Des études acoustiques telle que celles de Delattre et al. (1952) montrent que les formants de basse fréquence sont mieux perçus que les formants de haute fréquence. Hélas, malgré les efforts fournis durant ces deux dernières décennies, les connaissances à propos des représentations et des distances auditives sont insuffisantes à l'heure actuelle, elles ne permettent pas de justifier la valeur de λ.

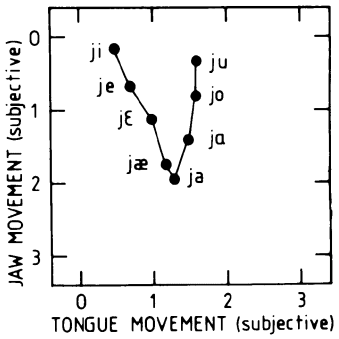


FIG. III.2: Estimations subjectives des mouvements de la langue et de la mâchoire (d'après Lindblom & Lubker (1985)).

Pour notre part, nous nous sommes inspirés des résultats expérimentaux de la TDF (cf. § I.4.3) dans laquelle les systèmes vocaliques préférentiels sont prédits lorsque λ est compris entre 0.2 et 0.3 et nous avons fixé pour ce premier jeu de simulations λ à 0.3.

III.3 Prédictions obtenues et comparaison avec les données d'UPSID

Nous avons réalisé des simulations sur la base du modèle ESPECE sur des sociétés d'APs avec des inventaires de 3 à 9 symboles. Nous avons effectué 40 simulations différentes de manière à pouvoir établir des statistiques correctes pour chaque taille de lexique. Chacune correspond à une semence donnée du générateur de nombres aléatoires. Au terme de chacune de ces simulations, nous avons calculé la moyenne des inventaires de la population en utilisant la routine d'appariement décrite en § II.6.2. Cette moyenne représente le lexique type des individus de la société, vu que l'ensemble des APs a sensiblement la même configuration après convergence. Puis, nous avons regroupé les moyennes similaires pour les simulations de chaque taille de lexique. Enfin, nous avons une nouvelle fois calculé la moyenne des lexiques communs regroupés, ce qui constitue un système vocalique généré par ESPECE.

Présentons, pour le moment, les résultats des prédictions du modèle ESPECE que nous discuterons par la suite. Les systèmes vocaliques résultant de ces simulations, ainsi que les systèmes répertoriés fréquemment dans la base UPSID_{317}, peuvent être lus sur les figures III.3 à III.10. Sur chaque figure, les lexiques moyens des simulations sont représentés par des «+», alors que les circonférences «O» indiquent le système vocalique obtenu en effectuant la moyenne de ces lexiques. Le pourcentage figurant à gauche de chaque système représente la proportion des simulations ayant cette configuration. Il convient de noter que ces pourcentages sont, en fait, des multiples de 2.5 % vu qu'ils ont été calculés sur la base 40 simulations. Aussi, nous nous garderons d'interpréter la présence de décimaux dans ces pourcentages comme une précision de l'ordre du dixième dans les statistiques.

III.3.1 Systèmes à 3 voyelles

Pour les lexiques à 3 voyelles, la société a opté systématiquement pour le système /i a u/ (cf. figure III.3). Ce dernier est rencontré dans 60 % des systèmes à 3 voyelles de la base UPSID_{317} (11 langues telles le *Groelandais*).

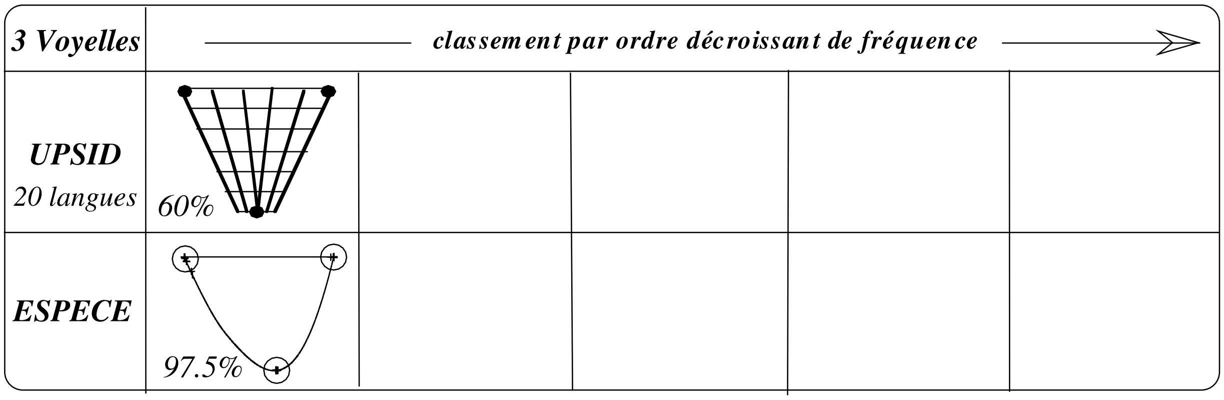


FIG. III.3: Différentes prédictions de systèmes vocaliques à 3 voyelles.

III.3.2 Systèmes à 4 voyelles

Lorsque les APs ont des inventaires de 4 symboles, nous observons deux systèmes différents (cf. figure III.4).

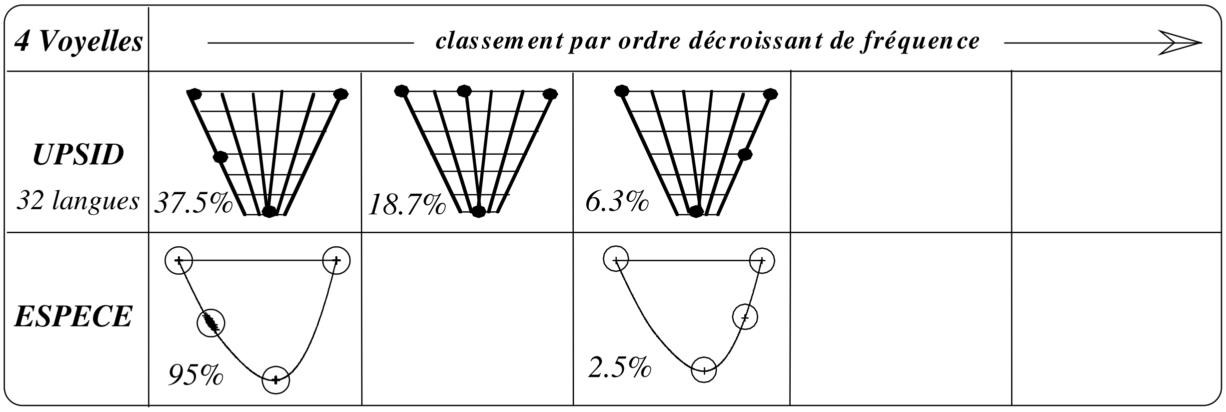


FIG. III.4: Différentes prédictions de systèmes vocaliques à 4 voyelles.

Le modèle ESPECE a abouti le plus souvent (95 % des cas) à une variante du système /i ˈeˈ a u/. Ce système est celui à 4 voyelles le plus commun dans les langues du monde (12 langues de la base UPSID$_{317}$ telles le *Shasta*). Nous retrouvons aussi une occurrence du système /i a ˈoˈ u/ ; ce système est observé dans la langue *Tiwi*.

III.3.3 Systèmes à 5 voyelles

Deux systèmes différents (cf. figure III.5) sont essentiellement dénombrés après convergence des sociétés avec des robots ayant cinq symboles.

Les systèmes à 5 voyelles sont les plus observés dans les langues du monde. Ils représentent à eux seuls 42.20 % des systèmes d'UPSID$_{317}$. Le système phare est /i ˈeˈ a ˈoˈ u/ qu'on retrouve dans pas moins de 88 langues parmi lesquelles l'*Espagnol* notamment. Ce système est aussi le fruit de 87.5 % des simulations d'ESPECE. Autrement, ESPECE a engendré des systèmes /i ˈeˈ a u ɨ/ (7.5 %). Ce système est observé dans 6 langues, citons l'*Acoma* à titre d'exemple.

III.3.4 Systèmes à 6 voyelles

Trois systèmes ont émergé des simulations d'ESPECE avec des lexiques de taille six (cf. figure III.6).

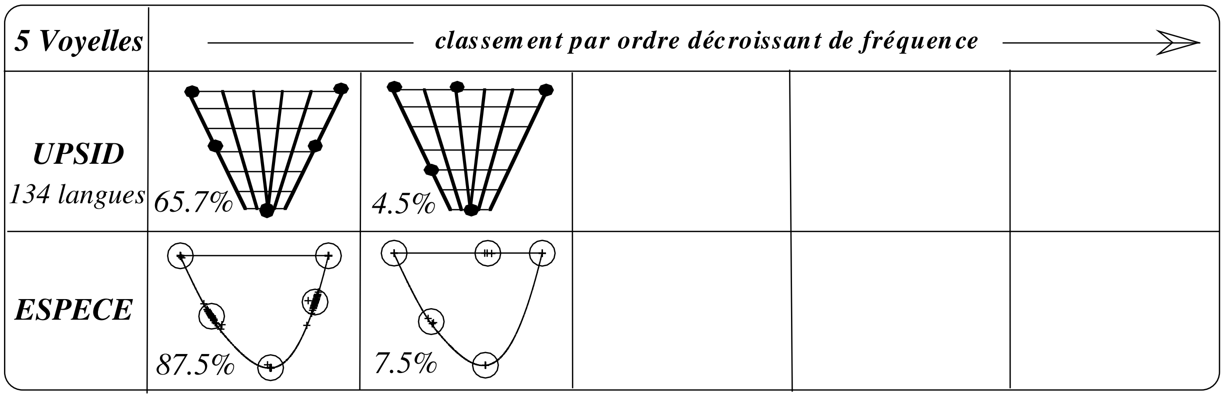


FIG. III.5: Différentes prédictions de systèmes vocaliques à 5 voyelles.

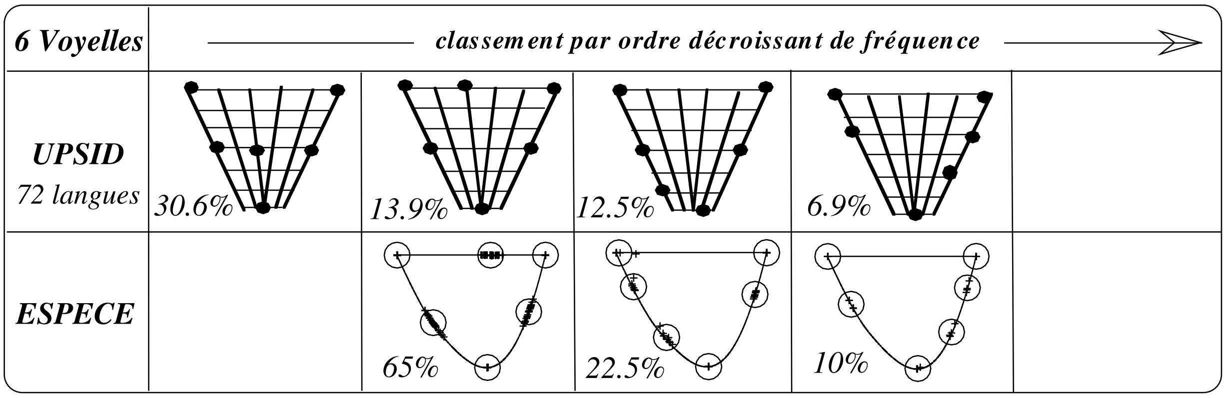


FIG. III.6: Différentes prédictions de systèmes vocaliques à 6 voyelles.

Notons que le système qui a le plus souvent émergé est également une variante du second système le plus fréquent (dix langues telles le *Lapon*) des langues ayant 6 voyelles. Ce système est /i ˈeˈ a ˈoˈ u ɨ/. En seconde position, nous retrouvons une variante du système /i e ɛ a ˈoˈ u/ qui est recensé dans neuf langues répertoriées dans la base UPSID$_{317}$ dont le *Chamorro*. Dans les autres cas, ESPECE a abouti au système /i ˈeˈ a ɔ o u/. Ce dernier est rencontré dans cinq langues. Le système le plus fréquent présente une voyelle centrale (le schwa /ə/) qui est une voyelle présentant quelques singularités, comme nous allons le décrire dans § III.3.8.

III.3.5 Systèmes à 7 voyelles

Pour des inventaires de taille sept, la société a convergé vers des variantes des trois systèmes suivants :

- le système /i ˈeˈ a ɔ o u ɨ/ représenté par la figure III.7 dans 42.5 % des simulations ;

ce système n'est pas observé dans les langues du monde (cf. figure III.8),

- le système /i e ɛ a ˈoˈ u ɨ/ (35 %) rencontré dans deux langues (l'*Awiya* et le *Bardi*),
- le système à 7 voyelles le plus fréquent, à savoir /i e ɛ a ɔ o u/ a émergé dans 22.5 % des simulations ; ce système est observé dans 20 langues dont l'*Albanais* notamment.

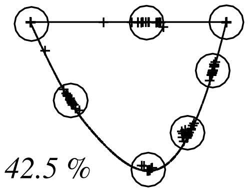


FIG. III.7: Système à 7 voyelles le plus prédit par ESPECE.

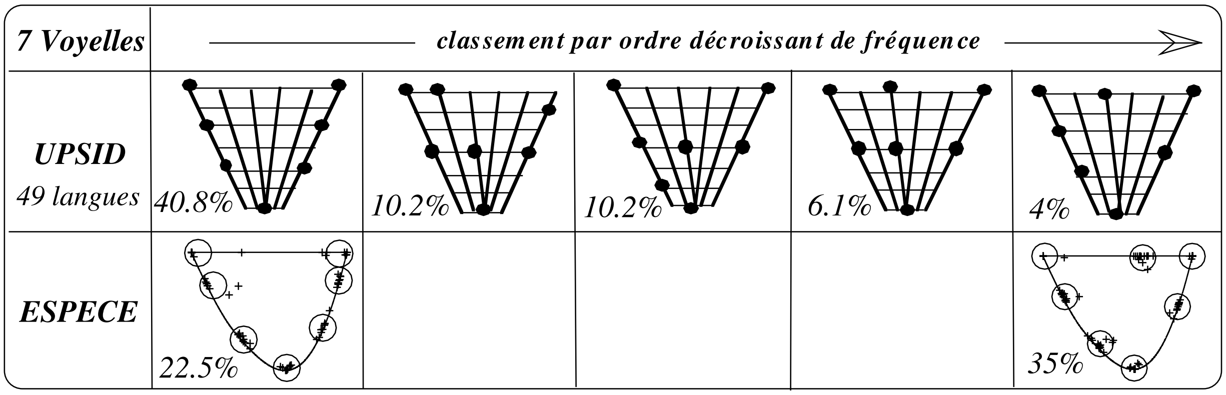


FIG. III.8: Différentes prédictions de systèmes vocaliques à 7 voyelles.

III.3.6 Systèmes à 8 voyelles

En fixant la taille des lexiques à huit, nous remarquons la prédominance d'un système avec des voyelles réparties sur tout le contour du triangle vocalique. Bien que ce système ne soit pas observé dans les langues du monde, il n'en demeure pas moins stable. Le système à 8 voyelles le plus fréquent (cf. figure III.9) présente une voyelle centrale assimilé au /ə/.

III.3.7 Systèmes à 9 voyelles

Nous remarquons, pour les lexiques de taille neuf, l'émergence de systèmes de voyelles distribuées également sur la périphérie du triangle vocalique pour les mêmes raisons

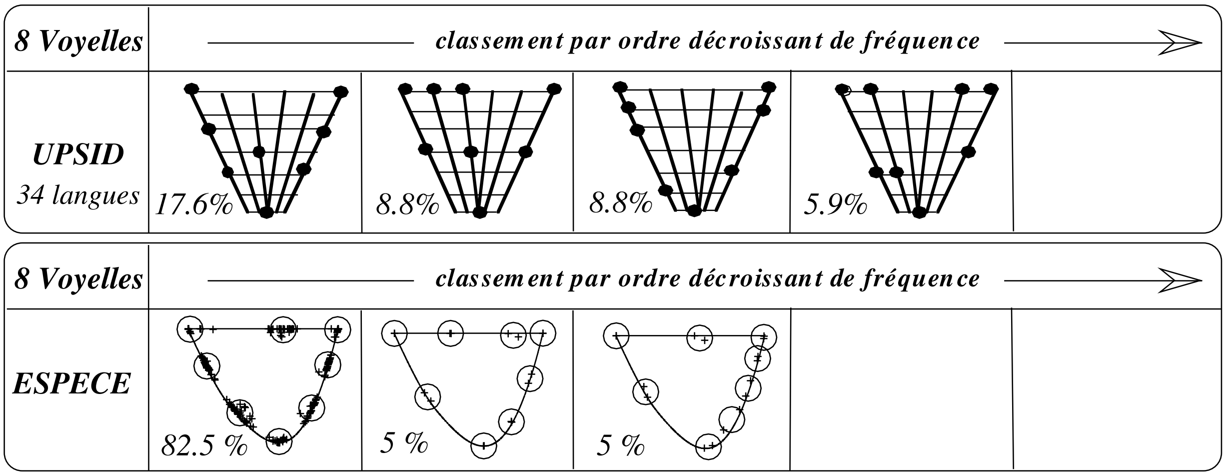


FIG. III.9: Différentes prédictions de systèmes vocaliques à 8 voyelles.

que celles invoquées pour les systèmes à huit voyelles. Le système le plus fréquent dans la base $UPSID_{317}$ (cf. figure III.10), ne contenant pas de voyelle périphérique haute, aurait été difficilement généré par ESPECE, surtout pour une valeur de λ fixée à «0.3».

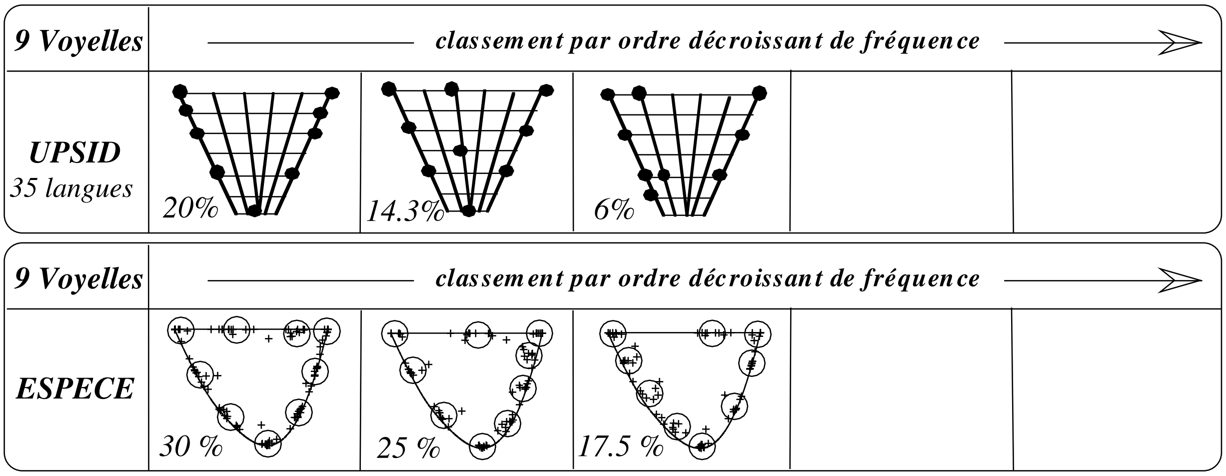


FIG. III.10: Différentes prédictions de systèmes vocaliques à 9 voyelles.

III.3.8 Discussion

Les prédictions d'ESPECE pour les systèmes à trois, quatre et cinq voyelles sont conformes aux tendances universelles des langues. Pour les systèmes à six voyelles, le modèle a produit des systèmes rencontrés dans la base $UPSID_{317}$. Il est important de noter que nos simulations ont produit des systèmes vocaliques dans lesquels les symboles sont essentiellement placés sur les périphéries /i/–/e/–/a/ et /a/–/o/–/u/. À partir de sept voyelles, la voyelle centrale fermée apparaît systématiquement. Ceci est dû au fait que les systèmes avec une telle voyelle sont stables dans le plan formantique (F_1,F_2). Les voyelles centrales (proches de /ə/) sont inexistantes, ce qui est en contradiction avec les statistiques sur les langues du monde.

Les théories calculatoires pour les prédiction des systèmes vocaliques (cf. § III.1) présentent seulement les systèmes optimaux. Par contre, le modèle ESPECE permet de générer des statistiques sur des systèmes. En effet, pour chaque taille de l'inventaire des agents, différents systèmes sont proposés avec un certain pourcentage comme c'est le cas pour les typologies. À titre d'exemple, ESPECE a généré trois systèmes à 6 voyelles rencontrés dans les langues du monde (cf. figure III.6).

Pour de grandes tailles d'inventaires (≥ 10), les sociétés éprouvent des difficultés pour converger vers un lexique unique. Nous ne présentons pas les résultats de ces simulations car en fait, les systèmes avec plus de neuf voyelles dans les langues du monde utilisent des traits supplémentaires. L'étude des caractéristiques de ces systèmes fera l'objet du chapitre IV.

Maintenant que nous avons présenté les remarques d'ordre général sur les prédictions obtenues, nous nous proposons d'aborder, plus précisément cette fois, les trois point suivants :

- l'occupation avant/arrière de la périphérie,
- le cas du schwa,
- l'influence du facteur de pondération λ sur les prédictions des systèmes à 7 voyelles.

l'occupation avant/arrière de la périphérie

L'examen des données typologiques montre que les langues du monde ont une nette préférence pour les systèmes symétriques[1] et, par suite, en comptant la voyelle basse, les langues ont tendance à opter pour un nombre impair de voyelles périphériques. Un second fait confirmé par ces données stipule que lorsqu'il y a asymétrie, les systèmes comprennent, généralement, plus de voyelles gauches (avants) que de voyelles droites (arrières).

Conformément à la TDF, le modèle ESPECE vérifie cette tendance à prédire plus souvent des systèmes avants que des systèmes arrières bien que les espaces perceptifs adoptés par ESPECE et la TDF aient des formes différentes. L'espace (F_1,F'_2) adopté par la TDF et schématisé par la figure III.11 est plus étiré à gauche. Nos APs évoluent, quant à eux, dans un espace acoustique (F_1,F_2), qui a une allure plus étirée à droite

[1]Inventaires ayant autant de voyelles avants que de voyelles arrières.

dans l'espace des Barks comme le montre la figure II.9. Ceci trouve une explication dans le rôle important joué, en plus de la dispersion, par le glissement le long de la périphérie du triangle vocalique (cf. § II.5.4) dans notre modèle. En effet, de par l'effet du glissement, les symboles projetés sur la périphérie arrière (droite) du triangle ont tendance à glisser plus rapidement vers le bas (zone du /a/) que ceux repoussés sur la périphérie avant (gauche).

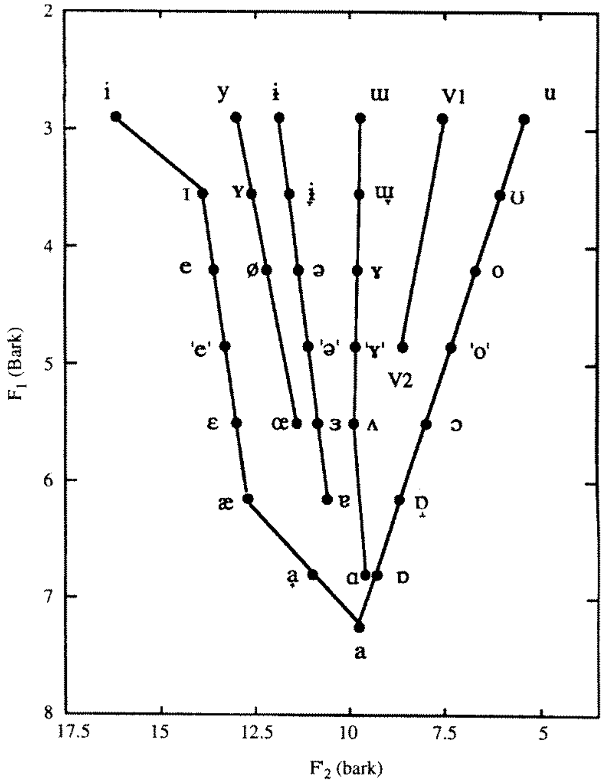


FIG. III.11: Prototypes vocaliques dans l'espace acoustique (F_1,F'_2) (d'après Schwartz et al. (1997*b*).

Le cas du schwa

La première difficulté des théories calculatoires (TD (cf. § I.4.2), TDF (cf. § I.4.3)) réside en l'impossibilité de prédire des systèmes avec une seule voyelle non périphérique en la voyelle centrale /ə/ (le schwa), et ce malgré leur occurrence dans les systèmes les plus fréquents à 6 et 8 voyelles. Le schwa est une voyelle instable au cours de l'évolution des langues, de plus courte durée par rapport aux autres voyelles. De plus, cette voyelle possède plusieurs allophones. Il en découle que les phonologues transcrivent régulièrement /ə/ à chaque fois qu'une voyelle centrale est rencontrée dans les langues du monde. En outre, Schwartz et al. (1997*a*) observent que le schwa est une voyelle «transparente» dans un système sonore, dans le sens où elle n'intervient pas dans la position des autres voyelles de l'inventaire. Toutes ces particularités ont amené Schwartz et al. (1997*a*) à supposer qu'un système, dans lequel figure le schwa, évolue à partir du système correspondant sans schwa, du fait d'une sorte de procédure de *relaxation systémique*, basée sur la réduction vocalique. Ce qui a conduit les auteurs à conclure que le schwa devrait être écarté des études sur les prédictions d'une théorie acoustico-perceptive des systèmes vocaliques.

Influence du facteur de pondération λ sur les prédictions des systèmes à 7 voyelles

Le système à 7 voyelles le plus commun dans les langues du monde est le système /i e ɛ a ɔ o i/. Or, ce système n'a été prédit que dans 22.5 % des simulations du modèle ESPECE. En effet, la majorité des systèmes prédits par ESPECE comportent une voyelle centrale fermée pouvant être assimilée à la voyelle /ɨ/.

Depuis les prédictions des modèles reprenant les fondements de la théorie de la Dispersion (cf. § I.4.2), et plus tard ceux de la TDF (cf. § I.4.3), il est établi qu'une valeur trop élevée du facteur de pondération λ fait apparaître des voyelles centrales fermées. En revanche, les systèmes périphériques sont favorisés en diminuant λ (cf. figure III.12).

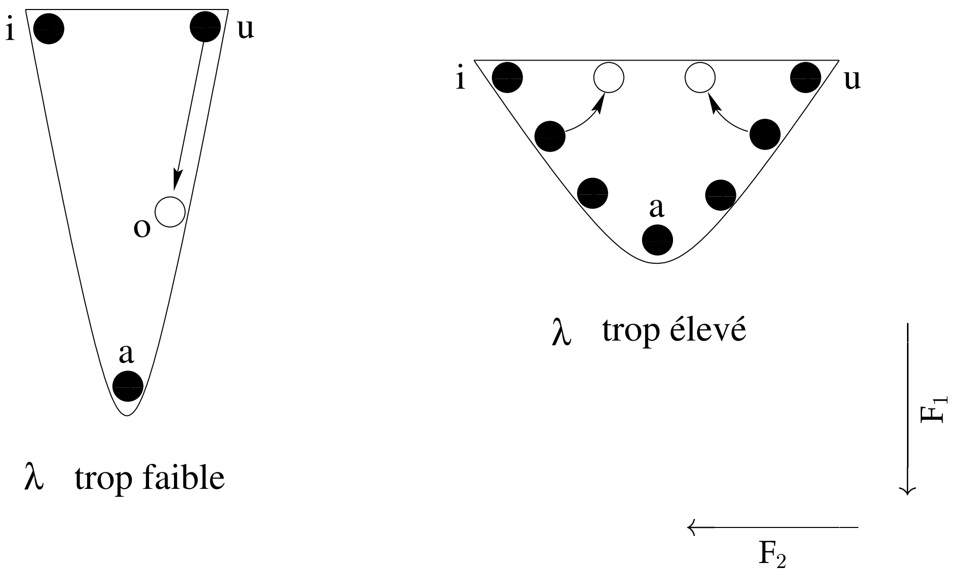


FIG. III.12: Effet du facteur λ : Un valeur trop élevée favorise les voyelles centrales, alors qu'une valeur trop petite pénalise la voyelle /u/.

En observant l'espace des phases de la TDF pour 7 voyelles, nous nous apercevons que ce système périphérique n'est prédit que pour des valeurs de λ comprises entre environ 0.12 et 0.22. Aussi, nous semble-t-il intéressant d'examiner les prédictions d'ESPECE pour des valeurs de λ plus petites que 0.3.

La figure III.13, ci-après, illustre les prédictions pour des valeurs de λ égales à 0.3 et à 0.2. Nous constatons que le système périphérique est celui qui émerge le plus lorsque λ vaut 0.2, ce qui est en concordance avec les résultats de la TDF.

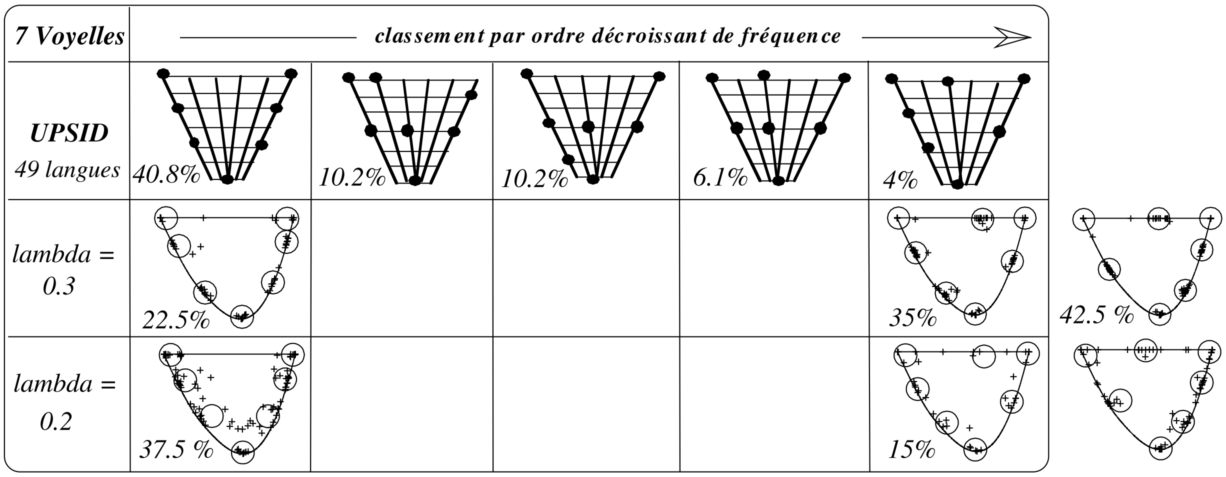


FIG. III.13: Influence du facteur λ sur les prédictions des systèmes à 7 voyelles.

La figure III.14 montre la stabilité des systèmes à 7 voyelles lorsque la population est initialisée par la configuration périphérique. Il est possible de constater que ce système n'est pas franchement stable pour une telle configuration.

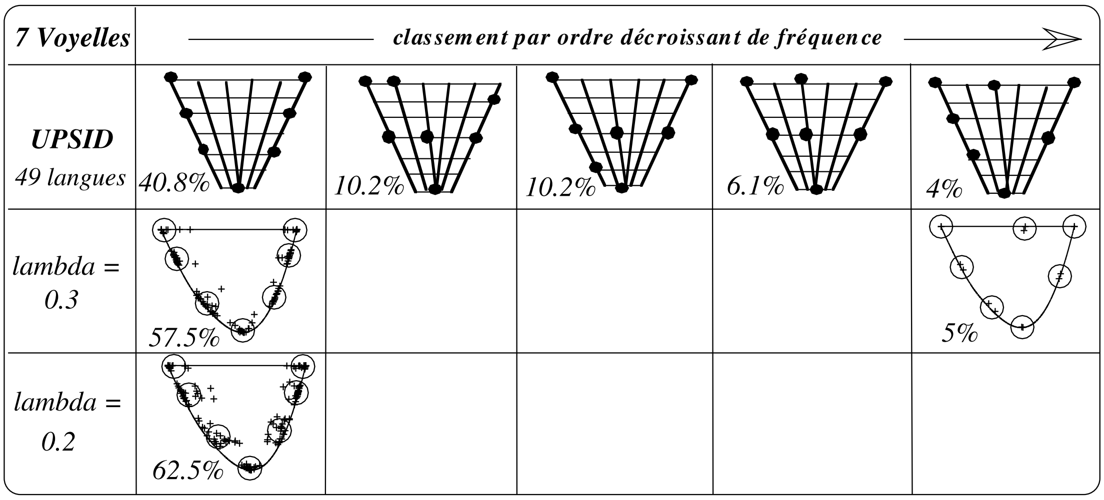


FIG. III.14: Stabilité des systèmes à 7 voyelles périphériques.

III.4 Variantes du modèle ESPECE

Les prédictions de systèmes vocaliques présentées ont été obtenues par le modèle ESPECE tel qu'il a été décrit dans § III.2. Nous nous proposons maintenant de tester plusieurs variantes de ce modèle. Plus précisément, nous décrirons , dans cette section, les résultats de quatre variantes : l'*initialisation aléatoire*, la *répulsion globale*, l'*attraction seule* et l'*espace acoustique en Hertz*.

Les prédictions présentées porteront sur des populations ayant des inventaires allant de 3 à 7 symboles. Nous avons préféré laisser de côté les lexiques de 8 et 9 symboles car leurs résultats sont peu significatifs lors des interprétations, du fait que leurs prédictions fournies par le modèle ESPECE ne reflètent pas les systèmes rencontrés dans les langues du monde.

III.4.1 L'initialisation aléatoire

Dans le modèle ESPECE, les symboles des agents de la société sont initialisés par des petites variations du neutre. Il serait opportun de savoir si cette initialisation a un impact quelconque sur les résultats, autrement dit, dans quelle mesure une initialisation différente induirait des modifications importantes sur les résultats. Dans ce sens, réalisons une batterie de simulations, mais cette fois, les symboles sont initialement générés aléatoirement dans le plan (F_1,F_2) délimité par le triangle vocalique.

Les prédictions du modèle ESPECE avec une initialisation aléatoire sont représentées sur les figures III.15 à III.19.

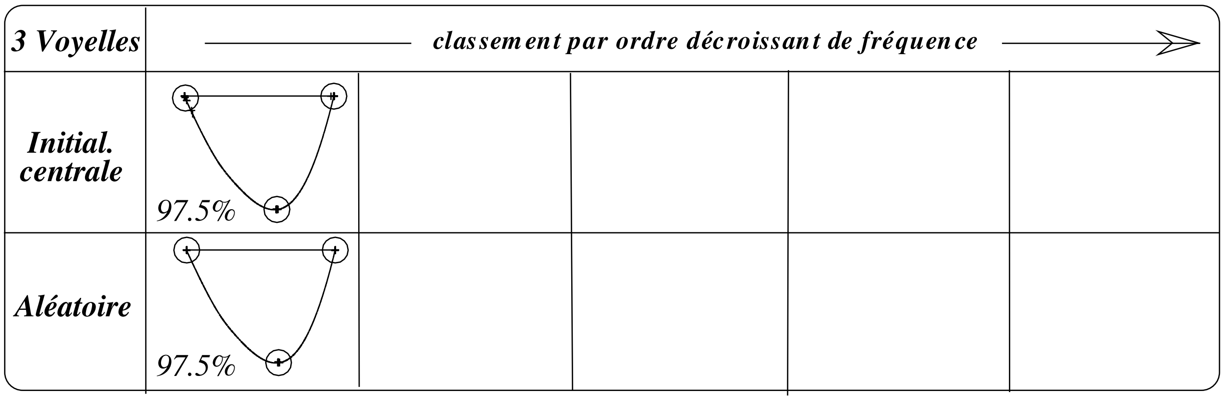


FIG. III.15: Distribution des voyelles des agents ayant un lexique de 3 voyelles dans le triangle vocalique en utilisant une initialisation aléatoire.

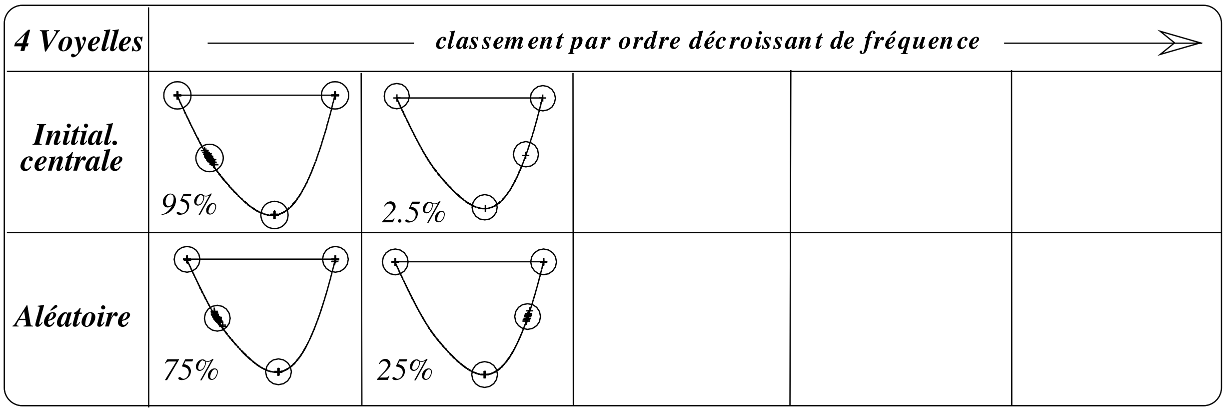


FIG. III.16: Distribution des voyelles des agents ayant un lexique de 4 voyelles dans le triangle vocalique en utilisant une initialisation aléatoire.

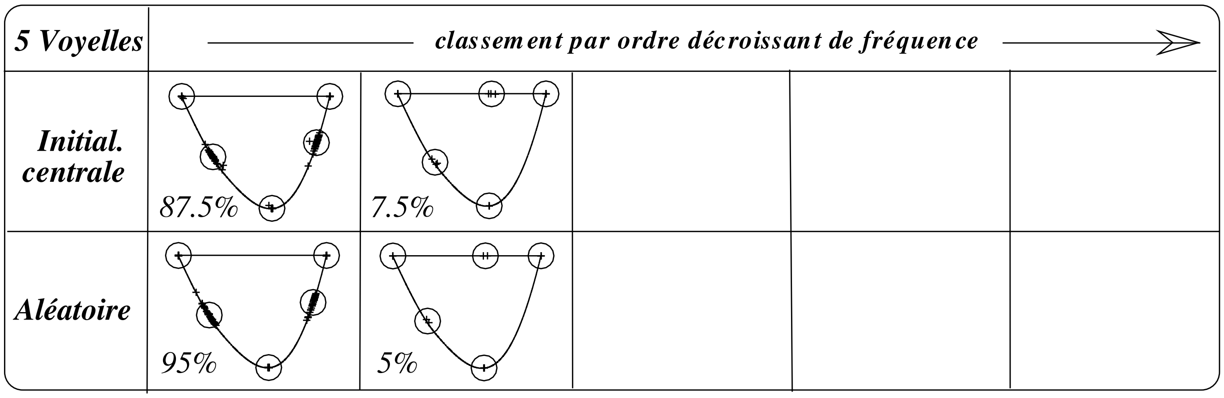


FIG. III.17: Distribution des voyelles des agents ayant un lexique de 5 voyelles dans le triangle vocalique en utilisant une initialisation aléatoire.

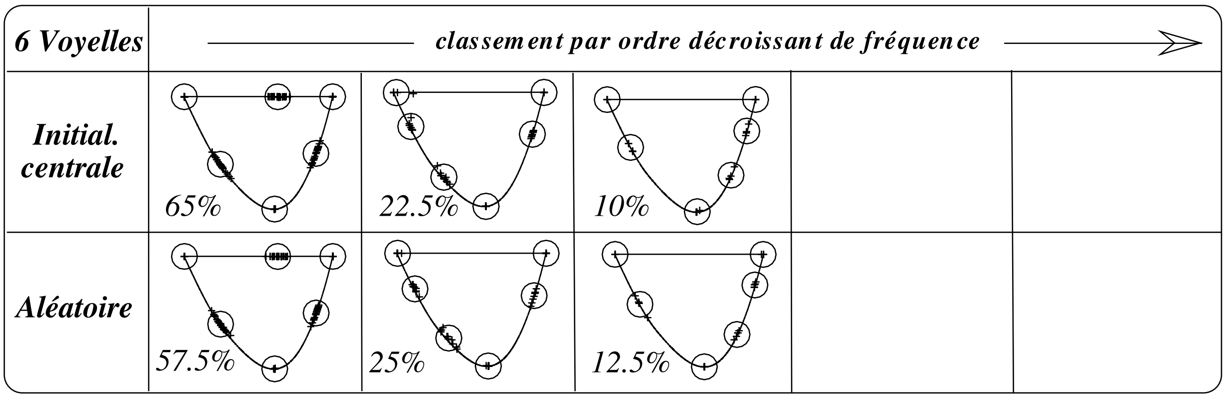


FIG. III.18: Distribution des voyelles des agents ayant un lexique de 6 voyelles dans le triangle vocalique en utilisant une initialisation aléatoire.

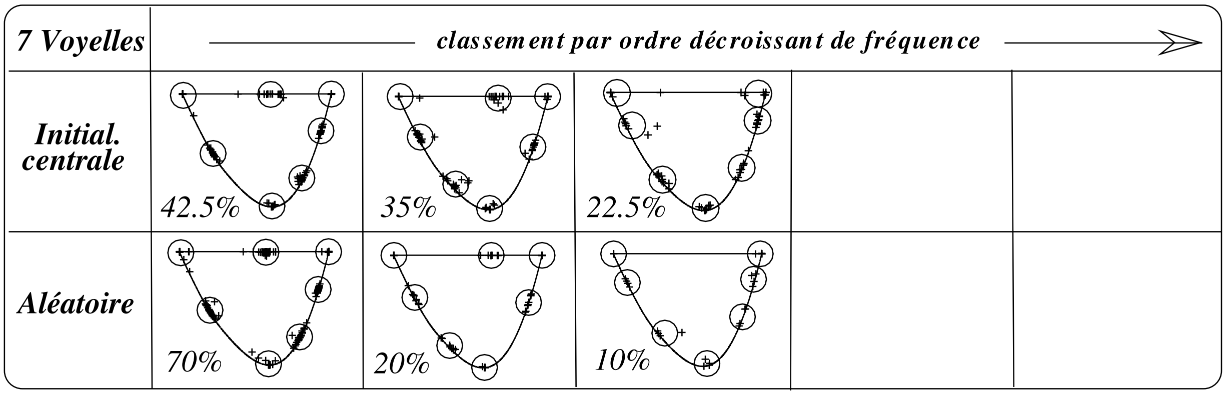


FIG. III.19: Distribution des voyelles des agents ayant un lexique de 7 voyelles dans le triangle vocalique en utilisant une initialisation aléatoire.

En observant les prédictions, nous pouvons noter qu'il n'y a pas de différences majeures entre ces systèmes et les systèmes émergeant du modèle ESPECE tel qu'il a été décrit auparavant. Cette constatation nous amène à dire que le mode d'initialisation des symboles des robots de la société n'a pas d'impact significatif sur les prédictions.

III.4.2 La répulsion globale

Comme son nom l'indique, la répulsion dans le modèle ESPECE consiste à repousser tous les symboles du lexique d'un AP auditeur de son symbole ayant subi une attraction (cf. § II.6.1). Nous nous sommes proposés d'étendre ce concept à celui de «répulsion globale». Dans un souci d'efficacité perceptive maximale, il s'agit d'effectuer, en plus de la répulsion par rapport au symbole attiré, une répulsion de tous les symboles, mis à part le symbole attiré, entre eux. La figure III.20 illustre bien ce concept : la répulsion que subit chaque symbole se résume à la résultante de la répulsion du symbole attiré (représentée sur la figure par une flèche pleine, le symbole attiré est schématisé par un carré noir) et de celles des autres symboles (représentées par des flèches non pleines).

Les résultats des simulations du modèle ESPECE avec répulsion globale sur des sociétés de 3, 4, 5, 6 et 7 symboles (cf. figures III.21 à III.25) montrent que ce sont les mêmes tendances de systèmes qui émergent, que ce soit en mettant en œuvre une répulsion simple ou une répulsion globale, à une réserve près sur le rapport avant/arrière pour les systèmes à 4 symboles,

En examinant de plus près les systèmes engendrés par ces simulations, nous observons que les *clusters* sont plus dispersés. Ce résultat trouve une interprétation dans le fait que dans ce mode de répulsion, les populations atteignent rapidement une configuration stable, qui n'est pas toujours identique.

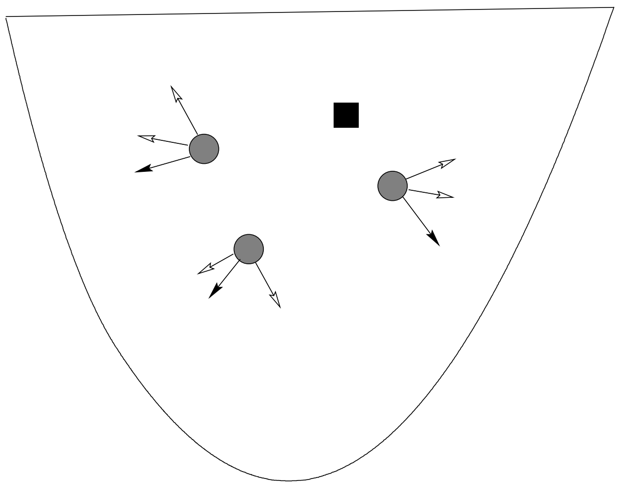


FIG. III.20: Schéma de la répulsion globale : Le carré noir représente le symbole attiré et les trois cercles schématisent les symboles à repousser ; chacun de ces trois symboles subit 3 répulsions : une de la part du symbole à attirer (illustrée par une flèche pleine) et les deux autres de la part des deux autres symboles (représentées par une flèche creuse).

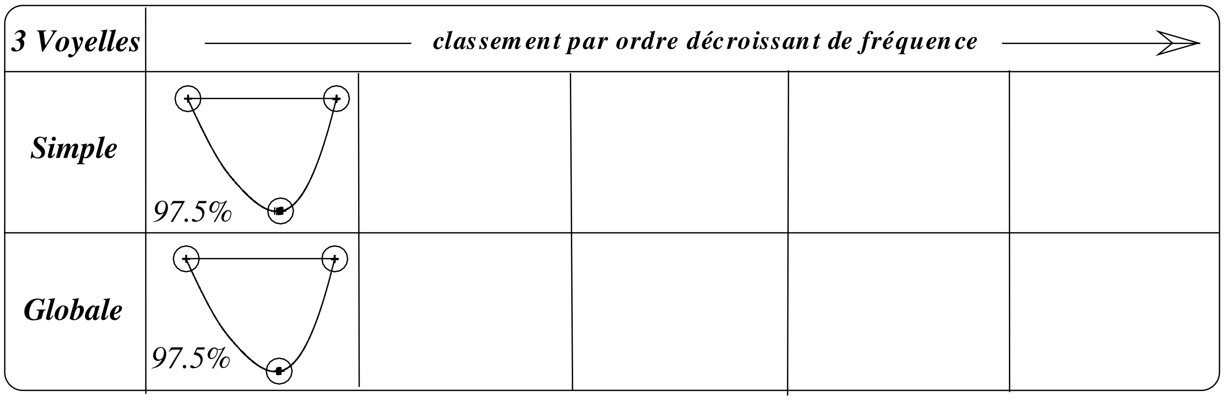


FIG. III.21: Distribution des voyelles des agents ayant un lexique de 3 voyelles dans le triangle vocalique en utilisant la répulsion globale.

En outre, lorsque la répulsion globale est mise en œuvre, le nombre de transactions nécessaire à une convergence est moindre. En revanche, la répulsion globale induit la réalisation d'opérations supplémentaires, ce qui influe sur le temps d'exécution des échanges. De ce fait, nous aboutissons à la conclusion que l'utilisation de la répulsion globale n'est pas un atout majeur du moment que la convergence des simulations du modèle ESPECE avec répulsion globale ou avec répulsion simple nécessitent, en moyenne, un temps de réalisation équivalent.

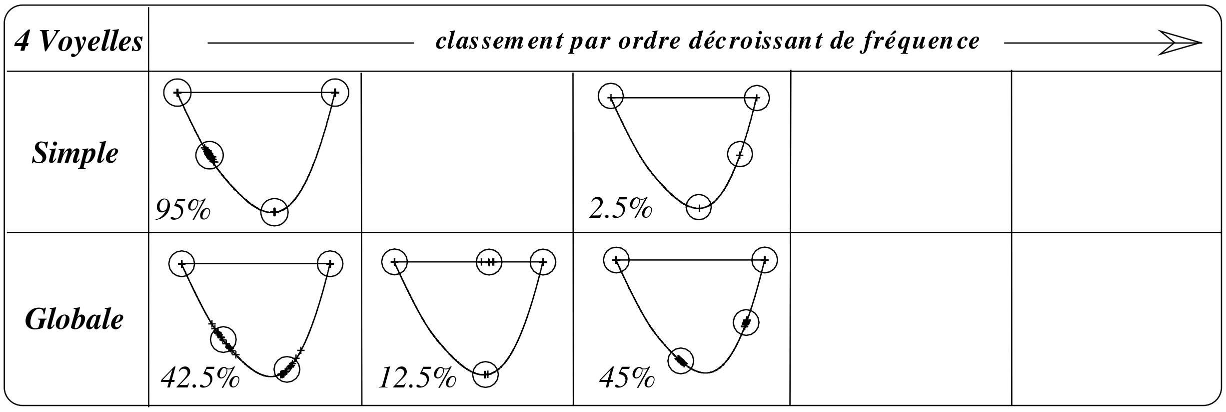


FIG. III.22: Distribution des voyelles des agents ayant un lexique de 4 voyelles dans le triangle vocalique en utilisant la répulsion globale.

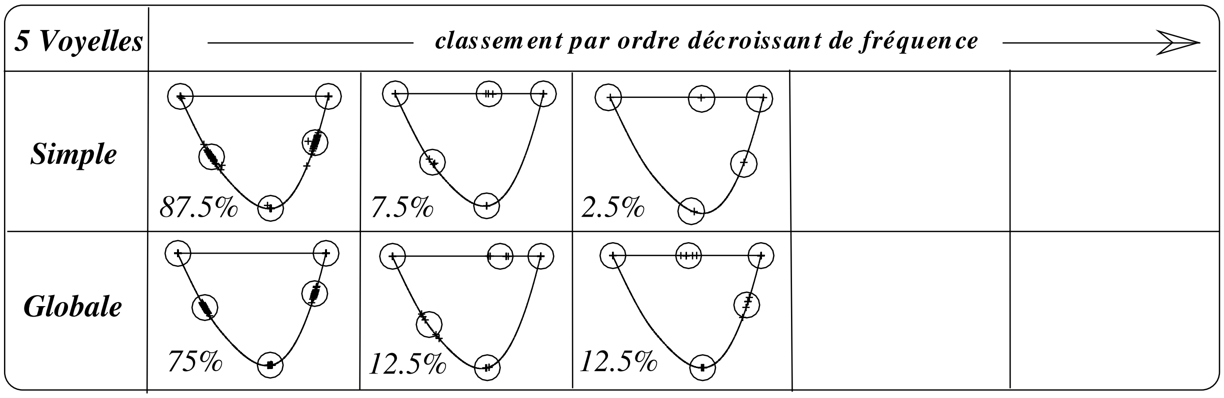


FIG. III.23: Distribution des voyelles des agents ayant un lexique de 5 voyelles dans le triangle vocalique en utilisant la répulsion globale.

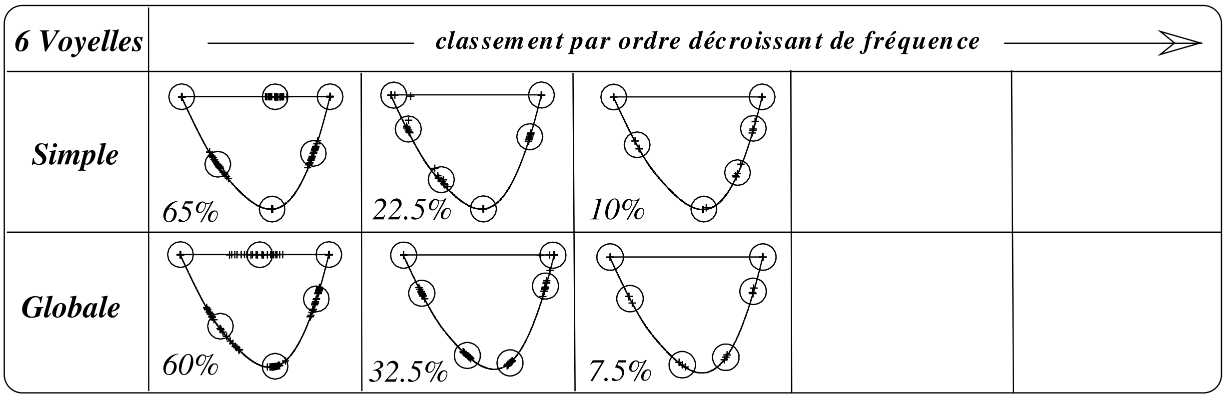


FIG. III.24: Distribution des voyelles des agents ayant un lexique de 6 voyelles dans le triangle vocalique en utilisant la répulsion globale.

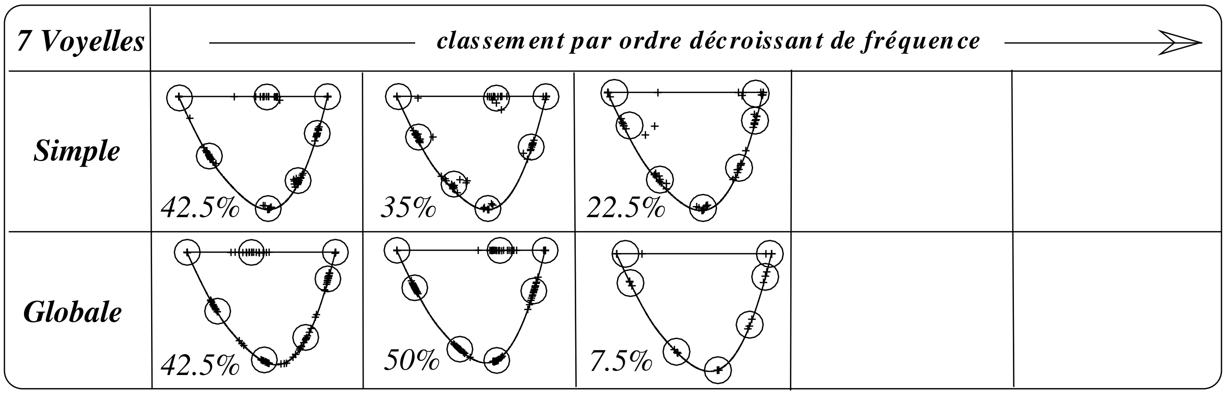


FIG. III.25: Distribution des voyelles des agents ayant un lexique de 7 voyelles dans le triangle vocalique en utilisant la répulsion globale.

III.4.3 Attraction seule et apprentissage

L'adaptation d'un agent auditeur, suite à la perception d'un symbole émis par son partenaire locuteur, s'appuie sur deux concepts qui sont, rappelons-le, l'attraction et la répulsion. Autant l'attraction possède une interprétation cognitive dans le sens où l'auditeur voulant répéter ce symbole se doit de rapprocher un de ses symboles vers la cible ; autant la justification de la répulsion et le mode de repousser sont discutables, si ce n'est que la répulsion induit le principe de dispersion globale classique depuis les travaux de Liljencrants & Lindblom (cf. § I.4.2).

Afin de vérifier si cette répulsion est vraiment fondamentale aux prédictions, nous nous sommes proposés de tester le modèle ESPECE sans utiliser de répulsion, et en adoptant une initialisation aléatoire. Les différentes populations ne convergeaient pas vers les mêmes configurations d'une part, et d'autre part, ces configurations ne reflétaient pas les préférences des langues du monde.

La figure III.26 montre l'exemple d'une simulation réalisée en procédant uniquement à des attractions.

Un modèle avec attraction seule serait suffisant dans le cas où nous voudrions effectuer un apprentissage. Un apprentissage signifie dans ce contexte qu'il existe un robot maître tenant toujours la fonction de locuteur, et qui veuille faire apprendre le lexique de ce robot aux autres individus de la population.

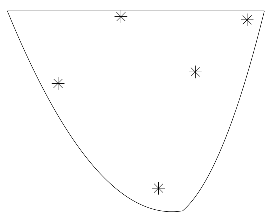


FIG. III.26: Résultat d'une simulation d'une population de 10 agents ayant chacun 5 symboles initialisés aléatoirement. Les transactions se limitent à de l'attraction uniquement.

III.4.4 Espace acoustique en Hertz

Nous avons opté pour l'espace perceptif en Bark pour faire évoluer les symboles de nos agents. Le Bark est une unité perceptive connue pour croître lentement dans la région de basse fréquence. Pour "couper court" à la critique de ne pas avoir tranché plutôt pour une autre échelle perceptive telle que l'*erb* (**e**quivalent **r**ectangular **b**andwidth) (Moore et Glasberg 1983) ou autres, nous proposons d'effectuer un jeu de simulations du modèle ESPECE dans un espace exprimé en Hertz, qui est une unité linéaire, contrairement aux unités perceptives. En d'autres termes, le calcul des distances perceptives, l'attraction ainsi que la répulsion seront réalisées dans l'espace (F_1,F_2) en Hertz, délimité par les équations (II.1), (II.2) et (II.3). Pour réaliser ces simulations, nous reprenons les mêmes paramètres que ceux du modèle décrit en § III.2, excepté pour :

- le facteur λ fixé, cette fois, à 0.2 ; nous avons constaté que la convergence était de meilleure qualité en fixant λ à cette valeur,
- la force de répulsion assignée à 5 000 ; il y a un facteur 10 000 entre cette force et celle à laquelle nous avons eu recours dans le cas des simulations dans l'espace des Barks pour les raisons suivantes :
  - il y a approximativement un facteur 100 entre une fréquence exprimée en Hertz et la même exprimée en Bark,
  - l'intensité de la répulsion (cf. équation II.10) est inversement proportionnelle à la distance perceptive ; de ce fait, pour obtenir une répulsion dans les mêmes ordres de grandeur que pour les simulations en Bark, la force de répulsion doit être égale à 100^2 (10 000).

Les prédictions obtenues sont illustrées sur les figures III.27 à III.31.

Les systèmes engendrés ne sont pas sans rappeler les prédictions générées en § III.3. Ce résultat nous laisserait à penser que notre modèle est peu sensible à l'échelle perceptive adoptée, du moment qu'il a produit des résultats semblables pour le cas extrême des fréquences en Hertz. Toutefois, nous maintenons notre choix de faire évoluer les

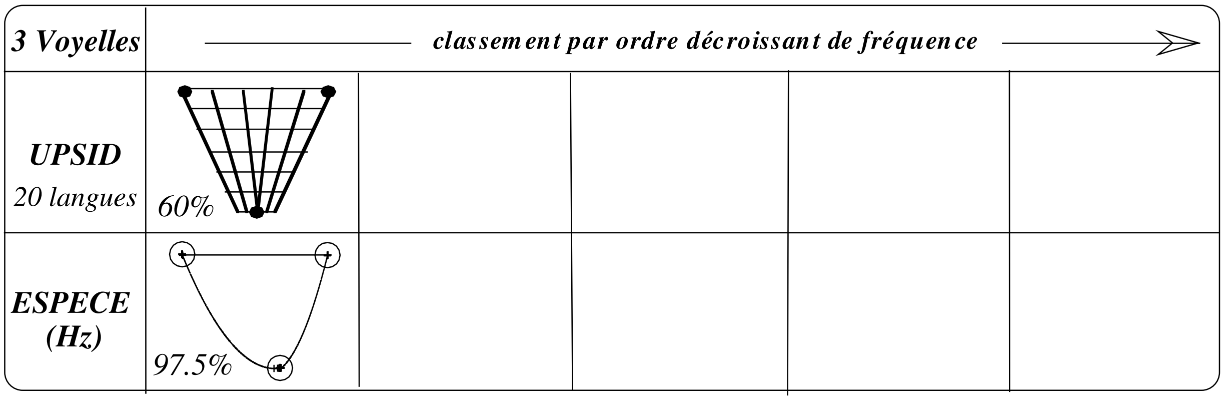


FIG. III.27: Prédictions des systèmes vocaliques à 3 voyelles réalisées dans un espace en Hz.

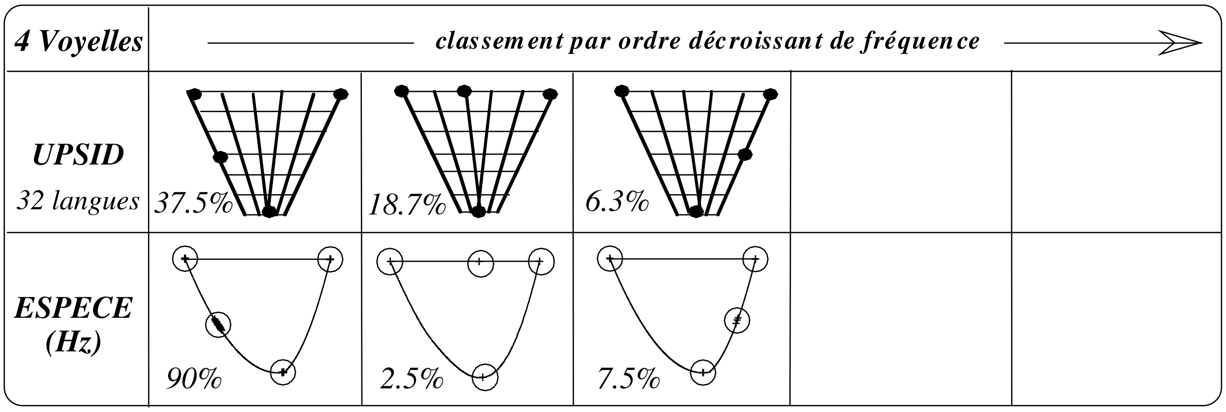


FIG. III.28: Prédictions des systèmes vocaliques à 4 voyelles réalisées dans un espace en Hz.

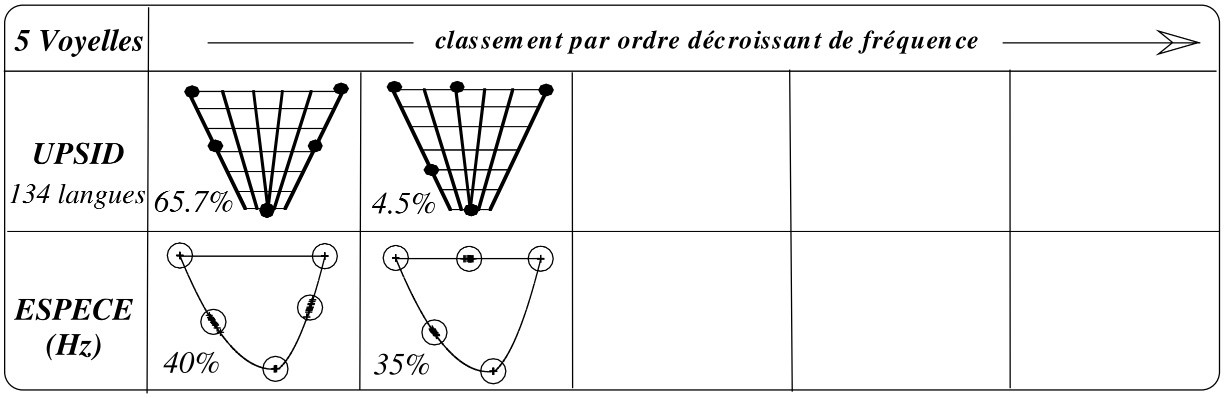


FIG. III.29: Prédictions des systèmes vocaliques à 5 voyelles réalisées dans un espace en Hz.

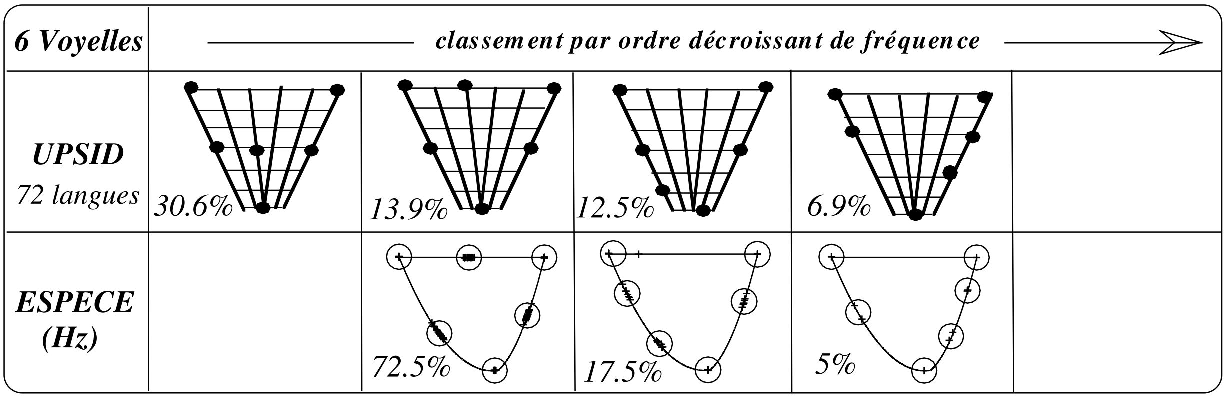


FIG. III.30: Prédictions des systèmes vocaliques à 6 voyelles réalisées dans un espace en Hz.

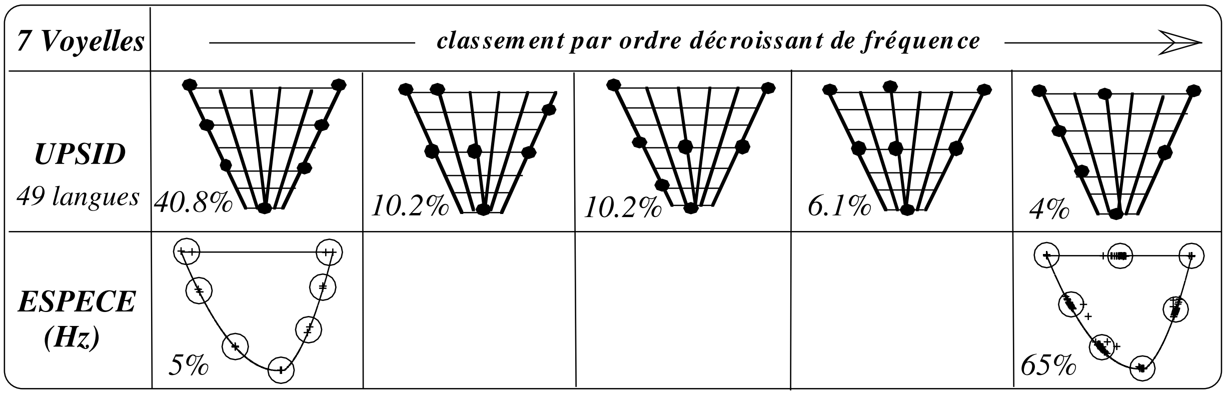


FIG. III.31: Prédictions des systèmes vocaliques à 7 voyelles réalisées dans un espace en Hz.

symboles dans une échelle perceptive par souci de cohérence avec les théories perceptives.

III.5 Étude de l'influence des paramètres sur les prédictions

Une question légitime que nous pourrions nous poser concerne l'influence potentielle des paramètres du modèle - la taille de la population, le nombre de transactions, la force de répulsion et le nombre de transactions avant de procéder à un remplacement - sur les prédictions.

Reprenons maintenant, un à un, ces paramètres et étudions le comportement du modèle selon les variations subies.

III.5.1 La taille de la population

Afin d'évaluer l'effet provoqué par la taille de la société sur les systèmes prédits, lançons des batteries de simulations avec des populations composées respectivement d'un seul agent et de vingt agents. Rappelons que dans le modèle ESPECE présenté, la société comportait dix agents. Les résultats des simulations avec ces différentes tailles de société sont présentés dans l'annexe C. À première vue, nous pouvons remarquer des différences entre les prédictions telles l'ordre des prédictions des systèmes à 6 et à 7 voyelles lorsque la société est constituée d'un seul agent. Néanmoins, il est aisé de s'apercevoir que ces dissemblances sont minimes, et que nous retrouvons, globalement, les mêmes tendances pour les différentes tailles. Il convient de noter, de plus, que la convergence est d'autant plus lente et gourmande en temps dès lors que la taille de la population augmente.

Remarquons, finalement, que le cas extrême où la société n'est composée que d'un seul agent revient, en réalité, à effectuer uniquement des répulsions. Par conséquent, le principe simulé dans ce cas n'est autre que le principe de la dispersion globale.

III.5.2 Le nombre de transactions

Nous nous proposons d'effectuer quelques simulations avec un nombre important de transactions (10 000). Les prédictions obtenues (cf. annexe D) sont pratiquement identiques à celles présentées auparavant, à quelques exceptions près. Cette constatation trouve une explication dans le fait que le facteur «somme des ajustements», introduit au début de la § III.3, devient quasiment stable au bout d'un certain nombre de transactions. En effet, il est aisé d'observer les plateaux sur la figure III.32. Ce résultat s'explique de par le changement subi par les agents de la société, au cours des mille premières transactions (environ), avant de les voir parvenir à une configuration stable, synonyme de convergence.

Par conséquent, nous pouvons dire que le nombre de transactions n'a qu'une infime portée sur les prédictions pour peu que ce nombre soit suffisamment important pour permettre la convergence.

III.5.3 La force de répulsion

Les prédictions présentées ont été simulées à l'aide du modèle ESPECE avec une force de répulsion égale à 0.5. Cette force pondère l'intensité avec laquelle seront repoussés les items lors de la phase de répulsion. Pour mesurer l'impact qu'a la valeur de

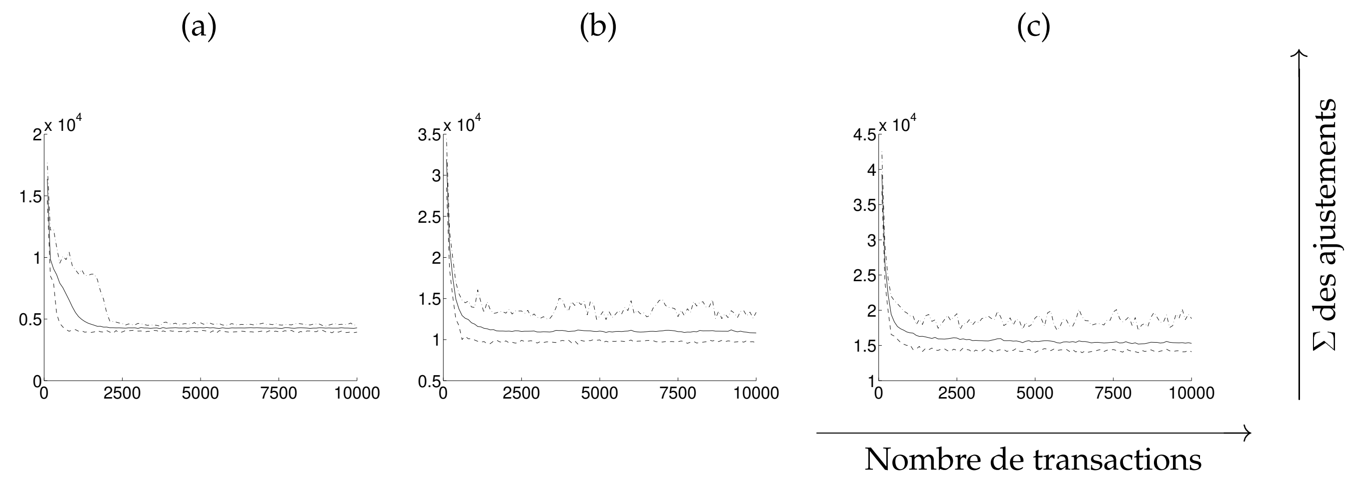


FIG. III.32: Évolution de la somme des ajustements au long des 10 000 transactions des 3 simulations présentées dans la figure III.1 : (a) avec 4 individus, (b) avec 6 individus et (c) avec 7 individus ; la courbe en trait plein représente la moyenne des valeurs de la somme des ajustements pour les 40 simulations, la courbe en pointillés schématise les valeurs minimales alors que les valeurs maximales sont illustrées par la courbe en traits discontinus.

cette force sur les prédictions obtenues, nous avons refait les mêmes simulations avec cette fois une force de répulsion plus petite (égale à 0.3) et une autre plus grande (égale à 1). Les systèmes prédits sont représentés dans l'annexe E. Globalement, les mêmes tendances sont retrouvées. Nous constatons un nombre plus grand de systèmes prédits lorsque la force de répulsion est petite. Ce grand nombre peut être expliqué par le fait que plus la force de répulsion est petite, plus les symboles, légèrement repoussés, ont le loisir de mieux explorer l'espace acoustique (F_1,F_2). Les populations auront ainsi la possibilité de converger vers un nombre plus important de configurations stables sur cet espace. En revanche, avec une force plus grande, la convergence est plus fastidieuse. Quelques éléments d'interprétation de cette observation résident dans le fait que les symboles sont, dans ce cas, fortement dispersés, ce qui pénalise leur regroupement. Par contre, lorsque les symboles sont soumis à une petite force de répulsion, nous avons constaté qu'ils se rassemblaient en clusters dès les premières transactions.

Notons que les prédictions sont complètement perturbées lorsque la force de répulsion est importante ($\gg 1$) du fait que les symboles sont rapidement projetés en dehors des limites du triangle vocalique, et se retrouvent instantanément cloîtrés dans les extrémités ; à savoir la zone du /i/, du /u/ et essentiellement la zone du /a/ qui fera l'objet de «disputes» entre plusieurs clusters.

III.5.4 Le remplacement

Le dernier paramètre dont il serait intéressant d'évaluer la portée sur les prédictions est le nombre de transactions avant de procéder à un remplacement. Rappelons juste que les remplacements ont été introduits dans le modèle ESPECE pour éliminer, à chaque fois, l'agent *intrus* de par sa configuration différente de celles des autres et qui, de ce fait, perturbe et pénalise la convergence de la société. En observant les prédictions du modèle ESPECE sans remplacement appliqué à la prédiction des systèmes vocaliques (cf. annexe F), nous retrouvons les mêmes tendances que celles du modèle avec remplacement. Nous remarquerons que pour certaines tailles de lexiques, le nombre de systèmes répertoriés pour le modèle sans remplacement est plus important. À titre d'exemple, nous dénombrons six systèmes à 7 voyelles différents alors que nous n'avions que trois systèmes lorsque nous procédions à des remplacements. L'obtention de ce nombre important de systèmes peut être due aux cas où la population n'a pu convenablement converger en raison d'un agent perturbateur. En outre, il est clair d'après les figures présentées dans l'annexe F que les systèmes engendrés sans avoir eu recours aux remplacements ne sont pas bien regroupés, c'est-à-dire que les *clusters* sont dispersés. Ce qui explique l'intérêt de maintenir des remplacements dans le modèle ESPECE.

III.5.5 Discussion

Nous avons passé en revue dans cette section les influences que pourraient avoir les différents paramètres sur les configurations engendrées par le modèle ESPECE appliqué à la prédiction des systèmes vocaliques. Nous pouvons retenir de ce qui précède que la taille de la population n'engendre pas de différences significatives sur les prédictions obtenues. En outre, nous pouvons dire sans trop prendre de risques qu'il est essentiel de maintenir le recours à des remplacements pour améliorer la qualité et accélérer les convergences.

Soulignons toutefois qu'il y a une partie d'arbitraire dans le jeu de paramètres que nous avons sélectionné pour faire simuler notre modèle. Nous n'irons pas jusqu'à prétendre que cet ensemble de valeurs des paramètres représente le jeu optimal. Bien évidemment, il reste à définir ce que serait un jeu optimal de paramètres. Il nous semble qu'un jeu ne serait dit optimal que s'il permettait de générer les prédictions les plus conformes aux langues du monde. Mais, ce jeu existe-il ? De toutes manières, la question d'optimalité du jeu de paramètres n'a un sens uniquement si le modèle servait à faire exclusivement de la prédiction de systèmes vocaliques. Or, ESPECE a été élaboré dans le but essentiel de voir si la possibilité théorique, que les langues seraient le fruit, depuis la nuit des temps d'échanges et de coopérations, ne serait pas dénuée de bon sens. Mentionnons enfin que nous avons opté pour une taille de population moyenne, une force de répulsion raisonnable et des remplacements toutes les cent transactions. La seule affirmation que nous pouvons avancer concerne le nombre de transactions : comme nous l'avons montré précédemment, au-delà d'un certain nombre de transactions, la population n'évolue plus. Il est donc inutile d'effectuer des simulations portant sur un gros nombre de transactions.

III.6 Conclusion

Nous avons abordé dans ce chapitre les caractéristiques du modèle ESPECE appliqué à la prédiction des systèmes vocaliques. Nous avons présenté ses divers paramètres, ainsi que les résultats de ses prédictions. Nous avons également présenté plusieurs variantes de ce modèle. Par ailleurs, nous avons discuté de l'influence de chacun des paramètres sur les résultats engendrés, et nous avons plaidé pour un choix de jeu de paramètres pour simuler notre modèle. D'autre part, nous avons comparé les systèmes vocaliques générés par les différentes simulations avec les données sur les langues du mondes. Nous avons constaté que les prédictions obtenues étaient proches des inventaires observés dans les langues.

Nous avons remarqué que le modèle ESPECE en version extrêmement affaiblie, c'est-à-dire avec un seul agent, est équivalent aux modèles reprenant les fondements des théories de la Dispersion (cf. I.4.2) et de la Dispersion-Focalisation (cf. I.4.3). Néanmoins, La différence fondamentale entre cette version d'ESPECE et les modèles cités réside dans le caractère des simulations : stochastique pour ESPECE et déterministe pour les autres méthodes. En effet, les méthodes de minimisation d'énergie emploient le procédé de la descente du gradient afin de trouver le système optimal. Par contre, ESPECE est basé sur des ajustements effectués au fur et à mesure des transactions.

Mentionnons au passage que la version d'ESPECE avec un seul agent, prive le modèle du côté interaction entre agents de parole. Par conséquent, il serait impossible de rendre compte dans un tel environnement des aspects réalistes de la négociation locuteur-auditeur (cf. théorie *Hypo & Hyper* (Lindblom, Guion, Hura et Moon 1995, Lindblom 1995) brièvement présenté dans l'annexe G), ainsi que de la prise en charge de la diversité anatomique au sein de la société d'APs[2]. Certes, ces points n'ont pas été pris en compte dans la présente étude. Cependant, il ne nous empêche pas d'envisager leur intégration dans le modèle dans le futur.

[2]Conduits vocaux différents par exemple.

Chapitre IV

Le modèle ESPECE intégrant le principe du MUAF

Résumé

Le cadre de simulation sera étendu, dans ce chapitre, afin de d'intégrer le principe du MUAF. Ce principe postule l'économie de l'usage de traits distinctifs dans les systèmes phonologiques complexes.

Après avoir énoncé les fondements théoriques de ce principe, nous présenterons le modèle décrit dans le chapitre II reproduisant, cette fois, ce principe. Les nouveaux protocoles d'échanges seront d'abord validés sur un exemple synthétique avant de les confronter, par la suite, à un espace perceptif à 3 dimensions.

Les résultats des simulations seront extensivement comparés aux systèmes vocaliques de la base UPSID_{317}. Nous terminerons ce chapitre par une discussion à propos de ces prédictions.

IV.1 Introduction

Déjà en 1956, Miller (1956) s'interroge sur les raisons qui font que le nombre de qualités vocaliques, qu'il qualifie de «magique», est égal à 7 ± 2. En effet, des études typologiques ont confirmé que les systèmes recrutaient essentiellement entre 3 et 9 qualités vocaliques différentes (cf. § I.3). Cette constatation suscite une importante question : *Quand* et *pourquoi* les systèmes vocaliques développent-ils un système secondaire ou utilisent-ils un *trait* supplémentaire ? La réponse classique est fournie par Lindblom & Maddieson (1988) qui notent que les articulations secondaires apparaissent lorsque le nombre de voyelles d'un système devient important, et par conséquent, trop large pour pouvoir être régi uniquement en termes de distances acoustiques dans l'espace formantique. Vallée (1994) suggère que le nombre maximum de *timbres vocaliques* correspondant à des voyelles *de base* que peut contenir un système est 9. Cette hypothèse peut être restrictive. Bien qu'elle soit vérifiée dans la majorité des langues répertoriées dans UPSID$_{317}$, il reste des langues dans cette base qui présentent plus de 9 timbres vocaliques, telles le *Français* (12 timbres vocaliques), ou à l'extrême le *Norvégien* (15 timbres vocaliques).

Avant de présenter les arguments qui ont poussé aussi bien Vallée (1994) que Crothers (1978) à avancer cette hypothèse et afin de rendre plus commode la compréhension de l'argumentation, donnons quelques définitions de la terminologie employée dans cette section.

IV.2 Définitions

Pour l'étude des voyelles, les articulations sont généralement subdivisées en deux types : l'articulation primaire qui qualifie la constriction la plus étroite du conduit vocal, et les articulations secondaires lorsque la voyelle est prononcée avec plusieurs composantes articulatoires simultanées. Ces articulations dites secondaires sont des articulations qui se superposent aux voyelles dites **de base** (appelées aussi voyelles **pleines** et réalisées avec un lieu et un mode d'articulation principaux). On parle alors de **dimensions articulatoires** secondaires par opposition aux voyelles de base. On dit aussi que les voyelles utilisent des **traits** supplémentaires. À chaque voyelle de base correspond des **qualités (ou timbres) vocaliques** différentes.

Dans la base de données UPSID$_{451}$ (Maddieson et Precoda 1990), chaque voyelle possède des valeurs spécifiques pour chacune des trois dimensions (hauteur, avancement et labialité) schématisées sur la figure IV.1.

Les voyelles peuvent, en plus, posséder une ou plusieurs modifications de produc-

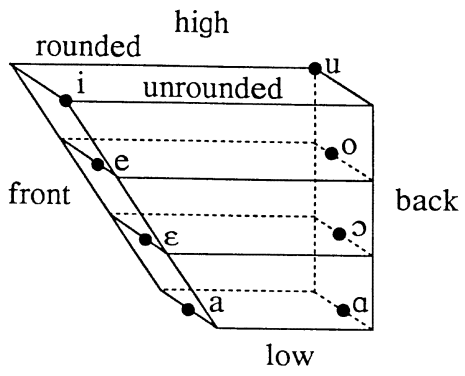


FIG. IV.1: Les dimensions des traits de base (d'après Ladefoged & Maddieson (1990)).

tion (traits supplémentaires). La nature de ces modifications est très variée dans la notation d'UPSID$_{451}$. Elles peuvent être liées :

- à la durée de la production longue *(long)* et ultra-brève *(overshort)*,
- à l'intervention d'un articulateur autre que le corps de la langue, la mâchoire où les lèvres ; la pointe de la langue pour les voyelles rétroflexées *(retroflexed)* et le voile du palais pour les voyelles nasalisées *(nasalized)*.
- à la modalité de production de son ou de la voix ; fricative *(fricative)*, aspirée *(breathy voiced)* et assourdie *(voiceless)*
- au rétrécissement dans des parties spécifiques du conduit vocal ; le pharynx pour les voyelles pharyngalisées *(pharyngealized)*, le larynx pour les voyelles laryngalisées *(laryngealized)* et le velum pour les voyelles vélarisées *(velar stricture)*.

Enfin, il existe également des modifications qui interviennent au même niveau que le trait de base d'avancement et que nous considérons comme des cas à part. Il s'agit des modifications rétractée *(retracted)* et avancée *(fronted)*.

IV.3 Limites universelles des systèmes vocaliques

Pour les systèmes de 3 à 9 voyelles, la tendance générale des systèmes est de posséder un nombre de voyelles identique au nombre de timbres. À chaque timbre correspond des qualités vocaliques différentes.

Par contre, au dessus de 9 voyelles, les systèmes, apparemment saturés, ne développent plus de nouveaux timbres afin d'augmenter leurs inventaires. Ils superposent, une, voire plusieurs complexités articulatoires afin d'établir des contrastes avec de nouvelles dimensions (ou traits) telles que la nasalité, la quantité, la pharyngalité, l'aspiration etc. Par exemple, le «Français» est catalogué dans certaines typologies (Vallée 1994) comme étant une langue possédant 16 voyelles : 10 voyelles pleines, 2 longues (utilisant la quantité comme dimension secondaire) et 4 nasales. Dans ce système figurent les voyelles /ɒ/ (p**â**te) et /ɒ̃/[1] (s**an**s) d'une part, et d'autre part, les voyelles /o/ (d**o**nner) et /õ/ (b**on**).

À cet universel, L'explication la plus probable est celle du contraste suffisant (cf. § I.4.2). En limitant le nombre de timbres distincts, les systèmes linguistiques évitent ainsi l'encombrement de l'espace vocalique et diminuent les probabilités de confusion dans la production et la perception des sons. Par conséquent, c'est la nécessité d'un contraste suffisant qui fait appel à des articulations plus complexes. De nombreux larges systèmes vocaliques diversifient leurs inventaires en ajoutant plutôt un ou plusieurs traits à des timbres de base.

IV.4 Définition du principe du MUAF

Afin de préserver la distinctivité perceptive dans de larges systèmes, il est nécessaire de recourir à de nouvelles dimensions articulatoires. Fait révélateur, ces dimensions ne sont pas ajoutées dans de petits systèmes, ce qui laisserait croire qu'une pénalité articulatoire est liée à leur utilisation. Cette pénalité «retarde» la sélection de voyelles complexes. Ohala (1980) qualifie ce phénomène de *«Maximum Use of Available distinctive Features (MUAF)»* (utilisation maximale des traits distinctifs disponibles) dans la structuration des inventaires phonétiques. Le principe du MUAF est invoqué afin d'expliquer d'une part pourquoi les langues, avec de larges inventaires vocaliques, ont tendance à employer des traits supplémentaires telles que la nasalité ou la quantité. D'autre part, il permet d'expliquer, l'absence dans les langues du monde, de systèmes vocaliques comme /i ẽ a̤ o̰ uˤ/, présentant pourtant un contraste perceptif plus grand que celui du système universellement favorisé /i e a o u/ (Maddieson 1986).

IV.5 Recalibration lexicale

Lindblom (1997) propose l'introduction de *contraintes développementales* afin d'expliquer pourquoi les systèmes phonologiques obéissent au principe du MUAF : l'acquisition de nouvelles formes serait plus aisée lorsque leurs commandes motrices associées chevauchent avec celles des structures préalablement acquises. Toutes autres caractéristiques (coût articulatoire et saillance perceptive) égales, l'apprentissage d'un nouveau jeu de commandes est d'autant plus simple que les éléments le constituant sont déjà

[1]Le symbole /◌̃/ signifie que la voyelle est nasalisée

présents dans l'ensemble des gestes déjà acquis. Ce principe constitue la base de la *Recalibration Lexicale* telle qu'elle a été définie par Lindblom (1997). La figure IV.2 illustre ces propos.

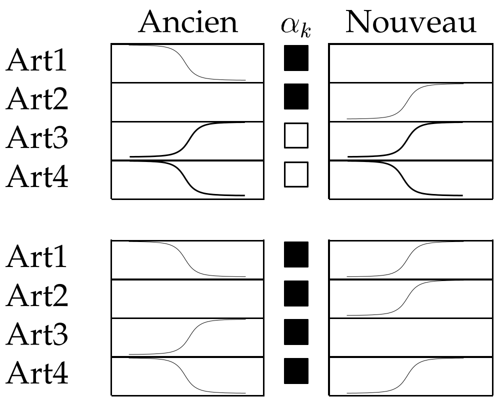


FIG. IV.2: Comparaison de deux hypothétiques scores moteurs : pour le premier, l'ancienne et la nouvelle forme partagent les trajectoires motrices de Art3 et Art4 (trajectoires en trait épais) ; alors que pour le second, l'ancienne et la nouvelle forme n'ont aucun mouvement en commun (d'après Lindblom (1997)).

Selon la théorie de la dispersion adaptative proposée par Lindblom en 1990 (cf. § I.4.2), un système phonétique optimal minimise, rappelons-le, l'équation (I.3) :

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\frac{A_{ij}}{d_{ij}})^2$$

avec :

d_{ij} : distance perceptive entre les phonèmes i et j,

A_{ij} : coût articulatoire entre les phonèmes i et j.

Afin d'implémenter la recalibration lexicale, Lindblom (1997) a affiné le terme articulatoire de cette équation. A_{ij} s'écrit désormais :

$$A_{ij} = A_i^* + A_j^* \tag{IV.1}$$

A_i^* est appelé *score post-lexical* et est défini par :

$$A_i^* = \sum_{k=1}^{M} \alpha_k(m_i) \int_0^T f_k(m_i, t)dt \tag{IV.2}$$

avec :

p	·	·	t	·	·	k	·	·
b	·	·	d	·	·	ɡ	·	·
m	·	·	n	·	·	ŋ	·	·

système MUAF

·	·	·	·	ʈ	·	·	q	ʔ
b	·	d̪	·	·	·	ɡ	·	·
·	ɱ	·	n	·	ɲ	·	·	·

système non-MUAF

FIG. IV.3: Effet de la recalibration lexicale (d'après Lindblom (1997)).

M : dimension de l'espace articulatoire,

m_i : score moteur (trajectoire des commandes motrices),

$f_k(m_i, t)$: «coût de lissage» de la trajectoire m_i à l'instant t pour l'articulateur k,

$\alpha_k(m_i)$: paramètre binaire contrôlant la présence de la trajectoire du $k^{ème}$ articulateur dans les gestes préalablement appris. Dans le cas affirmatif, $\alpha_k(m_i) = 0$ (schématisé par un carré blanc dans la figure IV.2), et de ce fait, le coût articulatoire du phonème i sera réduit favorisant de la sorte sa sélection. Autrement, $\alpha_k(m_i) = 1$ (schématisé par un carré noir dans la figure IV.2) et la recalibration lexicale ne sera pas mise en œuvre.

La recalibration lexicale simule l'effet du principe du MUAF. Si le développement du langage avait été caractérisé par un processus similaire à celui de la recalibration lexicale, les systèmes MUAF auraient été acquis plus aisément que les systèmes non-MUAF (cf. figure IV.3).

Dans la continuité des propos de Ohala (1980) et de Lindblom (1997), nous proposons une version révisée du modèle d'émergence ESPECE pour reproduire le phénomène du MUAF dans les systèmes vocaliques.

IV.6 Principes du modèle ESPECE révisé

L'idée est de reproduire l'effet suivant : lorsqu'un trait est utilisé, il est exploité systématiquement avant d'en exploiter un autre. À cet égard, nous avons imaginé un espace perceptif à trois dimensions : deux pour les formants et un pour le trait (cf. figure IV.4).

Chaque phonème est désormais caractérisé par une composante spectrale (formants) et par une composante supplémentaire (trait).

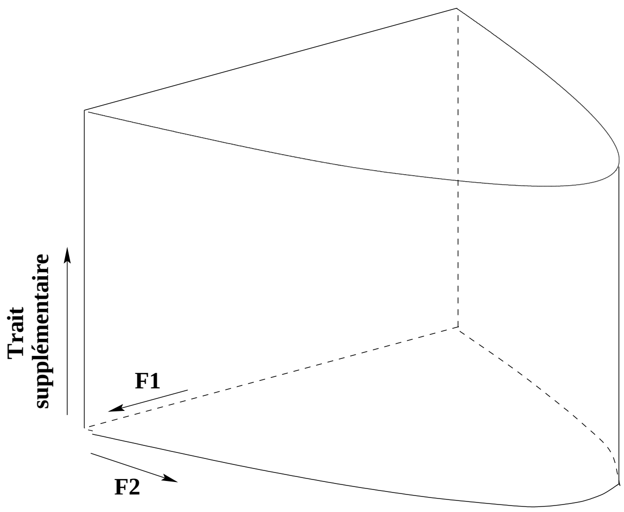


FIG. IV.4: Espace perceptif à trois dimensions.

Pour le modèle ESPECE appliqué à l'explication du principe du MUAF, nous suggérons de modifier le protocole de l'adaptation, et plus précisément, le processus de répulsion des différents items du lexique de l'AP auditeur. En effet, puisque nous avons désormais deux composantes (formants et trait), la répulsion s'effectuera uniquement sur une seule composante à la fois, en l'occurrence, celle où le gain en distinctivité est maximum.

Il nous reste à expliquer comment ce nouveau processus de répulsion s'apparente à la recalibration lexicale de Lindblom (1997), et constitue ainsi une implémentation du principe du MUAF dans le cadre du modèle ESPECE. Rappelons que l'effet de la recalibration lexicale est de favoriser l'acceptation des nouveaux items, qui partagent un maximum des composantes (motrices, dans le cas présent) avec les items déjà existants dans le lexique (cf. § IV.5). En d'autres termes, pour que deux items soient considérés distincts, il suffit qu'ils le soient au niveau d'une de ces composantes. Or, ceci est exactement ce que nous imposons dans la nouvelle version du modèle ESPECE, mais cette fois-ci au niveau de l'espace perceptif. En effet, lors d'une répulsion, seulement une dimension sera impliquée, toutes les autres restant immuables.

Dans l'objectif de tester ces idées, il nous semble judicieux de commencer par l'implémentation et la validation de ces protocoles sur une exemple synthétique avant de considérer l'espace perceptif à trois dimensions.

IV.7 Exemple synthétique

Il s'agit effectuer une simplification synthétique de l'espace perceptif à trois dimensions et de ne considérer qu'un espace à deux dimensions limité par un rectangle oblong (cf. figure IV.5). La particularité de ce rectangle est que sa *longueur* est quatre fois plus grande que sa largeur. Les traits sont représentés actuellement par l'axe vertical et l'axe horizontal. Soulignons que ces composantes sont strictement synthétiques et qu'elles n'ont aucune signification phonétique ou autre.

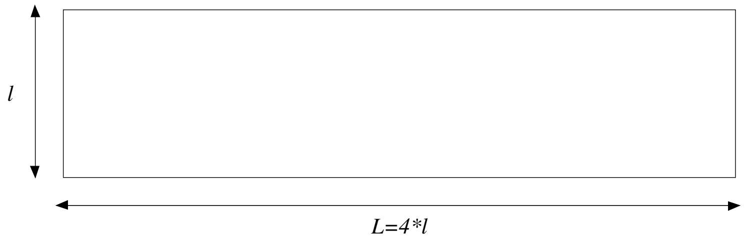


FIG. IV.5: Rectangle synthétique

Le nouveau protocole d'adaptation, dans lequel la répulsion est réalisée uniquement sur une dimension, est matérialisé par une métrique particulière que nous qualifions de distance MUAF, mais qui n'est en réalité rien d'autre qu'une distance de Chebychev[2]

La distance MUAF entre deux items a et b de coordonnées respectivement (x_a, y_a) et (x_b, y_b) a la forme suivante :

$$d(a,b) = max(|x_a - x_b|, |y_a - y_b|) \tag{IV.3}$$

Rappelons que cette distance est utilisée pour sélectionner, lors d'une transaction, l'item à attirer et ceux à repousser de l'agent récepteur.

Nous gardons le même protocole d'attraction que celui du modèle ESPECE initial. L'item attiré est déplacé à égale distance de l'item perçu et de son ancien emplacement. Par contre, le protocole de répulsion est révisé de sorte que les items ne soient plus repoussés que sur une seule dimension, celle où la distinctivité est la plus grande.

Autrement dit, pour réaliser une répulsion d'un item $a(x_a, y_a)$ par rapport à un autre $b(x_b, y_b)$, nous calculons premièrement la distance MUAF $d(a,b)$ (équation (IV.3)). Ensuite, nous altérons seulement la composante de a la plus différente de son homologue

[2]La distance de Chebychev a pour expression :

$$D(x,y) = \max_{i=1}^{m} |x_i - y_i|$$

.

de b. Soit ψ cette composante dictée par :

$$\psi = \underset{k}{argmax}(|k_a - k_b|) \tag{IV.4}$$

Autrement dit, ψ aura pour valeur soit x soit y.

L'expression de la répulsion largement inspirée de celle du modèle ESPECE est régie selon :

$$\psi_a \leftarrow \psi_a + I \tag{IV.5}$$

avec :

$$I = \xi * (\Delta_\psi/||\Delta||) * (1/d^2(a,b)) \tag{IV.6}$$
$$\Delta = a - b \tag{IV.7}$$
$$\Delta_\psi = \psi_a - \psi_b \tag{IV.8}$$

ξ étant la force de répulsion. Le terme Δ_ψ assure que la répulsion est faite dans la bonne direction.

Nous voulons par ce procédé qu'une dimension soit exploitée entièrement avant qu'une autre ne soit abordée.

Nous avons effectué des simulations de la version modifiée du modèle ESPECE pour le rectangle synthétique (cf. figure IV.5), mettant en oeuvre ce nouveau protocole d'adaptation.

Les simulations, au nombre de 40 pour chacune des tailles de lexique (de 3 à 10), ont porté sur une société de 10 agents. Le nombre d'échanges et la force de répulsion sont fixés respectivement à 5 000 et à 2. En outre, le nombre d'échanges constituant une époque est fixé à 100. Le facteur λ, n'ayant plus de raisons de pondérer quoi que ce soit dans ce contexte, est fixé à 1.

De même que pour le modèle ESPECE appliqué à la prédiction des systèmes vocaliques (cf. § III.3), chacune des 40 simulations pour chaque taille de lexique correspond à une semence différente du générateur de nombres aléatoires. À l'issue de chacune de ces simulations, la moyenne des inventaires de la population est également calculée à l'aide de la procédure d'appariement décrite en § II.6.2. Cette moyenne constitue la configuration type des agents après convergence.

Les résultats des simulations sont résumés dans les figures IV.6 et IV.7, constituées de «rectangles». Chaque rectangle représente les lexiques moyens regroupés, des simulations ayant des configurations similaires. Les lexiques sont illustrés par des «+».

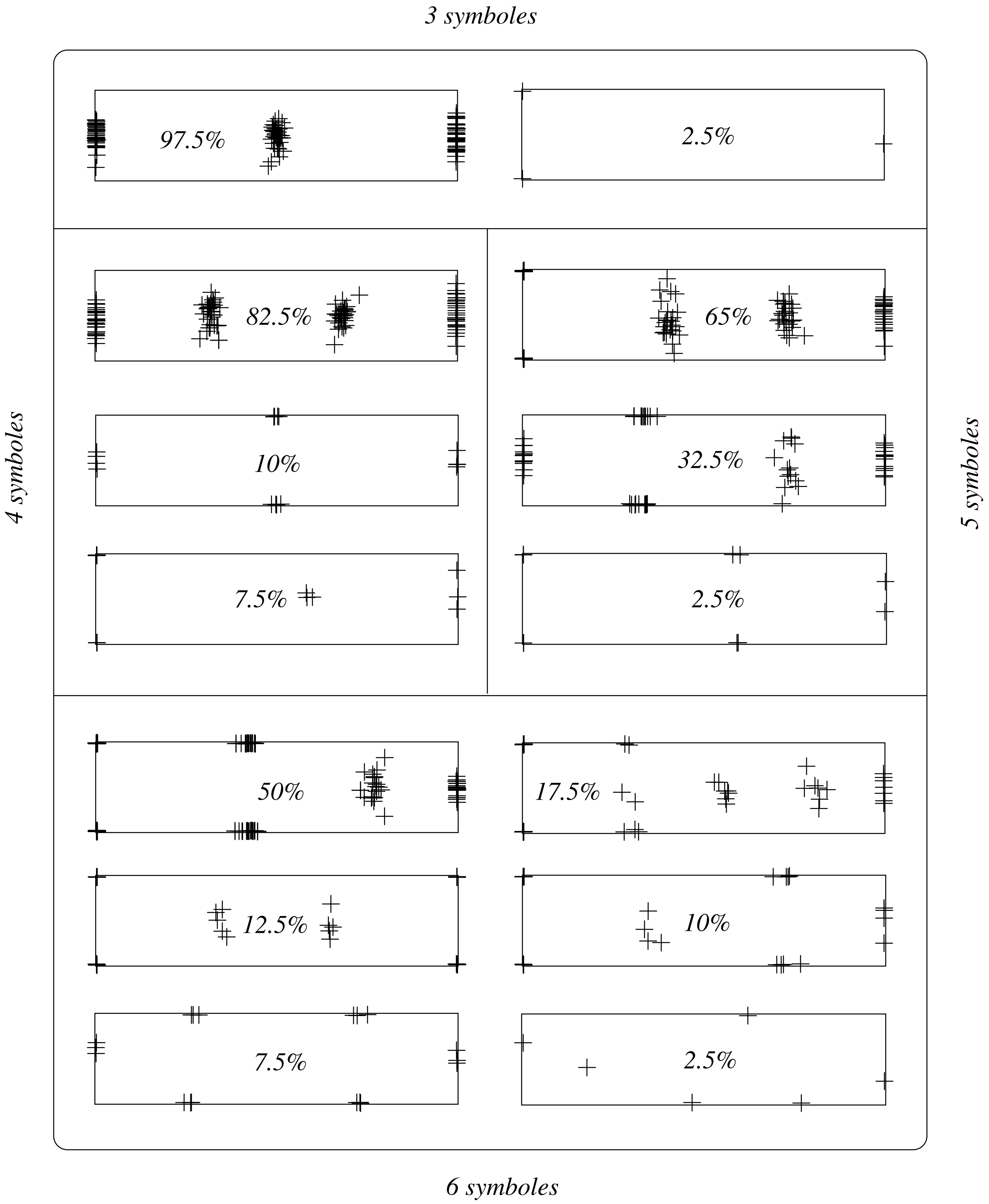


FIG. IV.6: Exploitation progressive de la seconde dimension dans des systèmes utilisant 3-, 4-, 5- et 6-symboles. Les pourcentages indiquent les statistiques sur 40 simulations effectuées.

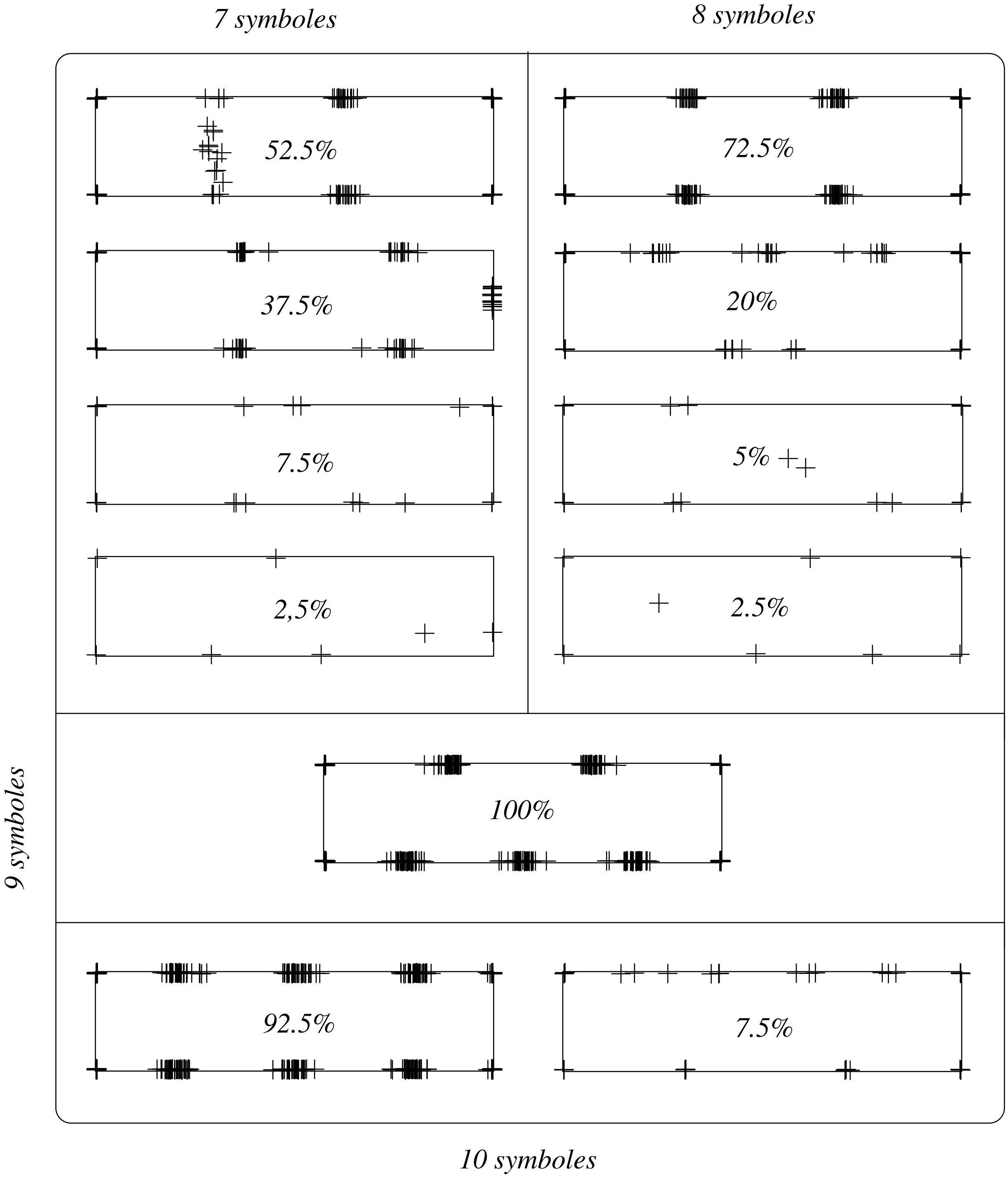


FIG. IV.7: Exploitation progressive de la seconde dimension dans des systèmes utilisant 7-, 8-, 9- et 10-symboles. Les pourcentages indiquent les statistiques sur 40 simulations effectuées.

Ces figures montrent clairement que pour des lexiques de petite taille (≤ 4), une seule dimension est explorée dans la grande majorité des cas. Puis, au fur et à mesure que la taille croît, la seconde dimension est utilisée. En outre, il est aisé de constater qu'il y a un dédoublement de 1, 2 puis 3 symboles lorsque la taille du lexique vaut respectivement 5, 6 puis 7 items. Notons que le système majoritaire à 8 items n'est rien d'autre qu'un dédoublement parfait du système à 4 items comme l'illustre si bien la figure IV.8.

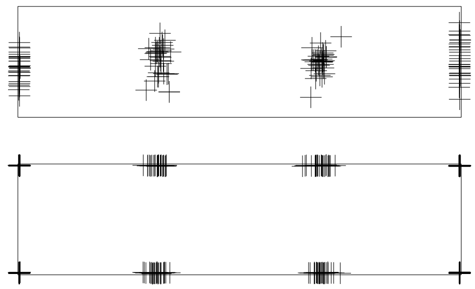


FIG. IV.8: Effet MUAF : dédoublement du système à 4 symboles pour le système à 8 symboles.

Il convient de noter que cette relation lie particulièrement les lexiques de taille 4 et 8 uniquement car le rectangle est 4 fois plus long que large. Aussi, pouvons-nous avancer, sans trop prendre de risques, que nous aurions observé, dans les systèmes à 10 items, un dédoublement parfait des systèmes à 5 items si la longueur avait été fixé à 5 fois la largeur.

Ce phénomène de dédoublement rappelle quelque peu un autre phénomène observé dans les inventaires vocaliques des langues du monde : celui concernant l'apparition du schwa (/ə/). En effet, il est bien connu que, dans les systèmes majoritaires, la voyelle centrale /ə/ apparaît dans les systèmes à 6 et à 8 voyelles et qu'elle laisse place dans les systèmes à 7 et à 9 voyelles à deux voyelles latérales (l'une antérieure et l'autre postérieure).

Comme nous l'avons mentionné en § IV.6, le nouveau protocole de répulsion dans ESPECE s'apparente aux principes de la recalibration lexicale de Lindblom (1997). Cependant, il convient de noter que le modèle ESPECE prédit une exploration aléatoire des traits non exploités, tandis que la recalibration lexicale ne prédit pas de variabilité dans les traits supplémentaires. À cet égard, une étude fine sur les données serait indispensable afin de connaître la manière dont sont réellement organisés les systèmes dans les langues du monde.

Le principe de l'exploration progressive de dimensions a été constaté lors des tests du modèle sur l'exemple synthétique du «rectangle». Nous nous proposons d'analyser maintenant l'impact de la métrique employée sur ces observations.

IV.8 Importance de la métrique

Nous nous proposons de reprendre les simulations réalisées en § IV.7 en gardant les mêmes paramètres mais en utilisant cette fois une métrique euclidienne au lieu de celle exprimée par l'équation (IV.3).

La distance entre deux symboles a et b respectivement de coordonnées (x_a, y_a) et (x_b, y_b) se calcule par la formule :

$$d^2(a,b) = (x_a - x_b)^2 + (y_a - y_b)^2 \tag{IV.9}$$

Les symboles subissent une répulsion classique semblable à celle employée par le prototype ESPECE appliqué aux prédictions vocaliques. L'intensité de cette répulsion s'écrit :

$$I = \xi * (\Delta/||\Delta||) * (1/d^2(a,b)) \tag{IV.10}$$

Avec

$$\Delta = a - b \tag{IV.11}$$

ξ étant la force de répulsion.

Les résultats des simulations, pour des tailles de 4 à 8 items, effectuées avec les mêmes paramètres que celles du rectangle (cf. § IV.7) sont représentés sur la figure IV.9.

Nous observons clairement en analysant ces rectangles que l'utilisation de la métrique euclidienne a mené à l'exploration de toutes les dimensions, et que l'effet «exploration progressive» a disparu. Cette constatation était prévisible du moment que cette métrique a pour effet de maximiser les distances inter-symboles comme c'est le cas pour les systèmes illustrés par la figure IV.10 (a), contrairement à ceux schématisés dans la figure IV.10 (b).

Nous venons de voir comment une métrique particulière associée à notre modèle d'émergence ESPECE pouvait simuler un intéressant phénomène que nous appellerons *MUAD (Maximal Use of Available Dimensions)*. Maintenant que ce prototype a été testé et validé sur un exemple synthétique, nous nous proposons d'utiliser ce concept dans une simulation «parole» proprement dite, et de réaliser une expérience similaire avec les voyelles.

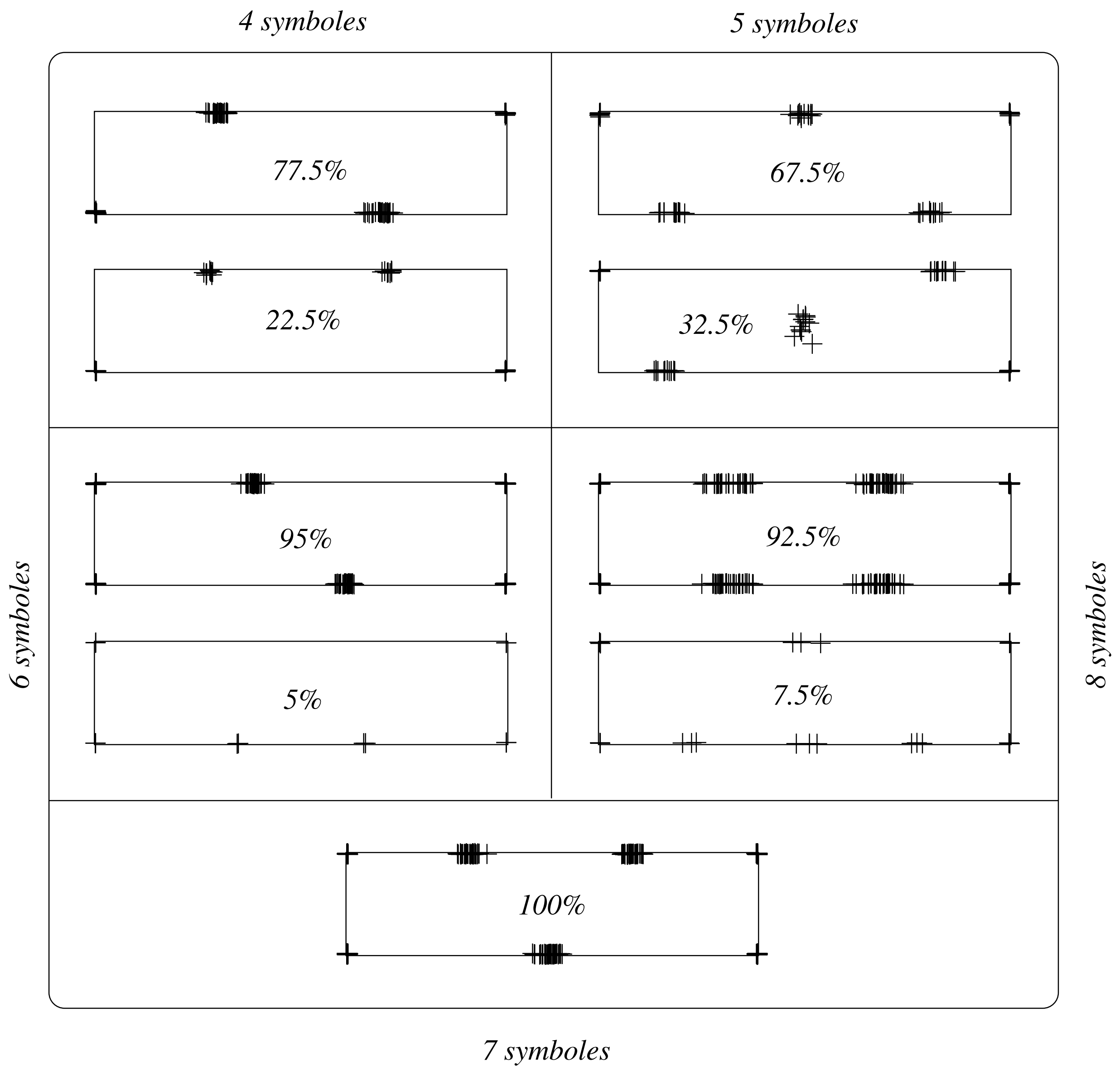
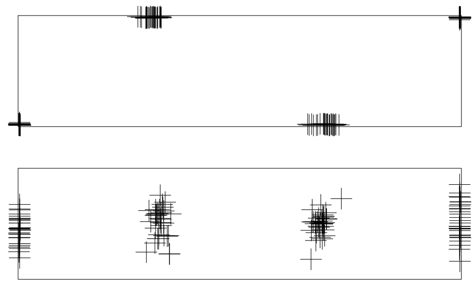


FIG. IV.9: Systèmes de 4-, 5-, 6- et 7- et 8-symboles engendrés en employant une métrique euclidienne. Les pourcentages indiquent les statistiques sur 40 simulations effectuées.

IV.9 Description du modèle tenant compte des deux composantes vocaliques

Pour le modèle ESPECE appliqué à la reproduction du principe du MUAF, nous avons modifié le protocole d'adaptation, et plus précisément, le processus de répulsion des différents items du lexique de l'AP auditeur. En effet, puisque nous avons désormais deux composantes (formants et trait), la répulsion s'effectuera uniquement sur une



(a)

(b)

FIG. IV.10: Résultats pour des répulsions basées sur deux métriques différentes : (a) Euclidienne et (b) MUAF.

seule dimension à la fois, en l'occurrence, celle où le gain en distinctivité est maximum.

Nous opérons maintenant dans un espace perceptif tridimensionnel (figure IV.4). Les symboles évoluant dans cet espace ont deux composantes : une composante acoustique représentée par les deux premiers formants et une composante supplémentaire pour le trait.

Le protocole d'adaptation du modèle ESPECE appliqué à l'apparition de traits supplémentaires dans les systèmes vocaliques, est une combinaison des principes de la simulation de la prédiction des voyelles et de celle du rectangle. Pour réaliser une répulsion d'un phonème i par rapport à un autre j, il faut commencer par calculer la distance formantique d_{form} exprimée en Bark et la différence entre les traits Δ_{trait} données par :

$$d_{form} = ((F_1^i - F_1^j)^2 + \lambda^2(F_2^i - F_2^j)^2)^{1/2} \qquad \text{(IV.12)}$$

$$\Delta_{trait} = trait_i - trait_j \qquad \text{(IV.13)}$$

Ensuite, comme ce fut le cas lors des simulations du rectangle, la répulsion s'effectuera selon la composante qui présente la plus grande différence afin d'accroître encore plus cette distinctivité. Par conséquent, après avoir comparé d_{form} et Δ_{trait}, nous identifierons la composante sur laquelle la répulsion sera accomplie.

L'intensité de la répulsion sur la composante formantique obéit à l'équation (II.10) employée lors de la prédiction des systèmes vocaliques (cf. § II.5.3). La répulsion s'accomplissant sur la dimension des traits est exprimée par :

$$trait_i \leftarrow trait_i + (1/\Delta_{trait}) \qquad \text{(IV.14)}$$

Comme ce fut le cas pour la répulsion formantique, l'intensité de cette répulsion est inversement proportionnelle à la différence de traits : la répulsion est d'autant plus importante que Δ_{trait} est faible.

La distance perceptive, utilisée pour sélectionner le symbole à attirer et les autres à repousser lors d'une transaction, entre les phonèmes i et j et est exprimée par :

$$d_{perc}^2 = d_{form}^2 + \Delta_{trait}^2 \quad \text{(IV.15)}$$

où :

d_{form} et Δ_{trait} sont données respectivement par les équations (IV.12) et (IV.13).

En plus des sept paramètres du modèle ESPECE, sera ajouté un huitième paramètre qui représente la hauteur du trait (valeur maximale que peut prendre la composante de la dimension trait).

IV.10 Expériences

Nous avons effectué des simulations du modèle ESPECE ainsi modifié. Nous avons repris le même jeu de paramètres que celui employé pour la prédiction des systèmes vocaliques à savoir :

- une taille de population égale à 10,
- un nombre de transactions s'élèvant à 3 000,
- une force de répulsion ayant pour valeur 0.5,
- un nombre de transactions avant de procéder à un remplacement fixé à 100,
- un facteur λ assigné à 0.3.

Les tailles de lexique considérées varient de 5 à 14 voyelles : il suffit de présenter, dans le contexte de cette expérience, les simulations des populations avec des individus possédant 5 symboles et plus pour sonder l'exploitation des composantes perceptives (formantique et trait) des populations. D'autre part, nous ne considérerons pas les populations avec des lexiques de taille supérieure à 14 car nous avons constaté que les langues de la base UPSID$_{317}$ de 15 voyelles et plus sont peu fréquentes. En outre, elles présentent toutes des systèmes vocaliques différents (cf. tableau IV.1). Il n'existe donc pas de système majoritaire pour ces tailles.

Pour chacune de ces tailles, l'expérience porte également sur la réalisation de 40 simulations pour chaque plage de variation du trait. Nous considérons quatre plages de variation pour tester l'influence qu'exercent les plages de variation du trait supplémentaire sur les résultats. Ces quatre plages sont fixées respectivement à : [0 1], [0 2], [0 3] et [0 4].

Décrivons maintenant les résultats de ces simulations en commençant par le cas où la taille des lexiques est fixée à 5 voyelles.

Nombre de voyelles	Fréquence dans les langues du monde	Nombre de systèmes différents	Fréquence du système majoritaire
3	18	8	10
4	15	13	2
5	74	25	34
6	42	29	10
7	29	22	7
8	25	24	2
9	26	23	3
10	21	15	7
11	11	11	1
12	16	11	5
13	2	2	1
14	9	8	2
15	6	6	1
16	9	9	1
17	4	4	1
18	4	4	1
19	3	3	1
20	2	2	1
24	1	1	1

TAB. IV.1: Fréquences d'apparition des systèmes vocaliques dans la base UPSID_{317}.

Les tableaux IV.2 jusqu'à IV.11 résument les statistiques sur 40 simulations pour chacune des quatre plages de variation du trait. Ces statistiques sont illustrées sous forme de graphique dans, respectivement, les figures IV.11 à IV.20.

Un système sera dit *(a+b)* si ce système présente *a* voyelles de base et *b* voyelles n'utilisant pas de nouveaux timbres mais exploitant un trait supplémentaire. Bien entendu, *"a+b"* est égale à la taille du lexique. Un exemple des différents systèmes à *(a+b)* est donné en annexe H. Les systèmes qualifiés d'«autres» donnent la proportion de systèmes n'ayant pas convergé.

IV.10.1 Systèmes à 5 voyelles

Plage du trait	% des syst. à (3+2)	% des syst. à (4+1)	% des syst. à (5+0)	% des autres
[0 1]	5 %	50 %	45 %	0 %
[0 2]	90 %	5 %	0 %	5 %
[0 3]	92.5 %	7.5 %	0 %	0 %
[0 4]	80 %	7.5 %	0 %	12.5 %

TAB. IV.2: Statistiques des systèmes à 5 voyelles pour des plages de variation du trait différentes. Un système à *(a+b)* signifie que ce système possède *a* voyelles de base et *b* voyelles utilisant un trait supplémentaire.

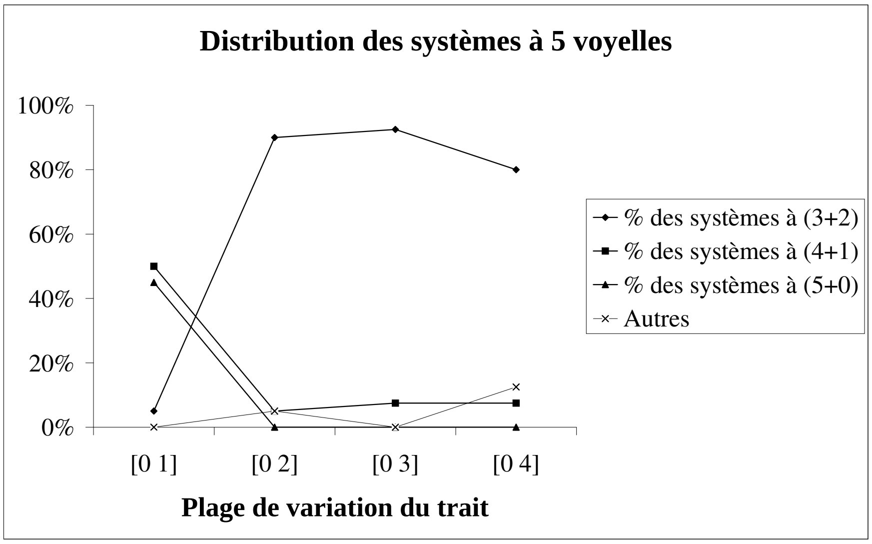


FIG. IV.11: Distribution des systèmes à 5 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales [0 1], [0 2], [0 3] et [0 4].

Au vu du tableau IV.2 et du graphique IV.11, les systèmes (5+0) exclusivement retrouvés dans les langues du monde ne sont prédits que lorsque le trait varie dans l'intervalle [0 1].

IV.10.2 Systèmes à 6 voyelles

Plage du trait	% des syst. à (3+3)	% des syst. à (4+2)	% des syst. à (5+1)	% des syst. à (6+0)	% des autres
[0 1]	0 %	12.5 %	80 %	5 %	2.5 %
[0 2]	72.5 %	25 %	2.5 %	0 %	0 %
[0 3]	92.5 %	2.5 %	2.5 %	0 %	2.5 %
[0 4]	70 %	15 %	5 %	0 %	10 %

TAB. IV.3: Statistiques des systèmes à 6 voyelles pour des plages de variation du trait différentes.

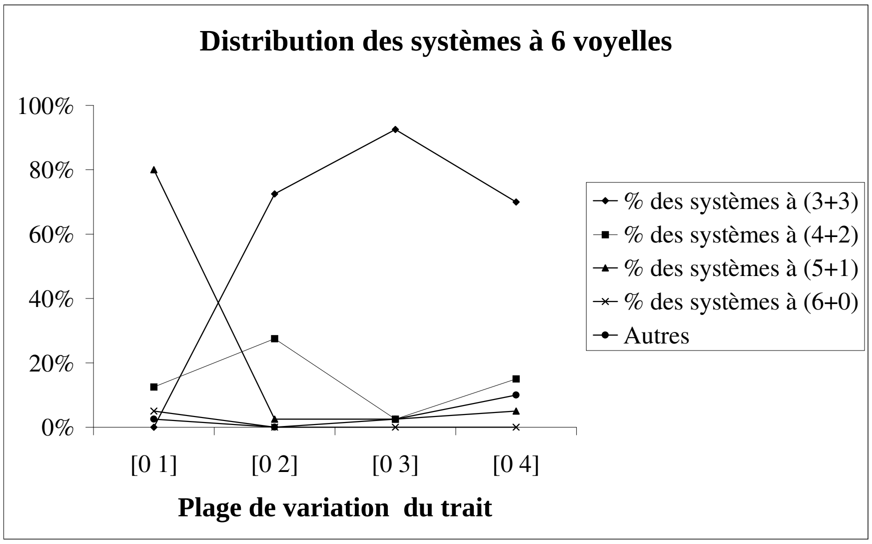


FIG. IV.12: Distribution des systèmes à 6 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Les langues à 6 voyelles présentent systématiquement des systèmes à (6+0). Nous retrouvons uniquement deux langues (le *Ngizim* et le *Khasi*) à (5+1). Le systèmes à (5+1) émergent le plus pour une petite plage du trait dans le modèle ESPECE.

IV.10.3 Systèmes à 7 voyelles

Plage du trait	% des syst. à (4+3)	% des syst. à (5+2)	% des syst. à (6+1)	% des autres
[0 1]	5 %	50 %	35 %	10 %
[0 2]	90 %	2.5 %	0 %	7.5 %
[0 3]	90 %	2.5 %	0 %	7.5 %
[0 4]	80 %	5 %	0 %	15 %

TAB. IV.4: Statistiques des systèmes à 7 voyelles pour des plages de variation du trait différentes.

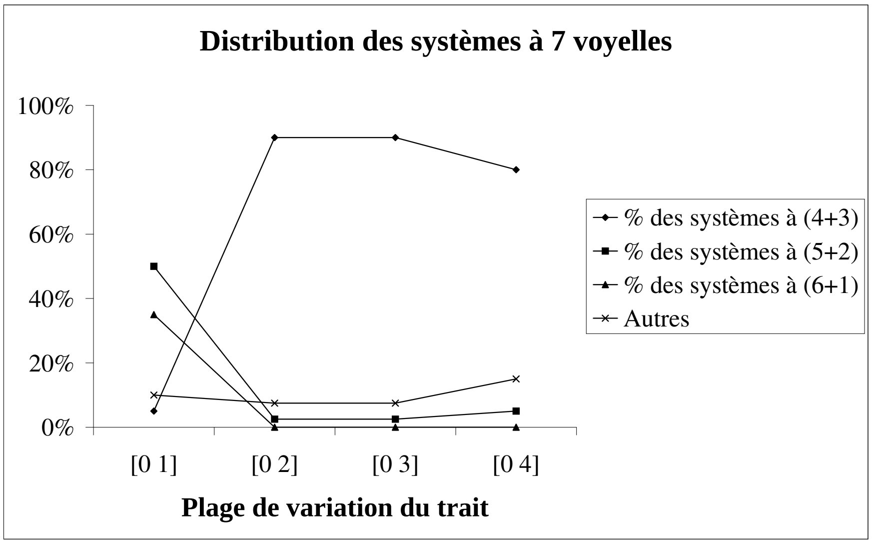


FIG. IV.13: Distribution des systèmes à 7 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

22 des 29 langues à 7 voyelles de la base UPSID_{317} possèdent 7 timbres vocaliques. Toutefois, nous retrouvons trois langues à (6+1) (le *Mandarin*, le *Tigré* et le *Mandchou*), une langue à (5+2) (l'*Atayal*) et trois langues à (4+3) (le *Bardi*, l'*Adzera* et le *Hupa*). Notons qu'ESPECE aboutit à ces 3 systèmes, en particulier lorsque le trait a la possibilité de varier entre 0 et 1.

IV.10.4 Systèmes à 8 voyelles

Plage du trait	% des syst. à (5+3)	% des syst. à (4+4)	% des syst. à (6+2)	% des autres
[0 1]	60 %	0 %	15 %	25 %
[0 2]	55 %	40 %	0 %	5 %
[0 3]	82.5 %	12.5 %	0 %	5 %
[0 4]	57.5 %	12.5 %	0 %	30 %

TAB. IV.5: Statistiques des systèmes à 8 voyelles pour des plages de variation du trait différentes.

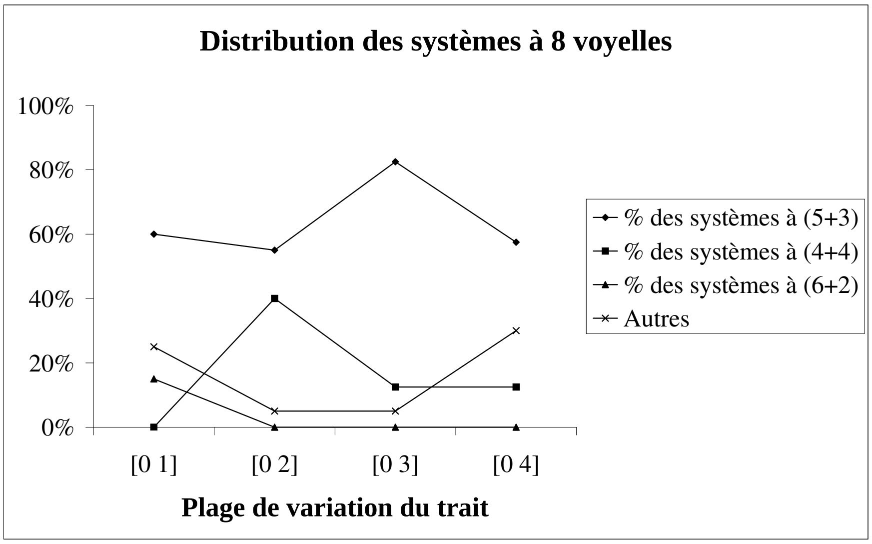


FIG. IV.14: Distribution des systèmes à 8 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Parmi les 25 langues à 8 voyelles, il existe 13 langues ne présentant pas 8 timbres vocaliques différents à savoir :

- 4 langues à (5+3) : *Dagomba, Yourake, Telefol et Dakota*,
- 6 langues à (4+4) : *Amahuaca, Paez, Uapichana, Mazatec, Cashinahua et Jivaro* ,
- 1 langue à (6+2) : *Wichita*,
- 2 langues à (7+1) : *Iai et Saban*.

C'est le système à (5+3) qui émerge le plus dans nos simulations.

IV.10.5 Systèmes à 9 voyelles

Plage du trait	% des syst. à (6+3)	% des syst. à (5+4)	% des syst. à (7+2)	% des autres
[0 1]	40 %	27.5 %	5 %	27.5 %
[0 2]	10 %	87.5 %	0 %	2.5 %
[0 3]	37.5 %	37.5 %	0 %	25 %
[0 4]	7.5 %	55 %	0 %	37.5 %

TAB. IV.6: Statistiques des systèmes à 9 voyelles pour des plages de variation du trait différentes.

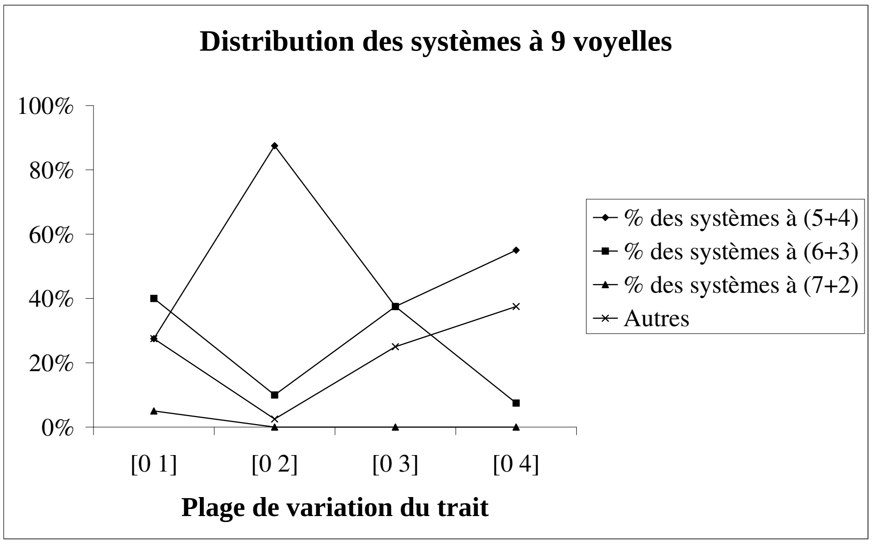


FIG. IV.15: Distribution des systèmes à 9 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Nous dénombrons 9 langues parmi les 26 répertoriées dans la base UPSID$_{317}$ qui ne soient pas à (9+0). Ces langues sont :

- *Lak et Nama* (6+3),
- *Brahoui, Songhaï, Chatino et Ocaina* (5+4),
- *Diegueno et Changchow* (7+2),
- *Karen* (8+1).

Le modèle ESPECE fait apparaître les systèmes à (5+4) et à (6+3) notamment lorsque la plage du trait n'est pas vaste.

IV.10.6 Systèmes à 10 voyelles

Plage du trait	% des syst. à (6+4)	% des syst. à (5+5)	% des syst. à (7+3)	% des autres
[0 1]	37.5 %	5 %	22.5 %	35 %
[0 2]	70 %	15 %	5 %	10 %
[0 3]	40 %	37.5 %	5 %	17.5 %
[0 4]	37.5 %	20 %	0 %	42.5 %

TAB. IV.7: Statistiques des systèmes à 10 voyelles pour des plages de variation du trait différentes.

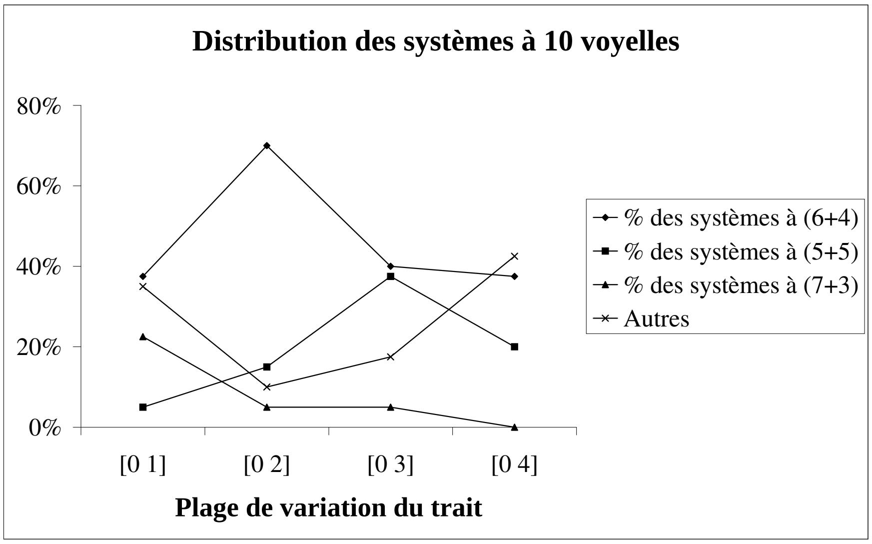


FIG. IV.16: Distribution des systèmes à 10 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

En examinant les systèmes de langues à 10 voyelles, nous retrouvons 12 langues à (5+5), 3 langues à (6+4), ainsi qu'à (7+3) et à (10+0). Les systèmes à (6+4) émergent le plus dans les prédictions du modèle ESPECE. Toutefois, des systèmes à (5+5) et à (7+3) sont également observés dans les prédictions.

IV.10.7 Systèmes à 11 voyelles

Plage du trait	% des syst. à (6+5)	% des syst. à (7+4)	% des syst. à (8+3)	% des autres
[0 1]	57.5 %	25 %	2.5 %	25 %
[0 2]	40 %	40 %	5 %	15 %
[0 3]	25 %	25 %	5 %	45 %
[0 4]	17.5 %	25 %	0 %	57.5 %

TAB. IV.8: Statistiques des systèmes à 11 voyelles pour des plages de variation du trait différentes.

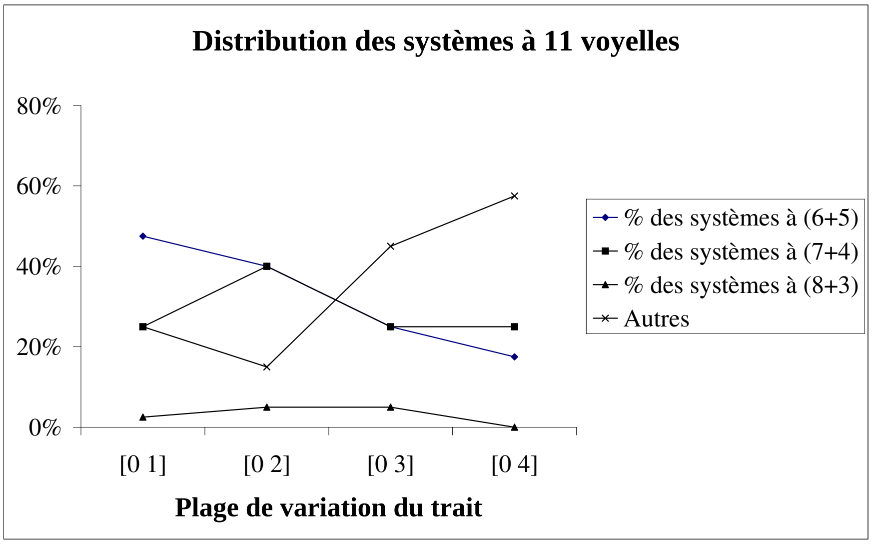


FIG. IV.17: Distribution des systèmes à 11 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

L'analyse de près des 11 langues à 11 voyelles révèle que ces langues se classent en :

- 4 langues à (6+5),
- 2 langues à (7+4) ainsi qu'à (11+0),
- 1 langue à (10+1), à (9+2) et à (8+3).

Les systèmes à (6+5) se trouvent être les plus prédits par ESPECE.

IV.10.8 Systèmes à 12 voyelles

Plage du trait	% des syst. à (6+6)	% des syst. à (7+5)	% des syst. à (8+4)	% des syst. à (9+3)	% des autres
[0 1]	20 %	25 %	5 %	2.5 %	47.5 %
[0 2]	12.5 %	35 %	37.5 %	0 %	15 %
[0 3]	2.5 %	12.5 %	12.5 %	0 %	72.5 %
[0 4]	2.5 %	20 %	2.5 %	0 %	75 %

TAB. IV.9: Statistiques des systèmes à 12 voyelles pour des plages de variation du trait différentes.

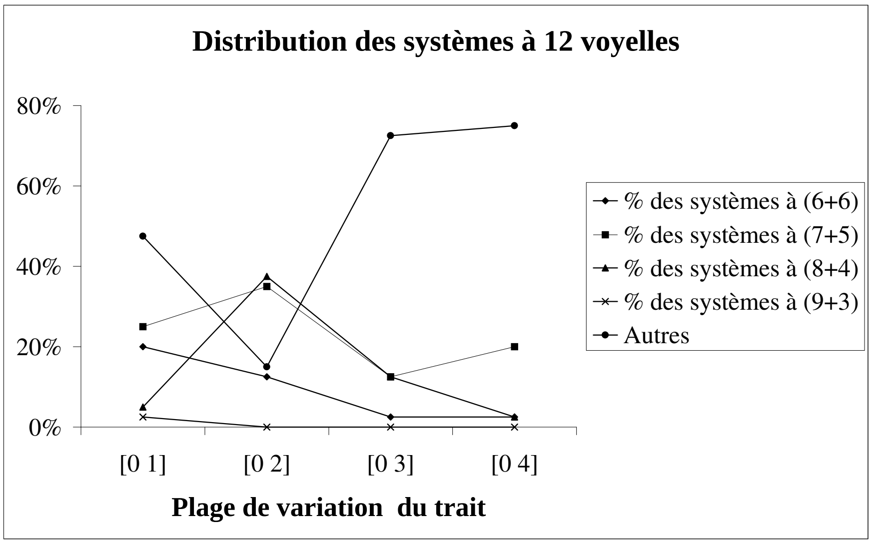


FIG. IV.18: Distribution des systèmes à 12 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Les 16 langues à 12 voyelles peuvent être classées en :

- 8 langues à (7+5),
- 4 langues à (6+6),
- 2 langues à (12+0),
- 1 langue à (8+4) et à (9+3).

Les systèmes à (7+5) puis à (6+6) émergent le plus du modèle ESPECE pour une petite plage de variation du trait. Soulignons toutefois que les systèmes à (8+4) et à (9+3) sont observés également.

IV.10.9 Systèmes à 13 voyelles

Plage du trait	% des syst. à (7+6)	% des syst. à (8+5)	% des syst. à (9+4)	% des autres
[0 1]	40 %	20 %	5 %	35 %
[0 2]	22.5 %	55 %	12.5 %	10 %
[0 3]	20 %	30 %	2.5 %	47.5 %
[0 4]	25 %	10 %	0 %	65 %

TAB. IV.10: Statistiques des systèmes à 13 voyelles pour des plages de variation du trait différentes.

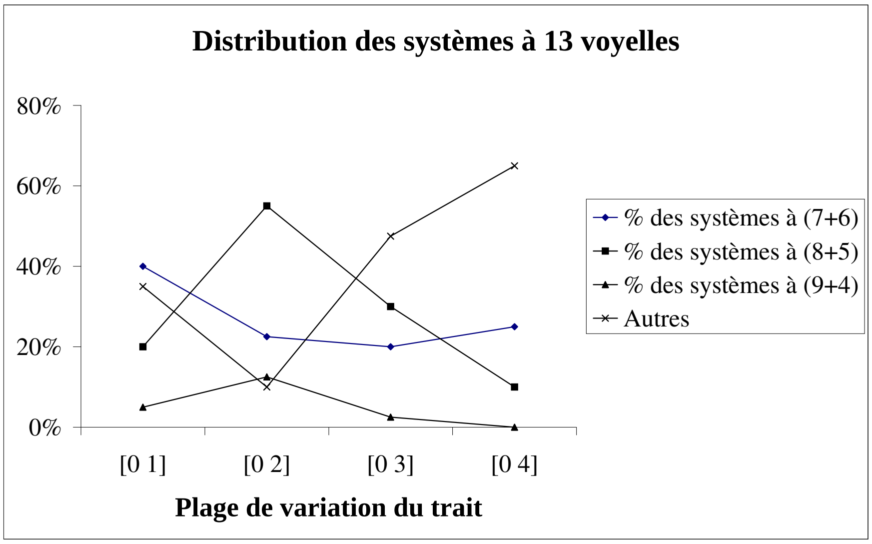


FIG. IV.19: Distribution des systèmes à 13 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Une des 2 langues à 13 voyelles possède un système à (7+6), l'autre présente le système (10+3).

IV.10.10 Systèmes à 14 voyelles

Plage du trait	% des syst. à (7+7)	% des syst. à (8+6)	% des syst. à (9+5)	% des syst. à (10+4)	% des autres
[0 1]	37.5 %	17.5 %	7.5 %	2.5 %	35 %
[0 2]	10 %	35 %	17.5 %	12.5 %	25 %
[0 3]	7.5 %	7.5 %	15.5 %	0 %	70 %
[0 4]	5 %	15 %	5 %	0 %	75 %

TAB. IV.11: Statistiques des systèmes à 14 voyelles pour des plages de variation du trait différentes.

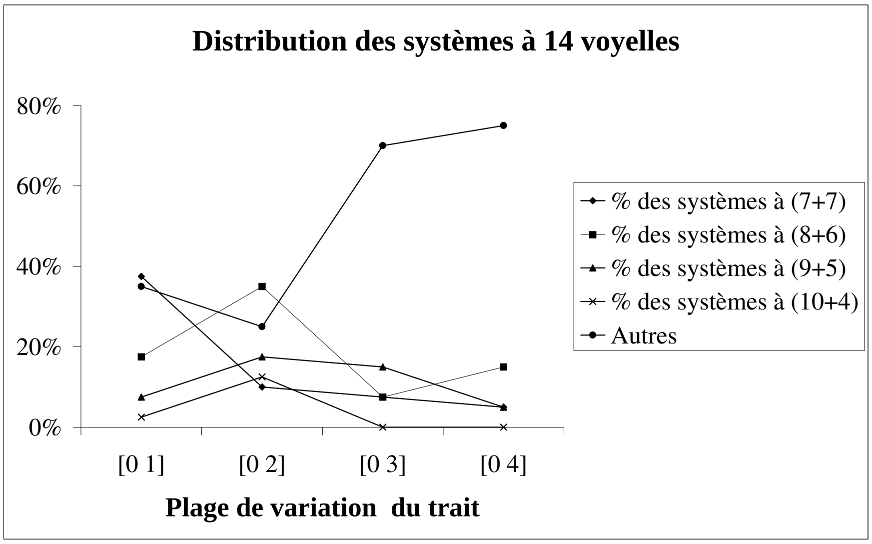


FIG. IV.20: Distribution des systèmes à 14 voyelles dans l'espace perceptif à 3 dimensions pour des plages de variation du trait égales à [0 1], [0 2], [0 3] et [0 4].

Nous dénombrons 4 systèmes à (7+7) et 2 à (8+6) dans 9 langues à 14 voyelles. Ces systèmes sont encore retrouvés dans les prédictions les plus élevées d'ESPECE, toujours lorsque la plage de variation est petite (<2).

IV.10.11 Discussion

Analyse des résultats

Les graphiques présentés montrent bien le phénomène de dédoublement de voyelles en exploitant la nouvelle composante offerte Il est aisé de remarquer que les systèmes, ayant de petites tailles de lexique, ont moins tendance à explorer cette seconde composante lorsque la plage de variation du trait est égale à 1. En revanche, le nombre de systèmes marginaux (n'ayant pas convergé) augmente lorsque la plage du trait est plus large. Ce résultat trouve une interprétation dans le fait que plus la plage du trait est vaste plus les symboles des individus explorent dès les premières transactions la nouvelle dimension, et ont du mal à opter pour une position commune. Nous déduisons donc qu'une hauteur maximale de trait importante (>3) pénalise la convergence. Le taux de «déchets» atteint son paroxysme pour de grandes tailles de lexique. En effet, l'espace perceptif se trouve saturé dans ce cas, et les individus éprouvent les plus grandes difficultés à s'entendre sur une configuration commune.

Par ailleurs, il était prévisible que l'exploitation d'une seule dimension deviendrait plus marquée pour de petites hauteurs, le cas extrême étant lorsque la hauteur est nulle, ce qui nous ramène à un espace perceptif à 2 dimensions exclusivement formantique.

La question légitime à laquelle nous sommes confrontés est de déterminer la plage de variation du trait qui rende compte le mieux des caractéristiques des langues du monde dans les prédictions du modèle ESPECE. Au vu des résultats présentés dans les tableaux IV.2 à IV.11, et dans les graphiques IV.11 à IV.20, nous pouvons d'ores et déjà éliminer les plages plus larges que $[0\ 2]$ car pour ces plages :

- les populations explorent systématiquement la seconde composante pour de petites tailles de lexiques (5 et 6 voyelles), ce qui n'est pas observé dans les langues du monde,
- les populations ont du mal à converger pour des tailles de lexique de plus de 9 symboles.

Certes, lorsque la plage est fixée à $[0\ 2]$, les prédictions de systèmes de plus de 6 voyelles sont observées dans les langues du monde, surtout pour les systèmes à 8 et à 9 voyelles. Cependant, les populations explorent également systématiquement la seconde composante pour des tailles de lexique de 5 et 6 voyelles.

Ceci nous amène à dire que les meilleures prédictions de notre modèle ont été obtenues pour la plus petite plage ($[0\ 1]$) notamment pour les systèmes à 5, à 6, à 11 et à 14 voyelles. Les prédictions des autres systèmes sont comparables à celles obtenues avec une plage de $[0\ 2]$.

En définitive, nous estimons que le plage de variation du trait doit être vraisemblablement inférieure à 1 pour permettre la réalisation de systèmes à 7, à 8 à 9 voyelles de base comme cela est majoritairement observé dans les langues du monde.

Digression autour du poids perceptif entre les deux composantes

Dans les simulations présentées dans ce chapitre, nous avons ajouté une dimension supplémentaire dans l'espace perceptif des voyelles par rapport à l'espace des timbres vocaliques de base (l'espace F_1–F_2). N'étant pas en mesure de justifier ni la nature exacte, ni la plage de variation de ce trait, des batteries de simulation ont été réalisées pour des tailles différentes du lexique et des hauteurs variées de la dimension supplémentaire. Bien évidemment, la variation de l'extension du trait (entre 1 et 4 dans notre cas) équivaut à la prise en compte des métriques différentes pour l'espace perceptif étendu (timbre de base + trait supplémentaire).

La question précise à laquelle nous sommes confrontés est la suivante : selon la nature du trait, peut-on inférer une métrique perceptive qui pondère convenablement les variations dans l'espace formantique et les variations le long de la dimension supplémentaire ? Nous essayerons humblement de donner des éléments de réponse à cette question en fonction des résultats des expériences perceptives disponibles dans la littérature. Deux cas de figure seront discutés : la nasalité et la durée (quantité).

La nasalité : Wright (1986) présente une étude basée sur l'analyse multidimensionnelle des similarités («multidimensional scaling» (Kruskal 1977)) sur la perception de la nasalité des voyelles. Dans son protocole d'expérience, les sujets devaient estimer la similarité entre des paires de voyelles dans une échelle de 0 à 100. Les stimuli sonores ont été produit par un locuteur en situation contrôlée et comprenaient les 8 voyelles orales périphériques de l'*Anglais Américain* ([i, e, ɛ, æ, ɑ, o, U, u]) et de leur correspondante nasalisée. Des précautions ont été prises pour assurer que la distinction de nasalisation s'effectuait uniquement par l'abaissement du velum, la configuration articulatoire de la langue et des lèvres restant la même.

L'analyse multidimensionnelle permet d'extraire les dimensions qui expliquent au mieux les jugements des similarités données par les sujets. Un des buts avoués est de représenter les données dans un espace de dimension réduite, ce qui est commode pour la visualisation. Cependant, les résultats peuvent éventuellement refléter une paramétrisation du signal au niveau du traitement auditif. Dans la figure IV.21, nous présentons le résultat de l'analyse multidimensionnelle de similarité de l'expérience de Wright (1986) dans les deux premières dimensions. Dans cet espace, nous observons, tout d'abord, que l'organisation topologique des voyelles respecte celle du plan F_1–F_2. En second lieu, il est clair que la distinction de «timbre» prime sur la distinction de nasalité dans cet espace.

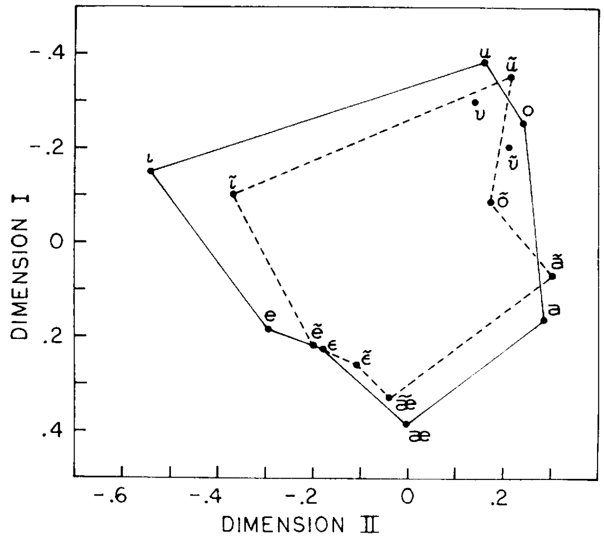


FIG. IV.21: Les deux premières dimensions de l'espace perceptif déterminé par l'analyse multidimensionnelle (d'après Wright (1986)).

Quand la troisième dimension principale est ajoutée, la distinction de la nasalité devient claire (cf. figure IV.22). D'après la figure, et en supposant que les échelles des trois axes sont les mêmes, il est possible de mesurer les distances entre les paires de voyelles orales/nasales et de comparer les poids relatives entre les dimensions I et II par rapport à la dimension III. Le résultats de nos mesures donne une moyenne de 0.29 pour la distance entre voyelles orales et nasales correspondantes selon la dimension III. Selon les dimensions I et II, les données ont des plages de variations autour de 0.8. Nous pouvons donc conclure que dans le cas de l'*Américain* (langue pour laquelle, disons au passage, la nasalité n'est pas de rôle de distinctivité) un poids relative situé entre 2 et 3 pour l'espace F_1–F_2 vis-à-vis du trait de nasalisation semblerait approprié.

Cette configuration de points dans l'espace à trois dimensions a amené l'auteur à proposer l'hypothèse du *cône tronqué*. Selon cette hypothèse, l'espace perceptif à trois dimensions possède une forme de cône tronqué (cf. figure IV.23). La nasalisation peut ainsi être interprétée comme une direction optionnelle permettant à l'espace vocalique de s'étendre, tout en maintenant un contraste perceptif suffisant lorsque le nombre de voyelles croît.

La durée : Abramson & Ren (1990) ont mené une étude sur la perception de la durée des voyelles du Thai. Notons au passage que le Thai est une langue pour laquelle la durée des voyelles a un rôle de distinctivité. Leur expérimentation a consisté à présenter à des sujets soit des voyelles courtes qu'ils allongeaient, soit des voyelles longues qu'ils excisaient. Leur objectif était d'étudier à partir de quelle variation sur la durée, la voyelle était perçu différemment de son timbre.

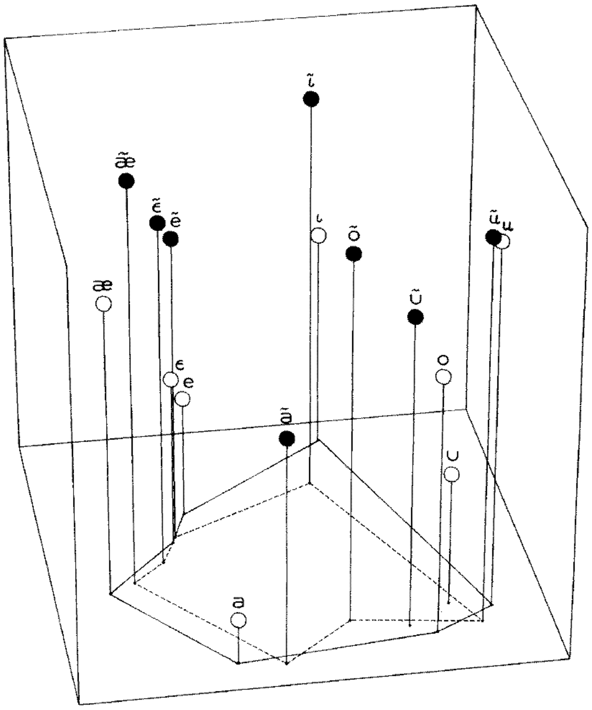


FIG. IV.22: Les trois dimensions de l'espace obtenu de l'analyse multidimensionnelle (d'après Wright (1986)).

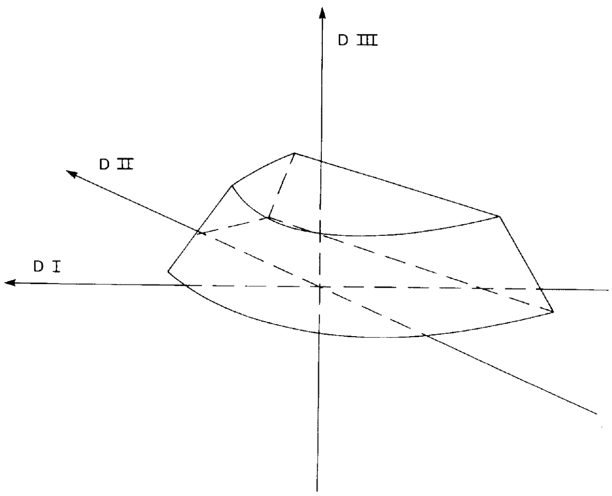


FIG. IV.23: La forme de l'espace perceptif vocalique selon l'hypothèse du cône tronqué (d'après Wright (1986)).

La figure IV.24 représente les résultats de l'expérimentation des variations sur /i/ et sur /i:/. La voyelle /i/ de référence a une durée de 35 ms, et ses deux premiers formants ont respectivement pour valeur 320 Hz et 2030 Hz. le /iː/ a, quant à lui, une durée de 105 ms, et ses deux premiers formants valent 295 Hz et 2115 Hz.

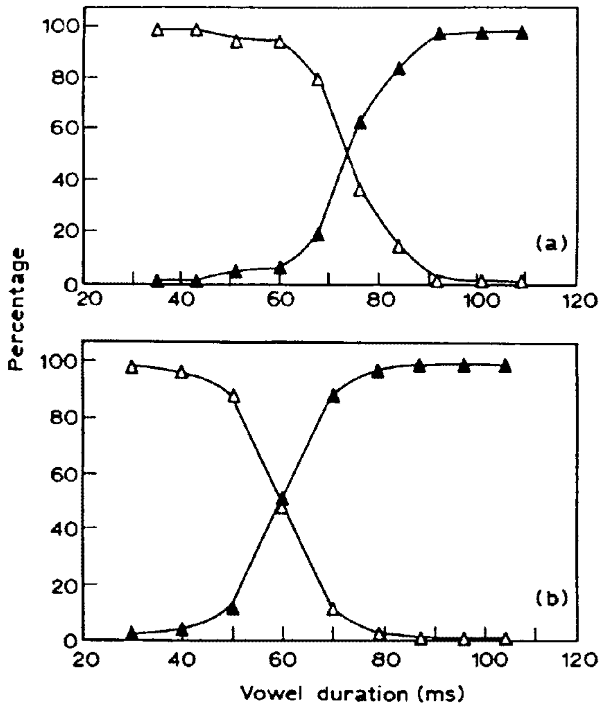


FIG. IV.24: Pourcentage de perception de voyelles courtes (△) ou longues (▲) pour des variantes de (a) /i/ allongé et (b) /iː/ écourté (d'après Abramson & Ren (1990)).

La figure IV.24 (a) représente les pourcentages de perception lorsque des variantes de /i/ sont présentées. Il est aisé de constater que pour une durée au delà de 75 ms, les sujets perçoivent /iː/. Par conséquent, nous déduisons de cette figure que lorsque le durée du /i/ est allongée de 40 ms ($= 75 - 35$), la voyelle perd sa distinctivité. De la même manière pour la figure IV.24 (b) qui illustre les pourcentages de perception du /iː/ excisé, nous constatons que le /iː/ perd sa distinctivité pour une durée plus petite que 60 ms. Autrement dit, lorsque /iː/ était excisé de plus de 45 ms ($= 105 - 60$), c'est la voyelle courte qui est perçue.

Il est possible de calculer la distance perceptive entre /i/ et /iː/ qui vaut 0.08 Bk. Nous pouvons conclure pour le /i/ qu'il y a compensation d'une distance perceptive de 0.08 Bk pour une durée entre 40 et 45 ms.

Gottfried & Beddor (Gottfried et Beddor 1988) ont mené une étude similaire avec les voyelles du Francais. La particularité de leur experience est qu'ils faisaient varier synthétiquement le spectre, tandis que la durée restait constante.

IV.11 Conclusion

Des études typologiques (Vallée 1994) ont montré qu'au-delà de 9 voyelles, les systèmes phonologiques de cardinal «N» avaient tendance à ajouter une nouvelle dimension. Dans la plupart des cas, cette dimension est soit la nasalité, soit la quantité. Par ailleurs, nous avons constaté que ces systèmes conservaient un nombre «N-5» de voyelles dans les dimensions de base pour les systèmes de moins de 13 voyelles, et un nombre «N-7» lorsque la taille du lexique vaut 14. Autrement dit, les langues de moins de 13 voyelles sont majoritairement des systèmes à (N-5 +5), tandis que les systèmes à (7+7) sont les plus observés dans les langues à 14 voyelles.

Nous avons élaboré une plate-forme de simulation visant à reproduire le principe du MUAF. Le concept du MUAF que nous avons testé permet ainsi de quantifier le rapport entre la composante formantique de base et la composante représentant un trait supplémentaire. Les simulations montrent que les systèmes *(a+b)* obtenus ne seront pas réalistes si cette composante supplémentaire est supérieure à 1. Nous pensons qu'elle doit même être inférieure à 1 pour permettre de meilleures prédictions de systèmes à 7, à 8 et à 9 voyelles. Toutefois, mentionnons que nous retrouvons dans nos prédictions, la plupart des inventaires des larges systèmes vocaliques, les plus fréquents dans UPSID_{317}.

En conclusion, nous estimons au vu des résultats, que le cadre de simulation proposé a atteint l'objectif escompté au départ, à savoir, la simulation de l'organisation des systèmes possédant un grand nombre de voyelles.

Conclusion et Perspectives

Le thème abordé dans cette étude relève de la justification des tendances universelles des systèmes phonologiques des langues du monde. La problématique soulevée dans ce manuscrit, est liée à la compréhension de la régularité des systèmes sonores des langues du monde malgré leur grande variété. Cette problématique constitue un des grands enjeux de l'étude de la Communication Parlée.

L'insuffisance des approches traditionnelles pour la prédiction des systèmes phonologiques est qu'elles n'abordent ni les aspects d'interaction parlée ni d'évolution qui caractérisent pourtant le langage. Les travaux effectués, dans le cadre de ce mémoire, ont eu pour objectif d'apporter quelques éléments concernant une approche évolutionniste en tant qu'alternative aux approches traditionnelles. Dans ce sens, le but etait d'analyser comment un code phonologique émergeait au sein d'un groupe d'agents par le jeu de la communication parlée.

Dressons pour commencer un bilan de nos travaux. La première étape a porté sur la conception d'une plate-forme d'échanges de voyelles au sein d'une communauté d'agents de parole. L'aboutissement à un lexique de voyelles commun à tous les agents est une conséquence des aspects de coopération mis en œuvre lors des échanges inter-agents.

Partant de ces principes, un modèle de simulation appelé ESPECE, basé sur la négociation lors des interactions inter-agents, a été développé. Ce modèle a été validé à l'issue d'une application à la prédiction des systèmes vocaliques. Les résultats des simulations se sont avérés satisfaisants suite à leur confrontation aux données obtenues à partir des inventaires phonologiques des langues du monde.

Ensuite, le cadre de simulation ESPECE a été étendu en vue d'intégrer le principe du MUAF. Ce principe permet d'expliquer l'apparition de traits supplémentaires dans les larges systèmes vocaliques. Les résultats générés ont été comparés avec les données existantes sur les langues du monde. Nous avons retrouvé, une nouvelle fois, dans nos prédictions, les principales tendances observées dans les systèmes de la base UPSID_{317} comportant cette fois, un nombre important de voyelles.

Selon nous, le principal apport de ce travail porte sur la consolidation de l'hypothèse selon laquelle le langage est un phénomène résultant essentiellement des communications entre les locuteurs qui l'emploient. Ces communications sont, bien entendu, assujetties à des contraintes fonctionnelles. En outre, une telle approche offre des perspectives dans la recherche de propriétés phonétiques des systèmes : l'expérience du MUAF est un exemple dans ce sens.

Toutefois, le système d'interactions par la communication parlée présenté dans cette thèse n'est pas achevé, et présente des limites évidentes, à savoir :

(i) les items des lexiques sont représentés uniquement en termes d'acoustique, même si l'espace des réalisations formantiques possibles est borné grâce à des considérations de production articulatoire,

(ii) et seules les voyelles y sont simulées.

À ce propos, nous pensons que l'introduction de ces limitations était nécessaire dans un premier temps pour l'élaboration d'un modèle solide. Les perspectives sont donc ouvertes aujourd'hui à des implémentations plus réalistes, avec l'introduction du modèle articulatoire et des transitions syllabiques de type consonne-voyelle.

Par ailleurs, le modèle ESPECE pourrait servir de point de départ à la construction d'un modèle robuste simulant des interactions locuteur-auditeur en situation naturelle. Un tel modèle devrait notamment prendre en compte :

- les aspects réalistes de la négociation locuteur-auditeur prônés par la théorie *Hypo & Hyper* (Lindblom et al. 1995, Lindblom 1995) (cf. annexe G) ; la perspective *H&H* suggère également la simulation des changements phonétiques (Ohala 1993, Labov 1994),
- la diversité anatomique des robots parlants ; autrement dit, envisager d'avoir au sein de la société, des agents de parole munis de conduits vocaux anatomiquement dissemblables ; ceci relève de la problématique classique de la normalisation (cf. § II.2.3).

À court terme, la configuration actuelle du modèle ESPECE pourrait être exploitée, avec un minimum de modifications, afin de simuler les effets géographiques et l'apparition de dialectes. Ces effets pourraient apparaître en plaçant les agents de parole d'une société sur une grille, et en restreignant les communications de chaque agent à son voisinage.

D'autre part, une multitude de simulations intéressantes pourraient être envisagées, telles l'observation de l'apprentissage de la parole (introduire de nouveaux agents de parole dans une société structurée), ou bien les effets du flux de l'immigration sur le langage des autochtones et des immigrants (intégrer un groupe d'agents, ayant un lexique commun, dans une société «parlant» une langue différente).

En conclusion, nous sommes conscients que nous n'avons pas révolutionné, avec notre modèle ESPECE, la recherche sur la justification des tendances universelles des langues du monde. Il nous semble, malgré tout, que l'effort de conception et de formalisation que nous avons entrepris dans cette étude, peuvent être bénéfique à un chercheur voulant adopter un approche locale pour la justification des régularités des systèmes sonores.

Enfin, force est de constater, au bout de tous ces développements, que le chemin à parcourir avant de trouver les explications sur le contenu des inventaires des systèmes de sons est encore important.

Bibliographie

Abercrombie, D. (1966), *Elements of General Phonetics*, Edinburgh University Press, Edinburg.

Abramson, A. S. et Ren, N. (1990), 'Dinstinctive vowel length : duration vs. spectrum in thai', *Journal of Phonetics* **18**, 79–92.

Abry, C., Boë, L.-J. et Schwartz, J.-L. (1989), 'Plateaus, catastrophes and the structuring of vowel systems', *Journal of Phonetics* **17**, 47–54.

Badin, P. et Fant, G. (1984), 'Notes on vocal tract computation', *STL-QPSR* **2–3**, 53–108.

Batali, J. (1998), Computational simulations of the emergence of grammar, J. Hurford, M. Studdert-Kennedy et C. Knight, eds, 'Approaches to the evolution of language : social and cognitive bases', Cambridge University Press.

Beautemps, D., Badin, P. et Laboissière, R. (1995), 'Deriving vocal-tract area functions from midsagittal profiles and formant frequencies : A new model for vowels and fricative consonants based on experimental data', *Speech Communication* **16**, 27–47.

Berrah, A.-R. (1997), Artificial life of a society of speech robots : Emergence of phonetic structures, 'Second German Workshop on Artificial Life', Dortmund, pp. 45–47.

Berrah, A.-R., Boë, L.-J. et Schwartz, J.-L. (1995), Emergent syllable using articulatory and acoustic principles, K. Elenius et P. Branderud, eds, 'XIIIth International Congress of Phonetic Science', Vol. 1, Stockholm, pp. 396–399.

Berrah, A.-R. et Laboissière, R. (1997), Phonetic code emergence in a society of speech robots : explaining vowel systems and the muaf principle, 'EuroSpeech97', Vol. 5, Rhodes-Grèce, pp. 2395–2398.

Berrah, A.-R., Glotin, H., Laboissière, R., Bessière, P. et Boë, L.-J. (1996), From form to formation of phonetic structures : An evolutionary computing perspective, T. Fogarty et G. Venturini, eds, '13th International Conference on Machine Learning Workshop on Evolutionary Computing and Machine Learning', Bari, pp. 23–29.

Bladon, A. et Lindblom, B. (1981), 'Modeling the judgement of vowel quality differences', *Journal of the Acoustical Society of America* **65**, 1414–1422.

Boë, L.-J., Schwartz, J.-L., Berrah, A.-R., Vallée, N. et Abry, C. (1996), La matérialité des structures sonores : 2-de la prédiction à l'ontogenèse, '21ème Journées d'Étude sur la Parole', Avignon, pp. 143–146.

Boë, L.-J., Perrier, P. et Bailly, G. (1992), 'The geometric vocal tract variables controlled for vowel production : Proposals for constraining acoustic-to-articulatory inversion', *Journal of Phonetics* **20**, 27–38.

Boë, L.-J., Perrier, P., Guérin, B. et Schwartz, J.-L. (1989), Maximal vowel space, 'EuroSpeech89', Vol. 2, Paris, pp. 281–284.

Botherel, A., Simon, P., Wioland, F. et Zerling, J.-P. (1986), *Cinéradiographie des voyelles et des consonnes du français. Recueil de documents synchronisés pour quatre sujets : Vues latérales du conduit vocal, vues frontales de l'orifice labial, données acoustiques*, Institut de Phonétique de Strasbourg, Strasbourg, France.

Brown, R. et Lenneberg, E. (1954), 'A study in language and cognition', *Journal of Abnormal and Social Psychology* **49**, 454–462.

Cangelosi, A. et Parisi, D. (1996), The emergence of a "language" in an evolving population of neural networks, Technical Report NSAL-96004, Istituto di Psicologia, C.N.R. Rome.

Chomsky, N. (1986), *Knowledge of Language : Its Nature, Origin and Use*, Greenwood press.

Chomsky, N. et Halle, M. (1968), *The Sound Pattern of English*, MIT Press, Cambridge, Ma.

Crothers, J. (1978), Typology and universals in vowel systems, J. Greenberg, C. Ferguson et E. Moravcsik, eds, 'Universals of Human Language', Stanford University Press, Stanford, pp. 93–152.

De Boer, B. (1997*a*), Generating vowel systems in a population of agents, 'Fourth European Conference On Artificial Life', MIT Press, Brighton.

De Boer, B. (1997*b*), Self organisation in vowel systems through imitation, 'ACL/EACL-97 SIGPHON workshop', Madrid.

Delattre, P., Liberman, A., Cooper, F. et Gertsman, J. (1952), 'An experimental study of the acoustic determinants of vowel color ; observations on one- and two-formant vowels synthesized from spectrographic patterns', *Word* **8**, 195–210.

Feng, G. (1983), Vers une synthèse par la méthode des pôles et zéros, '13ème Journées d'Étude sur la Parole', Bruxelles, pp. 155–157.

Glotin, H. (1995), 'La vie artificielle d'une société de robots parlants : Émergence et changements du code phonétique', Rapport de Stage du D.E.A de l'I.N.P. Grenoble.

Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Workingham, UK.

Gottfried, T. L. et Beddor, P. S. (1988), 'Perception of temporal and spectral information in french vowels', *Language and Speech* **31**, 57–75.

Greenberg, J. (1966), *Universals of language, 2nd edition*, MIT Press, Cambridge, Ma.

Hashimoto, T. et Ikegami, T. (1996), 'Emergence of net-grammar in communicating agents', *BioSystems* **38**, 1–14.

Heudin, J.-C. (1994), *La vie artificielle*, Hermès, Paris.

Jakobson, R., Fant, G. et Halle, M. (1952), *Preliminaries to Speech Analysis*, MIT Press, Cambridge, Ma.

Jospa, P. (1994), Formulation variationnelle du lien acoustico-articulatoire, '20ème Journées d'Étude sur la Parole', Trégastel, pp. 113–118.

Kirby, S. (1996), Function, selection and innateness : The emergence of language universals, Thèse de Doctorat, University of Edinburgh, Edinburgh, UK.

Kruskal, J. B. (1977), Multidimensional scaling and other methods for discovering structure, K. Enslein, A. Ralston et H. S. Wilf, eds, 'Statistical Methods for Digital Computers', Wiley, New York, chapter 12, pp. 296–339.

Laboissière, R. (1992), Préliminaires pour une Robotique de la Communication Parlée : inversion et contrôle d'un modèle articulatoire du conduit vocal, Thèse de Doctorat, Institut National Polytechnique, Grenoble, France.

Labov, W. (1994), *Principles of linguistic change : internal factors*, Blackwell, Cambridge, Ma.

Ladefoged, P. et Maddieson, I. (1990), 'Vowels of the world's languages', *Journal of Phonetics* **18**, 93–122.

Langton, C. (1995), *Artificial Life : an Overview*, MIT Press, Cambridge, Ma.

Liljencrants, J. et Lindblom, B. (1972), 'Numerical simulation of vowel quality systems : the role of perceptual contrast', *Language* **48**, 839–862.

Lindblom, B. (1975), Experiments in sound structure, 'VIIIth International Congress of Phonetic Sciences', Leeds. Aussi dans la *Revue de Phonétique Appliquée* **51**, 155-189. Université de l'Etat de Mons, Belgique.

Lindblom, B. (1986), Phonetic universals in vowel systems, J. Ohala et J. Jaeger, eds, 'Experimental Phonology', Academic Press, Orlando, pp. 13–44.

Lindblom, B. (1990), 'Models of phonetic variation and selection', *Phonetic Experimental Research. Institute of Linguistics. University of Stockholm* **11**, 65–100.

Lindblom, B. (1995), Approche intégrée de la production et de la perception, H. Méloni, ed., 'École thématique : fondements et perspectives en traitement automatique de la parole', pp. 1–8.

Lindblom, B. (1997), Systemic constraints and adaptive change in the formation of sound structure, J. Hurford, ed., 'Evolution of Human Language', Edinbrugh Univ. Press., Edinburgh.

Lindblom, B. et Engstrand, O. (1989), 'In what sense is speech quantal ?', *Journal of Phonetics* **17**, 107–121.

Lindblom, B. et Lubker, J. (1985), The speech homonculus and a problem of phonetic linguistics, V. Fromkin, ed., 'Phonetic Linguistics', Academic Press, Orlando, pp. 169–192.

Lindblom, B. et Maddieson, I. (1988), Phonetic universals in consonant systems, L. Hyman et C. Li, eds, 'Language, speech and mind', Routlege, Londres, pp. 62–78.

Lindblom, B. et Sundberg, J. (1971), 'Acoustical consequences of lip, tongue, jaw and larynx movement', *Journal of the Acoustical Society of America* **50**, 1166–1179.

Lindblom, B., Guion, S., Hura, S. et Moon, S. (1995), Is sound change adaptive ?, C. Fowler, ed., 'Sound Change, édition thématique de Revista di Linguistica'.

Lindblom, B., MacNeilage, P. et Studdert-Kennedy, M. (1984), Self-organizing processes and the explanation of languages universals, B. Butterworth, B. Comrie et O. Dahl, eds, 'Explanation of Language Universals', Mouton, pp. 181–203.

Lobanov, B. (1971), 'Classification of russian vowels spoken by different speakers', *Journal of the Acoustical Society of America* **49**, 606–608.

MacLennan, B. (1991), Synthetic ethology : an approach to the study of communication, C. Langton et . al, eds, 'Artificial Life II', Addison Wesley Pub. Co., Redwood City, pp. 631–658.

MacLennan, B. et Burghardt, G. (1994), 'Synthetic ethology and the evolution of cooperative communication', *Adaptive Behavior* **2**, 161–188.

Maddieson, I. (1986), *Patterns of Sounds*, Cambridge University Press, Cambridge, UK.

Maddieson, I. (1991), Investigating linguistic universals, 'XIIth International Congress of Phonetic Sciences', Vol. 2, Aix-en-Provence, pp. 346–354.

Maddieson, I. et Precoda, K. (1990), 'Updating upsid', *UCLA Working Papers in Phonetics* **74**, 104–111.

Maeda, S. (1988), 'Improved articulatory model', *Journal of the Acoustical Society of America* **84**(S1), S146.

Maeda, S. (1989), Compensatory articulation during speech : Evidence from the analysis of vocal-tract shapes using an articulatory model, W. Hardcastle et A. Marchal, eds, 'Speech Production and Speech Modelling', Academic Publishers, Kluwer, Netherlands, pp. 131–149.

Mantakas, M., Schwartz, J.-L. et Escudier, P. (1986), Modèle de prédiction du "deuxième formant effectif" f'_2 et application à l'étude de la labialité des voyelles avant du français, '15ème Journées d'Étude sur la Parole', Aix-en-Provence, pp. 157–161.

Martinet, A. (1970), *Économie des changements phonétiques : traité de phonologie diachronique*, Francke, A., Berne. 1ère édition : 1955.

Matsumoto, H. et Wakita, H. (1986), 'Vowel normalization by frequency warped spectral matching', *Speech Communication* **5**(2), 239–251.

Mattingly, I. (1966), 'Speaker variation and vocal tract size', *Journal of the Acoustical Society of America* **39**, 1219.

Miller, G. (1956), 'The magic number seven, plus or minus two : some limits on our capacity for processing information', *Psychological review* **3**, 81–97.

Moore, B. et Glasberg, B. (1983), 'Suggested formulae for calculating auditory-filter bandwidths and excitation patterns', *Journal of the Acoustical Society of America* **74**, 750–753.

Nordström, P. (1977), 'Female and infant vocal tracts simulated from male area functions', *Journal of Phonetics* **5**, 81–92.

Ohala, J. (1980), Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation, 'IXth International Congress of Phonetic Science', Vol. 3, Copenhague, pp. 181–185.

Ohala, J. (1993), The phonetics of sound change, C. Jones, ed., 'Historical linguistics problems and perspectives', Longman, London, pp. 237–278.

Perrier, P., Boë, L.-J. et Sock, R. (1992), 'Vocal tract area functions estimation from midsagittal dimensions with ct scans and a vocal tract cast : Modeling the transition with two sets of coefficients', *Journal of Speech and Hearing Research* **35**, 53–67.

Roudet, L. (1910), Éléments de phonétique générale, Paris, pp. 270–351.

Ruhlen, M. (1987), *A guide of the world's languages : classification*, Stanford University Press, Stanford, Ca.

Schroeder, M., Atal, B. et Hall, J. (1979), Objective measure of certain speech signal degradations based on masking properties of human auditory perception, B. Lindblom et S. Öhman, eds, 'Frontiers of Specch Communication Research', Academic Press, London, pp. 217–229.

Schwartz, J.-L., Boë, L.-J., Valée, N. et Abry, C. (1997*a*), 'Major trends in vowel system inventories', *Journal of Phonetics* **25**, 233–253.

Schwartz, J.-L., Boë, L.-J., Vallée, N. et Abry, C. (1997*b*), 'The dispersion-focalization theory of vowel systems', *Journal of Phonetics* **25**, 255–286.

Sedlak, P. (1969), 'Typological considerations of vowel quality systems', *Working Papers in Language Universals* **1**, 1–40.

Steels, L. (1996), Emergent adaptive lexicons, P. Maes et . al, eds, 'From Animals to Animats 4 : Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior', MIT Press, Cambridge, Ma.

Steels, L. (1997*a*), The origins of syntax in visually grounded robotic agents, M. Pollack, ed., 'Proceedings of the 15th International Joint Conference on Artificial Intelligence', Morgan Kauffman, Nagoya.

Steels, L. (1997*b*), Synthesising the origins of language and meaning using co-evolution and self-organisation, J. Hurford, ed., 'Evolution of Human Language', Edinburgh Univ. Press., Edinburgh.

Steels, L. (1997*c*), 'The synthetic modeling of language origins', *Evolution of Communication Journal* **1**(1), 1–35.

Steels, L. et Vogt, P. (1997), Grounding adaptive language games in robotic agents, 'Fourth European Conference On Artificial Life', MIT Press, Brighton.

Stevens, K. (1972), The quantal nature of speech : Evidence from articulatory-acoustic data, P. Denes et J. Davis, eds, 'Human Communication : a Unified View', McGraw-Hill, New York, pp. 51–66.

Stevens, K. (1989), 'On the quantal nature of speech', *Journal of Phonetics* **17**, 3–45.

Troubetskoy, N. (1939), 'Grundzüge des phonologie', *Travaux du cercle linguistique de Prague* **7**, 272p. Traduction Française par J. Cantineu sous le titre : Principes de Phonologie, Klincksieck, 1970, Paris, 394p.

Vallée, N. (1994), Systèmes vocaliques : de la typologie aux prédictions, Thèse de Doctorat, Université Stendhal, Grenoble, France.

Vallée, N., Boë, L.-J. et Schwartz, J.-L. (1997), Taxinomie phonologique et prédiction des tendances universelles des systèmes vocaliques des langues du monde, 'Journées d'Études Linguistiques «La voyelle dans tous ses états»', Université de Nantes, pp. 118–123.

Victorri, B. (1997), Débat sur la langue mère, O. Brossolet, ed., 'Pour la Science (édition française de Scientific American)', Berger-Levrault Graphique Toul, Paris, pp. 28–32.

Wakita, H. (1977), 'Normalization of vowels by vocal-tract length and its application to vowel identification', *IEEE Transactions on Acoustics, Speech and Signal Processing* **25**, 183–192.

Werner, G. et Dyer, M. (1991), Evolution of communication in artificial organisms, C. Langton et . al, eds, 'Artificial Life II', Addison Wesley Pub. Co., Redwood City, pp. 659–687.

Wright, J. (1986), The behavior of nasalized vowels in the perceptual vowel space, J. Ohala et J. Jaeger, eds, 'Experimental Phonology', Academic Press, Orlando, pp. 45–67.

Zahorian, S. et Jagharghi, A. (1991), 'Speaker normalization of static and dynamic vowel spectral features', *Journal of the Acoustical Society of America* **90**, 67–75.

Annexe A

Glossaire de quelques termes linguistiques employés

Allophone : The segments which a phoneme comprises or which represent it, are said to be members of that phoneme, or allophone of it (Abercrombie 1966) [p. 86].

Complementary distribution : When different segments are only found in differents environments, and never occur in the same environement, they are said to be in *complementary distribution* (Abercrombie 1966) [p. 86].

Grammaire : Arrangement des mots dans les phrases selon un ordre obéissant à des règles précises.

Phonème : A group of segments in complementary distribution constitutes a phoneme. A phoneme is a group of segments which are different from the point of view of general phonetic taxonomy but have the same function phonologically (Abercrombie 1966) [p. 86].

Segment : Unité minimale définie par certaines qualités phonétiques.

Syntaxe : Voir *Grammaire*.

Trait distinctif : Un trait distinctif (Jakobson, Fant et Halle 1952) est une caractéristique des structures sonores qui peut provoquer une différence sémantique entre deux mots. À titre d'exemple, les deux mots du Français *poule* et *boule* ne se distinguent que par le trait de voisement de la première consonne ce qui fait du voisement un trait distinctif en Français.

Annexe B

Glossaire de quelques termes employés dans le cadre d'ESPECE

Agent de parole (AP) : Un agent de parole est une entité abstraite possédant un lexique. Un AP et est capable d'agir sur lui-même suite à des communications avec d'autres AP.

Bark : Le Bark est une échelle perceptive.

ESPECE : Acronyme de «**E**mergence dans une **S**ociété d'agents de **P**arole par l'**E**ffet de la **C**oop**E**ration» ; ESPECE est un modèle de simulation basé sur la coopération lors des interactions entre agents de parole.

Individu : Voir *Agent*.

Inventaire : Ensemble de symboles possédé par un agent de parole.

Item : Configuration formantique d'une voyelle qui évolue au cours des échanges.

Lexique : Voir *Inventaire*.

MUAF : Maximum Use of Available distinctive Features.

Ontogenèse : Histoire de la formation d'un individu.

Phylogenèse : Histoire de l'espèce à laquelle appartient un individu.

Société : Communauté ou ensemble d'agents de parole.

Symbole : Voir *Item*.

Transaction : Interaction entre deux agents de parole.

Annexe C

Systèmes vocaliques pour différentes tailles de société

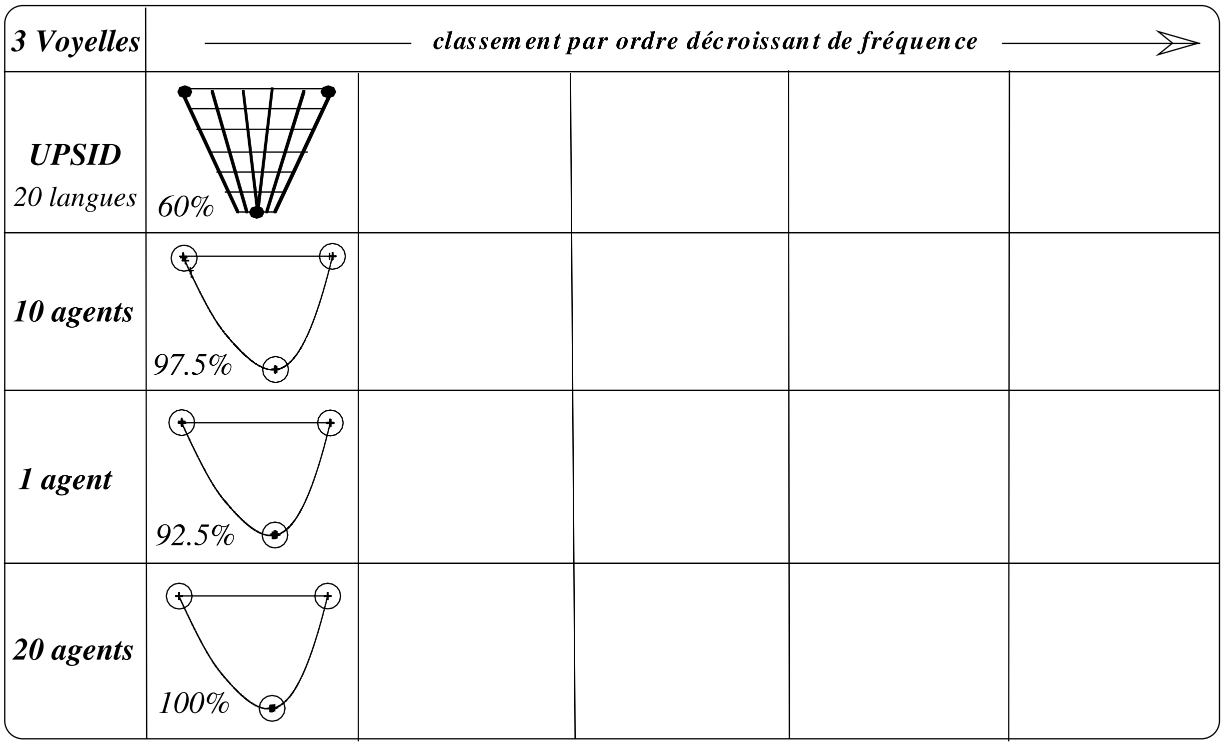


FIG. C.1: Distribution dans le triangle vocalique de voyelles de sociétés d'un, dix et vingt agents ayant un lexique de 3 voyelles.

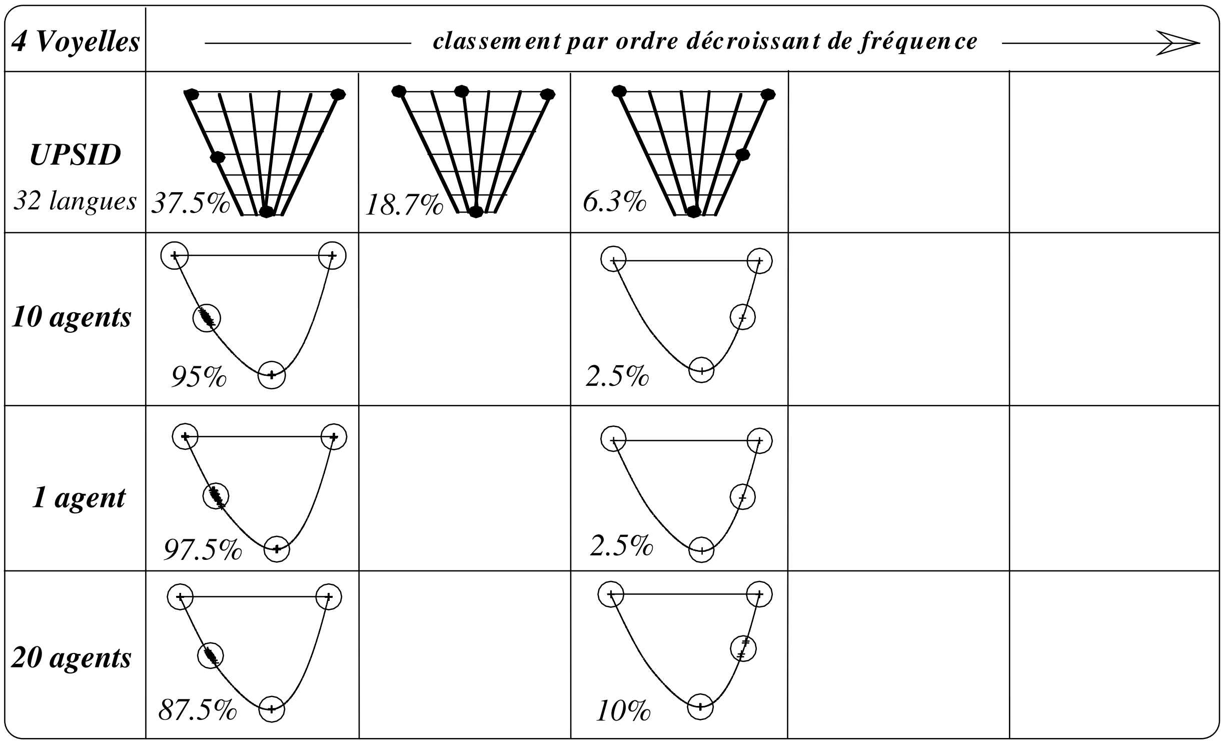


FIG. C.2: Distribution dans le triangle vocalique de voyelles de sociétés d'un, dix et vingt agents ayant un lexique de 4 voyelles.

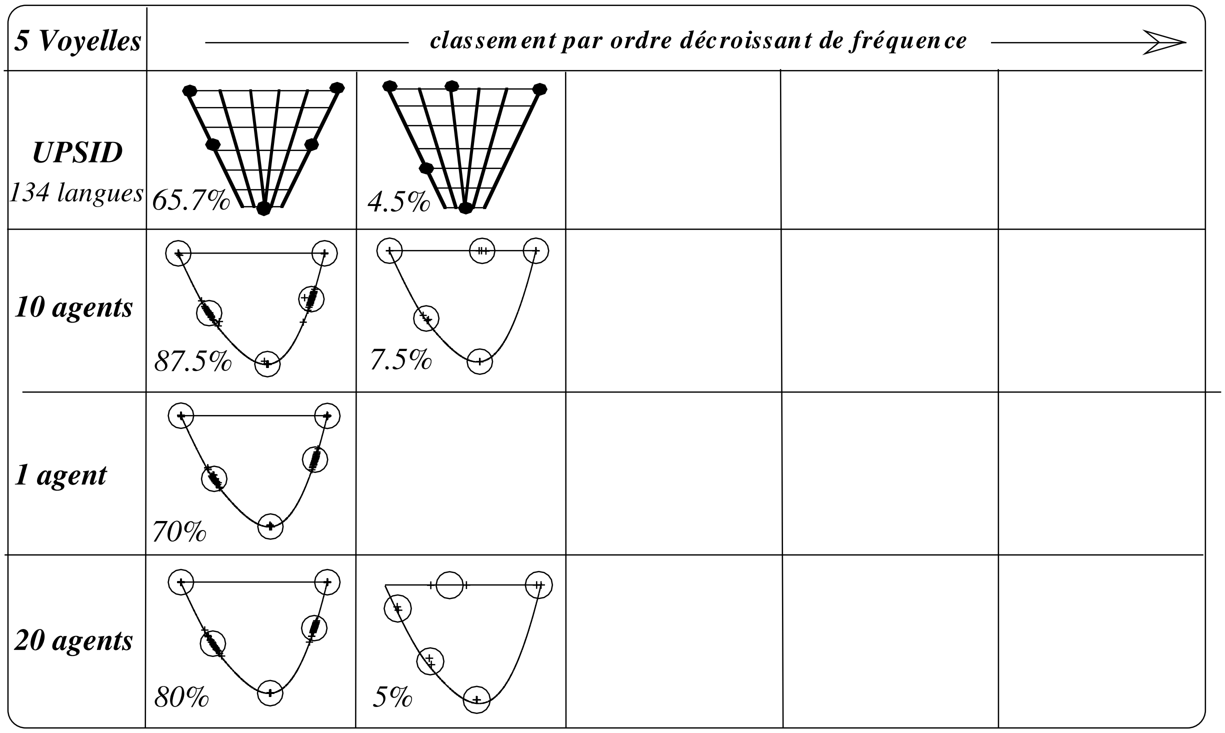


FIG. C.3: Distribution dans le triangle vocalique de voyelles de sociétés d'un, dix et vingt agents ayant un lexique de 5 voyelles.

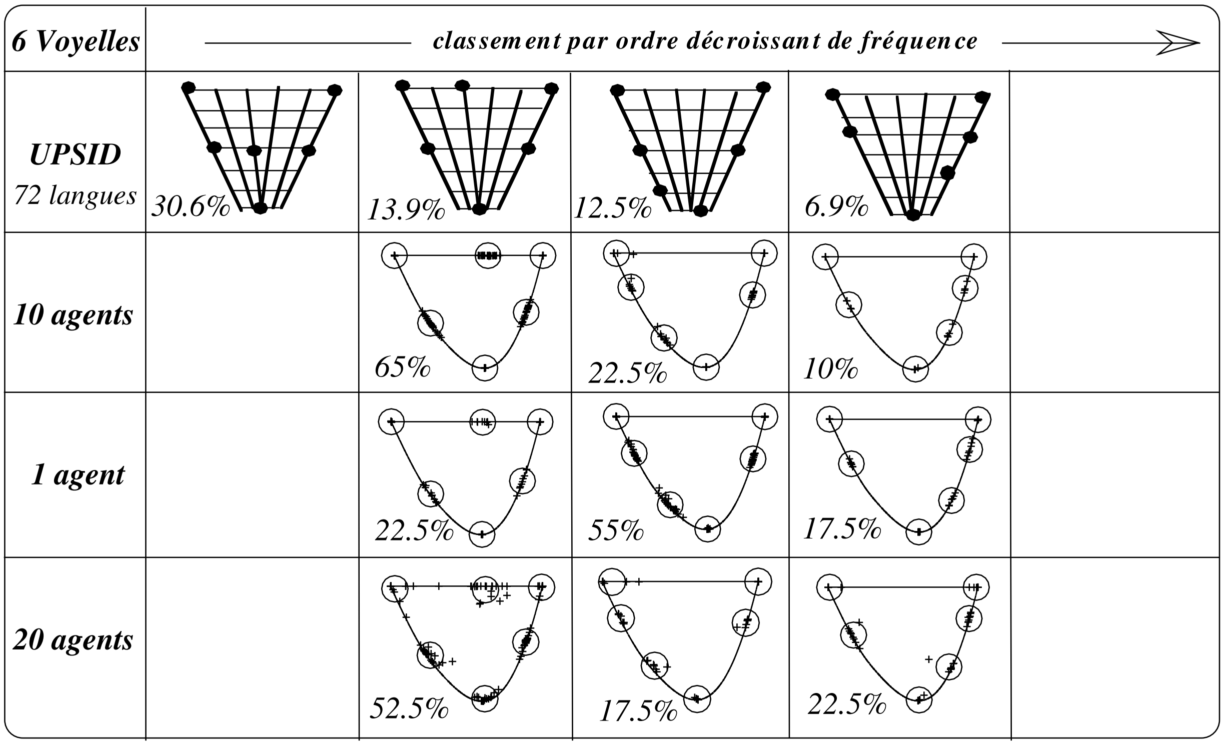


FIG. C.4: Distribution dans le triangle vocalique de voyelles de sociétés d'un, dix et vingt agents ayant un lexique de 6 voyelles.

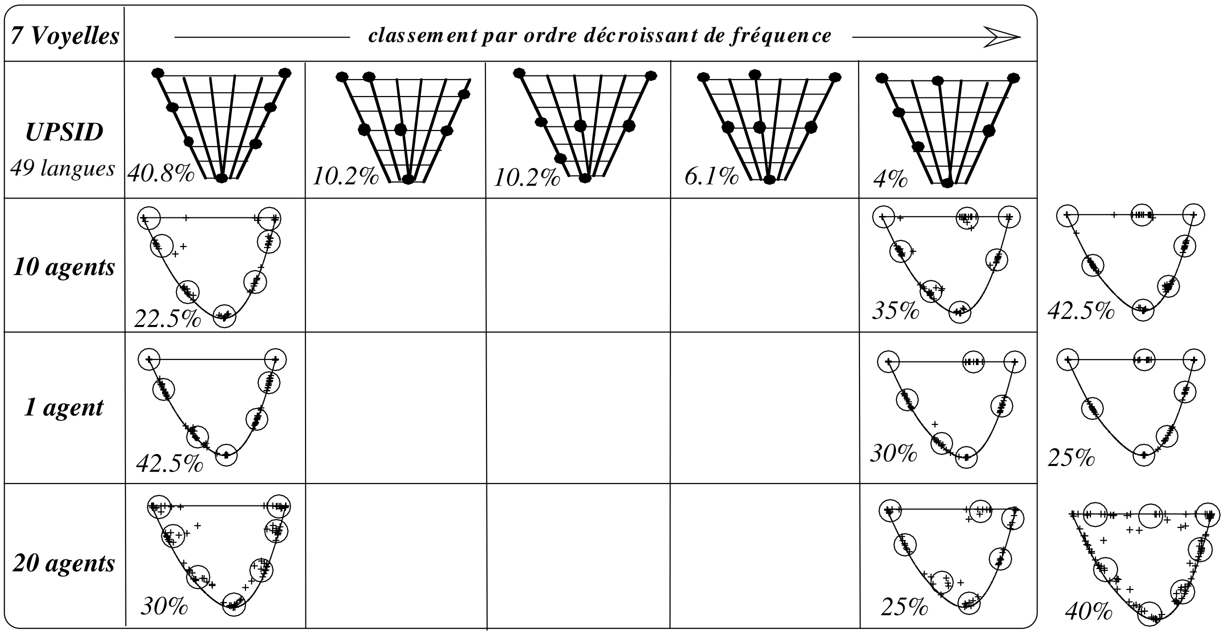


FIG. C.5: Distribution dans le triangle vocalique de voyelles de sociétés d'un, dix et vingt agents ayant un lexique de 7 voyelles.

Annexe D

Systèmes générés au bout d'un nombre important de transactions

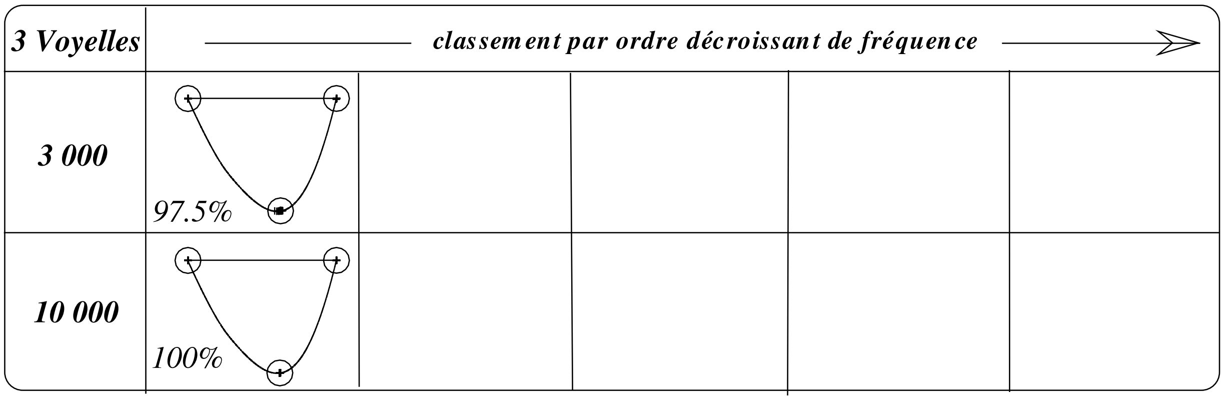


FIG. D.1: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 3 voyelles après 3 000 et 10 000 transactions.

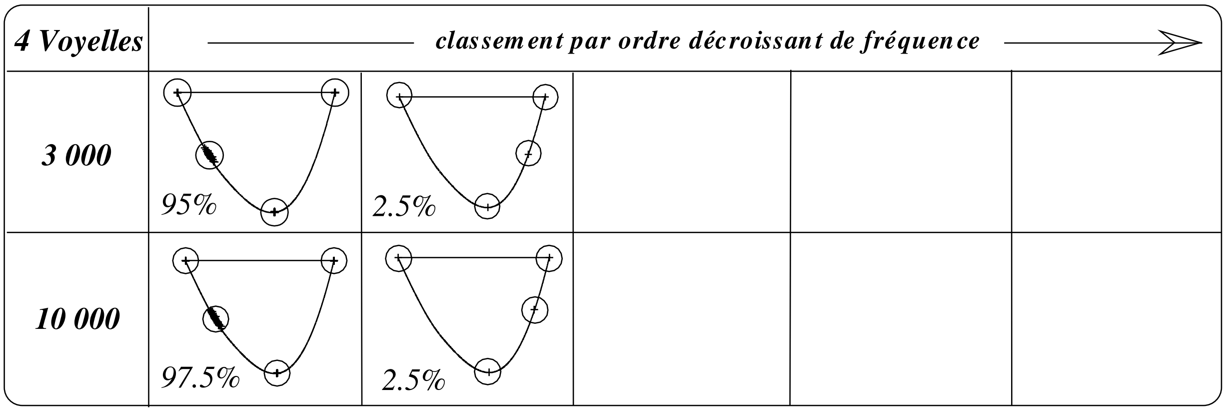


FIG. D.2: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 4 voyelles après 3 000 et 10 000 transactions.

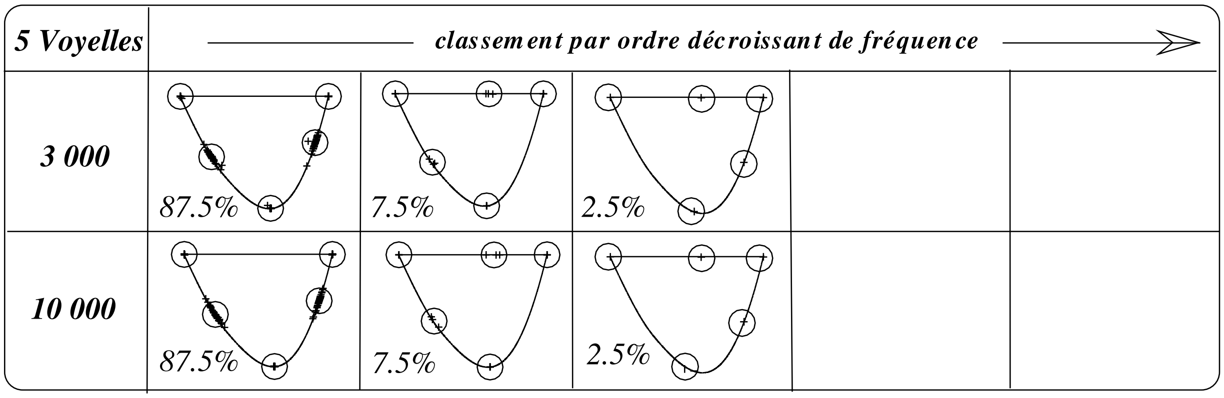


FIG. D.3: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 5 voyelles après 3 000 et 10 000 transactions.

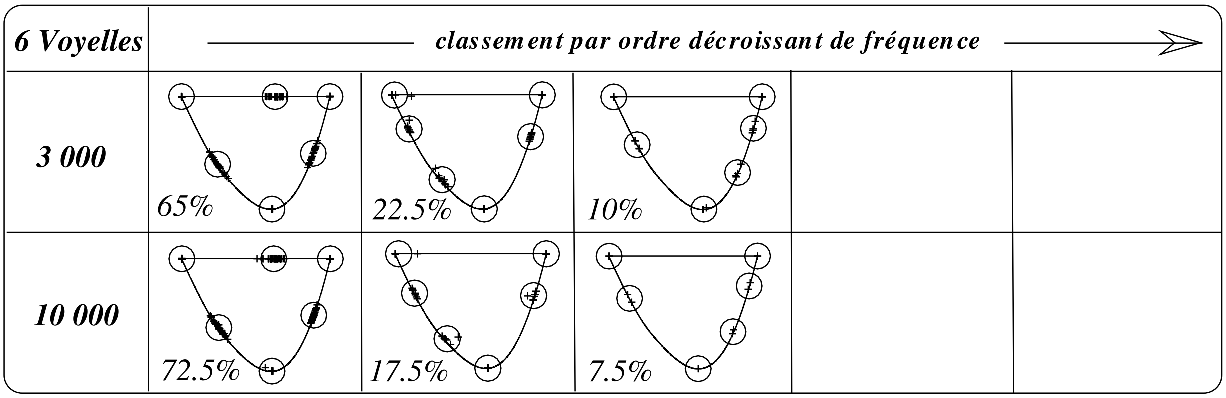


FIG. D.4: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 6 voyelles après 000 et 10 000 transactions.

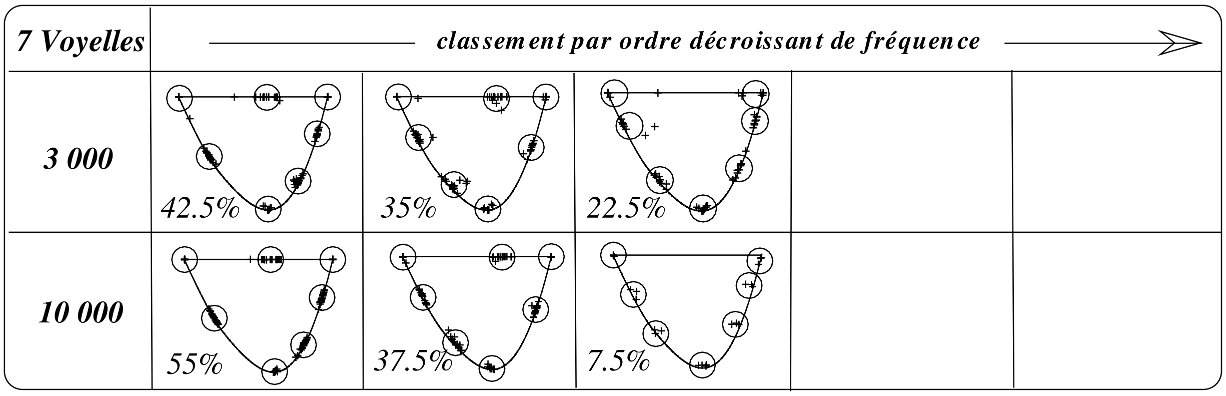


FIG. D.5: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 7 voyelles après 3 000 et 10 000 transactions.

Annexe E

Systèmes vocaliques générés avec différentes forces de répulsion

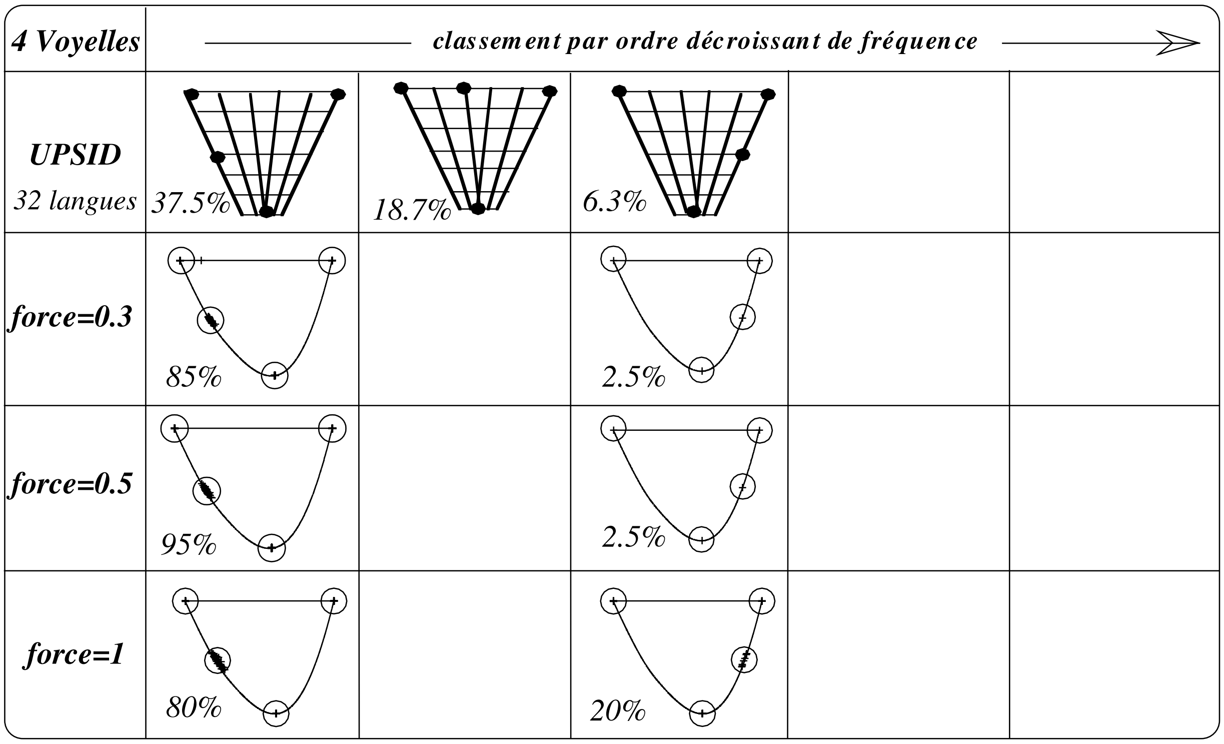


FIG. E.1: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 4 voyelles avec des forces de répulsion égales à 0.3, 0.5 et 1.

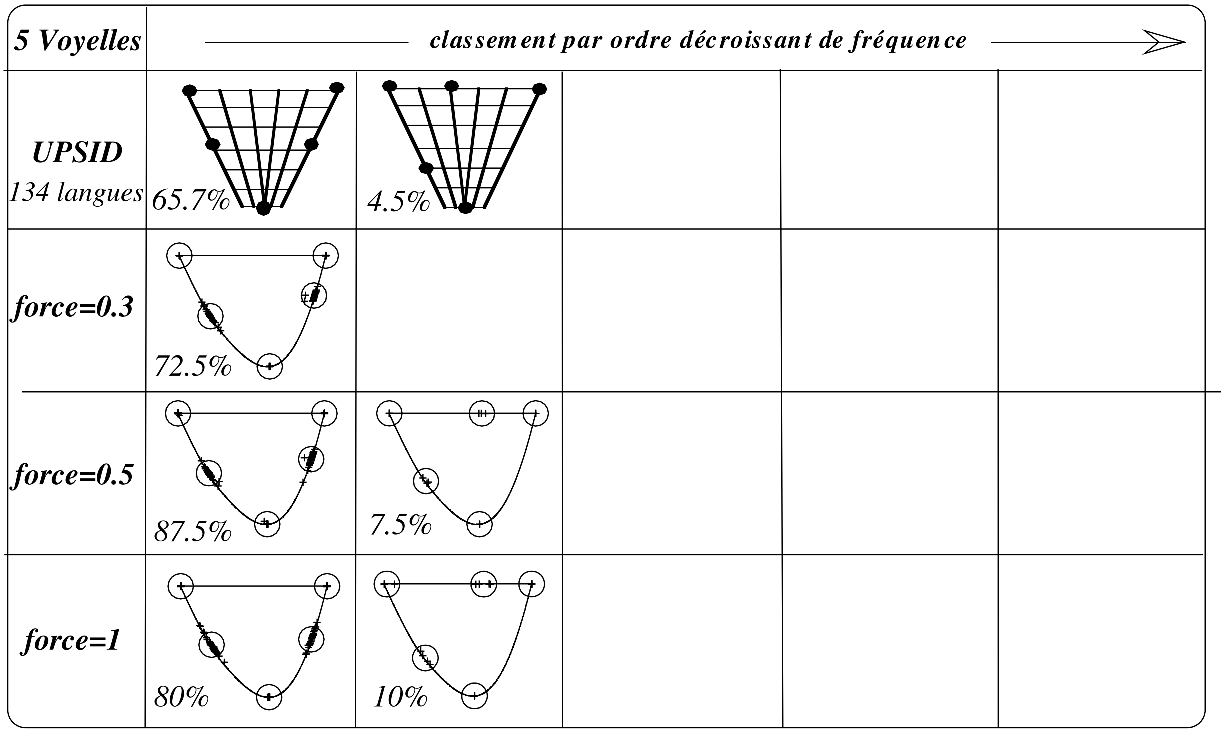


FIG. E.2: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 5 voyelles avec des forces de répulsion égales à 0.3, 0.5 et 1.

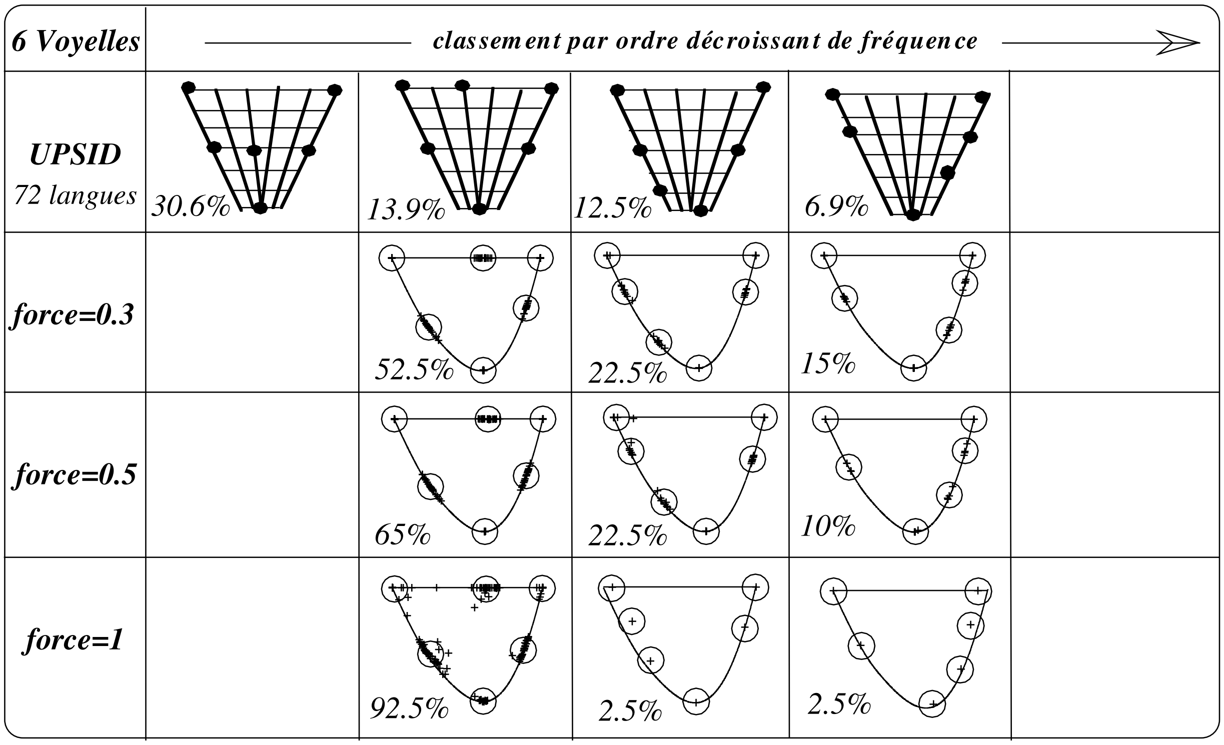


FIG. E.3: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 6 voyelles avec des forces de répulsion égales à 0.3, 0.5 et 1.

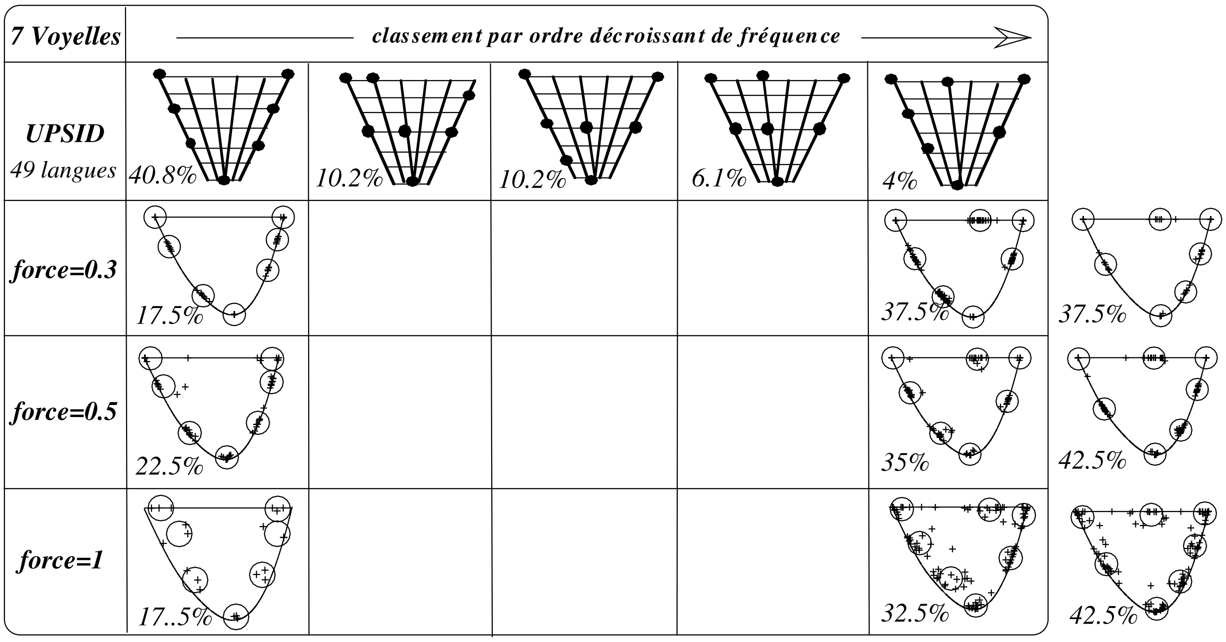


FIG. E.4: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 7 voyelles avec des forces de répulsion égales à 0.3, 0.5 et 1.

Annexe F

Systèmes vocaliques générés sans remplacement

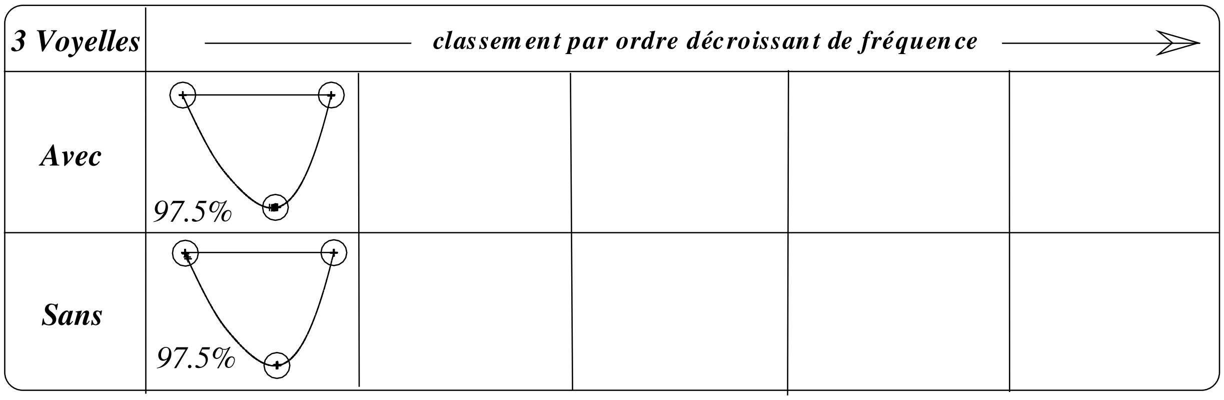


FIG. F.1: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 3 voyelles avec et sans remplacement.

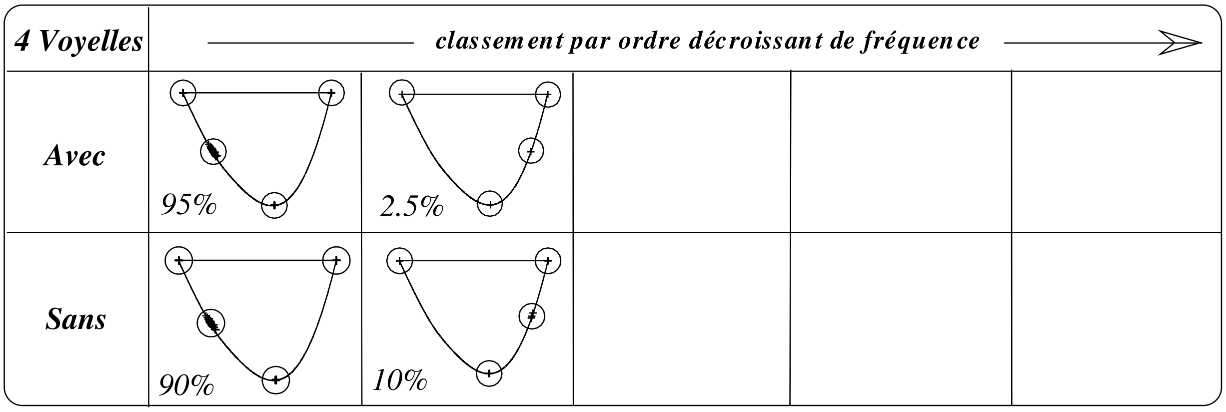


FIG. F.2: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 4 voyelles avec et sans remplacement.

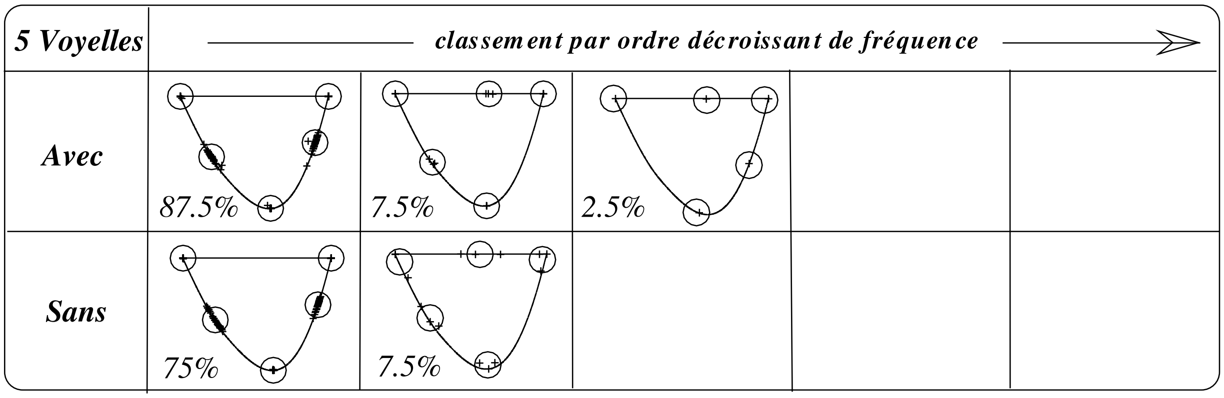


FIG. F.3: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 5 voyelles avec et sans remplacement.

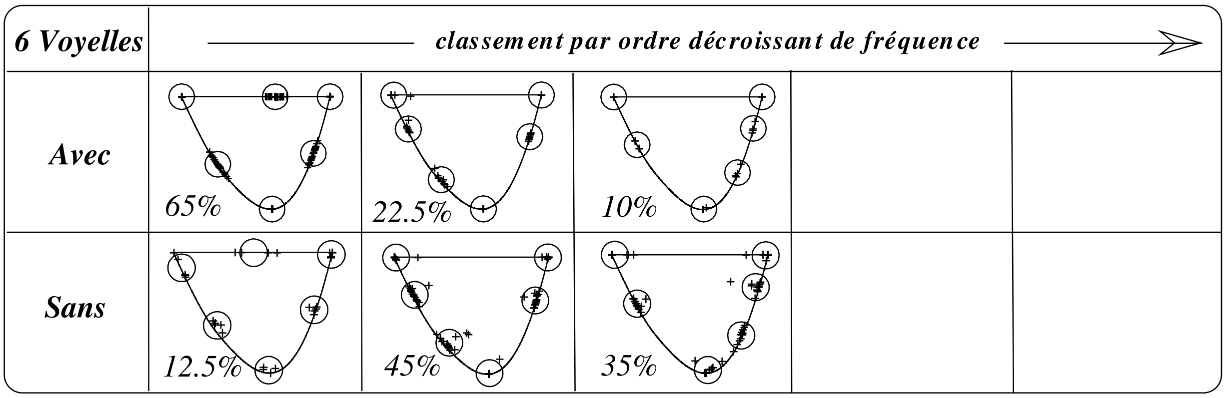


FIG. F.4: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 6 voyelles avec et sans remplacement.

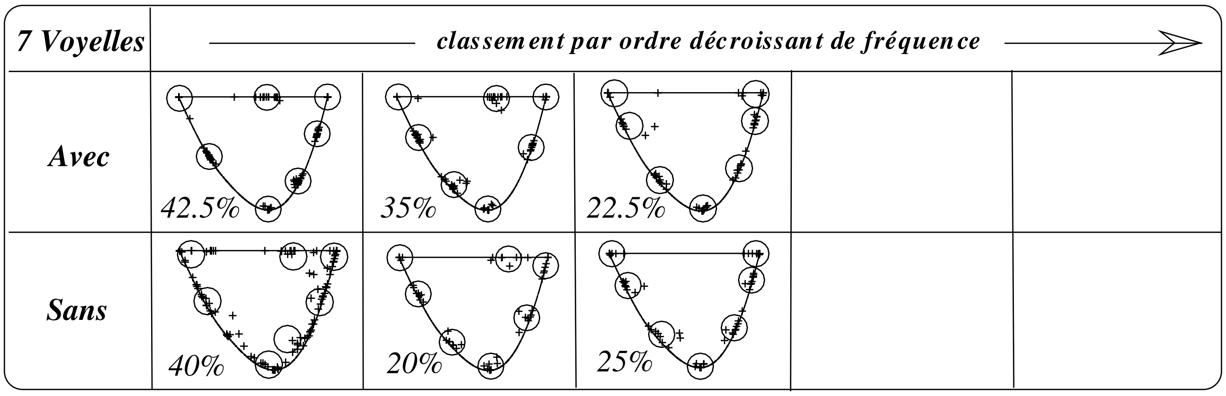


FIG. F.5: Distribution dans le triangle vocalique des voyelles d'une société de dix agents ayant un lexique de 7 voyelles avec et sans remplacement.

Annexe G

Brève présentation de la théorie Hypo & Hyper

La théorie *Hypo & Hyper* (Lindblom et al. 1995, Lindblom 1995) suggère que la variabilité des signaux de la parole est le résultat des *présuppositions phonétiques* du locuteur. Autrement dit, elle suppose que la probabilité d'une unité phonétique ne reste jamais constante, mais subit continuellement des fluctuations dynamiques. Comme illustration, considérons la prédictibilité du mot *oignons* dans les phrases suivantes :

- Le mot est '—'.
- Occupe-toi de tes '—'.

Il s'en suit que le signal n'est pas seul responsable de la formation de la sortie perceptuelle. Il ne doit pas contenir toute l'information, mais seulement assez pour produire ce résultat final par une interaction avec des connaissances mémorisées. Au moins, il faut que le locuteur s'assure que les patrons phonétiques soient suffisamment distinctifs pour une identification correcte du message.

Les points principaux peuvent être éclaircis au moyen d'un cas hypothétique idéal. Le locuteur fait une estimation continuelle de l'information qu'il faut mettre dans le signal, et à chaque instant, il adapte son articulation des phonèmes, des syllabes et des mots selon cette évaluation. Cette adaptation se manifeste le long d'une dimension avec des formes *'hyper'* articulées avec force à une extrémité, et avec des formes *'hypo'* produites moins énergiquement à l'autre bout. Comme le locuteur va d'une prononciation hypo à un style hyper, la durée et l'amplitude des gestes articulatoires ont tendance à devenir plus grandes. Le résultat est que la dépendance contextuelle des gestes articulatoires et acoustiques est minimale pour de la parole hyper mais maximale en hypo. On s'attend donc à ce que la coarticulation et le réduction caractérisent le style hypo, et que les voyelles et les consonnes hyper soient énoncées plus nettement et avec des propriétés phonétiques plus contrastives.

Selon cette hypothèse, la localisation exacte où se produit une certaine forme phonétique est déterminée par les suppositions que fait le locuteur en ce qui concerne le traitement perceptuel aussi bien que par son propre désir tacite de faire des simplifications articulatoires. C'est donc dans ce sens là qu'on peut dire que H&H est une théorie de présupposition des variations phonétiques.

Les changements phonétiques

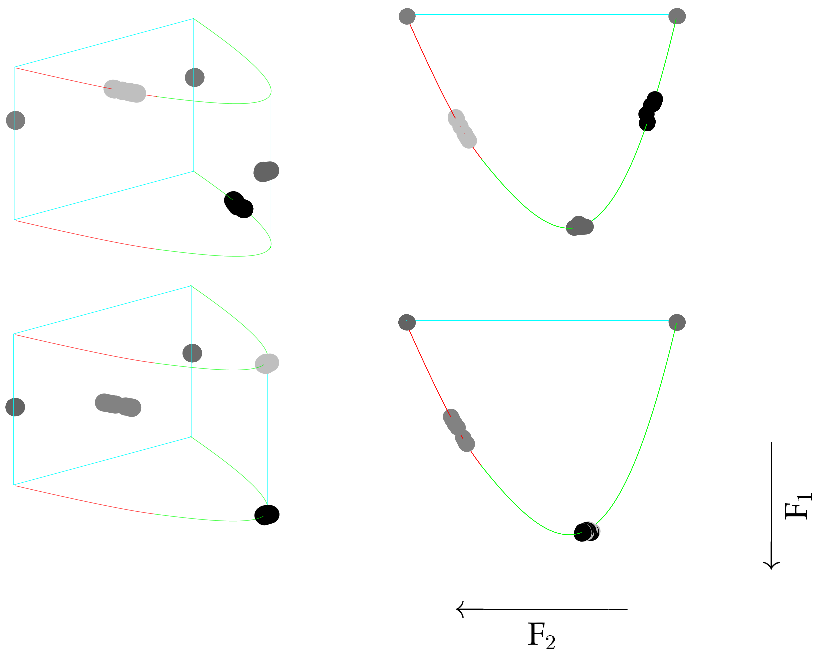
La perspective H&H suggère également des implications pour la compréhension des changements phonétiques historiques. Normalement, dans le traitement perceptuel, l'attention de l'auditeur se dirige vers le contenu du message plutôt que vers ses détails phonétiques. Pourtant, de temps en temps, la prononciation est au centre de l'attention. On peut suppposer, comme le fait Ohala dans ses travaux (Ohala 1993), que ce sont dans ces situations qu'une nouvelle prononciation se propose au locuteur.

L'auditeur joue un rôle primordial dans les théories de Ohala. Son observation de base est que les *misperceptions* sont à l'origine des changements phonétiques.

Annexe H

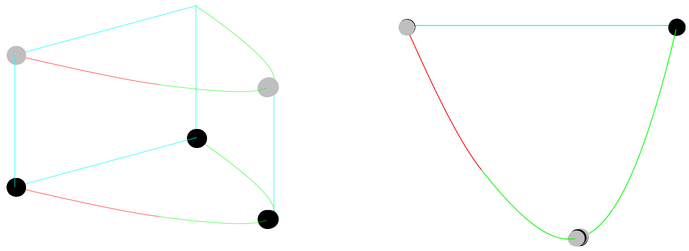
Exemples de systèmes vocaliques avec un trait supplémentaire

Ci-dessous quelques exemples de systèmes types générés lors de la simualtion du principe du MUAF. Sont représentés à droite les projections des systèmes dans l'espace perceptif tridimensionnel sur le plan F_1/F_2.

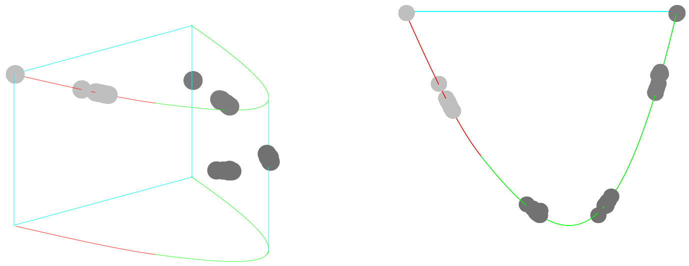


(5+0)

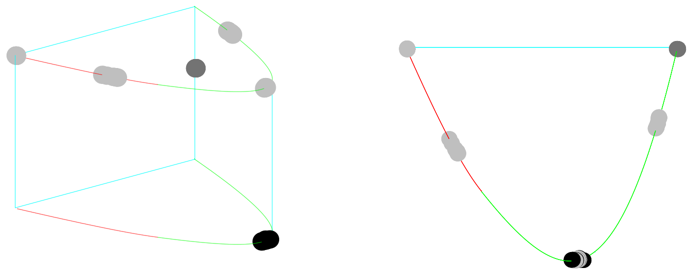
(4+1)



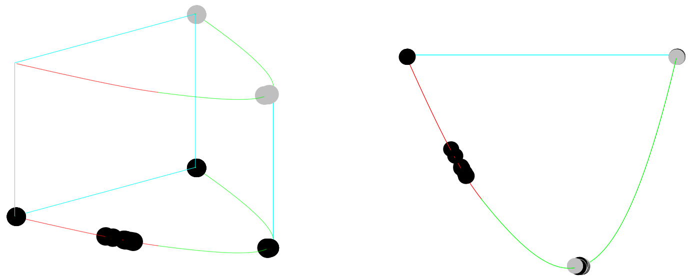
(3+2)



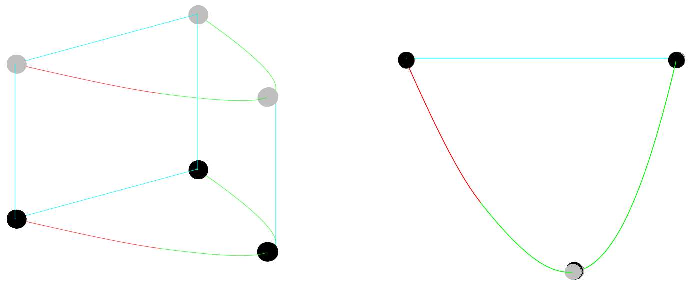
(6+0)



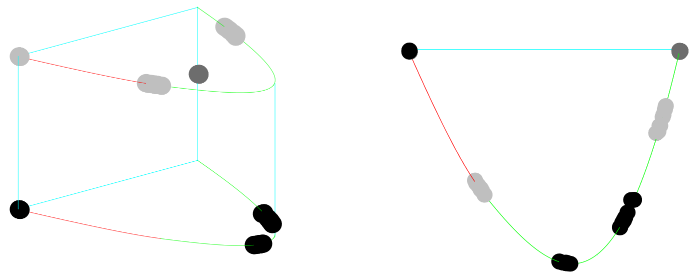
(5+1)



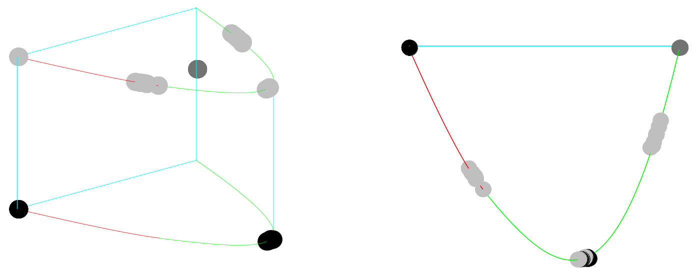
(4+2)



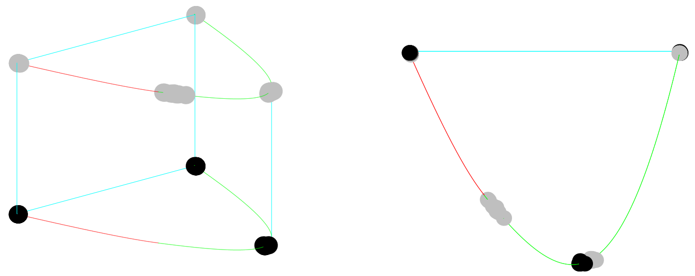
(3+3)



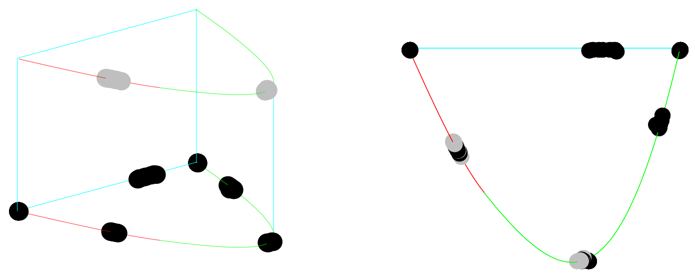
(6+1)



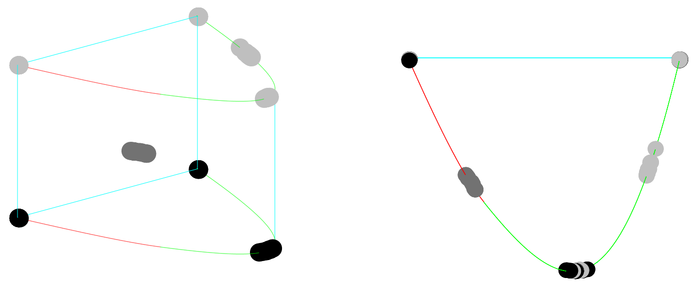
(5+2)



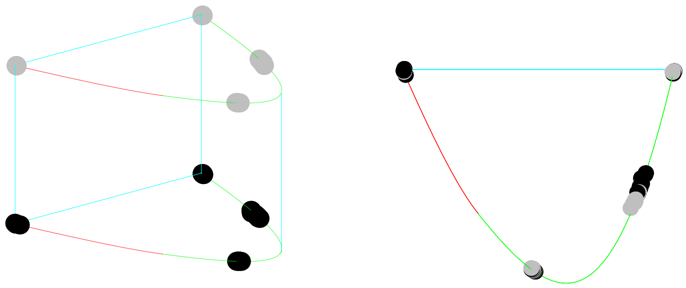
(4+3)



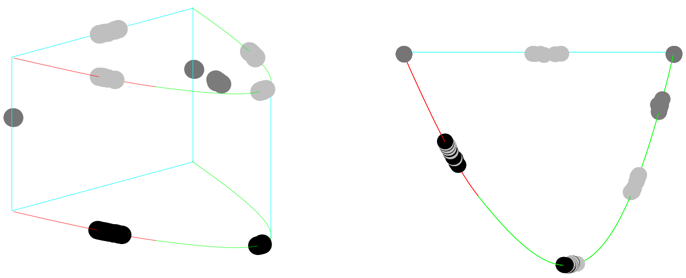
(6+2)



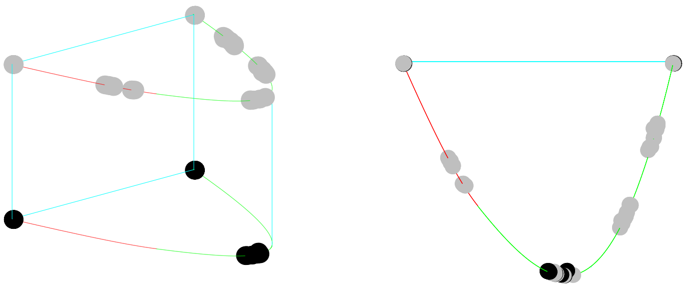
(5+3)



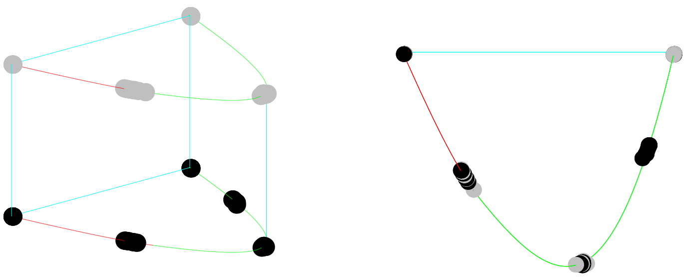
(4+4)



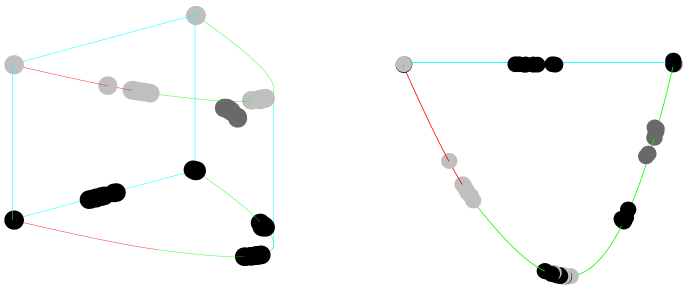
(7+2)



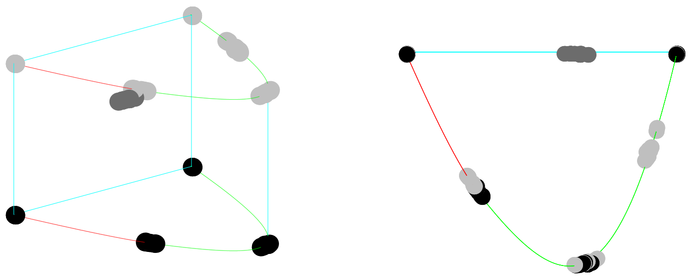
(6+3)



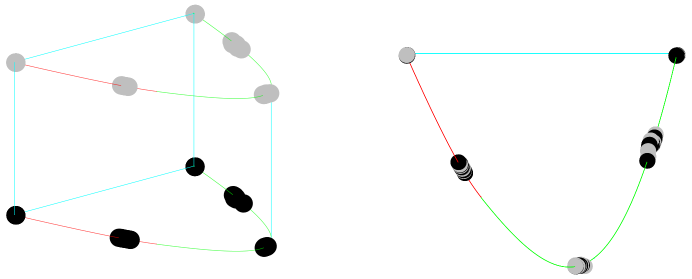
(5+4)



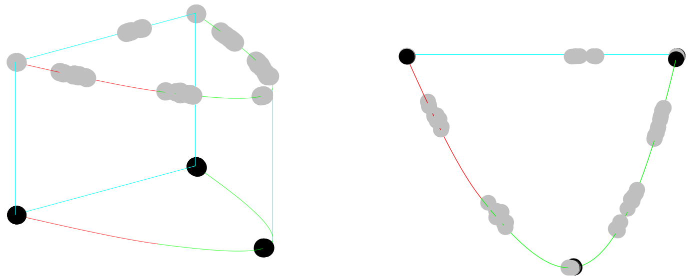
(7+3)



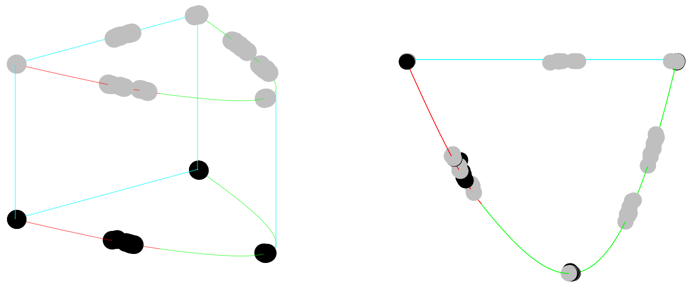
(6+4)



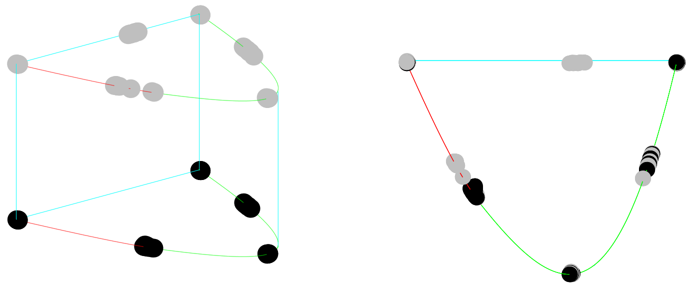
(5+5)



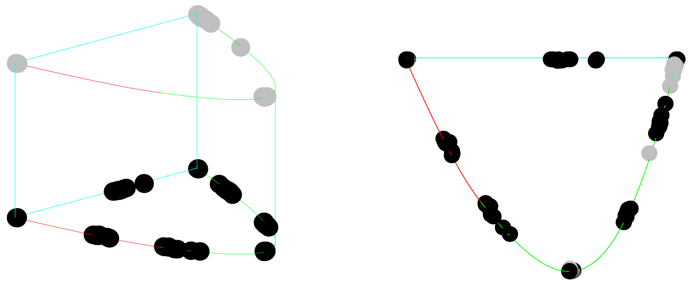
(8+3)



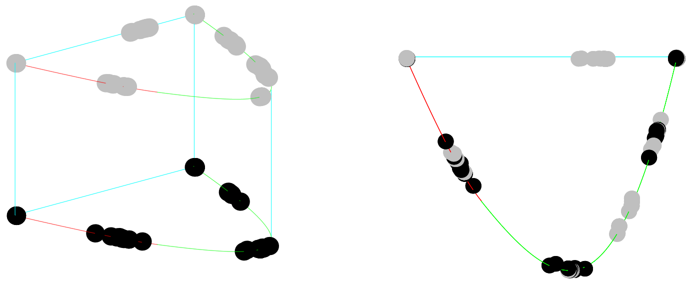
(7+4)



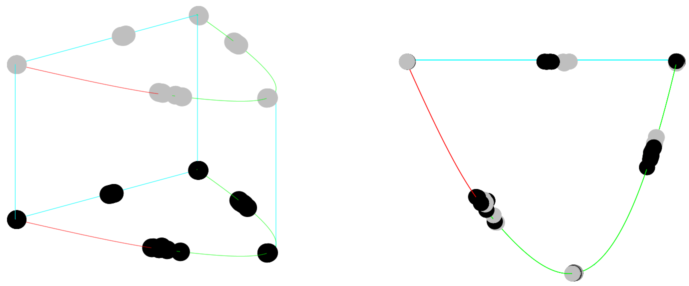
(6+5)



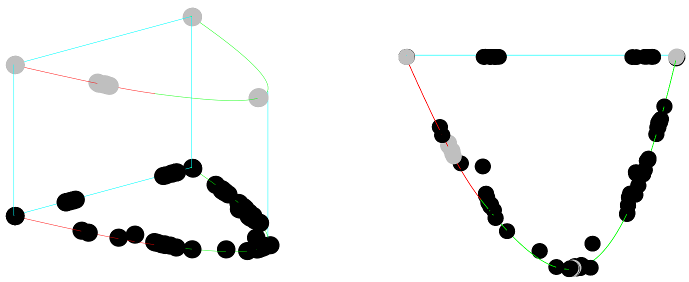
(8+4)



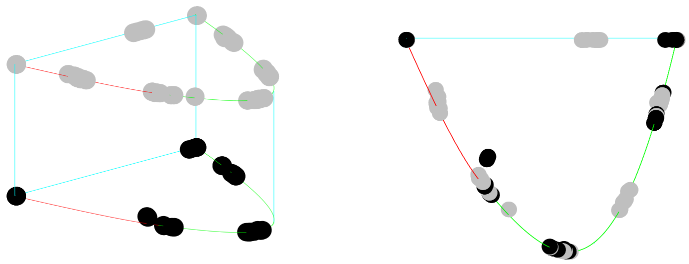
(7+5)



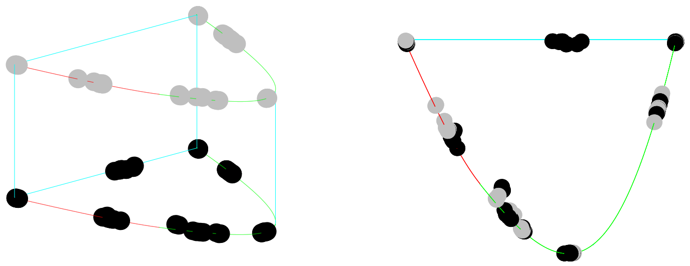
(6+6)



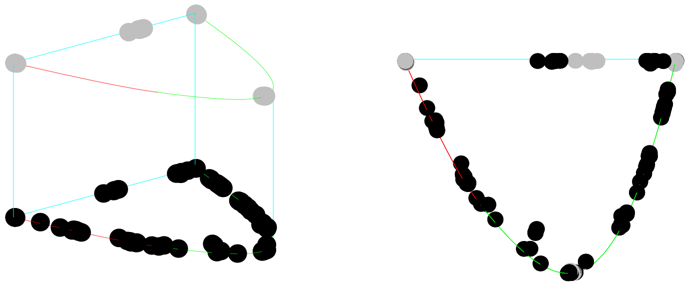
(9+4)



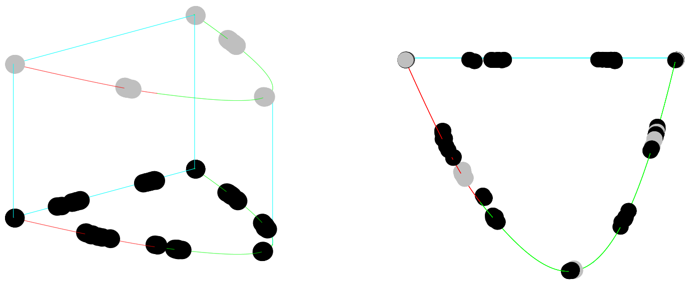
(8+5)



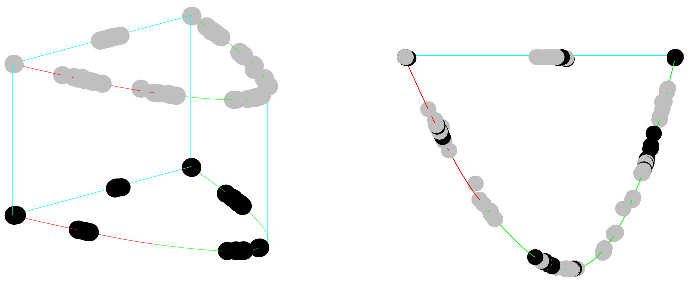
(7+6)



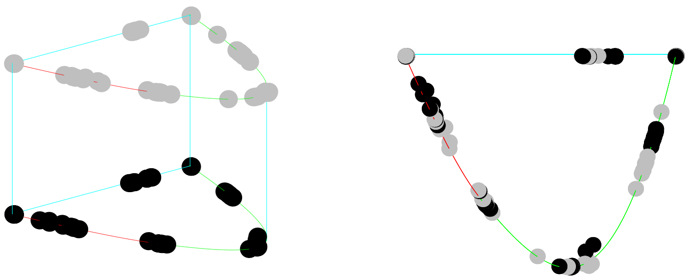
(10+4)



(9+5)



(8+6)



(7+7)