École des Hautes Études en Sciences Sociales

Année 1999

N° attribué par la bibliothèque

Thèse

pour l'obtention du grade de

Docteur de l'EHESS

Discipline : Sciences Cognitives

présentée et soutenue publiquement

par

Franck Ramus

le 25 novembre 1999

Rythme des langues et acquisition du langage

Directeur de thèse : Jacques Mehler

Jury : Nick Clements
Anne Cutler
Núria Sebastián-Gallés
Juan Segui

Cette thèse restera disponible aussi longtemps que possible sous forme électronique à l'adresse Internet suivante:
`http://www.ehess.fr/centres/lscp/persons/ramus/these/`, ou à défaut sur simple demande auprès de l'auteur: `franck.ramus@polytechnique.org`.

Ce n'est pas tous les jours qu'on a la chance de disposer d'une page entière d'un document officiel pour remercier les gens qui le méritent. J'en profite donc sans modération.

Mes plus chaleureux remerciements vont tout d'abord à Jacques Mehler, bien entendu pour avoir assuré la direction de cette thèse, mais aussi et surtout pour m'avoir fait entrer dans le monde de la recherche et des sciences cognitives dans les meilleures conditions qui soient.

Nick Clements, Anne Cutler, Núria Sebastián-Gallés et Juan Segui ont accepté de faire partie du jury ; qu'ils en soient remerciés.

Anne Christophe, Ghislaine Dehaene, Emmanuel Dupoux, Christophe Pallier, Sharon Peperkamp et tous les membres du LSCP, permanents ou de passage, m'ont aidé, motivé, critiqué, et m'ont fourni un environnement de travail stimulant et agréable.

Marina Nespor et Marc Hauser ont contribué à l'enrichissement de cette thèse au travers de fructueuses collaborations.

Cory Miller, Dylan Morris, Thierry Fourneyron et Sylvie Margules ont mis la main à la pâte en ce qui concerne le recueil de certaines de mes données.

Les discussions et échanges divers jouent souvent un rôle important dans la formation et l'évolution des idées. C'est le cas de ceux que j'ai eus avec Laura Bosch, Anne Cutler, Lila et Henry Gleitman, Peter Jusczyk, Jim Morgan, Thierry Nazzi, Ani Patel, Núria Sebastián-Gallés, et Roger Wales.

Jacques Mehler, Anne Christophe et Valérie Quiniou ont lu et annoté ma thèse dans son intégralité (un exploit qui sera sans doute rarement réédité !).

Brigitte Brunet, Michel Dutat, Susana Franck, Xavier Jeannin et Vireack Ul m'ont apporté leur précieuse aide technique et administrative.

Evelyn Hausslein m'a initié à l'expérimentation sur le nouveau-né ; les pédiatres, les sages-femmes, et le personnel soignant de la maternité Port-Royal m'ont apporté leur collaboration.

La DGA et le CNRS m'ont apporté leur soutien financier.

Enfin, je n'oublie pas que la psychologie cognitive n'est rien sans les sujets, et j'ai donc une pensée pour les 412 nouveau-nés et leurs 412 mamans assorties de quelques papas, les 268 étudiants et assimilés, et les 15 petits singes pas si bêtes qui ont bien voulu se prêter à mes drôles d'expériences.

Sommaire

1 Introduction : l'initialisation phonologique **1**

2 Le rythme de la parole **39**

3 La resynthèse de parole **65**

4 Discrimination de langues par des sujets adultes **81**

5 Discrimination de langues par des nouveau-nés **105**

6 Discrimination de langues par des singes **147**

7 Conclusion : le rôle du rythme dans l'acquisition du langage **165**

Annexes **181**

A Méthodes pour l'étude des nourrissons **181**

B Instructions et fichiers d'exemple **187**

C Expériences supplémentaires de discrimination de langues par des nouveau-nés **197**

Bibliographie **205**

Table des matières **221**

Table des figures **227**

Liste des tableaux **229**

Chapitre 1

Introduction : l'initialisation phonologique

A LA NAISSANCE, les êtres humains sont incapables de parler et n'ont pas l'air de comprendre ce qu'on leur dit. Pourtant, au bout de quelques années, ils maîtrisent (au moins) une langue, sa grammaire, sa prononciation, et ses quelques dizaines de milliers de mots. Comment y parviennent-ils ? Si l'on en croit les idées reçues, l'apprentissage d'une langue se résume à apparier des sons et des objets, et éventuellement à « se faire l'oreille » aux sonorités de la langue. Nous allons voir au contraire que le langage est complexe, au point que son apprentissage puisse paraître véritablement mystérieux.

Pour apprécier la difficulté de la question, mettons-nous temporairement dans la peau du bébé qui cherche à apprendre le français. Pour un bébé, le français, c'est du chinois... Lorsque nous écoutons du chinois, ou toute autre langue inconnue, la première chose qui nous frappe est que nous entendons un son continu qui ne s'interrompt que lorsque le locuteur hésite ou reprend sa respiration ; nous sommes parfaitement incapables de savoir quels mots celui-ci a prononcés. Cette observation met en évidence le fait que, contrairement à l'écriture, la parole ne comporte pas de « blancs » qui indiqueraient le début et la fin des mots. De même, lorsque le bébé entend « Regarde le chien », il entend en fait [ʀəgaʀdləʃjɛ̃][1]. Comment savoir quelles séquences de sons constituent des mots? C'est ce que l'on appelle le problème de la segmentation de la parole en mots. Même en supposant ce problème résolu, comment savoir que la séquence [ʃjɛ̃] est un nom, qui fait référence à un animal, plutôt qu'un verbe, qui fait référence à une action? De toute évidence, il n'est pas si facile que cela d'apparier des sons et des objets, ne serait-ce que parce qu'il faudrait déjà savoir *quels* sons apparier à *quels* objets.

1. En alphabet phonétique international.

L'apprentissage du langage est en fait truffé de ce genre de problèmes. C'est ce qui a motivé l'émergence d'une nouvelle approche, appelée *initialisation phonologique*. Le présent chapitre a pour but d'exposer cette approche, dans la mesure où elle sert de cadre théorique à toute la thèse. Nous allons commencer par expliquer plus en détail certains problèmes classiques rencontrés par les théories de l'acquisition du langage, puis nous montrerons comment l'initialisation phonologique peut contribuer à les résoudre.

1.1 Quelques problèmes classiques de l'acquisition du langage

Pour mieux apprécier ces problèmes, il est important de considérer la manière dont le langage est structuré. Pour correctement maîtriser sa langue maternelle, l'enfant devra en fait en apprendre :

la phonétique : l'ensemble des sons élémentaires de sa langue (les phonèmes) ;

la phonologie : les principes selon lesquels ces sons élémentaires s'assemblent en syllabes, et également les principes qui définissent l'accentuation, les tons, le rythme et l'intonation ;

la morphologie : les règles qui définissent les inflexions sur les mots, telles les marques du genre, du nombre, du temps des verbes, etc...

la syntaxe : les règles qui relient les mots entre eux, et qui permettent de déduire le sens d'une phrase du sens de ses mots ;

le lexique : les mots et leur signification (la sémantique).

Bien entendu, ces différents niveaux ne sont pas entièrement indépendants les uns des autres. Ainsi, la phonologie s'appuie sur la phonétique et en partie sur le lexique. Par ailleurs la morphologie, la syntaxe et le lexique sont étroitement liés. Pour comprendre l'acquisition du langage, il est donc important de déterminer comment et dans quel ordre ces différents types de connaissance sont appris par l'enfant.

1.1.1 Apprendre le lexique

Nous avons expliqué ci-dessus que le premier problème de l'acquisition du lexique consistait à segmenter la parole en mots. On pourrait objecter que les mères ne font pas des discours philosophiques à leurs enfants, mais au contraire leur disent des phrases très courtes et le plus souvent réduites à un seul mot. C'est à la fois vrai et faux. S'il est vrai que les phrases adressées par les mères à leurs jeunes enfants sont plus courtes que la moyenne,

elles ne se réduisent quasiment jamais à un seul mot (Ratner, 1996; van de Weijer, 1998). Encore pire, même lorsqu'un expérimentateur demande à des mères d'apprendre un mot particulier à leur enfant, elles ne le prononcent pas pour autant en isolation (Aslin, Woodward, LaMendola, & Bever, 1996). Le problème de l'extraction des mots du flux de parole est donc bien réel.

Supposons maintenant qu'il soit déjà résolu. Cela n'est pas suffisant pour apprendre le lexique. En effet, un lexique ne se résume pas à un catalogue de formes sonores, encore faut-il attacher un sens aux sons. Cette tâche peut sembler en apparence très simple : par exemple, la mère de l'enfant pointe le chien du doigt et dit « Regarde le chien ! » (cf. par exemple Locke, 1690). Soulignons quelques problèmes qui se posent à l'enfant avant de pouvoir apprendre le mot « chien » :

- Même en supposant que l'enfant a réussi à segmenter la phrase en les trois mots qui la constituent, encore faut-il qu'il associe le sens de « chien » au son correspondant, plutôt qu'au son de « regarde » ou de « le ». Ce qui suppose qu'il sache à l'avance que ces sons ne peuvent pas désigner un chien, ou bien qu'il soit déjà capable d'une analyse syntaxique primaire de la phrase : « regarde » est un verbe, « le » est un déterminant, « chien » est le seul nom de la phrase, donc c'est le seul mot qui puisse désigner un objet ou un être.
- Déduire le sens même du mot « chien » en observant la situation est également problématique, comme l'a fait remarquer Quine (1960). Car autant que l'enfant sache, sa mère pourrait très bien être en train de désigner le mur derrière le chien, le canapé sous le chien, la queue du chien, le mouvement de la queue du chien, le caniche, ce chien particulier, ou au contraire tous les animaux à poils noirs bouclés.
- Pire encore, tous les noms n'ont pas la qualité du mot « chien » de référer à un objet visible dans le monde. Ce problème est exacerbé dans le cas des enfants aveugles, pour lesquels *aucun* mot ne réfère à un objet visible dans le monde, et qui malgré tout acquièrent un lexique aussi important et avec la même rapidité que les enfants voyants (Landau & Gleitman, 1985). Sans aller chercher ce cas extrême, il est tout de même courant de parler de choses qui ne sont pas dans la pièce ou dans le champ visuel, même à un enfant.

L'apprentissage des verbes présente les mêmes problèmes en pire. En effet, un grand nombre de verbes fréquents ne correspondent à aucune action visible (notamment les verbes référant à une activité mentale : penser, croire, vouloir...) ou effectuée au moment de l'énonciation (les verbes à l'impératif réfèrent typiquement à des actions souhaitées plutôt qu'observées). Enfin, la nature même des verbes et des actions qu'ils décrivent rendent l'interpréta-

tion des situations particulièrement ambiguës. Pour reprendre un exemple de Gleitman (1990), un enfant peut entendre « le chien poursuit le chat » en observant la scène correspondante. Cependant, la même situation pourrait être décrite par « le chat fuit le chien ». Comment l'enfant peut-il savoir que le son [puʀsɥi] ne signifie pas « fuit »? Il lui faut connaître le sens de « chien » et de « chat » bien sûr, mais aussi savoir qu'en français l'ordre des mots est Sujet-Verbe-Objet. En effet, si l'ordre des mots était Objet-Verbe-Sujet[2], il devrait en déduire que le son [puʀsɥi] signifie « fuit ».

Ces quelques exemples illustrent le fait que l'apprentissage du sens des mots, et tout particulièrement des verbes, est un problème épineux du fait de la multiplicité des interprétations possibles des situations. Seule la maîtrise préalable de certaines règles de syntaxe de la langue semble pouvoir restreindre les choix possibles et guider l'enfant vers l'interprétation correcte. C'est ce que Gleitman (1990) appelle l'*initialisation syntaxique* du lexique (syntactic bootstrapping).

1.1.2 Analyser la syntaxe des phrases

Les théories de la syntaxe décrivent généralement la structure des phrases en termes de constituants (groupe nominal, groupe verbal...), de catégories syntaxiques (nom, verbe, adjectif, préposition...), de fonctions grammaticales (sujet, objet, complément...), de cas (nominatif, accusatif, datif...), etc.. Apprendre la syntaxe, c'est donc entre autres apprendre quels mots appartiennent à quelles catégories syntaxiques, et comment ils s'organisent pour former les constituants et remplir leur fonction grammaticale.

De même que pour l'apprentissage du sens des mots, les théories de l'acquisition de la syntaxe supposent que la parole est déjà segmentée en mots. En outre, elles supposent que l'enfant a à sa disposition le sens des mots et le sens global des phrases (Wexler & Culicover, 1980; Pinker, 1984; Radford, 1990). À partir de ces éléments, l'enfant peut alors essayer de déduire les règles grammaticales qui gouvernent les phrases qu'il entend.

Pour reprendre un exemple de Pinker (1984), considérons la phrase « le garçon jette une pierre », et supposons que l'enfant qui entend cette phrase l'a déjà segmentée en mots. Pour correctement analyser la phrase, il va lui falloir assigner à chaque mot sa catégorie grammaticale (nom, verbe), reconstituer les groupes verbaux et nominaux, et leur attribuer des fonctions (sujet, objet...). En l'absence de toute information sémantique, il pourrait parfaitement postuler que « le » est un nom, que « garçon » est un verbe,

2. Comme c'est le cas en Hixkaryana, une langue parlée au nord du Brésil (Comrie, 1989).

que « le garçon » est un groupe verbal, etc., et par conséquent en déduire que l'on forme les groupes verbaux en juxtaposant un nom puis un verbe. Il va de soi qu'en dérivant des règles de manière aussi peu contrainte, l'enfant ne risquerait pas de se mettre un jour à parler français. Pinker (1984) propose que ce sont les propriétés sémantiques des mots qui contraignent l'analyse que l'enfant peut faire de la phrase : c'est l'*initialisation sémantique* de la syntaxe (semantic bootstrapping). Selon lui, il faut donc supposer que l'enfant possède déjà les mots « garçon », « jette » et « pierre » dans son lexique, et qu'il détient le sens de la phrase, par exemple en observant la scène. Alors, le fait que les mots « garçon » et « pierre » réfèrent à des objets ou des personnes lui indique qu'il s'agit de noms. De même, le fait que « jette » réfère à une action lui indique qu'il s'agit d'un verbe (les correspondances nom/objet et verbe/action sont supposées innées). Enfin, l'observation que le garçon est le « jeteur » et la pierre le « jeté » lui indique que « le garçon » est sujet de la phrase et « une pierre » est complément d'objet[3]. Une telle analyse peut alors permettre à l'enfant d'apprendre que Sujet-Verbe-Objet est l'ordre correct des mots en français.

Ainsi, il semble que l'acquisition de la syntaxe ne soit pas entièrement autonome elle non plus. Au contraire, elle repose de manière cruciale sur le lexique et les propriétés sémantiques des mots.

1.1.3 Lexique et syntaxe : la poule et l'œuf ?

Nous sommes maintenant confrontés à une situation en apparence paradoxale. D'une part, l'acquisition du lexique comme de la syntaxe s'appuient sur une pré-segmentation de la parole en mots, or cette segmentation reste très problématique. D'autre part, l'acquisition du lexique et celle de la syntaxe semblent s'appuyer l'une sur l'autre, ce qui pose un problème de circularité.

Remarquons toutefois que la circularité n'est pas avérée a priori, mais peut n'être qu'une illusion du niveau de description grossier que nous employons (lexique et syntaxe). En y regardant de plus près, il se pourrait bien qu'une partie seulement (L1) de l'acquisition du lexique nécessite certaines connaissances syntaxiques (S1), et qu'une partie seulement (S2) de l'acquisition syntaxique nécessite certaines connaissances lexicales (L2). Dans le cas où d'une part L1 et L2 et d'autre part S1 et S2 sont disjoints, il n'y a là rien de paradoxal. Il n'y a circularité que si par exemple l'information syntaxique S1 nécessaire à l'apprentissage de l'information lexicale L1, nécessite elle-même

3. Nous n'introduisons pas ici la question des déterminants qui est nettement plus compliquée. Remarquons simplement que l'on voit beaucoup moins bien comment déduire leur rôle grammatical de leurs propriétés sémantiques.

L1 pour son apprentissage. Il est donc important d'expliciter quelles connaissances syntaxiques sont nécessaires à l'acquisition de quelles connaissances lexicales et vice-versa.

Malgré cette réserve, les théories de l'acquisition du lexique et de la syntaxe, au moins dans leur état actuel, semblent effectivement tomber sous le coup de la circularité. C'est du moins ce que suggèrent les exemples avec lesquels nous avons illustré les sections précédentes. En effet, nous avons vu notamment que 1) l'acquisition du sens des verbes nécessite de connaître l'ordre des mots, 2) l'acquisition de l'ordre des mots nécessite de connaître leur sens, en particulier celui des verbes. De même, 1) l'acquisition du sens des mots nécessite de connaître leurs catégories grammaticales, mais 2) l'apprentissage des catégories grammaticales n'est pas envisagée sans le recours au sens des mots. Étant donné que l'enfant doit apprendre *à la fois* le sens des mots, leurs catégories grammaticales et l'ordre dans lequel ils interviennent dans la phrase, le problème de savoir comment ce processus peut démarrer reste entier.

1.2 L'initialisation phonologique

1.2.1 Présentation de l'hypothèse

Les difficultés que rencontrent les théories « classiques » de l'acquisition du langage peuvent se résumer ainsi : l'enfant doit acquérir un grand nombre de connaissances abstraites sur sa langue maternelle ; pourtant, il n'a pour toute source d'information que l'onde acoustique de la parole. On ne peut donc pas se contenter de décrire l'acquisition du langage en restant au niveau des seules connaissances abstraites. Il est primordial de comprendre comment celles-ci peuvent être extraites du signal de parole. Cette démarche est non seulement nécessaire à une compréhension complète de l'acquisition du langage, mais elle peut également permettre de résoudre les paradoxes que nous avons mentionnés plus haut, en montrant comment les régularités présentes dans l'onde acoustique peuvent signaler des propriétés abstraites de la langue dont l'apprentissage semble problématique[4]. Pour reprendre un exemple concret, l'enfant doit apprendre à la fois l'ordre des mots et leur

4. Nous ne rentrerons pas ici dans les détails des langues des signes. Notons simplement que les propriétés abstraites à apprendre y sont probablement du même ordre, et doivent également pouvoir être apprises à partir du signal visuel. Si l'on souhaite alors avoir une théorie qui rende compte de l'acquisition des langues des signes aussi bien que des langues parlées, il sera sans doute nécessaire à terme de reformuler les régularités du signal en des termes plus généraux, indépendants de la modalité.

sens, or l'apprentissage de l'un semble reposer sur celui de l'autre, et vice-versa. Cette circularité peut être brisée si l'on montre que, dans le signal de parole, qui est la seule information directement accessible à l'enfant, il existe des indices qui permettent de déduire[5] par exemple des informations sur l'ordre des mots. Cette correspondance directe entre le signal de parole et de l'information grammaticale peut alors permettre d'*initialiser* l'acquisition à la fois du sens et de l'ordre des mots.

De fait, les mêmes auteurs qui ont proposé des théories de l'acquisition du lexique et de la syntaxe se sont également souciés de trouver des indices utiles dans le signal de parole. Gleitman et Wanner (1982) ont ainsi avancé l'hypothèse selon laquelle la prosodie de la parole, c'est-à-dire ses caractéristiques rythmiques et mélodiques, fournit un certain nombre d'indices concernant les frontières de mots et de constituants grammaticaux. Cette hypothèse a pris le nom d'*initialisation prosodique* (prosodic bootstrapping) (Pinker, 1984). Elle a été reprise et développée par Morgan (1986), qui a apporté une démonstration formelle que le problème de l'acquisition du langage voit sa complexité significativement réduite lorsque l'on fait l'hypothèse que la parole est perçue par l'enfant comme une série de constituants bien délimités par la prosodie (bracketed input hypothesis). Pinker (1984), qui se souciait notamment de savoir comment l'enfant pourrait découvrir le rôle grammatical des déterminants, étant donné leur faible contenu sémantique, a fait par ailleurs l'hypothèse que ceux-ci étaient signalés par des propriétés phonologiques particulières.

Ces différentes idées ont été regroupées par Morgan et Demuth (1996b) dans une formulation plus générale, l'*initialisation phonologique* :

> « La parole pourrait contenir des indices concernant certaines distinctions syntaxiques fondamentales, imposant des contraintes supplémentaires sur les analyses syntaxiques et sémantiques de l'enfant, signalant les domaines au sein desquels ces analyses peuvent être effectuées efficacement, et assurant que celles-ci vont démarrer dans la bonne direction. [...] Plusieurs types d'information sont présents dans la parole – phonétique, phonotactique,

5. AVERTISSEMENT : nous utilisons, ici et dans la suite, des termes « adultomorphiques » comme « déduire », « savoir », « supposer » ou encore « stratégie », à propos d'enfants de moins d'un an. En aucun cas nous ne suggérons que les nourrissons font des déductions et emploient des stratégies de manière consciente et volontaire. Il s'agit simplement d'alléger le texte et de suivre les usages de la psychologie du développement. Par « les enfants » nous entendons la plupart du temps « le cerveau des enfants ». On pourra traduire « les enfants emploient la stratégie X pour en déduire la propriété Y » par « le cerveau des enfants implémente la stratégie X pour apprendre/se sensibiliser à/s'ajuster à la propriété Y », étant entendu que ce que le cerveau implémente, il le doit soit à l'évolution de l'espèce humaine, soit aux apprentissages antérieurs de l'organisme.

> prosodique, stochastique – et chacun d'eux peut contribuer à des représentations syntaxiques élaborées des énoncés entendus. [...] Nous proposons le terme *initialisation phonologique* pour regrouper les hypothèses selon lesquelles les enfants peuvent dériver de l'information grammaticale rudimentaire de leurs analyses perceptuelles de la parole. » (Morgan & Demuth, 1996b, p. 2, traduction libre)

Remarquons que si la description est adéquate, le terme même d'initialisation phonologique est malheureux, car la phonologie consiste en des propriétés de la langue qui peuvent être relativement abstraites et non trivialement déductibles du signal de parole. Un terme plus approprié pourrait donc être « initialisation par le signal », ou bien « initialisation par les régularités de surface ». Ce point de terminologie étant fait, nous continuerons par la suite à utiliser le terme « initialisation phonologique » pour ne pas alourdir le texte.

Avant de passer en revue un certain nombre d'études qui viennent à l'appui de l'initialisation phonologique, nous allons maintenant effectuer une parenthèse épistémologique dans le but de clarifier le contexte théorique de l'hypothèse.

1.2.2 Le rôle de l'environnement dans l'acquisition du langage

L'idée selon laquelle le signal d'entrée est riche et structuré n'est pas nouvelle. Bien au contraire, c'était l'idée maîtresse du behaviorisme (cf. Skinner, 1957), qui l'a poussée à l'extrême, en soutenant que le comportement d'un organisme était entièrement explicable en termes des caractéristiques du stimulus d'entrée. Cette théorie a été sévèrement critiquée par Chomsky (1959), et a été abandonné au profit du paradigme cognitiviste, qui a au contraire accordé une importance centrale aux représentations mentales des organismes. De plus, Chomsky (1957) ayant argumenté de manière convaincante que certains aspects universels de la syntaxe étaient impossible à apprendre étant donnée la « pauvreté du stimulus », il est apparu inévitable de postuler l'existence de capacités et de représentations innées chez l'enfant. Chomsky a ainsi remis à la mode les explications dites « nativistes ». Le problème avec ce genre d'explications est leur apparente facilité, car si l'on ne prend pas soin de correctement justifier chaque postulat d'innéité, il devient alors possible d'expliquer tout et n'importe quoi[6]. Peut-être de ce fait les cognitivistes se sont-ils moins intéressés aux propriétés et régularités de l'environnement.

6. C'était le grand défaut de la psychologie des facultés, de Gall à McDougall.

Cette situation ne pouvait pas ne pas provoquer des réactions. Gibson (1979) a réaffirmé avec force son opposition au paradigme cognitiviste, et l'idée selon laquelle la perception peut être décrite sans référence à des processus mentaux, mais qu'elle capture « directement » les propriétés adéquates de l'environnement. Bien que Gibson ait suscité autant de critiques que Skinner (Fodor & Pylyshyn, 1981), ses travaux sur les indices visuels subtils qui sous-tendent l'appréciation des distances et le suivi des objets en mouvement (Gibson, 1950) restent emblématiques du fait que l'environnement est sans doute beaucoup plus riche que ce qu'un examen superficiel peut laisser supposer.

On peut voir dans le courant connexionniste actuel l'héritage de Gibson et Skinner. On y retrouve le refus du symbolisme et de l'innéité des représentations (voir Elman et coll., 1996, et Marcus, 1998, pour les débats les plus récents), et également l'idée qu'il y a beaucoup d'information pertinente à extraire de l'environnement. L'efficacité des modèles connexionnistes à extraire des régularités statistiques de leurs entrées a en effet entraîné un regain d'intérêt pour l'étude notamment des régularités statistiques de la parole (Seidenberg, 1997; Christiansen, Allen, & Seidenberg, 1998; Dominey & Ramus, sous presse).

On ne peut donc que constater que les préoccupations des connexionnistes et celles des partisans de l'initialisation phonologique convergent. Parlant de l'initialisation phonologique, Seidenberg (1997, p. 1601, traduction libre) avance même que

> « Les réseaux connexionnistes offrent une généralisation et une formalisation de cette notion en termes de satisfaction de contraintes probabilistes multiples. »

L'initialisation phonologique est-elle donc une proposition empiriste, behavioriste et connexionniste ? Pas nécessairement : à l'origine, elle se situe au contraire dans la lignée de la théorie de la grammaire universelle (Chomsky, 1976), et plus particulièrement de sa formulation appelée *Principes et Paramètres* (Chomsky, 1981). Dans cette théorie, les enfants possèdent de manière innée une grammaire universelle qui comporte des principes universels du langage, et des paramètres qui décrivent les variations possibles entre les langues. Le but de l'acquisition du langage est alors de fixer tous les paramètres dans l'état qui rend compte de la langue particulière à laquelle est exposé l'enfant. Dans ce contexte, l'hypothèse de l'initialisation phonologique propose que les régularités présentes dans la parole sont une source importante d'information permettant de fixer des paramètres dans le bon état, ou tout au moins de contraindre la manière dont ces paramètres vont être fixés par l'enfant. A contrario, la formulation connexionniste de cette hy-

pothèse laisse entendre que les régularités de la parole et de l'environnement sont suffisamment riches pour rendre compte de la totalité de l'acquisition du langage, sans même postuler des principes et des paramètres innés.

Bien qu'idéologiquement opposées, les deux positions s'accordent donc sur la constatation que le rôle des propriétés et régularités de la parole a été quelque peu négligé, et qu'il est souhaitable de les étudier de manière systématique. La question de savoir quelles composantes du langages sont innées (ou pas) reste largement ouverte, et la démarche à adopter est finalement indépendante des options idéologiques : il faut pousser l'hypothèse de l'initialisation phonologique jusqu'à ses limites, c'est-à-dire explorer sans relâche les propriétés de la parole, et voir jusqu'où l'on peut aller dans l'acquisition du langage avec ces informations. C'est que nous nous proposons de faire dans cette thèse pour une propriété phonologique particulière, le rythme de la parole.

Après cette digression qui aura permis de faire le point sur les enjeux théoriques de l'initialisation phonologique, nous allons maintenant détailler quelques cas précis dans lesquels cette hypothèse semble apporter des solutions concrètes aux paradoxes de l'acquisition du langage.

1.3 Quelques exemples d'initialisation phonologique

Dans cette section nous allons explorer trois exemples d'initialisation phonologique. Leur forme générale est du type « l'analyse de la propriété phonologique P1 permet aux enfants d'apprendre la propriété plus abstraite P2 ». Il ne constituent pas des explications définitives des apprentissages en question, mais plutôt des scénarios hypothétiques dont l'intérêt principal est leur plausibilité. Celle-ci se construit en trois étapes, que nous reprendrons dans chaque cas :

- trouver des indices phonologiques P1 qui sont corrélés avec les propriétés plus abstraites P2 dont on souhaite expliquer l'apprentissage ;
- montrer que les indices P1 sont perçus par les enfants à un âge suffisamment précoce (avant que les enfants donnent des signes de maîtriser P2) ;
- essayer de montrer que ces indices P1 sont effectivement utilisés par les enfants dans leur apprentissage de P2.

Mais tout d'abord, nous allons voir comment des indices phonologiques peuvent, comme l'avançaient Morgan et Demuth (1996b), signaler les domaines au sein desquels l'enfant doit effectuer son analyse du signal de parole.

1.3.1 Restreindre le domaine des analyses de la parole

Comme nous l'avons mentionné, la parole est continue, mais fort heureusement elle n'est pas infinie. Tous les locuteurs, mêmes les plus bavards, finissent tôt ou tard par s'arrêter de parler. Mieux que cela, ils font des pauses, ne serait-ce que pour reprendre leur respiration. Ainsi, si la parole n'est pas segmentée en mots, elle est au moins segmentée en unités plus grandes que les mots, que l'on appelle *énoncés phonologiques*. En fait, les énoncés phonologiques ne sont que la partie la plus saillante de la *hiérarchie prosodique*, que nous allons maintenant décrire succinctement.

1.3.1.1 La hiérarchie prosodique

La phonologie prosodique (Selkirk, 1984) décrit l'organisation de la parole au niveau supra-segmental, c'est-à-dire qu'elle s'attache aux unités plus grandes que le phonème (ou « segment »). Dans la théorie de Nespor et Vogel (1986), ces unités forment une hiérarchie, dans la mesure où chaque unité peut se décomposer en unités de plus petite taille, constituant ainsi plusieurs niveaux emboîtés les uns dans les autres. Nespor et Vogel (1986) proposent ainsi 7 types d'unités, dont la plus grande est l'énoncé phonologique. Celui-ci est constitué de groupes intonationnels, eux-mêmes constitués de groupes[7] phonologiques, et ainsi de suite, de groupes clitiques, de mots phonologiques, de pieds et enfin de syllabes. Une phrase d'exemple permettra d'illustrer les trois types d'unités qui nous intéressent plus particulièrement :

{[(Visiblement)]} {[(les anciens) (ennemis)]
[(sont) (allés)] [(à la pêche)]}

Les parenthèses ( ) signalent les groupes clitiques, les crochets [ ] les groupes phonologiques, et les accolades { } les groupes intonationnels. La phrase toute entière constitue un unique énoncé phonologique. L'existence de chacune de ces unités est motivée par la nécessité d'expliquer un certain nombre de phénomènes phonologiques que nous ne détaillerons pas ici. À titre d'exemple seulement, mentionnons que les phénomènes de liaisons entre les mots en français sont bien décrits au sein de la hiérarchie prosodique[8] : la liaison est systématique au sein des groupes clitiques (comme dans « les‿anciens »), optionnelle au sein des groupes phonologiques (elle se fait dans « anciens‿ennemis », mais pas obligatoirement dans « sont~allés »), mais rare entre deux groupes

7. Ici l'anglais « phrase » est traduit par « groupe », comme dans « groupe verbal/nominal » pour « verb/noun phrase », ceci afin d'éviter toute confusion avec le sens français de « phrase ». « Syntagme » aurait également été une traduction appropriée.

8. Mais n'oublions pas qu'ils sont aussi extrêmement variables selon l'accent régional ou social et le niveau de langue.

phonologiques (« allés$_\sim$à ») et interdite entre deux groupes intonationnels.

L'intérêt de la hiérarchie prosodique pour notre problématique est qu'elle établit un lien entre l'organisation grammaticale de la phrase et le signal de parole. En effet, les unités prosodiques sont définies principalement en fonction des propriétés grammaticales de leurs constituants. Par exemple, un groupe clitique est constitué d'un mot de contenu et des mots de fonction[9] non accentués qui s'y rattachent syntaxiquement. De même, un groupe phonologique est constitué d'une tête lexicale[10] et des mots qui se trouvent à sa gauche[11]. Ainsi, sans leur être identiques, les groupes clitiques et groupes phonologiques sont étroitement liés aux groupes grammaticaux constituant la phrase. D'autre part, les unités prosodiques sont plus ou moins saillantes dans le signal de parole. Ainsi, comme nous l'avons dit, les énoncés phonologiques sont délimités par des pauses. Les groupes intonationnels, lorsqu'ils ne sont pas délimités eux aussi par de véritables pauses, semblent néanmoins séparés les uns des autres. Ainsi, bien qu'il soit possible de prononcer la phrase d'exemple sans faire de pause, on peut néanmoins entendre une frontière prosodique entre « Visiblement » et « les ». C'est également vrai, dans une moindre mesure, pour les frontières de groupes phonologiques. Plus généralement, on peut considérer que plus les unités sont grandes, plus leurs frontières sont marquées dans le signal. Ainsi, dans notre phrase d'exemple, les frontières entre deux mots sont d'autant plus marquées qu'il y a de symboles)(|]}{ entre eux (6 entre « Visiblement » et « les » mais 0 entre « les » et « anciens » qui font partie du même groupe clitique).

1.3.1.2 Réalité acoustique et psychologique des frontières prosodiques

Au-delà de ces considérations intuitives sur les frontières entre unités prosodiques, les indices acoustiques qui signalent ces frontières ont été étudiés dans un certain nombre de langues. En-dehors des pauses qui marquent les frontières les plus importantes, on observe universellement un allongement des syllabes, une baisse de l'amplitude et une baisse de la fréquence fondamentale (F_0) de la parole avant une frontière prosodique (Bolinger, 1978; Fisher & Tokura, 1996). Celles-ci sont donc bel et bien marquées dans le signal. Il reste à savoir si elles sont perceptibles par le système auditif.

Streeter (1978) a montré que des sujets adultes sont capables de percevoir

9. Pour la distinction entre mots de fonction et mots de contenu voir la Section 1.3.3.1.

10. Une tête lexicale est le mot principal d'un constituant grammatical : le verbe d'un groupe verbal, le nom d'un groupe nominal, ou encore l'adjectif d'une proposition adjectivale.

11. Plus exactement du côté non-récursif de la tête.

des frontières de propositions en Anglais, en se basant à la fois sur la durée, l'amplitude et la fréquence fondamentale. Luce et Charles-Luce (1983) ont de plus montré que la durée est l'indice le plus important des trois. Cependant ces auteurs n'ont pas contrôlé précisément le type de frontières prosodiques qu'ils ont présenté à leurs sujets. À partir des exemples qu'ils donnent on peut estimer qu'il s'agissait pour l'essentiel de frontières de groupes intonationnels. Christophe, Dupoux, Bertoncini, et Mehler (1994) ont de plus montré qu'en français les frontières de groupes phonologiques, bien que plus subtiles, sont également perceptibles. Enfin, de Pijper et Sanderman (1994) ont montré que non seulement les frontières de groupes phonologiques et intonationnels étaient perceptibles en néerlandais, mais aussi que les sujets étaient capables de juger correctement de l'importance relative de ces frontières. Pour resituer par rapport à notre exemple, cela signifie que les sujets étaient plus ou moins capables d'estimer le nombre de symboles)(|[}{ présents entre deux mots. De plus, leurs jugements étaient corrélés avec l'importance des indices acoustiques présents dans les phrases. Ainsi, ces études montrent que les adultes sont sensibles aux frontières prosodiques, au moins pour les unités de taille égale ou supérieure au groupe phonologique.

Pour que les unités prosodiques aient un intérêt dans l'acquisition du langage, il faut cependant qu'elles soient perçues aussi par les nourrissons. C'est ce qu'ont montré Hirsh-Pasek et coll. (1987), en faisant écouter à des bébés de 7 mois des phrases entrecoupées par des silences. Les silences étaient placés soit aux frontières, soit au milieu de groupes intonationnels. Les nourrissons ont préféré écouter les phrases coupées aux frontières, montrant que dès 7 mois, ils sont sensibles à la cohésion des groupes intonationnels (des résultats similaires ont été obtenus par Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989). Utilisant la même technique (décrite à l'Annexe A.2), Jusczyk et coll., 1992 ont de plus montré que des enfants âgés de 9 mois sont sensibles également aux groupes phonologiques, et que ce résultat vaut même lorsque les phrases sont filtrées, ne préservant ainsi que les indices prosodiques (voir aussi Gerken, Jusczyk, & Mandel, 1994). Pour mieux comprendre comment en quelques mois de vie les enfants apprennent la forme des unités prosodiques de leur langue, Christophe et coll. (1994) ont présenté à des nouveau-nés de trois jours des paires de syllabes ne différant que par la présence ou non d'une frontière de groupe phonologique entre les deux (par exemple [mati] extrait de « mathé**mati**cien » ou de « panora**ma ty**pique »). Les nouveau-nés ont discriminé les bi-syllabes qui contenaient une frontière de ceux qui n'en contenaient pas, montrant ainsi que les indices acoustiques signalant les frontières de groupes phonologiques sont perceptibles dès la naissance (cf. technique de succions de haute amplitude, Annexe A.1).

Pour résumer, ces résultats convergent pour montrer que les frontières de

groupes phonologiques, de groupes intonationnels et d'énoncés phonologiques sont marquées dans le signal de parole par des indices tels que les pauses, les durées des segments et des syllabes, et la fréquence fondamentale, et que grâce à ces indices les frontières sont perceptibles dès la naissance et tout au long de la vie. Ceci permet aux enfants de repérer les unités prosodiques de leur langue avant la fin de leur première année de vie. Nous allons voir dans les exemples suivants que ce pré-découpage des phrases peut significativement simplifier un certain nombre d'apprentissages.

1.3.2 L'extraction des mots du flux de parole

Si le problème de l'extraction des mots n'est pas encore résolu, il vient de se simplifier considérablement grâce à la hiérarchie prosodique. En effet, nous savons maintenant qu'il n'est pas nécessaire de chercher les mots dans des phrases arbitrairement longues : le domaine de recherche se restreint en fait au groupe phonologique. En d'autres termes, les frontières de groupes phonologiques sont aussi des frontières de mots, ce qui réduit d'autant le nombre de frontières qu'il reste à trouver. Pour repérer les mots au sein des groupes phonologiques, l'enfant dispose encore d'un assez grand nombre d'indices. Ici, nous distinguerons les régularités distributionnelles, qui peuvent se déduire directement d'analyses statistiques sur le signal de parole, et les régularités lexicales, qui nécessitent d'être déjà familier avec la forme typique des mots et de leurs sons dans une langue donnée.

1.3.2.1 Régularités distributionnelles

Définition. Inspirée à l'origine par Harris (1954, 1955) et Hayes et Clark (1970), la notion de régularité distributionnelle a été introduite par Brent et Cartwright (1996) pour rendre compte de l'intuition selon laquelle

> « les séquences de sons qui se rencontrent fréquemment et dans des contextes variés sont de meilleurs candidats pour le lexique que celles qui se rencontrent rarement ou dans peu de contextes. » (p. 93)

Pour illustrer cette idée, reprenons la phrase « Regarde le chien », prononcée [RəgaRdləʃjɛ̃], et considérons une segmentation possible : [Rəg aRdl əʃjɛ̃]. Supposons de plus que l'enfant peut faire des statistiques sur un grand nombre d'énoncés qu'il a entendus. Il pourra avoir remarqué que les séquences [RəgaRd], [lə] et [ʃjɛ̃] reviennent relativement souvent, par comparaison notamment avec les séquences [aRdl], qui ne se rencontre que lorsque que « regarde » (ou « garde ») est suivi d'un mot commençant par [l], ou [əʃjɛ̃], qui ne se rencontre que lorsque « chien » est précédé de « le » ou « de » (et pas

lorsqu'il est précédé de « un » ou « mon »). De plus, les trois mots peuvent se retrouver dans un très grand nombre de contextes, puisque « regarde » peut être précédé et suivi de n'importe quel groupe nominal, « chien » peut être suivi ou pas par n'importe quel verbe, préposition, etc. Par comparaison, la séquence [Rəg] est nécessairement suivie de la séquence [aRd], sauf quelquefois comme dans « regain ». De même, [aRdl] ne peut être précédée que d'un ensemble assez limité de phonèmes (essentiellement [g], mais aussi beaucoup plus rarement [b], [f], [k], [d], [l]), et est nécessairement suivi de [ə], [a] ou [e]. Ainsi, les mots peuvent se rencontrer de manière générale dans un bien plus grand nombre de contextes que les séquences de sons incluses dans les mots [Rəg] ou chevauchant des frontières de mots [əʃjɛ̃].

Pour confirmer cette intuition, Brent et Cartwright (1996) ont écrit un algorithme qui mesure les régularités distributionnelles sur des corpus de parole transcrite phonétiquement (cf. également Cairns, Shillcock, Chater, & Levy, 1997 pour une approche similaire). Ils ont ainsi montré sur un corpus américain qu'il était possible de segmenter correctement 40 à 50% des mots en se basant exclusivement sur ces régularités. Bien que ce niveau de performance ne soit pas suffisant pour expliquer en totalité la segmentation de la parole en mots, il est clair que ce type d'information n'est pas négligeable et pourrait bien être l'un des multiples indices que les enfants ont à leur disposition.

Sensibilité aux régularités distributionnelles. S'inspirant d'un résultat préliminaire de Goodsitt, Morgan, et Kuhl (1993), Saffran et ses collègues ont manipulé systématiquement les régularités distributionnelles dans différents mini-langages artificiels. Une version typique d'un de ces langages était constituée de 4 mots: *tupiro*, *golabu*, *bidaku* et *padoti*. Ces mots ont ensuite été concaténés en un flux continu de syllabes, sans silences, sans intonation et sans répétition immédiate de mots. La question était de savoir si, en entendant le flux continu, il était possible de segmenter les mots. Une telle segmentation n'est en effet possible que sur la base des régularités distributionnelles : dans le flux de parole, [tu] peut être précédée de 3 syllabes différentes ([bu], [ku] et [ti]), mais est toujours suivie de [pi] ; [pi] est toujours précédée de [tu] et suivie de [ro] ; [ro] est toujours précédée de [pi], mais peut être suivie de trois syllabes différentes ([go], [bi] et [pa]). Ainsi, en remarquant, que [tu], [pi] et [ro] apparaissent toujours ensemble et dans le même ordre, et que [tupiro] apparaît dans des contextes variés, il est possible d'en déduire qu'il s'agit d'un mot, alors que [rogola], qui n'a pas les mêmes propriétés, n'en est

pas un[12]. Saffran, Newport, et Aslin (1996) ont fait écouter 21 minutes de ce mini-langage à des sujets adultes, et ceux-ci ont consécutivement reconnu les mots mieux que le hasard. De plus Saffran, Newport, Aslin, Tunick, et Barrueco (1997) ont également montré que la segmentation pouvait se faire de manière totalement passive et implicite, alors que les sujets effectuaient une tache qui n'avait rien à voir (du dessin) pendant l'écoute. Ce résultat a été obtenu à la fois chez des adultes et chez des enfants de 6 à 7 ans. Enfin, en utilisant la méthode d'orientation préférentielle de la tête (Annexe A.2), Saffran, Aslin, et Newport (1996) et Aslin, Saffran, et Newport (1998) ont fait écouter seulement 2 minutes du mini-langage à des enfants de 8 mois, et ont montré que ces enfants réagissaient ensuite préférentiellement aux mots de ce langage par rapport à des suites de syllabes chevauchant des frontières de mots (comme [rogola]). Ainsi, dès l'âge de 8 mois, les enfants sont sensibles aux régularités distributionnelles et peuvent s'en servir pour extraire des mots de la parole.

On pourra objecter que les études de Saffran et coll. confrontent les enfants à un langage très peu naturel. S'il est incontestable que ces expériences montrent leur sensibilité aux probabilités de transition, elles ne montrent pas nécessairement que cette capacité est utilisée par les enfants dans l'apprentissage du vrai langage. Mais Jusczyk et ses collègues ont de plus obtenu quelques résultats qui suggèrent que les enfants sont capables d'utiliser les régularités distributionnelles pour extraire les mots de la parole réelle. En utilisant la méthode d'orientation préférentielle de la tête (Annexe A.2), ils ont familiarisé des enfants avec des phrases contenant toutes le même mot monosyllabique dans différents contextes, et les ont ensuite testés avec ce même mot prononcé en isolation, ainsi que des mots n'apparaissant pas dans les phrases : les enfants ont préféré écouter le mot qu'ils avaient entendu dans les phrases, plutôt que ceux qui n'y figuraient pas (Jusczyk & Aslin, 1995). Ce résultat suggère que dès 7 mois et demi, les enfants sont capables de détecter dans la parole continue des syllabes qui reviennent fréquemment et dans des contextes variés.

Dans une autre étude, Jusczyk, Houston, et Newsome (sous presse) ont généralisé ce résultat à des mots bi-syllabiques : après avoir été familiarisés avec des phrases contenant des mots bi-syllabiques comme « kingdom », des enfants de 7 mois et demi ont préféré écouter ces mots en isolation plutôt que d'autres mots, ou même plutôt que des parties de ces mots familiers, comme le mot « king », qui apparaissait autant de fois que « kingdom » dans les

12. Cette idée a été formalisée par Saffran et ses collègues en utilisant la notion de probabilité de transition. Dans nos exemples, la probabilité de transition de [tu] à [pi] est de 1, comme celle de [pi] à [ro], alors que celle de [ro] à [go] n'est que de 1/3. Les frontières de mots correspondent ainsi aux minima de la suite des probabilités de transition.

phrases de familiarisation, mais toujours dans un contexte moins varié (suivi de [dom]). Dans une autre série d'expériences (Jusczyk, Hohne, & Bauman, sous presse), les chercheurs ont familiarisé des enfants américains de 9 mois avec le mot « night ». Dans la phase de test, ceux-ci n'ont pas reconnu le mot « night » lorsque celui-ci était présenté dans des phrases comportant toujours l'expression « night rate ». Néanmoins les enfants l'ont reconnu dans une autre condition où les phrases contenaient des expressions plus variées (« night time », « night schools », « night games »...). Ces enfants ont donc pu extraire le mot « night » lorsqu'il était présenté plusieurs fois dans des contextes variés, mais pas lorsqu'il était présenté le même nombre de fois dans un contexte fixe. Ainsi, dans les deux séries d'expériences, le comportement des enfants n'est interprétable que si l'on suppose que leur stratégie de segmentation des mots tient compte des régularités distributionnelles. Notons enfin que cette stratégie est d'autant plus puissante qu'elle ne nécessite aucune connaissance spécifique sur la langue à apprendre. Elle suppose seulement que l'enfant est capable de segmenter la parole en sons élémentaires, des phonèmes ou des syllabes, et de calculer au fur à mesure les probabilités de transition d'une unité à l'autre.

1.3.2.2 Régularités lexicales

Un autre type d'indice potentiellement utile pour segmenter les mots de la parole est l'ensemble des régularités propres aux mots de chaque langue. On pourrait croire a priori que ce type d'indice n'est pas pertinent dans l'acquisition du langage, dans la mesure où il faut d'abord connaître les mots pour découvrir leur propriétés typiques. Néanmoins, nous verrons à la fin de cette section qu'il est concevable de découvrir ces propriétés par d'autres moyens, sans nécessairement posséder au préalable les mots eux-mêmes.

Variations allophoniques. Il est bien connu que les phonèmes ont parfois une forme changeante : on dit qu'ils présentent plusieurs allophones. Par exemple, en anglais, la consonne [t] a trois allophones, le [t] standard, le [t^h] aspiré et le [$t^ʔ$] glottalisé. Il ne s'agit pas de phonèmes différents mais de variantes du même phonème qui se produisent dans des contextes différents. On pourrait craindre que cette variabilité supplémentaire des sons de la parole augmente encore la complexité de la perception et de l'apprentissage des mots. Il semble en fait que ce soit plutôt le contraire, car les différents allophones d'un phonème se distribuent en fonction de leur position dans le mot (Bolinger & Gerstman, 1957; Hockett, 1958). Par exemple, en anglais, le [t] est aspiré en début de syllabe, mais pas en fin de syllabe (Umeda & Coker, 1974). De plus il est souvent glottalisé en fin de mot. Quant au [r], il peut être

voisé ou non voisé selon sa position dans le mot et les consonnes voisines au sein de la même syllabe. Ainsi, les expressions « nitrate » et « night rate », bien que composés de la même séquence de phonèmes [najtrɛjt], diffèrent par les allophones de [t] et de [r] : dans « nitrate », [t] est aspiré et [r] est dévoisé, alors que dans « night rate » [t] est non-aspiré et [r] est voisé. Ces deux derniers indices signalent donc la frontière séparant les mots « night » et « rate ». De manière plus générale, les variations allophoniques sont autant d'indices potentiels pour déterminer certaines frontières de mots (Church, 1987).

Les nourrissons sont-ils sensibles à ces indices ? En utilisant la méthode de succions de haute amplitude (Annexe A.1), Hohne et Jusczyk (1994) ont montré que des bébés de 2 mois étaient capables de distinguer « nitrate » de « night rate », et ce sur la seule base d'indices allophoniques. Jusczyk et coll. (sous presse) ont de plus montré que des enfants de 10 mois et demi étaient capables de segmenter ces deux expressions à l'intérieur de phrases, sans les confondre l'une avec l'autre[13] (méthode d'orientation préférentielle, Annexe A.2). Ainsi, non seulement les nourrissons sont capables de percevoir les variations allophoniques très précocement, mais dès 10 mois et demi ils parviennent à s'en servir pour extraire des mots de la parole continue.

Contraintes phonotactiques. Ce sont l'ensemble des règles qui régissent l'organisation des sons dans les mots, dans une langue donnée. Suivant Brent et Cartwright (1996), on peut en distinguer au moins trois types:

Voyelles : Chaque mot doit posséder au moins une voyelle ;

Groupes de consonnes aux frontières : Dans chaque langue il existe un ensemble limité de groupes de consonnes qui peuvent se situer en frontière de mot. Par exemple, en français, aucun mot ne peut commencer par [gd], pas plus que se terminer par [nsk], contrairement au polonais (Gdansk).

Groupes de consonnes internes : Dans chaque langue il existe un ensemble limité de groupes de consonnes qui peuvent se situer à l'intérieur des mots. Par exemple, en français, aucun mot ne peut contenir le groupe [ʀdl].

Connaître les contraintes phonotactiques de sa langue peut permettre de rejeter un grand nombre de segmentations inexactes de la parole. Par exemple, si l'on considère la phrase « Garde le chien » [gaʀdləʃjɛ̃], sachant que chaque mot doit avoir au moins une voyelle, que [ʀdl] est un groupe impossible à la fois à l'intérieur et en frontière de mot, que [dl] est impossible en début

13. Dans une autre expérience, des enfants de 7 mois et demi ont échoué sur la même tâche. Voir cependant Mattys et Jusczyk (en préparation).

de mot, et que [ʃj] est impossible en fin de mot, il ne reste plus que deux segmentations possibles pour la phrase (dont la bonne !) : [gaʀd ləʃ jɛ̃] et [gaʀd lə ʃjɛ̃]. Brent et Cartwright (1996) ont implémenté les contraintes phonotactiques de l'anglais dans leur algorithme de segmentation automatique de la parole (voir section précédente), et ont montré que celles-ci permettaient d'élever les performances de la segmentation aux alentours de 70%, soit un gain considérable par rapport aux seules régularités distributionnelles.

Cette fois encore, les nourrissons semblent être sensibles à ce nouveau type d'indices. En utilisant des syllabes respectant la phonétique et les contraintes phonotactiques de l'anglais mais violant celles du néerlandais, et des syllabes présentant les propriétés inverses, Jusczyk, Friederici, Wessels, Svenkerud, et Jusczyk (1993) ont montré que des enfants américains et néerlandais de 9 mois (mais pas de 6 mois) préfèrent écouter les syllabes qui respectent les contraintes de leur langue. Toutefois, dans cette étude, les deux listes de mots différaient non seulement par leur phonotactique, mais aussi par leurs phonèmes. Friederici et Wessels (1993) ont néanmoins montré que des enfants néerlandais de 9 mois préféraient écouter des syllabes qui respectaient la phonotactique du néerlandais que des syllabes qui ne la respectaient pas, sans différences phonétiques cette fois. Par ailleurs, Jusczyk, Luce, et Charles-Luce (1994) ont montré qu'à partir de 9 mois, les enfants sont également sensibles à la *fréquence* avec laquelle les phonèmes sont employés en début et en fin de syllabe, préférant les syllabes composées de suites de phonèmes fréquentes en anglais, plutôt que peu fréquentes. Ce dernier résultat suggère que l'apprentissage par le nourrisson des contraintes phonotactiques de sa langue pourrait se faire par un mécanisme de calcul de fréquences relatives, les séquences interdites dans une position donnée étant des cas particuliers de fréquence égale à zéro. Enfin, au-delà de la simple préférence pour des séquences rencontrées fréquemment, Mattys, Jusczyk, Luce, et Morgan (1999) ont montré que les bébés américains de 9 mois utilisent effectivement les contraintes phonotactiques de l'anglais pour segmenter des mots de la parole continue (voir section suivante). Toutes ces expériences ont été réalisées grâce à la méthode d'orientation préférentielle (Annexe A.2).

Contours prosodiques typiques. Dans toutes les langues, le contour prosodique des mots obéit à des règles bien précises. Dans la plupart des langues étudiées, ces règles induisent des régularités de surface potentiellement utiles pour la segmentation en mots. Ainsi, Cutler et Carter (1987) ont montré que la grande majorité des mots anglais (et particulièrement les plus fréquents) commencent par une syllabe accentuée. C'est généralement le cas dans les langues germaniques ainsi qu'en russe, en hongrois... En fran-

çais, en revanche, c'est la dernière syllabe du mot qui est la plus accentuée ; en Espagnol et en Italien, la plupart des mots portent l'accent sur l'avant-dernière syllabe. C'est ce qui a conduit Cutler (1990) à proposer la *stratégie de segmentation métrique*[14] (metrical segmentation strategy), selon laquelle les adultes utilisent le contour prosodique prédominant dans leur langue pour reconnaître les mots plus facilement : en anglais, ils postulent inconsciemment des frontières de mots avant les syllabes accentuées (Cutler & Norris, 1988; Cutler, Mehler, Norris, & Segui, 1992). Bien évidemment, cette stratégie est aussi potentiellement utile aux enfants en quête de mots dans la parole (Cutler, 1996).

Cette hypothèse est confortée par les travaux de Jusczyk, Cutler, et Redanz (1993), qui ont montré que les enfants américains de 9 mois (mais pas ceux de 6 mois) préfèrent écouter des mots bi-syllabiques accentués en première (trochaïques) plutôt qu'en seconde syllabe (iambiques). En familiarisant des enfants de 9 mois avec des mots tri-syllabiques accentués en deuxième syllabe, Echols, Crowhurst, et Childers (1997) ont également montré que ceux-ci reconnaissent mieux les sous-séquences composées des deux dernières syllabes (trochaïques) que des deux premières (iambiques). En utilisant une technique de conditionnement (cf. Annexe A.4), Morgan (1996) a de plus montré que les enfants américains de 9 mois (mais pas ceux de 6) perçoivent les mots trochaïques comme plus cohésifs que les mots iambiques, lorsqu'il s'agit de mots nouveaux. Lorsqu'il s'agit de mots familiers, les trochaïques et iambiques ont la même cohésion. Ce résultat suggère que si les enfants utilisent une stratégie de segmentation métrique, cela ne les empêche pas pour autant d'apprendre des mots iambiques, s'ils sont rencontrés suffisamment souvent. Cette stratégie ne s'appliquerait donc qu'aux mots nouveaux.

Enfin, Jusczyk et ses collègues ont étudié en détail les stratégies employées par les enfants pour segmenter la parole continue[15] (Jusczyk et coll., sous presse). Ils ont tout d'abord familiarisé des enfants de 7 mois et demi avec des mots bi-syllabiques trochaïques, et les ont testés avec des phrases contenant (ou non) ces mots : les enfants ont préféré écouter les phrases contenant les mots familiers[16]. Ce résultat reste valable lorsque les enfants sont familiari-

14. La métrique est la structure accentuelle de la phrase.

15. Procédure d'orientation préférentielle de la tête (Annexe A.2).

16. Ce résultat a été répliqué en néerlandais (dont la plupart des mots sont également accentués en première syllabe) sur des enfants néerlandais de 9 mois. Des enfants américains de 9 mois, dans la même expérience, ont également reconnu les mots dans les phrases néerlandaises, malgré les différences phonétiques entre ces deux langues, ce qui montre la robustesse de leur stratégie de segmentation (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, sous presse).

sés avec les phrases et testés avec les mots en isolation. Un certain nombre d'expériences contrôles a de plus montré que les enfants ne reconnaissaient pas seulement les syllabes accentuées de ces mots, mais bien les deux syllabes présentées ensembles. Au même âge et suivant une procédure identique, les enfants ont échoué à reconnaître des mots iambiques, montrant que leur stratégie de segmentation n'est efficace que sur les mots trochaïques. Pourtant, ils ont été capables de reconnaître la deuxième syllabe de ces mots (la syllabe accentuée, comme « tar » dans « guitar »). Enfin, lorsque ces mots iambiques ont été présentés dans un contexte fixe (comme « guitar is », où « is » est une syllabe non accentuée), les enfants n'ont plus reconnu la deuxième syllabe seule [tar], mais celle-ci suivie de son contexte [tariz]. Ces expériences semblent montrer que les enfants de 7 mois et demi utilisent effectivement la stratégie de segmentation métrique, tout en respectant les régularités distributionnelles :

1° en plaçant les frontières de mots *avant* les syllabes accentuées ;

2° en rattachant à ces syllabes accentuées la syllabe suivante, si c'est toujours la même (contexte fixe).

Signalons pour finir que si cette stratégie est utilisée à 7 mois et demi, elle ne peut pas l'être éternellement, car il faut bien que les enfants anglophones apprennent un jour les mots iambiques. C'est ce qui semble être le cas dès 10 mois et demi, âge auquel les enfants américains segmentent correctement les mots iambiques (« guitar »), plutôt que leur syllabe accentuée présentée seule [tar] ou même suivie d'un contexte fixe [tariz].

Faut-il déjà connaître les mots pour découvrir les régularités lexicales ? Dans cette section, nous avons vu que les mots d'une langue n'ont pas une forme arbitraire, mais présentent au contraire un grand nombre de régularités phonétiques, phonotactiques et prosodiques. Il est justifié de se demander comment l'enfant peut découvrir ces régularités. Dans les modèles de reconnaissance de la parole chez l'adulte, on suppose généralement que ces régularités, si elles sont utilisées, sont compilées à partir du lexique dont dispose l'adulte (cf. par exemple Cutler et Norris (1988)). On ne peut bien sûr faire cette hypothèse pour l'enfant, sous peine d'introduire une nouvelle circularité. Alors, comment se rendre compte que les allophones et les groupes de consonnes ne sont pas les mêmes aux frontières et à l'intérieur des mots, si l'on ne connaît pas ces frontières ? Et comment savoir que la majorité des mots débutent avec une syllabe accentuée, si l'on ne sait pas où se trouve leur début ?

Une bonne partie de la solution va venir de la hiérarchie prosodique. Comme nous l'avons expliqué plus haut, si l'enfant n'a pas accès dès le dé-

part à toutes les frontières de mots, il au moins accès à certaines d'entre elles : au minimum les frontières d'énoncés phonologiques et, si l'on accepte l'argumentation déployée à la Section 1.3.1, les frontières de groupes phonologiques, qui sont aussi des frontières de mots[17]. Dans ces conditions, il est concevable que l'enfant analyse les régularités observables sur les frontières *dont il dispose*, à défaut de le faire sur toutes les frontières. La question devient alors : les régularités observables sur les frontières accessibles à l'enfant sont-elles les mêmes que celles observables sur l'ensemble des frontières, ou bien sont-elles différentes, au risque d'induire l'enfant en erreur ? La réponse peut en fait dépendre du type de régularité considéré.

Au moins dans le cas de la phonotactique, on peut répondre par l'affirmative grâce au travail de Brent et Cartwright (1996). Dans la version de leur algorithme qui utilisait les contraintes phonotactiques, celles-ci avaient été incorporées directement à l'algorithme sous forme de listes des groupes de consonnes autorisés en début et en fin de mots. Dans une version ultérieure, ces auteurs ont supprimé la liste et ont laissé l'algorithme compiler sa propre liste sur la base des frontières d'énoncés du corpus. Effectuant maintenant la tâche de segmentation avec cette nouvelle liste, l'algorithme a finalement atteint le même niveau de performance qu'auparavant (plus de 70%), prouvant ainsi qu'il suffit d'analyser les frontières d'énoncés phonologiques pour apprendre correctement les contraintes phonotactiques de l'anglais.

Pour ce qui est des régularités allophoniques, aucune étude systématique n'a été entreprise, mais Bourgeois (1991) a documenté pour au moins deux contrastes qu'ils étaient observables aux frontières : il s'agit du phénomène d'aspiration déjà décrit p. 17, et de la nasalisation des voyelles, lorsqu'elles précèdent une nasale appartenant à la même syllabe. Il est tentant d'extrapoler à partir de ces deux résultats. En effet, la recherche des allophones autorisés en frontières d'énoncés peut se faire exactement selon le même principe que celle des groupes de consonnes : il suffit pour cela de disposer d'une représentation phonétique des énoncés qui incluent les variations allophoniques. Cette condition paraît relativement plausible, dans la mesure où le tri entre les allophones et les phonèmes est un problème en soi pour l'enfant, dont il n'est pas certain qu'il soit résolu avant la fin de la première année de vie. Il est donc naturel de supposer que le nourrisson travaille sur une représentation phonétique pure incluant toutes les variantes allophoniques. Ainsi, il semble raisonnable, dans l'état actuel des connaissances de supposer que les variations allophoniques peuvent être apprises en observant uniquement les frontières de groupes phonologiques.

17. Moyennant, dans certaines langues, les phénomènes de re-syllabification et de liaison, dans le détail desquels nous ne rentrerons pas.

Dans le cas des régularités prosodiques, on peut prédire un relatif succès de cette stratégie : dans les langues où une majorité de mots présentent le même contour accentuel, il est probable que les mots se situant en frontière d'unité prosodique aient la même régularité. Ainsi, l'enfant anglais pourrait remarquer que les unités prosodiques débutent généralement par une syllabe accentuée et finissent par une syllabe non accentuée, alors que l'enfant français remarquerait le contraire. Il reste néanmoins à vérifier que les régularités prosodiques sont bien généralisables aux frontières d'unités prosodiques. Un problème potentiel pourrait venir des mots monosyllabiques : beaucoup de phrases anglaises commencent par exemple par des pronoms (typiquement des mots monosyllabiques non accentués), et/ou finissent par un mot monosyllabique accentué (par exemple la phrase « he looks at the dog » commence par une syllabe non accentuée et se termine par une syllabe accentuée). Ce problème ne semble pas toucher le français, car les mots de fonction sont le plus souvent accentués lorsqu'ils sont situés en fin d'unité prosodique.

1.3.2.3 Exploitation de l'ensemble des régularités

Nous avons donc vu qu'il existe des régularités distributionnelles qu'il est possible d'extraire directement en analysant le signal non segmenté, et des régularités lexicales qui reflètent les propriétés des mots dans une langue donnée. Ces régularités lexicales semblent être au moins partiellement dérivables de celles observées aux frontières prosodiques. Faut-il s'inquiéter de ce « partiellement » ? En fait, il n'y a pas lieu de s'attendre à ce que le succès de cette stratégie soit total, car d'une part le processus d'apprentissage des mots est largement incrémental, et d'autre part l'utilisation conjointe de toutes les régularités multiplie leur puissance. Jusqu'à présent, nous nous sommes en quelque sorte placés dans l'hypothèse la plus défavorable, qui est que l'enfant ne connaît strictement aucun mot et qu'il doit tous les trouver en même temps à l'aide d'un seul indice phonologique à la fois. Si cette hypothèse de travail est utile pour exploiter au mieux les données de l'initialisation phonologique, la réalité est sans doute très différente.

En effet, chaque mot appris par l'enfant sert à trouver de nouvelles frontières. Par exemple, le prénom de l'enfant (très fréquent et souvent prononcé en isolation), les mots « maman », « papa » (très fréquents et apparaissant dans de nombreux contextes) sont appris extrêmement tôt par l'enfant[18] (4

18. Il a d'ailleurs été proposé que ces tous premiers mots constituent la base de la stratégie de segmentation métrique. Ceci repose sur l'observation qu'en anglais, ces mots sont pratiquement tous bi-syllabiques accentués en première syllabe (mommy, daddy, baby, doggie, les différents diminutifs des prénoms...). Cela permettrait d'atténuer l'objection mentionnée à la section précédente.

mois et demi pour le prénom, Mandel, Jusczyk, & Pisoni, 1995 ; 6 mois pour « mommy » et « daddy », Tincoff & Jusczyk, 1999). Dès lors, lorsque l'enfant entend « regarde maman », il peut segmenter « maman » d'office et a gagné une nouvelle frontière à analyser, dont il pourra par exemple déduire que [ʀd] est une suite de consonnes possible en fin de mot. Il pourra également noter que « maman » est accentué en dernière syllabe, de même que la séquence [ʀəgaʀd] qui reste à étudier. Si l'enfant entend à nouveau cette séquence dans des contextes différents, les régularités distributionnelles lui suggéreront qu'il s'agit bien d'un mot. Ainsi de suite, chaque mot appris contribue, à chaque fois qu'il est entendu dans un nouvel énoncé, à découvrir de nouveaux mots. Plus l'enfant possédera de mots, plus il lui sera facile de découvrir les autres.

Ce que reflète également l'exemple ci-dessus, c'est que les différents types de régularités semblent coopérer et apporter des informations complémentaires les unes aux autres. Cette intuition a été étayée rigoureusement au moins dans le cas des régularités distributionnelles et des contraintes phonotactique. En testant leur algorithme sur chaque type de régularité en l'absence de l'autre, Brent et Cartwright (1996) ont montré que leurs contributions à la performance de la segmentation étaient indépendantes et additives. Autrement dit, sur les 70% de frontières de mots trouvées, celles qui n'ont pas été trouvées grâce aux régularités distributionnelles ont été indiquées indépendamment par les contraintes phonotactiques, et vice-versa. On pourrait donc espérer améliorer encore l'algorithme en ajoutant les régularités allophoniques, et en autorisant la formation progressive d'un lexique.

Si la plupart des études ont cherché dans un premier temps à isoler chaque type de régularité pour en évaluer la contribution de manière indépendante, quelques études se sont penchées sur l'interaction des différents types de régularités. Déjà, dans la section précédente, nous avions noté que dans des expériences qui cherchaient à isoler les régularités prosodiques (Jusczyk et coll., sous presse) ou allophoniques (Jusczyk et coll., sous presse), des régularités distributionnelles étaient également présentes, et étaient effectivement utilisées par les enfants pour segmenter les mots. Morgan et ses collègues (Morgan, 1994; Morgan & Saffran, 1995) ont manipulé les différents types de régularité de manière plus systématique : en utilisant une procédure de conditionnement (cf. Annexe A.4), ils ont étudié les contributions respectives des régularités distributionnelles et prosodiques, en les présentant de manière conjointe ou décorrélée. Ils ont trouvé que les enfants de 6 mois segmentent la parole en s'appuyant essentiellement sur les régularités prosodiques, sans se préoccuper des régularités distributionnelles. À 9 mois en revanche, les enfants ont utilisé les deux types de régularités lorsqu'ils convergeaient vers une même segmentation, mais pas lorsqu'ils étaient dissociés. Ce résultat suggère que la *corrélation* des différentes régularités est un indice à part entière, qui

semble prendre de l'importance à partir de 9 mois.

En s'appuyant sur les résultats de segmentation de mots bi-syllabiques obtenus par Jusczyk et coll. (sous presse), Mattys et coll. (1999) ont étudié l'interaction entre les régularités prosodiques et phonotactiques. Ils ont pour cela utilisé des mots bi-syllabiques, accentués soit en première, soit en deuxième syllabe, et présentant un groupe de consonnes interne soit fréquent, soit peu fréquent en anglais. L'idée est que suivant la stratégie de segmentation lexicale, les mots accentués en deuxième syllabe sont segmentés comme deux mots. De plus, s'ils comportent un groupe de consonnes interne peu fréquent, cela suggère également qu'il s'agit de deux mots. En manipulant la place de l'accent et la fréquence des groupes de consonnes dans le même mot, on peut donc rendre les indices compatibles ou contradictoires. Les auteurs ont trouvé que les enfants de 9 mois préfèrent écouter les mots dont les deux types de régularités suggèrent la même segmentation, plutôt que des segmentations contradictoires. Comme dans les expériences de Morgan et Saffran (1995), la corrélation des deux indices est importante dès l'âge de 9 mois. De plus, en leur donnant le choix entre des mots où les régularités prosodiques et phonotactiques étaient en conflit, Mattys et ses collègues ont trouvé que les enfants préféraient les mots trochaïques, indépendamment des régularités phonotactiques. Ainsi, bien qu'à 9 mois les enfants soient sensibles à la phonotactique et à la manière dont celle-ci est corrélée aux régularités prosodiques, ces dernières restent tout de même l'indice majeur pour la forme typique des mots.

Pour résumer les études exposées ci-dessus et en extrapolant un peu, on peut dire que la sensibilité pour les régularités prosodiques apparaît vers 6 mois, celle pour les régularités distributionnelles à 7 mois et demi, celle pour les régularités phonotactiques à 9 mois, et celle pour les régularités allophoniques vers 10 mois et demi. Dès 9 mois, les enfants semblent capables d'intégrer différents types de régularités et sont sensibles à leurs corrélations. Néanmoins, lorsque les indices sont ambigus ou en conflit, ce sont les régularités prosodiques qui l'emportent.

1.3.2.4 Conclusion

Parmi les premières étapes de l'acquisition du langage, le domaine de la segmentation de la parole en mots est sans aucun doute celui qui a suscité le plus de recherches. Celles-ci ont largement porté leurs fruits et, d'un problème qui semblait inextricable à la section 1.1.1 (et il y a seulement 20 ans), nous sommes passés à un problème qui, s'il est loin d'être encore totalement résolu, est cependant sur la bonne voie : sa solution est connue

dans les grandes lignes[19], et les zones d'ombre portent maintenant plutôt sur des détails. Cet exemple est donc emblématique de l'intérêt de rechercher les régularités présentes dans le signal, comme le préconise l'hypothèse de l'initialisation phonologique. Dans les sections qui suivent, nous allons examiner d'autres domaines où cette approche a un rôle à jouer, et qui, bien que moins étudiés jusqu'à présent, sont tout aussi prometteurs.

1.3.3 L'apprentissage des catégories grammaticales

À la Section 1.1.1 nous avons vu que l'apprentissage du sens des mots serait grandement facilité si l'enfant avait accès à l'avance aux catégories grammaticales des mots (c'était l'initialisation syntaxique). Néanmoins, à la Section 1.1.2, nous avons également vu que l'apprentissage de la syntaxe des phrases, et notamment des catégories grammaticales, était difficilement envisageable sans le recours au sens des mots (c'était l'initialisation sémantique), conduisant ainsi à une situation paradoxale. Nous allons maintenant voir comment l'initialisation phonologique peut nous sortir de ce mauvais pas.

1.3.3.1 Mots de fonction et mots de contenu

Définition. Les mots de contenu sont tous les mots ayant un réel contenu sémantique : les noms, les verbes, les adjectifs et les adverbes. Il existe dans chaque langue un très grand nombre de mots de contenu (plusieurs dizaines de milliers), et ce nombre est en principe illimité : de nouveaux mots apparaissent tous les jours en fonction des nouveaux concepts que nous avons besoin d'exprimer. Pour cette raison on les appelle aussi les mots de la classe ouverte. Les autres mots sont des mots de fonction, ou mots grammaticaux. Il s'agit des pronoms, déterminants, prépositions, auxiliaires, etc.[20] Ces mots ont un contenu sémantique faible ou indéfinissable, et sont en nombre très limité et fixe (on les appelle aussi les mots de la classe fermée). Les mots de fonction servent à relier les mots de contenu et à véhiculer la structure syntaxique de la phrase. Enfin, la distinction entre mots de contenu et mots de fonction est valable dans toutes langues[21].

19. On pourrait ajouter : pour l'apprentissage de l'anglais, et sans doute des autres langues indo-européennes. Pour des langues très différentes comme les langues agglutinantes, où la notion même de mot est à revoir, les données manquent, même s'il est peu probable que les conclusions soient radicalement différentes.

20. On y inclut aussi les morphèmes de fonction, tels ceux qui marquent le nombre, le genre et autres propriétés grammaticales, et qui sont généralement attachés aux mots de contenu.

21. Dans les langues agglutinantes où la notion de mot est légèrement différente, les mots de fonction se retrouvent sous la forme de morphèmes de fonction attachés aux mots de

Contribution à l'acquisition du langage. Plusieurs problèmes de l'acquisition du langage peuvent se trouver considérablement simplifiés s'il s'avère que les enfants peuvent apprendre précocement les mots de fonction.

Comme nous l'avons mentionné plus haut, tout mot, une fois appris, facilite l'acquisition des autres mots. C'est bien sûr vrai aussi des mots de fonction. Si à un stade précoce, l'enfant a appris les mots de fonction les plus fréquents (comme le déterminant « le »), la segmentation de toutes les phrases les contenant (comme « Regarde le chien ») en sera simplifiée. Le bénéfice est d'autant plus grand que, si les mots de fonction sont en nombre limité, leur fréquence est considérable : ils constituent seulement 0,1% du dictionnaire, mais 40% des phrases (Kucera & Francis, 1967) ! Ainsi, un début de lexique comprenant seulement une cinquantaine de mots de fonction peut permettre à l'enfant de placer correctement plus de 40% des frontières de mots[22].

Le problème d'une telle stratégie est qu'elle conduit à sur-segmenter la parole : par exemple, à extraire l'article « la » des mots « lapin » ou « maladie », d'où une segmentation et une analyse de la phrase complètement fausses. C'est là que la hiérarchie prosodique (Section 1.3.1) a son rôle à jouer. En effet, les mots de fonction se situent le plus souvent aux extrémités des groupes phonologiques (voir ci-dessous). Cette contrainte permet de faire un tri significatif entre les vrais mots de fonction et les mêmes syllabes situées à l'intérieur d'autres mots. On peut dès lors segmenter systématiquement les mots de fonction situés en bordure de groupes phonologiques, et de ce fait restreindre encore plus le domaine d'analyse, éventuellement jusqu'au mot phonologique.

A la Section 1.1.1, nous avons également vu qu'un des problèmes de l'apprentissage du sens des mots résidait dans l'attribution à chaque forme sonore de mot du sens qui lui revient. Savoir reconnaître certains mots de fonction peut éviter de leur attribuer un sens inapproprié : toujours dans le même exemple, si l'enfant sait que « le » est un mot de fonction, il ne lui attribuera pas le sens de « chien ». Il n'aura plus qu'à déterminer les rôles respectifs du verbe et du nom[23].

Enfin, l'enfant pourrait détecter les liens entre les mots de fonction et les constituants grammaticaux auxquels ils appartiennent, c'est-à-dire re-

---

contenu.

22. Chaque mot de fonction a deux frontières, ce qui porte le pourcentage théorique aux alentours de 80, moins les cas (nombreux) où le mot est déjà en bordure d'unité prosodique, et moins les cas où plusieurs mots de fonction sont contigus.

23. Cet argument suppose en outre que l'enfant n'a pas seulement isolé les mots de fonction comme un ensemble de formes sonores reconnaissables, mais aussi qu'il « sait » quel rôle grammatical ils peuvent (ou ne peuvent pas) jouer. Cette connaissance pourrait être innée, ou due à un apprentissage qui reste à préciser.

marquer quels mots de fonction peuvent appartenir à quels constituants. Il pourrait en cela être aidé de certains universaux du langage, comme le fait qu'au sein d'un groupe intonationnel, il ne peut y avoir qu'un seul groupe verbal, mais éventuellement plusieurs groupes nominaux (Nespor, Guasti, & Christophe, 1996). Il pourrait donc en déduire que les mots de fonction apparaissant plusieurs fois dans un même groupe intonationnel appartiennent à des groupes nominaux, et que ceux qui n'apparaissent jamais qu'une fois sont associés à des groupes verbaux (Christophe, Guasti, Nespor, Dupoux, & van Ooyen, 1997).

Caractérisation phonologique. De nombreux auteurs ont signalé que les mots de fonction et de contenu possèdent généralement des propriétés phonologiques assez différentes (Selkirk, 1984; Nespor & Vogel, 1986). Par exemple, en anglais, la plupart des mots de fonction sont monosyllabiques et non accentués, alors que les mots de contenu comportent toujours un accent tonique. En serbo-croate, les mots de contenu comportent toujours un ton haut sur l'une de leurs syllabes, mais pas les mots de fonction. En japonais de Tokyo, les mots de fonction peuvent perdre leur ton haut dans certaines circonstances, les mots de contenu jamais (Selkirk, 1996 ; voir Peters & Strömqvist, 1996; Morgan, 1996; Hung & Peters, 1997 pour des exemples en suédois, chinois mandarin, taïwanais...). De manière plus générale, Selkirk (1996) propose deux propriétés universelles des mots de fonctions :

1° Contrairement aux mots de contenu, les mots de fonction n'engendrent pas systématiquement leurs propres mots prosodiques.

2° Les mots de fonction ont une prosodie plus variable que les mots de contenu.

Morgan, Shi, et Allopenna (1996) proposent une autre généralisation (compatible avec la précédente) : les mots de fonction sont « minimaux », à la fois dans la production et dans la perception de la parole. En se basant avant tout sur l'anglais, le turc et le mandarin, ces auteurs remarquent notamment que :

- Les mots de fonction ont une complexité minimale, à la fois en termes de nombre de syllabes (plutôt une que plusieurs), et en termes de complexité syllabique (souvent V ou CV, avec la voyelle non diphtonguée et souvent réduite). De plus, ils tendent à se simplifier encore plus dans certains contextes (par exemple la contraction de l'auxiliaire dans « he's », ou du déterminant dans « l'avion » ou encore « je l'vois »).
- Ils utilisent souvent un répertoire de phonèmes réduit par rapport aux mots de contenu (notamment des voyelles plus centrales), et ces phonèmes sont plus sujets à des phénomènes d'assimilation ou d'harmonie

que dans les mots de contenu.

- En termes de mesures acoustiques, les mots de fonction tendent à avoir une moindre durée et une moindre amplitude, ainsi qu'un contour intonatif plus plat et plus bas que les mots de contenu.
- Enfin, les mots de fonction présentent également une régularité distributionnelle, qui est qu'ils se situent souvent en bordure d'unités prosodiques (dans notre phrase d'exemple p. 11, trois mots de fonction sur quatre sont en début de groupe phonologique). On peut ajouter que les mots de fonction sont les mots qui se rencontrent à la fois le plus fréquemment, et dans les contextes les plus variés : l'article « le » peut être suivi de n'importe lequel des noms masculins de la langue française, alors qu'un nom masculin sera quasiment toujours précédé d'un déterminant ou d'un adjectif pré-nominal, qui sont en nombre très limité. Ce sont donc les meilleures candidats à une segmentation par analyse des probabilités de transition.

Bien entendu, ces propriétés générales se traduiront dans chaque langue par un ensemble d'indices spécifiques. Mais globalement, on peut considérer comme plausible que les enfants remarquent assez tôt ces petits mots qui reviennent tout le temps, souvent en bordure d'unité prosodique (donc plus facilement repérables), et qui ont des propriétés assez constantes.

Notons enfin que chacune des propriétés sus-mentionnées, prise séparément, peut n'être qu'un prédicteur assez faible du fait qu'un mot est de fonction ou de contenu. Pour l'anglais par exemple, Morgan et coll. (1996) ont trouvé que la nature des voyelles permettait de classifier correctement au mieux 63% des mots, et la durée des voyelles 64%. Néanmoins, le fait que ces deux indices soient corrélés est un indice supplémentaire. En entraînant des réseaux de neurones formels sur les deux indices à la fois, ces auteurs ont ainsi vu le taux de classification correcte dépasser 80%. Ils ont depuis généralisé ce résultat avec le même succès à une dizaine d'indices différents, et à deux autres langues, le mandarin et le turc (Shi, Morgan, & Allopenna, 1998).

Perception par le nourrisson. En utilisant une procédure de succions de haute amplitude (cf. Annexe A.1), Shi, Werker, et Morgan (1999) ont récemment montré que des nouveau-nés âgés de quelques jours, qu'ils soient de milieu anglophone ou pas, étaient capables de discriminer une liste de mots de fonctions d'une liste de mots de contenu anglais, même lorsque les mots étaient appariés en nombre et complexité des syllabes, et en fréquence d'apparition des exemplaires. Ce résultat montre que les nouveau-nés sont capables de distinguer les deux catégories sur la base des indices restants

(durée, amplitude, contour intonatif), sans nécessiter une expérience particulière avec l'anglais. Cette expérience ne montre cependant pas que les nouveau-nés peuvent déjà *catégoriser* ces mots en mots de fonction et mots de contenu, mais qu'ils possèdent un pré-requis, l'aptitude à distinguer les deux catégories.

La sensibilité aux mots de fonction en tant que tels ne semble apparaître que bien plus tard (mais étonnamment tôt cependant). En ce qui concerne la capacité à segmenter les mots, nous avons vu que Jusczyk et ses collègues en ont démontré les débuts aux alentours de 7 mois et demi, mais il s'agissait toujours de noms. Des expériences récentes utilisant les mêmes techniques (Shady et Jusczyk, en préparation) semblent montrer qu'il faut attendre 10 mois et demi pour que les enfants américains segmentent les mots « a » et « the » de la parole continue.

Shafer, Gerken, Shucard, et Shucard (1992), Shafer, Shucard, Shucard, et Gerken (1998) ont par ailleurs montré que des enfants de 11 mois avaient des potentiels évoqués différents lors de l'écoute de phrases où les mots de fonction avaient été remplacés par des syllabes arbitraires, par rapport à l'écoute de phrases normales. Un résultat similaire a également été obtenu par Shady, Jusczyk, et Gerken (1998) sur des enfants de 10 mois et demi en utilisant la technique d'orientation préférentielle de la tête (Annexe A.2). Ces auteurs ont de plus répliqué ce résultat en utilisant des faux mots de fonction qui avaient des propriétés phonologiques similaires aux mots d'origine. Ce résultat suggère qu'à 10 mois et demi, les enfants connaissent déjà un certain nombre de mots de fonction, plutôt que seulement leurs propriétés phonologiques générales. En revanche, dans une autre expérience où ce sont les mots de contenu qui ont été remplacés par des faux mots similaires, les enfants ne se sont pas orientés préférentiellement vers les vraies phrases. Ce résultat se comprend en rappelant que les mots de contenu sont très nombreux, et à 10 mois et demi les enfants sont de toute évidence prêts à en apprendre de nouveaux. Cela semble donc moins être le cas pour les mots de fonction. Shady et ses collègues ont ensuite *permuté* les mots de fonction au sein des phrases test (typiquement les auxiliaires étaient échangés avec les déterminants). Les enfants de 10 mois et demi n'ont pas réagi non plus à cette manipulation, montrant que s'ils connaissaient un certain nombre de mots de fonction au point de pouvoir s'étonner de leur remplacement par des syllabes sans signification, ils ne savaient pas pour autant quel était leur rôle exact dans la phrase. Cette sensibilité à la position des mots de fonction dans la phrase n'a été montrée qu'à partir de 16 mois (Shady et coll., 1998).

En résumé, ces études montrent que les mots de fonction semblent effectivement être appris sur la base de leurs propriétés phonologiques et distributionnelles distinctives. De ce fait, au moins un certain nombre de ces mots

peuvent être appris très précocement, avant même le premier anniversaire de l'enfant. Cette découverte rend par conséquent assez plausible l'hypothèse selon laquelle un apprentissage précoce des mots de fonction peut servir à initialiser d'autres apprentissages grammaticaux.

1.3.3.2 Noms et verbes

Être capable de classifier les mots de la langue en deux catégories rudimentaires, les mots de contenu et les mots de fonction, est sans aucun doute un grand pas dans l'acquisition du langage. Ce n'est pourtant pas suffisant pour apprendre correctement le sens des mots, ni pour effectuer l'analyse syntaxique des phrases. Dans ces deux domaines, il serait d'un grand secours de pouvoir raffiner un peu les catégories, et notamment au sein des mots de contenu, de savoir distinguer les noms des verbes.

Dans cet esprit, Kelly a abondamment documenté les différences phonologiques qui distinguent les noms des verbes en anglais (cf. Kelly, 1992, et Kelly, 1996, pour un panorama). Il cite notamment :

- 90% des mots bi-syllabiques accentués en première syllabe sont des noms, et 85% de ceux qui sont accentués en deuxième syllabe sont des verbes ;
- Les noms contiennent généralement plus de syllabes que les verbes, et à nombre de syllabes égal, ils ont plus de phonèmes et sont plus longs (en durée) ;
- Les noms tendent à avoir des voyelles plus basses que les verbes ;
- Les noms comportent plus souvent des consonnes nasales que les verbes.

Considérant l'étendue des contrastes phonologiques perceptibles par les nourrissons, que nous avons abondamment décrite dans les sections précédentes, il est à peu près certain qu'au cours de la première année de vie (si ce n'est à la naissance), les nourrissons anglophones doivent être capables de discriminer les noms et les verbes sur la base de leurs propriétés phonologiques. Savoir quand ils commencent à distinguer leurs propriétés grammaticales est une toute autre question, pour laquelle nous ne disposons actuellement que d'une limite supérieure. Cassidy et Kelly (1991) ont en effet testé des enfants américains âgés de 4 ans dans une tâche d'apprentissage explicite de mots nouveaux, et ont montré que ceux-ci étaient plus susceptibles d'associer un mot nouveau à une action s'il comportait une syllabe, et à un objet s'il en comportait trois. Ce résultat montre que ces enfants étaient capables de s'appuyer sur la régularité concernant le nombre de syllabes pour assigner une catégorie grammaticale au mot nouveau. En testant des sujets américains adultes, Kelly et ses collègues ont montré que de manière générale, ceux-ci

utilisaient l'ensemble des indices sus-cités pour déterminer si un mot nouveau était un nom ou un verbe (Kelly, 1988; Cassidy & Kelly, 1991; Kelly & Bock, 1988).

En résumé, ces études montrent que les enfants de 4 ans et les adultes américains sont globalement sensibles aux propriétés phonologiques caractéristiques des noms et des verbes. Faute d'études sur des enfants plus jeunes, on ne peut pas savoir si cette sensibilité émerge suffisamment tôt pour pouvoir jouer un rôle significatif dans l'acquisition des catégories grammaticales. Un reproche que l'on peut également faire aux études mentionnées est d'avoir toujours mis les sujets dans une situation de choix forcé entre nom et verbe. Or si l'on suppose que l'enfant a avant tout accès à la distinction grossière entre mots de contenu et mots de fonction, encore faut-il qu'au sein des mots de contenu il soit également capable de distinguer les noms et les verbes des autres mots de contenu, les adjectifs et les adverbes. S'il s'avérait que les propriétés phonologiques des adjectifs et des adverbes recouvrent celles des noms et des verbes, l'enfant aurait bien du mal à s'y retrouver. Comment peut-il y parvenir? Les adjectifs et les adverbes possèdent-ils des propriétés phonologiques qui les distinguent des noms et des verbes? Ces questions sont actuellement sans réponse.

Enfin, force est de constater qu'un travail similaire à celui de Kelly et ses collègues n'a été fait dans aucune autre langue. On peut d'ores et déjà dire qu'un des principaux indices pour l'anglais, la place de l'accent, est sans effet en français puisque l'accent y est toujours en dernière syllabe. Les autres indices restent à explorer. Ainsi, faute d'être validée dans d'autres langues que l'anglais, on ne peut pour l'instant pas considérer que l'acquisition des catégories noms/verbes a de manière générale une solution phonologique. L'idée est cependant séduisante et mériterait d'être testée dans d'autres langues.

1.3.4 L'organisation syntaxique de la phrase

1.3.4.1 Les constituants syntaxiques

Un grand nombre d'éléments ont déjà été mentionnés qui sont susceptibles de faciliter l'apprentissage de la syntaxe par l'enfant. Comme nous l'avons dit, les théories classiques de l'acquisition de la syntaxe supposent que la parole est déjà segmentée en mots. Le fait que les régularités phonologiques offrent à l'enfant la possibilité de relativement bien segmenter la parole en l'espace de 12 à 18 mois rend tout leur intérêt à ces théories, que nous ne décrirons pas plus en détail ici.

Un autre élément d'un grand secours est la hiérarchie prosodique. Celle-ci permet à l'enfant d'avoir un accès direct à des unités de la parole qui

sont pertinentes du point de vue syntaxique. Le degré d'adéquation entre les unités prosodiques et les constituants syntaxiques est cependant l'objet d'un débat. Un des exemples typiques utilisés pour alimenter le débat est illustré les deux phrases suivantes :

1° [Les enfants] # [mangent] [du pain].

2° [Ils # mangent] [du pain].

Ces deux phrases se composent d'un groupe nominal suivi d'un groupe verbal ; la frontière syntaxique séparant les deux groupes est marquée #. Dans la première phrase, cette frontière syntaxique correspond à une frontière de groupes phonologiques (marqués par les crochets). Dans la seconde, la réduction du groupe nominal à un simple pronom fait qu'il ne constitue pas un groupe phonologique à part entière, mais se rattache à celui du verbe. En conséquence, la frontière syntaxique n'est plus signalée par une frontière prosodique. Gerken et coll. (1994) ont exposé des enfants américains de 9 mois à des phrases anglaises du type 2, en insérant des pauses soit à la frontière syntaxique, soit à la frontière prosodique. Les enfants ont préféré écouter les phrases coupées à la frontière prosodique, montrant ainsi qu'à cet âge ils sont plus sensibles à la cohérence prosodique de la phrase qu'à sa cohérence syntaxique. Faut-il par conséquent abandonner l'idée de l'initialisation prosodique de la syntaxe? Dans les débats récents (Morgan & Demuth, 1996a), la question divise les spécialistes du domaine entre ceux qui pensent que la prosodie n'est pas un bon indicateur de la syntaxe (Venditti, Jun, & Beckman, 1996; Fernald & McRoberts, 1996) et ceux qui pensent que la corrélation prosodie-syntaxe, bien qu'imparfaite, est néanmoins suffisamment fiable en moyenne pour être utile (Steedman, 1996; Fisher & Tokura, 1996; Jusczyk & Kemler-Nelson, 1996; Gerken, 1996).

Malheureusement aucune étude quantitative n'est venue à ce jour offrir des données permettant de réellement juger. Portons néanmoins au crédit des partisans de l'initialisation phonologique trois arguments intéressants :

- S'il est vrai que toutes les frontières syntaxiques ne sont pas nécessairement marquées par une frontière prosodique, à l'inverse *toutes* les frontières d'unités prosodiques (égales ou supérieures au groupe phonologique) signalent des frontières syntaxiques. Il s'agit là d'un lien suffisamment systématique pour ne pas être négligé.
- Le fait que les unités prosodiques ne semblent pas concorder parfaitement avec les unités syntaxiques pourrait tout simplement indiquer que les théories de la syntaxe en usage actuellement ne sont pas adaptées. Steedman (1996) propose à cet effet une théorie alternative de la syntaxe, dans laquelle les unités syntaxiques coïncident réellement avec les unités prosodiques. Il reste néanmoins à évaluer si cette nouvelle théo-

rie est à même de décrire l'ensemble des phénomènes syntaxiques aussi efficacement que les théories classiques.

– L'argument de la multiplicité des indices, qui a déjà prouvé son efficacité dans les discussions sur la segmentation en mots et sur la distinction mots de fonction/mots de contenu, reste valable dans le présent contexte. La prosodie n'est pas le seul indice disponible, et dans les cas où elle fait faux bond, d'autres peuvent lui suppléer, notamment ce que l'enfant peut déjà avoir appris sur les catégories grammaticales. Par exemple, dans le type de phrases (n ° 2) qui suscite le débat, l'enfant peut s'épargner une analyse erronée s'il sait déjà que « ils » est un mot de fonction associé à un groupe verbal (en effet, « ils » n'apparaît pas plus d'une fois dans un groupe intonationnel, cf. Section 1.3.3.1), et a fortiori s'il sait que « mange » est un verbe.

1.3.4.2 L'ordre des mots

Pour terminer, mentionnons le seul cas à notre connaissance pour lequel un lien direct et non ambigu entre la phonologie et la syntaxe ait été avancé. Il s'agit à l'origine d'une proposition de Mazuka (1996), selon laquelle un paramètre important gouvernant l'ordre des mots peut être déterminé à l'aide d'indices prosodiques. Cette proposition a depuis été reprise et développée par Nespor et coll. (1996).

L'ordre des mots est bien évidemment un ingrédient essentiel de la structure syntaxique des phrases et, comme nous l'avons dit à la Section 1.1.1, c'est aussi un élément important pour attribuer leur sens aux mots. Pour reprendre l'exemple que nous avions utilisé p. 4 « le chien poursuit le chat », si l'enfant sait que « poursuit » est le verbe, et que dans sa langue le verbe précède son complément, il peut en déduire que le complément du verbe est « le chat », est donc que l'animal poursuivi (le chat) correspond au son [ləʃa]. Dans la situation inverse, où il n'aurait pas la scène sous les yeux, mais où il connaîtrait le sens des mots, il peut en déduire lequel poursuit lequel. C'est en fait exactement l'expérience qu'ont réalisée Hirsh-Pasek et Golinkoff (1996) sur des enfants de 17 mois, en leur faisant écouter des phrases équivalentes à « le chien poursuit le chat », et en leur présentant simultanément sur des écrans les deux scènes possibles (le chien et le chat échangeant les rôles dans la deuxième scène). Les enfants ont regardé préférentiellement l'écran diffusant la bonne scène, celle qui correspondait à l'ordre des mots en anglais. Notons de plus que l'expérience de Shady et coll. (1998) suggère également que les enfants de 16 mois sont sensibles à la position des mots de fonction. Il est donc important de comprendre comment cette connaissance au moins partielle de l'ordre des mots est acquise dans ces délais.

Le paramètre de direction de la tête (Chomsky, 1981) divise les langues du monde en deux catégories : celles qui ont la tête à gauche (notamment la plupart des langues indo-européennes) et celles qui ont la tête à droite (comme le turc, le japonais). Avoir la tête à gauche signifie que la tête lexicale d'un constituant syntaxique se situe à son début : par exemple, en français, le verbe se situe au début du groupe verbal (comme le verbe « mangent » dans le groupe « mangent du pain » cf. p. 33). Les langues qui ont la tête à droite présentent bien sûr la configuration inverse.

La propriété d'avoir la tête à gauche ou à droite est avant tout une propriété des constituants syntaxiques, tels que groupes nominaux, groupes verbaux, etc. En termes prosodiques, cette propriété se reflète au niveau de l'unité correspondante, le groupe phonologique. Elle se manifeste par le phénomène de proéminence : il s'agit de l'accentuation qui est portée sur un mot du groupe phonologique. Dans les langues ayant la tête à gauche, la proéminence se situe à droite du groupe phonologique (le mot « pain » est le plus accentué au sein du groupe « mangent du pain »). Dans les langues ayant la tête à droite, la proéminence se situe à gauche. Ainsi, il peut suffire à l'enfant de remarquer la proéminence en vigueur dans sa langue pour en déduire l'état du paramètre de direction de la tête (Nespor et coll., 1996).

Nous avons déjà vu (Section 1.3.1.2) que dès la naissance, les nourrissons sont sensibles aux frontières de groupes phonologiques, et qu'au plus tard à 9 mois, ils sont sensibles à leur cohésion. Guasti, Nespor, Christophe, et van Ooyen (sous presse) rapportent de plus une expérience suggérant que des nourrissons de 2 mois sont sensibles à la proéminence. Dans ces expériences, les enfants ont discriminé des phrases françaises de phrases turques, phrases qui avaient été appariées syllabe pour syllabe, et qui ne différaient que par la place de leur proéminence (voir Section 5.1.2 p. 110 pour plus de détails).

Dès 2 mois, les nourrissons ont donc probablement à leur disposition l'information nécessaire à la détermination du paramètre de direction de la tête. Il reste bien sûr à tester s'ils utilisent effectivement cette stratégie. D'autres expériences seront nécessaires dans ce but. La corrélation entre l'indice prosodique et le paramètre syntaxique est néanmoins suffisamment remarquable et universelle pour que cette information ne soit pas totalement négligeable dans l'acquisition du langage.

1.4 À la recherche de nouveaux indices phonologiques

Comme nous l'avons vu, l'initialisation phonologique est à même de fournir une classe d'explications convaincantes, permettant de comprendre comment des propriétés abstraites du langage peuvent être apprises par l'enfant. Ce type de démarche est encore relativement nouveau, et les quelques exemples que nous en avons donné ci-dessus n'épuisent pas, loin sans faut, l'ensemble des apprentissages pour lesquels on peut espérer que l'initialisation phonologique va jouer un rôle. Il est donc important de continuer à étudier les propriétés phonologiques de la parole, leurs liens avec d'autres propriétés plus abstraites, et leur perception par les jeunes enfants.

Parmi les propriétés phonologiques à explorer, le rythme de la parole nous paraît être un bon candidat. En effet, le rythme, de même que par exemple la proéminence, est une propriété qui prend un nombre limité de valeurs : les linguistes ont classé les langues du monde en trois groupes selon leurs propriétés rythmiques. Il est plausible que les langues appartenant à un même groupe partagent également d'autres propriétés, qui restent à déterminer. Notamment, il a été postulé que le type de rythme pourrait être lié à l'unité de segmentation de la parole utilisée par les locuteurs de chaque langue (Cutler, Mehler, Norris, & Segui, 1986; Otake, Hatano, Cutler, & Mehler, 1993; Cutler, 1996)[24]. Par ailleurs certaines études suggèrent que le rythme de la parole est perçu par les enfants, et ce dès la naissance. Ceci se traduit par leur capacité à discriminer des langues, si et seulement si celles-ci ont des types de rythme différents (Mehler et coll., 1988; Nazzi, Bertoncini, & Mehler, 1998). Ces expériences autorisent néanmoins d'autres interprétations, et nous ne pouvons donc à ce stade tenir pour acquis que les nouveau-nés perçoivent effectivement le rythme de la parole.

Le but de la présente thèse va donc être d'étudier le rythme de la parole, et notamment:

- de déterminer dans quelle mesure les nouveau-nés peuvent percevoir certaines propriétés rythmiques de la parole ;
- de préciser les mécanismes et les origines de cette capacité ;
- de comprendre quel est son rôle pour les étapes ultérieures de l'acquisition du langage.

24. Ici, Cutler utilise le mot « segmentation » dans un sens différent de celui que nous avons utilisé jusqu'à présent. Dans ce contexte, il s'agirait d'une procédure de découpage de la parole en unités phonologiques comme le pied pour l'anglais, la syllabe pour le français ou la more pour le japonais, ces unités servant à leur tour dans la segmentation de la parole en mots.

Dans le chapitre 2, nous ferons le point des connaissances sur le rythme de la parole, et nous plaiderons pour une approche particulière du rythme que nous étaierons à l'aide de mesures acoustico-phonétiques effectuées en huit langues. Sur cette base, nous proposerons un modèle computationnel de la perception du rythme de parole, dont nous détaillerons les prédictions à l'aide de simulations sur ordinateur.

Au chapitre 3, nous présenterons une technique, la resynthèse de parole, qui nous permettra de produire les stimuli que nous utiliserons pour tester la perception du rythme.

Au chapitre 4, en nous appuyant sur la technique de resynthèse, nous testerons un certain nombre de prédictions de notre modèle de la perception du rythme, en menant des expériences de discrimination de langues sur des sujets adultes.

Au chapitre 5, nous étudierons en détail la perception du rythme de la parole chez le nouveau-né, à l'aide également d'expériences de discrimination de langues.

Au chapitre 6, nous poserons la question de savoir à quel point la capacité des nouveau-nés à percevoir le rythme est spécifique à l'espèce humaine. Nous tenterons d'y répondre à l'aide d'expériences de discrimination de langues chez le singe.

Enfin, au chapitre 7, nous discuterons l'ensemble des résultats obtenus, et nous émettrons une hypothèse quant au rôle précis du rythme dans l'acquisition du langage.

Chapitre 2

Le rythme de la parole

Platon définissait le rythme comme « l'ordre du mouvement »[1]. En étant plus explicite, on peut dire que le rythme est l'impression créée par des unités qui se répètent dans le temps. C'est en musique que la notion de rythme est la plus évidente : les notes en sont les unités, et celles-ci se répètent à intervalles réguliers ou multiples les uns des autres. De plus, certaines unités sont plus saillantes et plus régulières (comme les coups de grosse caisse par exemple), et ont par conséquent une contribution plus importante au rythme.

Peut-on parler de rythme dans la parole? La notion est déjà moins claire, car les unités de la parole sont moins évidentes, et aucune ne semble se répéter avec la régularité du métronome. Un certain nombre de linguistes ont néanmoins tenté de décrire le rythme de la parole, en se basant tout d'abord sur les impressions subjectives qu'ils avaient en l'écoutant.

2.1 Revue des théories

2.1.1 La théorie de l'isochronie

2.1.1.1 Historique

La première observation rapportée fut que différentes langues peuvent avoir différents types de rythme. Lloyd James (1940) a remarqué que le rythme de l'anglais ou du néerlandais était comparable à celui d'un message en Morse, alors que celui de l'espagnol ou de l'italien évoquait plutôt une mitraillette. Pike (1945) a étendu cette observation à d'autres langues et a introduit la notion d'*isochronie*, en proposant que dans certaines langues, les

1. Dans *Les Lois*, Livre II, passage sur l'éducation.

syllabes se répétaient à intervalles réguliers, alors que dans d'autres c'étaient les accents toniques. Il a par conséquent appelé les premières les *langues syllabiques* et les secondes les *langues accentuelles*. Abercrombie (1967) a de plus postulé que dans toutes les langues du monde le rythme se basait soit sur l'isochronie des syllabes, soit sur celle des accents toniques. De nombreux chercheurs ont depuis ajouté des langues dans les deux classes, et Ladefoged (1975) a même proposé l'existence d'un troisième type de rythme se basant sur la more[2] : en japonais par exemple, le rythme reposerait sur l'isochronie des mores et non des syllabes. Le tableau 2.1 récapitule la répartition dans les trois catégories des langues étudiées à ce jour.

On peut remarquer que les classes rythmiques sont partiellement corrélées aux relations génétiques entre les langues, les langues germaniques étant généralement accentuelles alors que les langues latines sont plutôt syllabiques. La continuité historique s'applique bien sûr au rythme comme aux autres propriétés de la langue, mais il est bien clair que cette notion n'est pas restreinte aux seules langues européennes, et les classes de rythme transcendent en fait largement les grands groupes linguistiques.

Mentionnons enfin que cette hypothèse des classes de rythme a été reprise pour expliquer certains résultats psycholinguistiques (Cutler et coll., 1992; Mehler et coll., 1993; Cutler & Otake, 1994; Cutler, 1996; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998) et la capacité des nourrissons à discriminer des langues (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996; Nazzi et coll., 1998), que nous détaillerons à la Section 5.1.

2.1.1.2 L'isochronie en échec

Puisque les trois types de rythme ont été définis par le fait que certains événements reviennent à intervalles réguliers, il s'agit là d'une hypothèse que l'on peut vérifier en mesurant les intervalles en question dans le signal de parole. De nombreux linguistes se sont attelés à cette tâche, sans réellement parvenir à confirmer l'hypothèse.

Ceux qui se sont intéressés aux langues accentuelles ont montré qu'en anglais, les intervalles séparant les accents toniques n'étaient pas particulièrement réguliers, mais que leur durée était en fait proportionnelle au nombre

2. La more est une unité inférieure à la syllabe. Elle peut être constituée d'une voyelle, d'une consonne suivie d'une voyelle, de l'allongement d'une voyelle, d'une consonne placée après la voyelle, ou encore de la première moitié d'une géminée. Ainsi, les syllabes [too] ou [ton] comptent deux mores : [to] et [-o] ou [-n]. N.B. : les auteurs ultérieurs n'ont pas nécessairement repris cette distinction supplémentaire, classant les langues moraïques parmi les langues syllabiques. Dans la suite, nous mentionnons donc alternativement deux ou trois classes, selon les auteurs cités.

TAB. 2.1 – *Répartition des langues dans les classes de rythme.*

Langue	Référence
Langues accentuelles	
Anglais	Pike (1945)
Arabe	Abercrombie (1967)
Néerlandais et les langues germaniques en général	Ladefoged (1975)
Russe	Abercrombie (1967)
Polonais	Rubach et Booij (1985)
Thaï	Luangthongkum (1977), cité par Dauer (1983)
Langues syllabiques	
Espagnol	Pike (1945)
Français	Abercrombie (1967)
Italien	Bertinetto (1981)
Catalan	Mehler et coll. (1993)
Grec	Arvaniti (1994)
Telegu	Abercrombie (1967)
Yoruba	Abercrombie (1967)
Hindi	O'Connor (1973), cité par Dauer (1983)
Langues moraïques	
Japonais	Ladefoged (1975)
Tamoul	Steever (1987)
Alyawarra, Aranda et Paiute	Haraguchi et Tanaka (1992) cités par Nazzi (1997)

Ce tableau, non exhaustif, renvoie à la plus ancienne référence (à notre connaissance) à chaque langue.

de syllabes qu'ils comportaient (Bolinger, 1965; Lea, 1974; O'Connor, 1965; Shen & Peterson, 1962).

Pour ce qui est des langues syllabiques, la notion d'isochronie des syllabes a été contredite par Wenk et Wiolland (1982), qui ont proposé qu'en français l'unité rythmique était plutôt de l'ordre du groupe phonologique. Borzone de Manrique et Signorini (1983) ont de plus montré que les durées de syllabes n'étaient pas particulièrement régulières en espagnol. Ce résultat a été confirmé par Hoequist (1983), qui a de plus montré que les durées des syllabes en japonais n'étaient pas proportionnelles au nombre de mores qu'elles

comportaient : dans les deux langues, les durées des syllabes découlent assez directement de leur structure.

D'autres chercheurs encore ont étudié plusieurs langues à la fois pour mieux en évaluer les différences. Roach (1982) a choisi le français, le telegu et le yoruba pour les langues syllabiques, et l'arabe, l'anglais et le russe pour les langues accentuelles. Il n'a trouvé de différences entre les deux groupes ni en termes de durées des syllabes, ni en termes d'intervalles entre les accents. Dauer (1983) a également mesuré ces intervalles en anglais d'une part, et en espagnol, italien et grec d'autre part, et a conclu qu'ils n'étaient pas plus réguliers en anglais que dans les langues syllabiques, et que dans les quatre langues leur durée était proportionnelle au nombre de syllabes qu'ils comportaient.

Ainsi, les études empiriques ont infirmé plutôt que confirmé l'hypothèse de l'isochronie. Cet échec ne remet cependant pas nécessairement en cause l'idée selon laquelle les langues peuvent présenter différents types de rythme, dont nous allons voir maintenant qu'il est possible d'en rendre compte indépendamment.

2.1.2 Alternatives

2.1.2.1 Une nouvelle approche du rythme de parole

Alors que pour Pike (1945) et Abercrombie (1967) le rythme était une primitive du langage qui divisait les langues du monde en un nombre limité de catégories, Dasher et Bolinger (1982) ont proposé que le rythme n'était qu'une propriété dérivée d'autres propriétés phonologiques de la langue. Dauer (1983) a étayé ce point de vue en mettant en évidence les propriétés phonologiques typiquement associées aux différents types de rythme, dont les principales sont :

la structure syllabique : les langues accentuelles possèdent une plus grande variété de types de syllabes que les langues syllabiques. Par conséquent, elles ont des syllabes plus complexes[3]. De plus, dans les langues accentuelles l'accent tonique est lié à la structure syllabique, se plaçant le plus souvent sur les syllabes les plus complexes. Dans les langues syllabiques, en revanche, l'accent et la structure syllabique sont indépendants.

3. Cette déduction repose bien sûr sur le fait que l'inventaire des syllabes d'une langue commence toujours par les syllabes les plus simples. Par exemple, une langue qui admet le type de syllabe CCVC admet également les types CVC et CV (les syllabes de type V et VC peuvent éventuellement être des exceptions à cette règle). Cf. Blevins (1995).

la réduction vocalique : dans les langues accentuelles, les syllabes non accentuées comportent généralement un inventaire vocalique réduit, parfois à une seule voyelle, le schwa [ə]. Ces voyelles réduites sont significativement plus courtes ou même parfois totalement absentes.

Ces phénomènes se combinent pour donner l'impression que dans les langues accentuelles, certaines syllabes sont beaucoup plus saillantes que d'autres, puisqu'elles sont plus complexes et accentuées, par rapport à d'autres syllabes plus simples, non accentuées et réduites. Dans les langues syllabiques, l'absence de corrélation entre la structure syllabique, l'accent et la nature des voyelles rend les syllabes plus égales les unes aux autres. On comprend maintenant pourquoi on peut avoir l'impression que le rythme est porté par les seules syllabes accentuées dans les langues accentuelles, alors qu'il serait porté par toutes les syllabes dans les langues syllabiques. Ainsi, cette nouvelle approche du rythme de parole permet à la fois de comprendre les intuitions des linguistes qui ont défini les types de rythme, et les phénomènes phonologiques qui en sont la source.

2.1.2.2 Les langues intermédiaires

Une conséquence de cette nouvelle approche est que les types de rythme ne sont plus des catégories absolues, mais sont en fait dépendants des propriétés phonologiques de chaque langue et de leur corrélation. Il devient alors envisageable qu'une langue possède des propriétés phonologiques telles qu'elle ne tombe plus dans l'une des classes de rythme pré-définies.

C'est ce qu'a suggéré Dauer (1987), en déclarant qu'il n'y a en fait pas de classes de rythme, mais que le rythme de la parole se reflète plutôt par un continuum, sur lequel les langues se placent en fonction de l'ensemble de leurs propriétés phonologiques. Sur ce continuum, une langue donnée pourrait donc être plus ou moins accentuelle ou syllabique, ou encore intermédiaire.

Nespor (1990) a étayé cette hypothèse à l'aide d'exemples précis de langues vraisemblablement intermédiaires, comme le catalan et le polonais. Le catalan est une langue très proche de l'espagnol et qui a une structure syllabique assez similaire. Elle a donc été le plus souvent décrite comme une langue syllabique. Pourtant, elle présente le phénomène de réduction vocalique, qui est caractéristique des langues accentuelles. Le polonais a la configuration opposée : cette langue a des syllabes très complexes, comme les langues accentuelles, mais n'a pas la réduction vocalique. Ces deux langues, présentant des propriétés phonologiques mixtes, sont donc intermédiaires sur le continuum de Dauer (1987). De ce fait, leur position au sein des classes de rythme a été l'objet de débats (Hayes & Puppel, 1985; Mascaró, 1976; Rubach & Booij, 1985; Wheeler, 1979).

Remarquons toutefois que la découverte de langues intermédiaires n'implique pas qu'il faille abandonner les classes de langues au profit d'un continuum uniforme. Il se pourrait simplement qu'il y ait plus de classes que celles qui ont été proposées jusqu'à présent (cf. Auer, 1993 et Levelt & van de Vijver, 1998 pour des propositions concernant 5 classes de langues). Le problème du modèle phonologique proposé par Dauer (1987) est qu'il ne permet pas de trancher entre les deux alternatives.

2.1.2.3 Perspectives

La question est néanmoins d'importance si l'on se rappelle notre motivation pour étudier le rythme de parole. Si les propriétés rythmiques des langues se distribuent en un nombre fini de classes, il est concevable qu'elles puissent servir d'indice pour l'apprentissage d'autres propriétés du langage, celles-ci étant généralement discrètes (au sens mathématique du terme). Si au contraire les propriétés rythmiques s'étalent de manière uniforme, il est plus difficile d'imaginer quelle contribution elles pourraient avoir dans un scénario d'initialisation phonologique.

Ce qui nous manque pour trancher la question, c'est en fait une quantification des propriétés phonologiques qui donnent naissance au rythme. Dans le modèle de Dauer (1987), il est implicite que le rythme est *la somme* des différentes propriétés phonologiques, mais la manière de calculer cette somme n'est pas définie. Par exemple, Dauer ne dit pas si la complexité syllabique et la réduction vocalique contribuent également au rythme, et sinon, dans quelles proportions. En conséquence, il est impossible de déterminer si une langue comme le polonais est plutôt plus accentuelle ou plus syllabique, ou exactement intermédiaire. Il est également impossible de dire si le polonais et le catalan présentent le même rythme, ou bien si l'une est plus accentuelle que l'autre. Il faudra pourtant bien répondre à ces questions si l'on veut évaluer quel type d'information l'enfant polonais ou catalan pourrait tirer de son analyse du rythme. Pour y parvenir, nous allons donc proposer dans la section suivante une implémentation du modèle phonologique du rythme.

2.2 Expérience 1 : Mesures acoustico-phonétiques en huit langues

2.2.1 Justification de la démarche

Dauer (1987) avait déjà remarqué que « ni la syllabe, ni l'accent tonique n'ont une définition phonétique générale, ce qui dès lors rend impossible une

définition purement phonétique du rythme de la parole. Toutes les études empiriques ont nécessité de déterminer à l'avance la place des accents (lorsqu'il y en a) et la définition d'une syllabe dans la langue en question pour pouvoir démarrer » (pp. 447-448, traduction libre).

Dans la mesure où notre but est d'étudier la perception du rythme chez le nouveau-né, et où l'on ne peut supposer que celui-ci possède déjà des connaissances spécifiques à sa langue maternelle, il est clair que nous devons éviter cet écueil et tenter justement d'évaluer les propriétés rythmiques de la parole de la manière la plus universelle possible. Nous partons pour cela de l'hypothèse formulée par Mehler et coll. (1996), selon lesquels la perception du nourrisson est centrée sur les voyelles. En effet, selon eux, « les voyelles portent l'essentiel de l'énergie du signal de parole, durent plus longtemps que la plupart des consonnes, et sont plus stables. Elles portent aussi l'accent et indiquent si une syllabe est forte ou faible » (p. 112, traduction libre). Par ailleurs, il semble que les nouveau-nés soient plus attentifs aux voyelles qu'aux consonnes (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988), et qu'ils soient sensibles au nombre de syllabes (et donc de voyelles) dans un mot, indépendamment de leur structure ou de leur accentuation (Bertoncini, Floccia, Nazzi, & Mehler, 1995; Bijeljac-Babic, Bertoncini, & Mehler, 1993; van Ooyen, Bertoncini, Sansavini, & Mehler, 1997). Mehler et coll. (1996) avancent donc que le nouveau-né perçoit la parole avant tout comme une succession de voyelles de durées et d'énergies variables, entrecoupées de périodes de bruit non analysées (les consonnes). C'est ce qu'ils appellent le modèle TIGRE (Time-Intensity Grid Representation). En nous basant sur cette hypothèse, nous allons donc tenter de montrer que l'on peut donner une définition phonétique du rythme qui puisse reposer uniquement sur une segmentation grossière de la parole en consonnes et voyelles [4].

2.2.2 Méthode

Nous avons sélectionné un ensemble de phrases d'un corpus multilingue créé à l'origine par Thierry Nazzi (1997 ; Nazzi et coll., 1998), et étendu pour les besoins de notre étude à l'espagnol, au catalan et au polonais [5]. Ce corpus se compose de 54 phrases écrites en français, et traduites approximativement dans chacune des langues par l'une des locutrices enregistrées. Il s'agit

4. On pourrait encore arguer que la notion même de consonne et de voyelle n'est pas nécessairement universelle. Nous pensons néanmoins qu'une segmentation de la parole équivalente à la nôtre pourrait être obtenue en re-formulant l'hypothèse du TIGRE en des termes moins spécifiques, par exemple à partir d'une courbe de sonorité universelle.

5. Nous remercions Laura Bosch et Núria Sebastián-gallés de nous avoir fourni les enregistrements espagnols et catalans.

de phrases déclaratives, typiques de bulletins d'informations radiophoniques. Elles ont été lues par 4 locutrices natives de chaque langue dans une cabine insonorisée et digitalisées à 16 kHz[6]. Huit langues ont été choisies au sein de ce corpus: trois langues classifiées comme accentuelles (anglais, néerlandais et polonais), quatre langues dites syllabiques (français, espagnol, catalan et italien) et une langue dite moraïque (japonais). Parmi ces huit langues, deux ont des propriétés rythmiques qui font l'objet de débats et sont jugées comme intermédiaires par Nespor (1990) (catalan et polonais). Dans chaque langue, nous avons choisi 5 phrases par locutrice, soit 20 phrases par langue et 160 phrases en tout, de telle manière à ce qu'à travers les langues les phrases soient appariées en nombre de syllabes et soient de durées similaires en moyenne. Les phrases choisies comprenaient de 15 à 19 syllabes, et duraient environ 3 secondes en moyenne.

Nous avons ensuite effectué une segmentation phonétique des 160 phrases, en identifiant chaque phonème et en l'alignant par rapport au signal de parole à l'aide d'un logiciel de visualisation du signal. L'inventaire phonétique de chaque langue a été utilisé pour identifier les phonèmes des phrases correspondantes. Chaque phonème a ensuite été classé comme voyelle ou consonne. Seules les semi-voyelles ont nécessité un choix spécifique pour le classement : les semi-voyelles pré-vocaliques (comme dans « fier » [f**j**ɛʀ] ou « fillette » [fi**j**ɛt]) ont été classées comme des consonnes ; les semi-voyelles post-vocaliques (comme dans « fille » [fi**j**]) ont été classées comme des voyelles[7].

Enfin, pour rendre compte de l'hypothèse selon laquelle le nouveau-né a accès uniquement à la distinction entre consonnes et voyelles (et silence), et pas à des distinctions au sein de ces catégories, nous définissons comme objets de notre étude :

les intervalles vocaliques : toute séquence ininterrompue de voyelles (du début de la première voyelle de la séquence à la fin de la dernière) ;

les intervalles consonantiques : toute séquence ininterrompue de consonnes ; remarquons que l'on pourrait également les nommer « intervalles inter-vocaliques », pour continuer à ne faire référence qu'aux voyelles[8].

6. Notons de plus que ces phrases ont été lues normalement, c'est-à-dire comme si elles s'adressaient à des adultes, et donc pas sur le mode du « motherese » (ou « mamanais », le ton typiquement utilisé par les parents pour s'adresser aux enfants).

7. Ce choix a été dicté par les considérations suivantes : dans notre corpus, l'immense majorité des semi-voyelles post-vocaliques appartiennent à des diphtongues tombantes, marquées comme des voyelles suivies de semi-voyelles. Ces diphtongues sont présentes essentiellement en anglais et en néerlandais, où elles sont considérées comme des voyelles. Il paraissait donc naturel de les classer comme telles. Cf. note 11 pour une discussion de ce choix a posteriori.

8. A ceci près que les éventuels silences ne comptent ni pour les intervalles vocaliques,

A partir des durées des intervalles vocaliques et consonantiques que nous avons mesurées, nous définissons maintenant trois variables prenant chacune une valeur par phrase :

$\%V$ la proportion des intervalles vocaliques au sein de la phrase, calculée comme la somme des durées des intervalles vocaliques divisée par la durée totale de la phrase[9] ;

ΔV l'écart-type des durées d'intervalles vocaliques au sein de la phrase ;

ΔC l'écart-type des durées d'intervalles consonantiques au sein de la phrase.

2.2.3 Résultats

Le Tableau 2.2 donne le nombre de mesures et les valeurs moyennes de nos trois variables dans chaque langue.

TAB. 2.2 – *Nombre de mesures et valeurs moyennes des trois variables par langue.*

Langues	N_v [b]	N_c [c]	$\%V$ [a] (Écart-type)	ΔV [a] (Écart-type)	ΔC [a] (Écart-type)
Anglais	307	320	40.1 (5.4)	4.64 (1.25)	5.35 (1.63)
Polonais	334	333	41.0 (3.4)	2.51 (0.67)	5.14 (1.18)
Néerlandais	320	329	42.3 (4.2)	4.23 (0.93)	5.33 (1.50)
Français	328	330	43.6 (4.5)	3.78 (1.21)	4.39 (0.74)
Espagnol	320	317	43.8 (4.0)	3.32 (1.00)	4.74 (0.85)
Italien	326	317	45.2 (3.9)	4.00 (1.05)	4.81 (0.89)
Catalan	332	329	45.6 (5.4)	3.68 (1.44)	4.52 (0.86)
Japonais	336	334	53.1 (3.4)	4.02 (0.58)	3.56 (0.74)

[a] Les valeurs sont multipliées par 100 pour faciliter la lecture.
[b] Nombre d'intervalles vocaliques.
[c] Nombre d'intervalles consonantiques.

On peut maintenant visualiser la situation des 8 langues dans un espace à 3 dimensions, où chaque dimension représente une variable. Les Figures 2.1, 2.2 et 2.3 montrent la projection de chaque langue sur les plans ($\%V$,ΔC), ($\%V$,ΔV) et (ΔV,ΔC).

Le plan ($\%V$,ΔC) reflète assez fidèlement à l'image traditionnelle des classes de rythme, en séparant clairement les trois groupes. On peut tester la fiabilité de la séparation en effectuant une analyse de variance (ANOVA)

ni pour les intervalles consonantiques.

9. $\%C = 1 - \%V$ et n'est donc pas une variable indépendante.

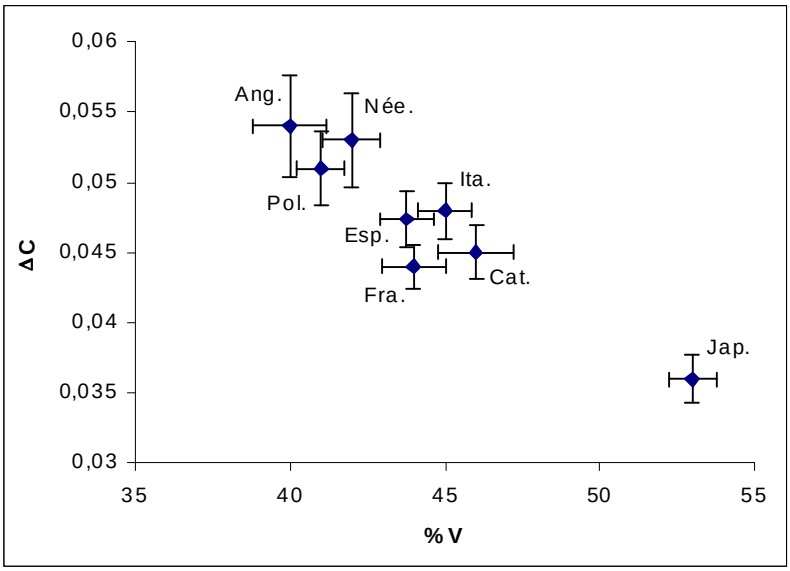

FIG. 2.1 – *Projection des huit langues sur le plan* (%V,ΔC)*. Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

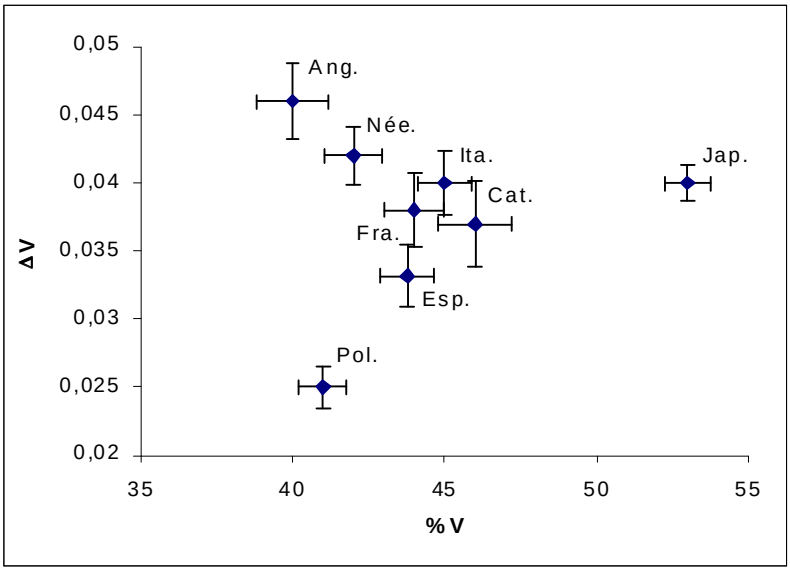

FIG. 2.2 – *Projection des huit langues sur le plan* (%V,ΔV)*. Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

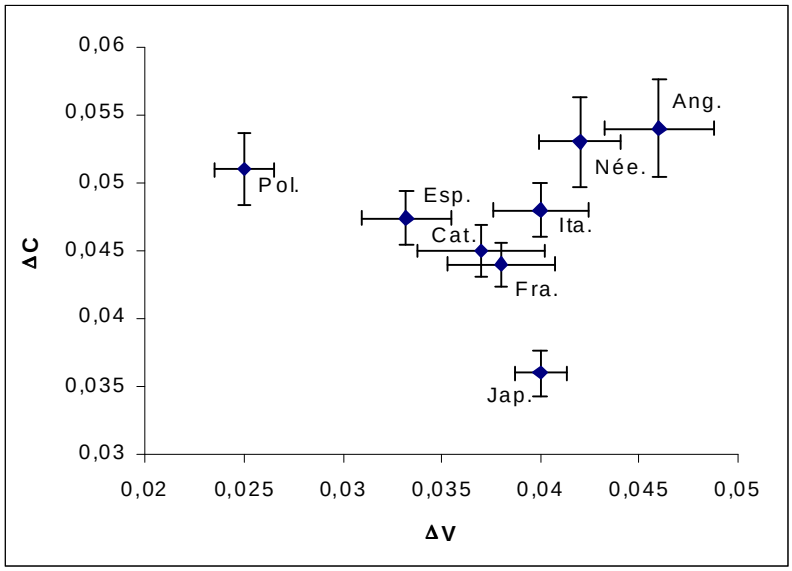

FIG. 2.3 – *Projection des huit langues sur le plan* (ΔV,ΔC). *Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

prenant en compte le facteur « classe de rythme », qui prend trois valeurs : langues accentuelles (anglais, néerlandais, polonais), syllabiques (français, espagnol, italien, catalan) et moraïques (japonais). On trouve un effet significatif de ce facteur à la fois pour $\%V$ et pour ΔC ($p < 0.001$), mais pas pour ΔV. De plus, un test de Tukey (comparaisons multiples) montre que chaque classe est significativement différente des deux autres ($p < 0.001$ pour chaque comparaison avec $\%V$, $p \leq 0.001$ pour ΔC).

Ainsi, les variables $\%V$ et ΔC semblent toutes les deux pouvoir rendre compte de la notion traditionnelle de classe de rythme. On peut d'ailleurs remarquer qu'elles sont fortement corrélées ($r = -0.93$, $p < 0.01$). ΔV, en revanche, n'est pas significativement corrélée avec les deux autres variables, et suggère d'autres différences qui ne sont pas prises en compte dans les classes habituelles : notamment, cette variable met le polonais en exergue par rapport à toutes les autres langues, ce qui est compatible avec l'hypothèse de Nespor (1990). Le catalan, autre langue potentiellement intermédiaire, n'est cependant singularisé par aucune de nos variables.

2.2.4 Discussion

Puisque l'un de nos buts était d'implémenter le modèle phonologique du rythme à l'aide d'une segmentation phonétique, il est important d'évaluer si

les résultats obtenus sont interprétables du point de vue de ce modèle.

C'est clairement le cas en ce qui concerne $\%V$ et ΔC, qui sont directement reliés à la structure syllabique. En effet, les syllabes complexes comportent plus de consonnes que les syllabes simples. Ainsi, plus une langue admet des syllabes complexes, plus la proportion moyenne des consonnes augmente, et donc plus $\%V$ diminue[10]. En même temps, les langues qui ont des syllabes complexes ont également les syllabes plus simples (voir note 3), donc la proportion de consonnes y est plus variable que dans les langues où toutes les syllabes sont simples, et par conséquent ΔC augmente. On comprend maintenant pourquoi $\%V$ et ΔC sont corrélées. En résumé, les langues qui ont une structure syllabique complexe ont un $\%V$ plus faible et un ΔC plus élevé que les langues qui ont une structure syllabique simple, et cette analyse est confirmée par l'adéquation entre la Figure 2.1 et la structure syllabiques des langues en question[11].

En ce qui concerne ΔV, l'interprétation est plus délicate. En effet, plusieurs facteurs sont susceptibles d'influencer simultanément la variabilité des intervalles vocaliques :

la réduction vocalique (anglais, néerlandais, catalan)

l'allongement contrastif[12] des voyelles (japonais)

l'allongement de voyelles dans certains contextes (italien)

l'existence de voyelles intrinsèquement longues (voyelles « tense » et diphtongues en anglais et en néerlandais, voyelles nasales en français)

Les deux premiers facteurs avaient déjà été décrits comme pouvant contribuer au rythme de la parole (Dauer, 1987), mais les deux suivants n'ont jamais été mentionnés. Comme ΔV reflète la somme de ces quatre phénomènes, il n'est pas certain que cette variable soit réellement liée au rythme de parole. De fait, l'axe ΔV semble moins relié aux classes de rythme habituelles. Il reflète néanmoins les propriétés phonologiques sus-mentionnées : les deux langues ayant les plus basses valeurs de ΔV (polonais, espagnol) sont celles qui ne présentent aucun des phénomènes phonologiques en question ; celles qui ont les valeurs les plus élevées (anglais, néerlandais) sont celles qui présentent

10. Dans le même esprit, il a été montré, pour le chinois mandarin et celui de Shanghaï, qu'une langue ayant des syllabes plus complexes a également des syllabes plus longues en moyenne (Duanmu, 1994).

11. On peut maintenant évaluer a posteriori l'effet de la décision que nous avons prises concernant les semi-voyelles. Si les diphtongues tombantes (du type [fij]) avaient été considérées comme des séquences VC plutôt que VV, les valeurs de %V en anglais et en néerlandais s'en seraient trouvées diminuées, ce qui aurait accentué la séparation entre les langues syllabiques et les langues accentuelles. Ainsi, si ce choix a un certain caractère arbitraire, on peut estimer qu'au moins il n'a pas biaisé les résultats *en faveur* de l'hypothèse des classes de rythme.

deux de ces phénomènes. Par rapport aux deux autres variables, ΔV suggère avant tout que le polonais est une langue différente à la fois des langues syllabiques et des langues accentuelles, comme le suggérait Nespor (1990). Il reste néanmoins à vérifier par des expériences de perception du rythme si cette différence phonologique se traduit effectivement par une différence rythmique perceptible. C'est ce que nous ferons au Chapitre 4. Nous allons maintenant essayer de clarifier les prédictions qu'il est possible de faire à partir de nos mesures.

2.3 Un modèle de la perception du rythme, simulations et prédictions

À la section précédente nous avons réalisé une implémentation du modèle phonologique du rythme qui permet de retrouver, à partir d'une segmentation de la parole en consonnes et voyelles, les propriétés phonologiques supposées être à l'origine des différents types de rythme. Nous avons de fait observé que nos mesures semblent pouvoir rendre compte des classes de rythme traditionnelles. Il est donc tentant de supposer que la perception du rythme de parole peut se réduire à la segmentation de la parole en intervalles consonantiques et vocaliques, et au calcul de variables telles que $\%V$ ou ΔC. Pour évaluer un tel modèle, il est nécessaire de le formuler précisément et d'essayer d'en tirer des prédictions testables sur le plan expérimental. Les données expérimentales dont nous disposons sont des expériences de discrimination de langues chez le nouveau-né. L'aptitude du modèle à prédire quelles paires de langues sont discriminées (ou pas) par les nouveau-nés serait un bon argument en sa faveur, ainsi qu'en faveur de l'hypothèse selon laquelle les nouveau-nés discriminent les langues grâce au rythme. On pourrait croire qu'il suffit de regarder la Figure 2.1 pour prédire que deux langues sont distinguables sur la base du rythme si et seulement si elles appartiennent à deux classes rythmiques différentes. On ne peut toutefois pas analyser les mesures de manière aussi directe, car l'issue dépend du recouvrement entre les différentes langues, du nombre de phrases considérées, et de la tâche dans laquelle la discrimination est mesurée. Nous allons donc être conduits à simuler le plus fidèlement possible les expériences de discrimination de langues telles qu'elles sont réalisées.

2.3.1 Simulation 1 : Discrimination de langues par les adultes, basée sur $\%V$

À notre connaissance, aucune expérience passée de discrimination de langues chez des sujets adultes n'a réellement isolé le rythme de la parole, et n'est donc en mesure de fournir des données pertinentes pour notre problématique (voir Section 3.1). La réalisation de telles expériences est néanmoins le but des chapitres 3 et 4. Dans cette section nous allons donc anticiper ces expériences et tenter d'en prédire les résultats.

2.3.1.1 Modèle de la tâche

La tâche utilisée au Chapitre 3 est une tâche de catégorisation de phrases avec entraînement : les sujets écoutent des phrases extraites de deux langues une par une, et doivent deviner de quelle langue il s'agit. Pendant une première phase ils sont entraînés sur un ensemble de 20 phrases, puis dans une seconde phase ils sont testés sur 20 nouvelles phrases, prononcées par des locutrices différentes. (cf. section 3.3.1)

Cette procédure est formellement analogue à une régression logistique (Hosmer & Lemeshow, 1989). Étant donné un certain nombre de couples de valeurs d'une variable numérique V et d'une variable binaire L, cette procédure statistique trouve la valeur seuil de V qui sépare le mieux les valeurs de L. Elle peut être appliquée à une partie des données, et la valeur de V ainsi déterminée peut être utilisée pour prédire les valeurs de L sur une autre partie des données, ce qui revient à simuler les phases d'entraînement et de test.

Pour chaque paire de langues possibles ($L1$,$L2$), nous simulons donc l'expérience de discrimination correspondante à l'aide d'une régression logistique. Les variables L et V prennent une valeur par phrase : la langue de la phrase ($L1$ ou $L2$) et sa valeur de $\%V$. Nous avons choisi $\%V$ plutôt qu'une autre variable car elle a une variance moindre (cf. Tableau 2.2) et est donc susceptible de donner des résultats plus fiables. La procédure est appliquée sur la moitié des phrases seulement, soit 10 phrases par langue, prononcées par 2 locutrices pour chaque langue. La valeur seuil trouvée par la régression est alors utilisée pour classifier l'autre moitié des phrases, prononcées par les deux autres locutrices de chaque langue. Le pourcentage de ces phrases correctement classifiées donne donc le score prédit en phase de test. Nous avons de plus effectué chaque simulation deux fois, en échangeant entre-temps les phrases d'entraînement et celles de test : cette précaution supplémentaire permet d'éviter d'éventuels effets d'asymétrie entres les phrases d'entraînement et de test. Le score rapporté est la moyenne des deux scores obtenus.

TAB. 2.3 – *Simulations des expériences de discrimination de langues chez l'adulte basées sur %V pour les 28 paires de langues.*

	Ang.	Née.	Pol.	Fra.	Ita.	Cat.	Esp.
Née.	57.5						
Pol.	50	57.5					
Fra.	60	60	65				
Ita.	65	62.5	65	55			
Cat.	65	62.5	65	57.5	35		
Esp.	62.5	57.5	62.5	50	50	37.5	
Jap.	92.5	92.5	95*	90	90	87.5	95*

Les scores sont les pourcentages de phrases de test correctement classifiées par les régressions logistiques effectuées sur les phrases d'entraînement. La variable prédictive est %V, et le niveau du hasard est 50%.
** Pour ces paires de langues l'une des deux régressions n'a pas convergé, c'est-à-dire que la solution de la régression n'était pas unique. Cela se produit lorsque la variable prédictive (%V) permet de complètement séparer les phrases des deux langues (100% de classifications correctes dans la phase d'entraînement). Dans ce cas, c'est le score de la régression qui a convergé qui est indiqué dans le tableau.*

2.3.1.2 Résultats

Les simulations correspondant aux 28 paires de langues possibles ont été effectuées, et les scores prédits en phase de test sont rapportés dans le Tableau 2.3. Il s'agit des pourcentages de phrases de test correctement classifiées, le niveau du hasard étant de 50%.

Il apparaît que les scores suivent de très près les classes de rythme, les scores étant presque toujours supérieurs pour les paires de langues appartenant à des classes différentes (scores $\geq 60\%$) que pour celles appartenant à la même classe (scores $< 60\%$). La seule exception à cette tendance est la paire néerlandais/espagnol, pour laquelle le score de discrimination atteint seulement 57.5%. Ces résultats permettent d'évaluer la proportion de phrases qu'il est possible de correctement classifier sur la base de $\%V$. Comme le suggérait la Figure 2.1, le succès de la catégorisation est donc étroitement lié au fait que les deux langues appartiennent ou non à la même classe rythmique. Nous disposons maintenant d'un ensemble de prédictions détaillées que nous testerons en partie au Chapitre 4.

2.3.2 Simulation 2 : Discrimination de langues par les nouveau-nés, basée sur $\%V$

2.3.2.1 Expériences passées

De nombreuses études ont déjà exploré les capacités de discrimination de langues chez les nourrissons, impliquant diverses paires de langues et différents types de stimuli, et des sujets de différents âges. Elles seront passées en revue au Chapitre 5. En attendant, le Tableau 2.4 recense les principaux résultats obtenus chez des nouveau-nés. Nous ne mentionnons pas ceux obtenus chez des nourrissons plus âgés, car d'autres facteurs interviennent alors : à partir de 2 mois, les bébés ne semblent plus discriminer qu'entre leur langue maternelle et une autre langue (Mehler et coll., 1988; Christophe & Morton, 1998). De plus, certaines expériences ont montré que d'autres indices que le rythme sont accessibles à cet âge, comme l'intonation, la phonétique ou la phonotactique (Christophe & Morton, 1998; Guasti et coll., sous presse; Nazzi, Jusczyk, & Johnson, soumis).

TAB. 2.4 – *Revue des expériences de discrimination de langues chez le nouveau-né.*

Paire de langues	Disc.[a]	Stimuli	Référence
Français/russe	Oui	Naturels[b] et filtrés[c]	Mehler et coll. (1988)
Anglais/italien	Oui	Naturels	(Mehler et coll., 1988 ; cf. ré-analyse par Mehler & Christophe, 1995)
Anglais/espagnol	Oui	Naturels	Moon, Cooper, et Fifer (1993)
Anglais/japonais	Oui	Filtrés	Nazzi et coll. (1998)
Anglais/néerlandais	Non	Filtrés	Nazzi et coll. (1998)
Anglais+néerlandais/ espagnol+italien	Oui	Filtrés	Nazzi et coll. (1998)
Anglais+espagnol/ néerlandais+italien ou anglais+italien/ néerlandais+espagnol	Non	Filtrés	Nazzi et coll. (1998)

[a] Cette colonne indique si le résultat obtenu a été interprété comme une discrimination ou non.
[b] Les stimuli utilisés étaient des phrases naturelles et non dégradées.
[c] Les phrases ont subi un filtrage passe-bas à 400 Hz.

Remarquons qu'aucune de ces expériences prises individuellement ne prou-

ve que les nouveau-nés ont discriminé les langues grâce au rythme de la parole. En effet, de nombreux autres indices sont présents dans la parole naturelle qui sont susceptibles d'avoir aidé la discrimination. Même dans la parole filtrée, l'intonation des phrases est bien préservée et constitue un indice majeur. Ainsi, l'hypothèse selon laquelle le rythme est l'indice crucial pour le nouveau-né repose essentiellement sur le fait qu'ils ne semblent discriminer des langues que lorsque celles-ci appartiennent à des classes rythmiques différentes. De plus, dans l'expérience opposant des paires de langues (Nazzi et coll., 1998), ils discriminent les paires si et seulement si elles sont congruentes avec les classes de rythme (anglais+néerlandais contre espagnol+italien) mais pas lorsqu'elles sont mixtes (anglais+espagnol contre néerlandais+italien, ou l'autre combinaison).

Pour s'assurer définitivement que les nouveau-nés discriminent les langues sur la base du rythme, il est donc important d'effectuer de nouvelles expériences en s'assurant qu'aucun autre indice que le rythme n'est disponible (ce sera l'objet des Chapitres 3 et 5). En attendant, puisque nous disposons d'un modèle de la perception du rythme de parole, nous allons examiner s'il est possible de simuler correctement les expériences déjà effectuées, et de prédire les résultats des expériences à venir. Dans la suite, nous considérerons notamment toutes les expériences mentionnées dans le tableau 2.4, à l'exception de la paire français/russe, étant donné que le russe ne fait pas partie de notre corpus.

2.3.2.2 Modèle de la tâche

Les expériences citées au Tableau 2.4 reposent toutes sur des procédures d'habituation/déshabituation (cf. Annexe A.1) : pendant une première phase, l'enfant se familiarise avec l'une des deux langues, c'est-à-dire qu'il s'en forme une représentation [13]. Au changement de langue, l'enfant réagit spontanément au fait que les phrases ne correspondent plus à cette représentation, ce qui se manifeste par ses succions. Un aspect crucial de cette procédure est que le moment du changement de langue n'est pas signalé à l'enfant. Contrairement aux expériences que l'on peut faire sur des adultes, il n'y a donc aucun apprentissage préalable des deux catégories. Pour cette raison, ni les comparaisons usuelles entre groupes de données (qui présupposent les catégories), ni les procédures d'apprentissage supervisé (comme à la section précédente avec la régression logistique) ne peuvent rendre compte de la tâche de l'enfant dans une procédure d'habituation/déshabituation. Nous allons donc devoir

13. Dans le cas où il s'agit de sa langue maternelle, cette représentation existe éventuellement déjà avant le début de l'expérience, mais cela ne change rien à notre propos.

modéliser plus fidèlement cette tâche, afin d'effectuer des simulations comparables aux expériences sur les nouveau-nés.

Nous avons retenu pour ce modèle les hypothèses et caractéristiques suivantes :

- Une expérience se déroule en un certain nombre d'étapes, chaque étape consistant en la présentation d'une phrase. Les étapes sont indexées par n.
- À l'étape n, l'enfant se représente le rythme de la phrase entendue par sa valeur $\%V_n$.
- À l'étape n, l'enfant forme également un prototype P_n des phrases qu'il a entendues au cours de l'expérience. Nous définissons ce prototype comme la moyenne des valeurs de $\%V$ des phrases entendues depuis le début[14] :
$$P_n = \frac{1}{n}\sum_{i=1}^{n} \%V_i$$
- L'enfant a un niveau de stimulation S_n qui est modulé par les événements qui se produisent dans son environnement. Toutes choses égales par ailleurs, nous considérons que durant une expérience les seuls événements à se produire sont l'apparition d'une phrase à chaque étape. Nous définissons la stimulation S_n à l'étape n comme la distance sur l'axe $\%V$ entre la nouvelle phrase et l'ancien prototype, soit :
$$S_n = |\%V_n - P_{n-1}|$$
- Enfin, nous supposons un lien causal entre la stimulation de l'enfant à un instant donné et les succions qui s'ensuivent, tel qu'une augmentation de la stimulation entraîne une augmentation du taux de succion dans certains délais. Cette hypothèse étant faite, nous ne modélisons pas plus avant ce lien, et nous analyserons directement l'évolution de la stimulation de l'enfant.

Pour résumer, nous supposons que l'enfant se représente le rythme des phrases entendues comme des points sur l'axe ΔV. Il est d'autant plus stimulé par une phrase qu'elle est plus éloignée sur cet axe de l'ensemble des phrases déjà entendues. Il réagit donc à la nouveauté des phrases qu'il entend. Les hypothèses ci-dessus ne sont bien sûrs pas arbitraires, mais découlent directement de celles qui sont faites pour analyser et interpréter les taux de succion recueillis dans les expériences.

14. Pour être plus réaliste, on pourrait implémenter une mémoire limitée, le prototype étant calculé par exemple sur les 10 dernières phrases entendues. Ici, comme le nombre total de phrases entendues est faible (10 par phase), ce point ne paraît pas essentiel.

Nous adoptons pour les simulations le même dessin expérimental que pour les expériences, à savoir 40 sujets, divisés en deux groupes : expérimental (changement de langue et de locuteur) et contrôle (changement de locuteur seulement). L'ordre de présentation des langues et des locutrices est contrebalancé à travers les sujets, constituant ainsi 8 sous-groupes, et l'ordre de présentation des phrases est tiré de manière aléatoire. Les sujets appartenant à un même sous-groupe ne diffèrent donc que par l'ordre des phrases entendues, la variabilité individuelle n'étant pas modélisée. Pour chaque sujet, la simulation se déroule comme suit :

1° La phase d'habituation, comprenant 10 phrases prononcées par 2 locutrices dans la langue d'habituation, jouées dans un ordre aléatoire. P_n et S_n sont calculés à chaque étape.

2° Passage automatique de la phase d'habituation à la phase de test.

3° La phase de test, comprenant 10 nouvelles phrases prononcées par 2 nouvelles locutrices dans la langue de test, jouées dans un ordre aléatoire.

Les courbes de stimulation S_n sont alors comparées entre le groupe contrôle et le groupe expérimental.

On peut discuter deux différences entre les présentes simulations et les expériences effectuées sur le nouveau-né. Premièrement, dans les expériences, le passage à la phase de test se fait lorsque les taux de succion de l'enfant ont atteint un certain critère d'habituation. Il s'agit de s'assurer 1° que le changement de phase se produit toujours à un stade comparable de la succion de l'enfant, 2° que l'enfant a la possibilité d'augmenter sa succion à nouveau après le changement. Dans nos simulations, ce critère n'a pas de sens, car il ne concerne que le lien entre la stimulation et la succion, que nous ne modélisons pas. Après les 10 phrases de la phase d'habituation, tous les sujets simulés ont atteint le même état : leur prototype P_{10} est égal à $\frac{1}{10}\sum_{i=1}^{10} \%V_i$, et ne se modifiera pas de manière significative en présentant les mêmes phrases à nouveau. Il est donc logique de terminer la phase d'habituation après la dixième phrase. Deuxièmement, les expériences de discrimination de langues mentionnées au Tableau 2.4 utilisent généralement 40 phrases par langue, au lieu de 20 dans nos simulations. Néanmoins, nous présenterons au Chapitre 5 des expériences effectuées également avec 20 phrases par langue, sans que cela semble avoir d'incidence sur les effets observés. On peut par ailleurs remarquer que l'utilisation d'un plus grand nombre de phrases dans les simulations ne pourrait que rendre les prototypes de chaque langue plus précis, et par conséquent augmenter les chances d'observer un effet de discrimination. L'utilisation de 20 phrases par langue dans nos simulations est donc conservatrice.

2.3.2.3 Résultats

Nous avons effectué les simulations pour les 28 paires de langues de notre corpus. Pour chaque simulation, la discrimination a été évaluée en comparant le niveau de stimulation pendant la phase de test entre les groupes expérimental et contrôle[15]. La variable dépendante est le niveau moyen de stimulation dans la phase de test ($\frac{1}{10}\sum_{n=11}^{20} S_n$) et la variable indépendante est le groupe (expérimental ou contrôle). Pour éviter de faire une hypothèse sur la distribution des niveaux de stimulations, nous avons utilisé pour la comparaison un test non-paramétrique de Mann-Whitney. Le Tableau 2.5 montre les niveaux de significativité de ce test pour toutes les simulations. Ces valeurs sont directement comparables les unes aux autres dans la mesure où elles concernent des données similaires en tous points.

TAB. 2.5 – *Simulations des expériences de discrimination de langues chez le nouveau-né sur 28 paires de langues.*

	Ang.	Née.	Pol.	Fra.	Ita.	Cat.	Esp.
Née.	**0.18**						
Pol.	1	0.84					
Fra.	<0.001	0.02	0.18				
Ita.	**<0.001**	0.006	0.02	0.68*			
Cat.	<0.001	0.01	0.007	0.51	0.04*		
Esp.	**0.006**	0.21	0.04	0.97*	1	0.68*	
Jap.	**<0.001**	**<0.001**	<0.001	<0.001	<0.001	<0.001	<0.001

Les valeurs données représentent la significativité (p) du test de Mann-Whitney pour chaque paire de langue. Les cases en gras sont celles pour lesquelles des données expérimentales sont disponibles. Les cases étoilées sont celles pour lesquelles la stimulation du groupe contrôle était supérieure à celle du groupe expérimental (voir texte).

Les étoiles dans le tableau signifient que pour la paire de langues en question, la stimulation en phase de test du groupe contrôle était supérieure à celle du groupe expérimental. Cette différence est même significative dans le cas de la paire catalan/italien. Cela ne signifie pas, bien sûr, que notre simulation prédit une discrimination dans ce cas. La valeur de p donnée par

15. Comme nous l'avons expliqué à la section précédente, tous les sujets simulés ont atteint le même prototype au moment du changement. Il n'est donc pas nécessaire (ni souhaitable) de prendre en compte le niveau de stimulation avant le changement, par le biais d'une soustraction ou d'une analyse de covariance, comme cela se fait dans l'analyse des taux de succion.

le test de Mann-Whitney (« 2-tail ») correspond à l'hypothèse nulle (groupe expérimental = groupe contrôle), mais nous ne prédisons une discrimination que lorsque l'hypothèse alternative (groupe expérimental >groupe contrôle) est acceptée. Ainsi, pour ces 4 paires, nous ne prédisons aucune discrimination, et les 4 valeurs de p correspondant à notre hypothèse sont en réalité très proches de 1. Pour comprendre comment le groupe contrôle peut avoir une stimulation supérieure au groupe expérimental, il faut se rappeler que le groupe contrôle subit lui aussi un changement : il passe de deux locutrices à deux autres locutrices de la même langue. Dans le cas de la paire catalan/italien, cela signifie que les locutrices étaient très différentes les unes des autres, et le groupe contrôle a de fait discriminé les deux paires de locutrices, alors que le groupe expérimental n'a pas discriminé les deux langues. Ce cas de figure ne peut évidemment se produire que dans le cas de langues très proches, tellement proches que les différences entre certaines locutrices d'une même langue peuvent être supérieures aux différences entre certaines locutrices de langues différentes. C'est donc sans ambiguïté que l'on peut conclure qu'il n'y a pas de discrimination de langues dans les cas étoilés, même pour la paire catalan/italien.

Afin de mieux visualiser les données, nous donnons aussi les courbes de stimulation pour trois simulations représentatives. La figure 2.4 montre la simulation de la paire anglais/japonais, la figure 2.5 la paire anglais/espagnol et la figure 2.6 la paire espagnol/italien, illustrant respectivement des effets de discrimination important ($p < 0.001$), moyen ($p = 0.006$) et nul ($p = 1$).

Chaque figure donne la stimulation moyenne pour tous les sujets des deux groupes à chaque étape de la simulation. Notons que la stimulation n'est pas définie à l'étape 1 ($S_1 = |\%V_1 - P_0|$ et P_0 n'est pas défini), et n'apparaît donc pas sur les figures. Le changement de phase a lieu entre les étapes numérotées 10 et 11.

Les simulations sont en accord avec les résultats des expériences sur les nouveau-nés (en gras dans le Tableau 2.5), et soutiennent globalement l'hypothèse des classes de rythme. En effet, la discrimination est significative ($p < 0.05$) lorsque les langues appartiennent à des langues de classes différentes, mais pas lorsqu'elles appartiennent à la même classe. Deux exceptions notables sont les paires polonais/français et espagnol/néerlandais, pour lesquelles nous ne prédisons pas de discrimination bien qu'elles recouvrent deux classes rythmiques. Aucune expérience de discrimination de langues n'ayant été réalisée avec ces paires, nous ne pouvons dire à l'heure actuelle s'il y a lieu de remettre en question l'hypothèse des classes de rythme, ou bien de croire qu'il s'agit d'effets marginaux dus à la petite taille de notre échantillon de locutrices et de phrases. Il sera nécessaire d'effectuer des mesures sur un corpus plus grand et de nouvelles expériences de discrimination de langues

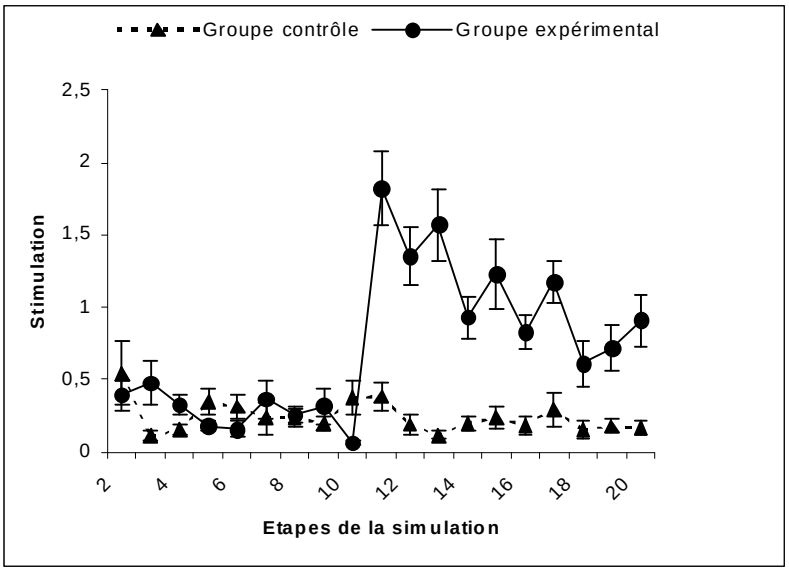

FIG. 2.4 – *Simulation de la discrimination anglais/japonais par des nouveau-nés. Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

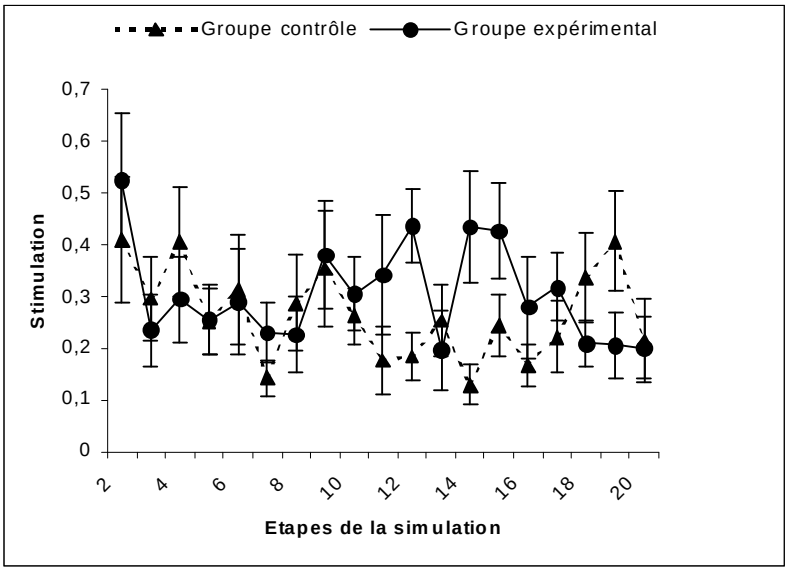

FIG. 2.5 – *Simulation de la discrimination anglais/espagnol par des nouveau-nés. Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

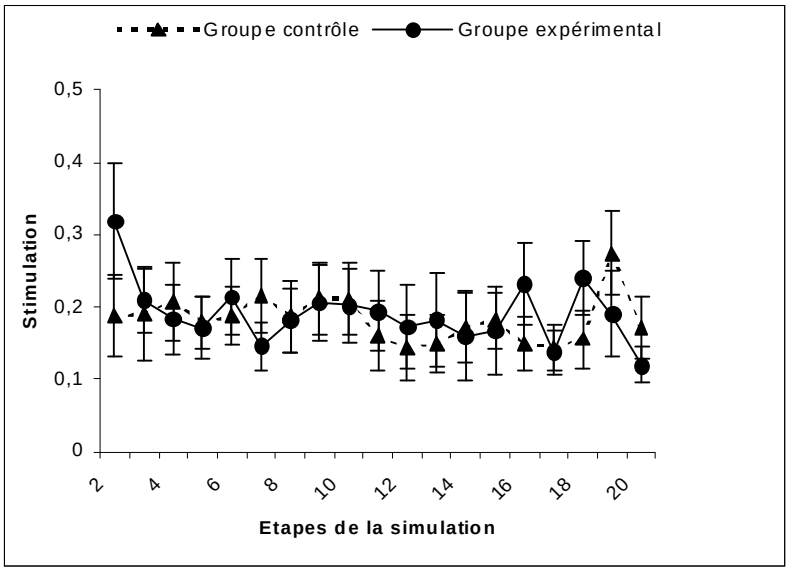

FIG. 2.6 – *Simulation de la discrimination espagnol/italien par des nouveau-nés. Les barres d'erreurs représentent* ± 1 *erreur standard de la moyenne.*

pour le savoir.

2.3.2.4 Discrimination entre groupes de langues

Nazzi et coll. (1998) ont de plus montré que les nouveau-nés sont capables de discriminer des groupes de langues (cf. Tableau 2.4). Nous avons donc également simulé ces expériences, en essayant de suivre le dessin expérimental d'aussi près que possible. Ainsi, nous avons choisi deux locutrices pour chacune des langues (anglais, hollandais, espagnol et italien), et nous avons formé des groupes de phrases (anglais+néerlandais, espagnol+italien, anglais+italien, néerlandais+espagnol, anglais+espagnol, néerlandais+italien). Les sujets du groupe expérimental ont subi le changement d'anglais+néerlandais à espagnol+italien (compatible avec les classes rythmiques), et les sujets du groupe contrôle ont le subi le changement d'anglais+italien à néerlandais+espagnol, ou bien d'anglais+espagnol à néerlandais+italien (incompatible avec les classes rythmiques). L'ordre d'apparition des groupes de langues a été contrebalancé à travers les sujets, et au sein de chaque groupe de langues les phrases ont été tirées pour chaque sujet dans un ordre aléatoire, indépendamment de la langue.

Remarquons que dans ce dessin expérimental il n'y a pas de groupe contrôle à proprement parler, puisque le soi-disant groupe contrôle est exposé à des groupes de langues différents du groupe expérimental. Ainsi, contrai-

rement aux simulations précédentes, il n'y a pas de raisons de croire que les deux groupes auront formé des prototypes identiques à la fin de l'habituation. En évaluant les stimulations des deux groupes pendant la phase de test, il est donc nécessaire de prendre en compte la stimulation pendant l'habituation. Pour ce faire, nous soustrayons à la stimulation moyenne en phase de test la stimulation moyenne en phase d'habituation. La variable dépendante est donc $\frac{1}{9}(\sum_{n=11}^{19} S_n - \sum_{n=2}^{10} S_n)$ [16].

Nous avons simulé 32 sujets, comme dans l'expérience de Nazzi et coll. (1998). En testant l'effet de groupe sur la variable dépendante à l'aide d'un test de Mann-Whitney, nous trouvons que celui-ci est significatif ($p = 0.01$), montrant que le niveau de stimulation augmente plus lors du passage à un groupe de langues appartenant à une classe rythmique différente, que lors du passage entre deux groupes sans cohérence rythmique. Cette expérience est donc également simulée avec succès par notre modèle.

2.4 Conclusions et perspectives

Le rythme de parole peut être considéré sous deux angles différents : celui des linguistes qui ont décrit les différences de rythme entre les langues, et celui des psycholinguistes qui l'ont utilisé pour formuler des hypothèses sur l'acquisition du langage. Si c'est cette deuxième approche qui motive notre recherche, il n'en reste pas moins que la validité de l'hypothèse linguistique est cruciale pour les théories psycholinguistiques qui reposent dessus. De ce point de vue, l'impossibilité de trouver une confirmation expérimentale à l'hypothèse de l'isochronie était un problème sérieux.

En nous appuyant sur un modèle phonologique du rythme élaboré par Dasher et Bolinger (1982) et Dauer (1983, 1987), nous avons montré qu'il était néanmoins possible de rendre compte empiriquement des classes de rythme, en effectuant une segmentation de la parole en consonnes et voyelles et en mesurant les intervalles vocaliques et consonantiques. Cette validation expérimentale des classes de rythme n'est bien sûr pas suffisante dans la mesure où elle repose sur seulement huit langues, choisies parmi celles étudiées par les linguistes qui ont promu l'hypothèse. Il est parfaitement possible qu'en ajoutant plus de langues sur la Figure 2.1, la séparation entre les trois classes se dissolve dans une distribution uniforme. Il se pourrait également que d'autres groupes de langues se forment. Par exemple, Levelt et van de Vijver (1998) ont proposé 5 classes de langues de différentes complexités syllabiques. Trois de ces classes semblent correspondre aux traditionnelles langues accentuelles, syllabiques et moraïques. Sur les deux autres, l'une aurait une complexité

16. Nous ne considérons que 9 points par phase car S_1 n'est pas défini.

syllabique intermédiaire entre les langues syllabiques et moraïques, l'autre aurait des syllabes encore plus simples que les langues moraïques (strictement CV). Puisque nous ne possédons pas les langues en question dans notre corpus, nous ne pouvons évaluer la proposition de 5 classes de langues, mais force est de constater, en regardant la Figure 2.1, qu'il semble y avoir de la place pour une classe supplémentaire entre le catalan et le japonais, et bien sûr aussi au-delà du japonais. Nos données ne nous permettent donc pas de conclure définitivement sur la typologie des types de rythme, mais sont au moins compatibles avec l'hypothèse traditionnelle de trois classes, et laissent ouverte la possibilité qu'il y en ait un plus grand nombre.

Il nous semble que ce travail mériterait d'être poursuivi et approfondi dans plusieurs directions :

- Augmenter la taille du corpus pour chaque langue. Même si les résultats obtenus avec 20 phrases par langue sont déjà convaincants, on aimerait s'assurer qu'il se généralisent effectivement à des échantillons plus représentatifs de chaque langue.
- Augmenter le nombre de langues. Il est bien évident que notre corpus est extrêmement limité dans ce domaine. De grandes familles de langues en sont totalement absentes. Il sera indispensable de les intégrer à nos statistiques pour déterminer de manière définitive s'il y a des classes de rythme, et dans ce cas en quel nombre.
- Étendre le corpus à de la parole spontanée. On peut en effet toujours craindre que la parole lue n'ait pas une prosodie représentative de la langue.
- Étendre le corpus à de la parole adressée à des enfants. Cela permettrait bien sûr de vérifier que nos résultats restent valables dans ce contexte, mais aussi de tester l'hypothèse selon laquelle la séparation entre les classes rythmiques de langues serait plus grande dans le « motherese ». En effet, il a déjà été montré que certains indices prosodiques susceptibles de faciliter l'acquisition du langage sont plus accentués dans ce mode de parole (Kemler Nelson et coll., 1989).
- Développer la segmentation automatique de la parole. Ce serait un moyen d'augmenter radicalement la taille du corpus à moindres frais. De plus, la plausibilité de notre modèle de l'extraction du rythme de la parole serait confortée par la mise en évidence effective d'un algorithme universel de segmentation de la parole en consonnes/voyelles, qui n'est pour l'instant qu'une simple hypothèse.

Nos mesures ont effectivement suggéré un modèle de la perception du rythme de parole qui, aussi simpliste qu'il soit, est à même de simuler les expériences de discrimination de langues à la fois chez l'adulte et chez le

nouveau-né. Ces simulations, basées sur la variable $\%V$, ont abouti aux prédictions correctes dans le cas des expériences déjà effectuées sur le nouveau-né. On peut se demander ce qu'il en serait avec l'une des autres variables que nous avons définies. On peut avancer que les résultats obtenus avec ΔC auraient été assez proches dans la mesure où les deux variables sont fortement corrélées (cf. Figure 2.1), mais ΔC aurait peut-être prédit moins de discriminations, car cette variable a une variance plus élevée que $\%V$ (cf. Tableau 2.2). En ce qui concerne ΔV, on peut penser que cette variable pourrait prédire des discriminations supplémentaires entre le polonais et les autres langues accentuelles. L'opportunité d'inclure cette variable devra en fait se juger sur la base de résultats expérimentaux visant à mieux cerner la manière dont le rythme du polonais est perçu ; la question est donc reportée à la Section4.3.

De manière plus générale, nos simulations fournissent un grand nombre de prédictions qu'il nous appartient maintenant de tester à l'aide de nouvelles expériences sur le nouveau-né et sur l'adulte. Ce sera l'objet des trois prochains chapitres.

Chapitre 3

La resynthèse de parole

Si l'on souhaite tester rigoureusement l'hypothèse selon laquelle il est possible de discriminer des langues sur la base du rythme de parole, et seulement du rythme, il est essentiel de ne fournir aux sujets que les indices rythmiques dans les tâches de discrimination. La présence de tout autre indice susceptible de permettre la discrimination suffit en effet à introduire le doute quant à celui effectivement utilisé par les sujets. Par ailleurs, nous avons fait au précédent chapitre un certain nombre d'hypothèses que nous souhaitons justifier, notamment que le nouveau-né perçoit la parole comme une succession de consonnes et de voyelles, et que l'appréciation du rythme d'une langue peut se réduire à des statistiques sur les intervalles vocaliques et consonantiques. Nous devons donc être en mesure de produire des stimuli adaptés à l'évaluation de ces hypothèses. Une revue des différentes méthodes déjà utilisées pour isoler la prosodie de la parole suffira à nous convaincre qu'il est nécessaire de développer une nouvelle technique.

3.1 Revue des techniques décrites dans la littérature

Bond et Fokes (1991) ont superposé du bruit à la parole dans le but de réduire les indices non-prosodiques. La nature exacte des indices préservés par ce procédé semble cependant difficilement analysable. D'autres chercheurs sont parvenus à ne préserver que l'intonation de la parole, en synthétisant un ton copiant la courbe de fréquence fondamentale de phrases naturelles (de Pijper, 1983; Maidment, 1976, 1983; Willems, 1982). En modulant un ton obtenu de manière similaire par l'enveloppe énergétique des phrases, Ohala et Gilbert (1979) ont produit des stimuli susceptibles de reproduire à la fois le rythme et l'intonation des phrases.

D'autres chercheurs sont parvenus par d'autres méthodes à des stimuli

comparables, isolant plus ou moins bien le rythme et l'intonation. Lehiste et Wang (1977) ont pour cela utilisé une méthode d'inversion spectrale, changeant le signe d'un échantillon sur deux. de Pijper et Sanderman (1994) ont utilisé de la synthèse LPC[1], en normalisant tous les coefficients de manière à donner aux les voyelles un son de schwa et à rendre les consonnes méconnaissables.

Enfin, un certain nombre d'auteurs ont utilisé le filtrage passe-bas de la parole (Schaffer, 1984; den Os, 1988; Mehler et coll., 1988; Dehaene-Lambertz, 1995; Nazzi, 1997; Nazzi et coll., 1998) pour en extraire la prosodie au sens large[2]. C'est notamment cette technique qui a la plus été utilisée dans les expériences sur les nouveau-nés.

Le filtrage passe-bas revient à éliminer toutes les fréquences qui dépassent une certaine fréquence de coupure, généralement 180 ou 400 Hz. L'effet obtenu sur la parole est à peu près le même qu'écouter une conversation à travers un mur. La plupart des phonèmes, et donc des mots, deviennent inintelligibles, mais la prosodie globale de la parole semble préservée. Le problème est qu'il n'est pas possible de déterminer plus précisément quelles propriétés de la parole sont préservées ou éliminées par le filtrage. Celles-ci dépendent de la fréquence de coupure, mais pas de manière uniforme. En effet, les phonèmes diffèrent par la manière dont leur énergie se répartit sur la gamme des fréquences. Par conséquent, le filtrage à une fréquence de coupure donnée ne filtre pas tous les phonèmes de manière égale. Dans des phrases filtrées à 400 Hz, il est d'ailleurs souvent possible de reconnaître des syllabes ou même des mots, mais pas tous. Il semble donc que le filtrage ne soit pas la technique adéquate lorsque l'on souhaite contrôler avec précision les propriétés de la parole que l'on veut isoler. En première approximation seulement, on peut dire que le filtrage à 180 Hz, et dans une moindre mesure à 400 Hz, isole à la fois le rythme et l'intonation de la parole.

den Os (1988) s'est efforcée de dissocier le rythme de l'intonation, en ajoutant une étape de traitement avant le filtrage : elle a tout d'abord resynthétisé ses phrases à l'aide de la méthode LPC, en rendant la fréquence fondamentale constante à 100 Hz. Cette opération a eu pour effet de reproduire les phrases avec une intonation monotone. Elle a ensuite appliqué un filtrage passe-bas à 180 Hz, pour ne finalement préserver que le rythme des phrases. Cette méthode tombe donc sous le coup des critiques que nous venons de faire au filtrage, auxquelles on peut ajouter que la synthèse LPC produit de la parole de qualité assez médiocre. Néanmoins, nous reprendrons

1. Linear Prediction Coefficients.

2. Des variantes ont également été exploitées : filtrage passe-bande (de 70 à 200 Hz) pour Lehiste et Wang (1977), association du filtrage passe-bas et de l'inversion spectrale pour Kreiman (1982).

par la suite l'idée d'éliminer l'intonation en rendant constante la fréquence fondamentale[3].

Enfin, la resynthèse de parole a déjà été utilisée pour dé-lexicaliser des phrases, et n'en conserver que la prosodie. C'est ce qu'a fait Anne Christophe dans une expérience où des sujets devaient discriminer des phrases turques et françaises (voir Guasti et coll., sous presse). Ces phrases ont été resynthétisées en remplaçant toutes les voyelles par [ə] (schwa), toutes les occlusives par [p], les fricatives par [ʃ], les nasales par [m], les liquides par [l] et les semi-voyelles par [j]. Cette transformation préserve la durée des phonèmes, la structure des syllabes, leur rythme, et une partie de la phonotactique. La courbe de fréquence fondamentale a été également reproduite, préservant donc également l'intonation des phrases. Une méthode très similaire a également été utilisée par Pagel, Carbonell, et Laprie (1996), pour évaluer la perception de frontières prosodiques sans donner d'indices sémantiques. Si dans ces expériences les indices préservés dépassent largement le seul rythme de la parole, la technique de resynthèse a néanmoins l'avantage d'être très souple, ce qui va nous permettre de mieux contrôler les propriétés de la parole qui nous intéressent. Nous allons maintenant expliquer en détail en quoi consiste cette technique.

3.2 La resynthèse

La resynthèse de parole peut se définir en termes très généraux. Il s'agit 1° de mesurer un certain nombre de paramètres acoustiques, phonétiques et/ou prosodiques sur des énoncés de parole, et 2° d'utiliser un algorithme de synthèse prenant ces paramètres en entrée pour synthétiser de nouveaux énoncés reproduisant les paramètres des anciens. La nature précise des paramètres que l'on peut mesurer et reproduire dépend bien sûr de l'algorithme de synthèse que l'on utilise.

L'originalité de notre approche réside dans le fait de modifier sélectivement les paramètres mesurés avant de les utiliser pour la synthèse. On peut ainsi entièrement contrôler les paramètres que l'on souhaite éliminer ou préserver. Pour explorer les possibilités de cette démarche, nous avons resynthétisé en plusieurs versions un ensemble de phrases anglaises et japonaises. Ces deux langues ont été choisies parce qu'elles présentent des différences

3. Après avoir obtenu ces phrases purement rythmiques, den Os (1988) les a utilisées dans une tâche de discrimination de langues où les sujets étaient familiers avec l'une des langues et disposaient d'une transcription écrite des phrases. Étant donnée la présence de ces indices supplémentaires, on ne peut donc pas retenir cette expérience comme montrant une discrimination sur la base du rythme seul.

rythmiques maximales, à la fois selon les linguistes et selon nos propres mesures.

3.2.1 Phrases d'origine

Nos phrases sont prises dans le corpus de Thierry Nazzi déjà utilisé au chapitre précédent, comprenant donc des phrases anglaises et japonaises, lues par 4 locutrices par langue. Nous avons sélectionné 5 phrases par locutrice, soit un total de 20 phrases par langue. Entre les deux langues, ces phrases étaient appariées en nombre de syllabes (16.2 en moyenne dans chaque langue) et avaient des fréquences fondamentales moyennes comparables[4] (229 Hz $\pm$15.3 pour l'anglais, 233 Hz $\pm$15.9 pour le japonais, $F(1,38) < 1$). Leurs durées moyennes, en revanche, n'étaient pas rigoureusement égales (2752 ms $\pm$219 pour l'anglais, 2627 ms $\pm$122 pour le japonais, $F(1,38) = 4.98$, $p = 0.03$), mais voir la discussion Section 3.4.2.

3.2.2 Traitement général

Nous avons choisi d'utiliser l'algorithme MBROLA[5] (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996), qui permet de synthétiser par diphones des phrases en en spécifiant les phonèmes, leurs durées et la courbe de fréquence fondamentale pendant le même temps[6]. Par conséquent, ce sont ces mêmes paramètres qu'il faut mesurer sur les phrases d'origine. Chaque phrase est donc resynthétisée en passant par les étapes suivantes :

1° La segmentation phonétique décrite au chapitre précédent, visant à identifier chaque phonème, sa position et sa durée, est effectuée.

2° La fréquence fondamentale est extraite par intervalles de 5 ms, en utilisant le logiciel BLISS, de John Mertus.

3° Les deux types d'information sont alignés dans un fichier texte comprenant, sur chaque ligne, un code pour le phonème à synthétiser, sa durée, et les points de fréquence fondamentale correspondants ; voir l'Annexe B.1 pour un exemple d'un tel fichier.

4. Nous avons défini la fréquence fondamentale moyenne d'une phrase comme la moyenne de tous ses points de fréquence fondamentale non nulle.

5. MBROLA est disponible librement sur Internet à l'adresse suivante : `http://tcts.fpms.ac.be/synthesis/mbrola.html`.

6. Un diphone est une suite de deux phonèmes. La synthèse par diphones consiste à concaténer des diphones préalablement enregistrés. Elle nécessite donc une base de diphones de la langue que l'on souhaite synthétiser. Par exemple, pour synthétiser le mot « bateau » [bato], le programme cherche les diphones [ba], [at] et [to] dans une base du français, les concatène en utilisant certaines procédures pour lisser les transitions, et leur applique la fréquence fondamentale voulue.

4° L'une des transformations détaillées à la section suivante est appliquée au fichier.

5° Le fichier résultant est donné en entrée du programme MBROLA. Pour la synthèse, nous utilisons une base de diphones du français fournie avec le programme, ceci afin de rester neutre par rapport aux deux langues considérées.

3.2.3 Transformations

Nous avons appliqué quatre types de transformations aux phrases, correspondant à différents niveaux d'information que nous souhaitions préserver[7]. Nous avons donné à chaque transformation un nom rappelant les phonèmes qu'elle utilise :

saltanaj **:** Dans cette transformation directement inspirée de celle décrite par Guasti et coll. (sous presse), nous remplaçons chaque phonème par un exemplaire constant de même manière d'articulation. Ainsi, nous remplaçons les fricatives par [s], les voyelles par [a], les liquides par [l], les occlusives par [t], les nasales par [n] et les semi-voyelles par [j]. L'intonation est copiée sur la courbe de F_0 mesurée. La transformation des phonèmes et l'utilisation d'une base de diphones française empêche donc de reconnaître les phonèmes et les mots de la langue. Néanmoins, une partie de la phonotactique est préservée (par exemple la présence ou non de groupes de consonnes), de même que le rythme et l'intonation.

sasasa **:** Ici toutes les consonnes sont transformées en [s] et les voyelles en [a], et la courbe de F_0 est préservée. De ce fait, il n'y a plus aucune information phonotactique dans la phrase, seuls sont préservés le rythme et l'intonation.

sasasa plat **:** C'est la même transformation que précédemment, à ceci près que la courbe de F_0 mesurée n'est pas utilisée, elle est remplacée par une fréquence fondamentale constante à 230 Hz (c'est ce qu'avait fait den Os, 1988 avec la synthèse LPC). Ainsi, l'intonation est éliminée (plus précisément elle est monotone ou plate), donc le rythme reste le seul indice présent.

aaaa **:** En remplaçant tous les phonèmes par [a] et en appliquant la courbe de F_0 mesurée, on obtient un long [a] variant continûment en fréquence fondamentale. Cette transformation reproduit donc la courbe mélodique de la phrase, c'est-à-dire l'intonation, à l'exclusion de toute autre

7. Des exemples de phrases resynthétisées suivant ces différentes transformations sont disponibles sur Internet à l'adresse suivante :
`http://www.ehess.fr/centres/lscp/persons/ramus/resynth/ecoute.htm`.

information.

3.3 Expériences 2a–2d : Discrimination anglais-japonais à l'aide des phrases resynthétisées

Maintenant que nous disposons de stimuli dont les propriétés ont été soigneusement contrôlées, il est important d'évaluer si des sujets sont effectivement capables de percevoir et d'analyser ces propriétés. L'anglais et le japonais présentent de nombreuses différences, notamment du point de vue de la phonotactique et du rythme, qui découlent de la structure des syllabes. Par ailleurs, bien que nous ne connaissions pas d'étude comparée systématique de l'intonation en anglais et en japonais, il est à peu près certain que ces deux langues présentent des types d'intonation différents, ne serait-ce que parce que l'anglais est une langue du type « tête à gauche », alors que le japonais est au contraire « tête à droite » (cf. Section 1.3.4.2). Ces deux langues diffèrent donc par tous les paramètres que nous avons préservés dans nos stimuli ; nous prédisons par conséquent que si notre méthode de resynthèse reproduit suffisamment fidèlement les paramètres en question, il devrait être possible de discriminer les phrases anglaises et japonaises quelle que soit la transformation employée lors de la resynthèse. Nous allons nous en assurer en effectuant, pour chacune des quatre transformations, une expérience de discrimination sur des sujets français adultes.

3.3.1 Méthode

Les quatre expériences seront désignées par le type de stimuli utilisés, respectivement *saltanaj*, *sasasa*, *sasasa plat* et *aaaa*. Dans la suite nous présentons les quatre expériences comme quatre conditions de la même expérience, ne différant que par le type de stimuli. Un même sujet n'a participé qu'à une seule des conditions.

Le protocole expérimental a été programmé sur ordinateur à l'aide du langage EXPE[8] (Pallier, Dupoux, & Jeannin, 1997). Les instructions précisaient aux sujets qu'ils allaient être entraînés à reconnaître des phrases « modifiées acoustiquement » en deux langues, le Sahatu et le Moltec. Les instructions (données en Annexe B.2) étaient écrites de telle manière à faire croire aux sujets qu'il s'agissait de langues réelles inconnues d'eux.

8. Ce programme est disponible librement sur Internet à l'adresse suivante : `http://www.ehess.fr/centres/lscp/expe/expe.html`.

Les phrases ont été divisées en deux ensembles égaux, chacun comportant 10 phrases de chaque langue prononcées par 2 locutrices de chaque langue. L'un d'eux a constitué le corpus d'entraînement et l'autre le corpus de test. Les sujets ont été entraînés à discriminer les phrases sur le corpus d'entraînement, puis testés sur le corpus de test, afin d'évaluer si ce qu'ils avaient appris en phase d'entraînement était spécifique à ces phrases seulement, ou généralisable à d'autres phrases prononcées par d'autres locutrices.

La phase d'entraînement se composait de une à trois sessions d'entraînement, chaque session comportant l'écoute des 20 phrases d'entraînement. Au début de cette phase, une phrase de chaque langue était tirée au sort parmi les phrases d'entraînement, et était jouée à titre d'exemple préliminaire. Ensuite, les 20 phrases d'entraînement étaient jouées une par une dans un ordre aléatoire. Après l'écoute de chaque phrase, le sujet devait deviner la langue en question et répondre en entrant S ou M au clavier (pour Sahatu et Moltec). La bonne réponse était alors immédiatement révélée, permettant ainsi au sujet d'essayer de s'améliorer. Les sujets qui atteignaient 70% de bonnes réponses à la fin d'une session d'entraînement passaient directement à la phase de test[9]. Les autres subissaient une nouvelle session d'entraînement avec les mêmes phrases. Le nombre total de sessions d'entraînement pour un même sujet était limité à 3, après quoi celui-ci passait d'office à la phase de test, indépendamment de son score.

Dans la phase de test, les sujets entendaient les 20 phrases de test une par une dans un ordre aléatoire. Ils répondaient S ou M après chaque phrase et la bonne réponse leur était alors révélée, afin qu'ils ne risquent pas d'intervertir les deux langues. Dans cette phase, chaque phrase n'était entendue qu'une fois par un même sujet.

3.3.2 Participants

Soixante-quatre étudiants ont participé volontairement et gratuitement, dont 50 hommes et 14 femmes. Ils étaient tous locuteurs natifs du français, et leur âge moyen était 22 ans. Ils ont été recrutés en résidence universitaire des grandes écoles parisiennes et testés dans leurs propres chambres à l'aide d'un ordinateur portable. Les 16 premiers sujets ont été testés dans la condition *saltanaj*, les 16 suivants dans la condition *sasasa*, et ainsi de suite *aaaa* et *sasasa plat*.

9. Dans la condition *saltanaj*, ce seuil était de 75%, et a été baissé à 70% dans les conditions suivantes pour permettre à plus de sujets de terminer l'expérience rapidement.

3.3.3 Résultats

Les pourcentages de réponses correctes sont résumés, session par session, dans le Tableau 3.1. Ce tableau montre que le nombre de sujets décroît au cours des 3 sessions d'entraînement dans la mesure où ceux qui atteignent 70% passent directement en phase de test. On observe de plus que les scores avoisinent 60% pour la plupart des sessions, sauf pour l'expérience 2c dans laquelle les scores ne décollent guère de 50%.

TAB. 3.1 – *Exp. 2a–2d : Pourcentages moyens de réponses correctes dans les différentes sessions des 4 conditions. Entre parenthèses : le nombre de sujets.*

	Entr. 1	Entr. 2	Entr. 3	Test
Exp. 2a : *saltanaj*	61.8 (16)	59.6 (14)	61.2 (12)	66.9 (16)
Exp. 2b : *sasasa*	54.2 (16)	63.1 (13)	66.1 (9)	65.0 (16)
Exp. 2c : *aaaa*	49.7 (16)	55.0 (14)	54.1 (11)	50.9 (16)
Exp. 2d : *sasasa plat*	62.5 (16)	55.5 (10)	55.6 (8)	68.1 (16)

Les analyses qui suivent ont été restreintes à la phase de test, pour juger de l'aptitude des sujets à généraliser leurs critères de classification à de nouvelles phrases prononcées par de nouvelles locutrices. Plutôt que d'analyser directement les taux de réponses correctes, nous avons converti les réponses des sujets en taux de détection et de fausses alarmes (au sens de la théorie de la détection des signaux), afin d'avoir une mesure plus fiable de la discrimination qui tienne compte des biais de réponse. Nous avons pris comme taux de détection le pourcentage de phrases japonaises correctement reconnues, et comme taux de fausses alarmes le pourcentage de phrases anglaises erronément classifiées comme japonaises. A partir de ces deux taux nous pouvons calculer les scores de discrimination A' et les mesures de biais de réponse B''_D [10], présentés dans le Tableau 3.2.

Un test de normalité de Kolmogorov permet de s'assurer que les distributions de A' sont normales ($p > 0.36$ pour toutes les conditions). Les scores

10. Notons D le taux de détection et F le taux de fausse alarmes. A' et B''_D sont donnés par :

$$\text{Si } D \geq F, \text{ alors } A' = \frac{1}{2} + \frac{(D-F)(1+D-F)}{4D(1-F)}$$

$$\text{si } D < F, \text{ alors } A' = \frac{1}{2} - \frac{(F-D)(1+F-D)}{4F(1-D)}$$

$$B''_D = \frac{(1-D)(1-F) - DF}{(1-D)(1-F) + DF}$$

Cf. Snodgrass et Corwin (1988) et Donaldson (1992).

TAB. 3.2 – *Exp. 2a–2d : Statistiques par condition. A' est comparé à 0.5 et B''_D à 0.*

	Détection	Fausses alarmes	A'	B''_D
Exp. 2a : *saltanaj*	0.69	0.35	0.71 [a]	−0.11
Exp. 2b : *sasasa*	0.65	0.35	0.68 [b]	−0.02
Exp. 2c : *aaaa*	0.56	0.54	0.52	−0.18 [c]
Exp. 2d : *sasasa plat*	0.70	0.34	0.72 [a]	−0.18

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$.

A' sont comparés à 0.5, qui représente le niveau du hasard, à l'aide d'un test de t. Ils s'avèrent être significativement supérieurs à 0.5 dans les conditions *saltanaj* [$t(15) = 4.47$, $p = 0.0004$], *sasasa* [$t(15) = 3$, $p = 0.009$] et *sasasa plat* [$t(15) = 4.15$, $p = 0.0009$], mais pas dans la condition *aaaa* [$t(15) < 1$].

Pour savoir si les sujets ont des résultats réellement inférieurs dans la condition *aaaa*, nous effectuons des comparaisons multiples entre les diverses conditions en appliquant une correction de Bonferroni. La condition *aaaa* est significativement différente des conditions *saltanaj* ($p = 0.01$) et *sasasa plat* ($p = 0.02$), et marginalement différente de la condition *sasasa* ($p = 0.026 \times 6$ dû à la correction de Bonferroni). Aucune des autres comparaisons n'est significative. En résumé, les sujets ont discriminé les deux langues dans toutes les conditions sauf la condition *aaaa*, dans laquelle ils semblent avoir répondu au hasard. La Figure 3.1 illustre la comparaison des scores A' entre les 4 conditions.

En ce qui concerne l'indice de biais de réponse B''_D, celui-ci n'est significativement différent de 0 que dans la condition *aaaa* ($p = 0.046$), indiquant que les sujets ont répondu significativement plus souvent Sahatu (japonais) que Moltec (anglais). Ceci suggère que lorsque les sujets sont en situation d'échec et n'ont pas de stratégie de réponse particulière, ils ont tendance à adopter une réponse par défaut [11].

11. La raison pour laquelle cette réponse par défaut est plutôt sahatu que moltec est difficile à cerner : il pourrait s'agir d'une préférence pour le côté gauche (S par rapport à M sur le clavier), ou bien l'intonation de l'anglais pourrait être plus facilement assimilable à celle du japonais que le contraire (mais nous ne connaissons aucune donnée dans ce sens).

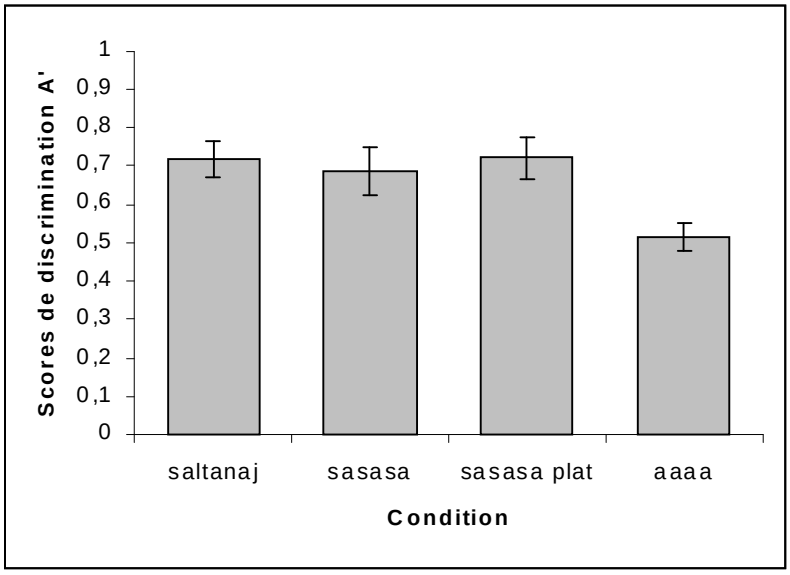

FIG. 3.1 – *Exp 2a–2d : Scores A' moyens par condition. Les barres d'erreur représentent ± 1 erreur standard de la moyenne.*

3.4 Discussion

3.4.1 Rôles respectifs du rythme et de l'intonation – Expérience 2e

Les résultats présentés dans la section précédente suggèrent premièrement que le rythme de la parole est suffisant pour discriminer l'anglais du japonais (condition *sasasa plat*). Ce résultat avait été prédit par les auteurs qui conjecturaient que c'était là la source des discriminations observées chez le nouveau-né (Mehler et coll., 1996; Nazzi et coll., 1998), ainsi que par nos simulations (Section 2.3), mais n'avait jamais été obtenu dans les études antérieures, faute d'avoir correctement isolé le rythme des autres indices potentiels. Deuxièmement, nos résultats suggèrent également que le rythme est nécessaire à la discrimination, dans la mesure où les sujets discriminent les deux langues dans toutes les conditions où le rythme est présent (*saltanaj*, *sasasa* et *sasasa plat*), mais pas dans la seule où le rythme est absent (*aaaa*). On pressent néanmoins que cette conclusion n'est pas généralisable, dans la mesure où nous n'avons éliminé le rythme que dans une condition assez particulière. Il est probable que la discrimination serait possible dans des conditions où de l'information lexicale et/ou phonotactique serait présente sans indice rythmique. Il n'est en tous cas pas question de prétendre que le

rythme est le seul indice valable pour discriminer des langues.

Il reste néanmoins à comprendre pourquoi les sujets ont échoué à discriminer les langues sur la base de l'intonation. La première possibilité est que les intonations de l'anglais et du japonais ne sont pas distinguables. Cela poserait dans ce cas un grave problème à la théorie selon laquelle la proéminence de l'intonation pourrait servir à fixer le paramètre de direction de la tête (Nespor et coll., 1996). Une deuxième possibilité est que l'intonation est bien plus difficile à apprécier, juger et comparer que le rythme, au moins dans une tâche de catégorisation comme la nôtre. Pourtant, d'autres chercheurs ont montré qu'il était possible de discriminer les intonations de deux langues : c'est ce qu'ont fait Maidment (1976, 1983) pour l'anglais et le français, et Willems (1982) et de Pijper (1983) pour l'anglais et le néerlandais. On peut cependant remarquer que dans ces expériences, les sujets étaient toujours locuteurs natifs de l'une des langues, et leur tâche s'apparentait plus à de la reconnaissance de la langue familière qu'à de la catégorisation pure.

Expérience 2e. Afin de savoir s'il s'agit là d'une différence cruciale, nous avons recruté 16 sujets anglophones natifs (4 hommes et 10 femmes, 10 américains, 4 anglais, 1 de Singapour, 1 d'origine inconnue) par petites annonces. Ils ont été payés pour leur participation à cette expérience ainsi qu'à une autre. Leur âge moyen était de 29 ans. Ils ont été testés dans l'expérience *aaaa* dans les mêmes conditions que les sujets français, à ceci près que les instructions leur précisaient que les deux langues étaient l'anglais et le sahatu, et qu'ils devaient différencier les phrases grâce à leur intonation (voir Annexe B.3). Ils ont obtenu un score de discrimination A' moyen de 0.61 (± 0.14), significativement supérieur au hasard [$t(15) = 3.25$, $p = 0.005$]. Ils n'ont pas eu de biais de réponse [$B''_D = 0.09$, $t(15) < 1$]. Par ailleurs la différence entre ce groupe et le groupe correspondant de sujets francophones a été testée à l'aide d'un test de t pour deux échantillons [$t(30) = -1.91$, $p = 0.065$] et d'un test non paramétrique de Mann-Whitney ($p = 0.045$, 2-tail). On peut donc considérer que les deux groupes de sujets se sont comportés de manière significativement différente. Ainsi, il est effectivement possible de distinguer les intonations anglaises et japonaises, mais cela nécessite l'utilisation de connaissances préalables sur l'intonation d'une des deux langues au moins. Ce résultat suggère qu'il est difficile de se représenter l'intonation des phrases, lorsqu'elles ne sont pas dans la langue maternelle. A contrario, les sujets français n'ont eu aucun mal à se représenter le rythme des phrases anglaises et japonaises, alors même que ces langues n'appartiennent pas à la même classe rythmique que le français.

3.4.2 Considérations méthodologiques

Certains détails méthodologiques nécessitent d'être discutés avant de tirer des conclusions plus générales.

Premièrement, nous avions mentionné que les phrases anglaises et japonaises n'étaient pas de durées rigoureusement égales. Les phrases anglaises étant plus longues d'environ 5%, on pourrait arguer que les sujets ont pu effectuer la discrimination en classifiant les phrases les plus longues comme anglaises et les plus courtes comme japonaises. Il serait possible d'effectuer une régression logistique sur les durées pour vérifier l'efficacité d'une telle stratégie, mais cela n'est pas nécessaire grâce à l'échec des sujets dans la condition *aaaa*. En effet, si les sujets avaient effectué la discrimination sur la base de la durée des phrases, ils auraient du faire de même et réussir à discriminer les langues dans cette condition, dans laquelle les phrases sont de mêmes durées que dans les autres conditions. Puisqu'ils ne l'ont pas fait, cette stratégie n'est plausible pour aucune des quatre conditions.

On peut également se demander si la base de diphones et les phonèmes choisis n'auraient pas pu introduire un biais. Cette question ne se pose en fait que pour la condition *saltanaj*, dans la mesure où les stimuli *sasasa* ou *aaaa* ne pourraient guère être différents avec une autre base de diphones. Les six phonèmes français choisis [s,a,l,t,n,j] existent également en anglais et en japonais, et nous avons vérifié que la transposition des phonèmes ne produisait pas de groupes de consonnes illégaux dans la langue concernée. Tous les diphones résultants étaient légaux en français, ce qui conditionnait d'ailleurs une synthèse correcte. On peut éventuellement objecter que dans certains cas la transposition des phonèmes induisait un changement dans la syllabation. Par exemple, le syntagme « the truck » [θətrʌk] est transformé en [satlat]. [tl] est un groupe de consonnes possible en anglais mais seulement à travers une frontière de syllabes (comme dans « butler »), et ceci est vrai en français également. Ainsi, la transformation déplace la frontière de [θə.trʌk] à [sat.lat]. Cela ne pose problème bien sûr que si l'on utilise ces stimuli dans une tâche qui requiert une syllabation précise. Dans notre cas, où la discrimination était possible grâce à des différences phonotactiques et rythmiques massives, il n'y a pas lieu de supposer que cela ait une quelconque influence.

Enfin, on peut être surpris de ce que d'une part, les scores des sujets soient relativement bas (au maximum 68% de réponses correctes, alors que nos simulations en prédisent 92.5%), et d'autre part, que leurs scores ne soient pas d'autant plus élevés qu'il y a d'indices présents dans les stimuli. Dans nos simulations, la régression logistique a une résolution qui n'est limitée que par la précision des valeurs de $\%V$, et trouve de fait le critère optimal de séparation entre les deux langues. Force est de constater qu'au bout de

3 sessions d'entraînement, nos sujets ne le trouvent pas. On peut imputer cela à la difficulté de la tâche, qui nécessite de se concentrer pendant une dizaine de minutes sur des stimuli très appauvris et relativement monotones. De plus, les régularités à découvrir ne sont pas ponctuelles, mais s'étalent sur la durée d'une phrase, ce qui demande sans doute une attention d'autant plus soutenue. On peut penser que dans ces conditions la variabilité des stimuli augmente également la difficulté de la tâche. C'est ce qui pourrait expliquer pourquoi les scores ne sont pas supérieurs dans la condition *sasasa* que dans la condition *sasasa plat* : si le rythme est l'indice le plus fiable, la présence supplémentaire de l'intonation qui, nous l'avons vu, n'est pas bien analysée par nos sujets, ne constitue pas une aide mais peut au contraire les distraire ou les détourner de l'indice principal. En ce qui concerne la condition *saltanaj*, on peut conjecturer que les indices phonotactiques sont en fait redondants avec le rythme [12], et n'apportent donc pas d'information supplémentaire.

La question la plus importante qui reste en suspens est de savoir *comment* les sujets ont effectué la discrimination. Tout d'abord, on peut toujours craindre d'avoir introduit par inadvertance dans les stimuli des indices discriminants qui n'auraient rien à voir avec l'hypothèse testée. Certes, les stimuli *sasasa* préservent rigoureusement les intervalles vocaliques et consonantiques, mais on aimerait être sûr que c'est bien ce genre d'indices que les sujets ont utilisé pour effectuer la tâche. Il est évidemment très difficile d'avoir accès à ce genre d'information. Après l'expérience, il avait été demandé aux sujets de préciser la stratégie et les indices qu'ils avaient utilisés. Malheureusement, la plupart des sujets n'ont pas été capables de verbaliser de manière suffisamment précise ce qu'ils avaient fait, et leurs compte-rendus n'ont globalement pas été informatifs. Nous proposons néanmoins d'accéder à cette information en analysant les erreurs des sujets.

3.4.3 Comparaison de l'expérience et de la simulation

Dans la simulation présentée en section 2.3.1, la régression logistique trouve la valeur seuil de $\%V$ qui sépare au mieux les phrases des deux langues. Bien entendu, les phrases qui, du fait du recouvrement entre les deux langues, se trouvent du « mauvais » côté de la valeur seuil sont mal classifiées par la simulation. Si les sujets humains utilisaient un critère de décision équivalent à $\%V$, ils devraient donc échouer sur ces mêmes phrases. Cette prédiction nous invite à comparer plus en détail les erreurs des sujets et celles de la simulation.

Signalons tout d'abord que les mesures et simulations du chapitre précé-

12. Un groupe de consonnes en *saltanaj* équivaut à un [s] plus long en *sasasa*.

dent avaient été planifiées indépendamment des expériences ci-dessus, et par conséquent ce ne sont pas tout à fait les mêmes phrases qui ont été utilisées. Elles ont toutefois été choisies dans le même corpus, et par chance 26 de ces phrases se trouvent être communes à l'expérience et à la simulation de la discrimination anglais/japonais. Dans la simulation, trois phrases seulement avaient été mal classifiées[13], dont deux utilisées également dans l'expérience. Il apparaît que ces deux phrases (anglaises) sont également les plus mal classifiées par les sujets : respectivement 38% et 43% de réponses correctes. Cela signifie que les sujets les ont plus souvent attribuées au japonais qu'à l'anglais, mimant ainsi la régression logistique basée sur $\%V$. Cette similitude est frappante, mais ne repose bien sûr que sur deux phrases. On peut néanmoins généraliser la démarche.

Bien que dans la simulation, le nombre de phrases mal classifiées soit faible, le critère de décision, lui, est connu pour toutes les phrases : il s'agit de la valeur de $\%V$. Si l'on suppose que les sujets humains catégorisent également les phrases sur leurs valeurs de $\%V$, on doit alors prédire que leurs taux d'erreurs suivent ces valeurs : les phrases dont la valeur de $\%V$ est typique de leur langue devraient être plus aisément classifiées que celles qui ont des valeurs atypiques. Cette prédiction est testée par la Figure 3.2, qui donne, pour chaque phrase de l'expérience, son score moyen de classification correcte par les sujets en fonction de sa valeur de $\%V$.

Cette figure montre tout d'abord la séparation presque parfaite des phrases anglaises et japonaises sur l'axe $\%V$, ce qui explique les 92.5% de réponses correctes dans la simulation. On observe par ailleurs que les phrases anglaises ont des valeurs $\%V$ beaucoup plus étalées que les phrases japonaises. De fait, la structure syllabique du japonais étant plus contrainte, elle n'autorise que peu de variabilité. Mais la constatation la plus remarquable est que les phrases anglaises sont d'autant plus mal classées quelles ont une grande valeur de $\%V$, ce qui signifie que tout se passe comme si les sujets jugeaient effectivement les phrases anglaises sur leur valeur de $\%V$, et utilisaient un seuil de décision moyen d'environ 0.47. Cette observation est confirmée par une régression linéaire ($R = -0.82$, $p < 0.001$). Ces données viennent donc directement à l'appui du modèle de la perception du rythme que nous avons proposé à la Section 2.3.

Cependant, la corrélation entre $\%V$ et les scores de catégorisation ne tient pas du tout dans le cas du japonais ($R = 0.004$, $p = 0.99$). La raison principale en est sans doute le plus faible étalement des valeurs de $\%V$ pour le japonais, qui réduit d'autant la probabilité d'observer une corrélation[14]. Re-

13. Il y avait 92.5% de réponses correctes sur 40 phrases.

14. On pourrait également arguer que si en anglais les variations de $\%V$ reflètent prin-

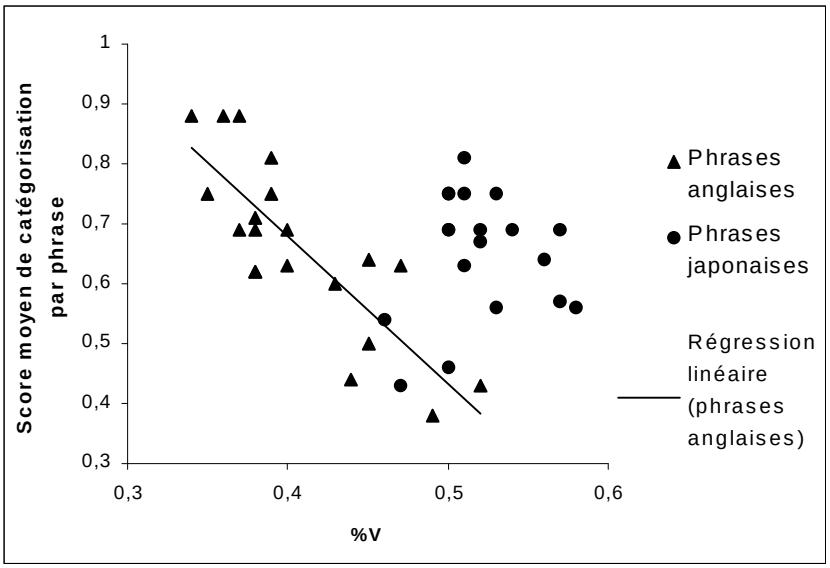

FIG. 3.2 – *Discrimination anglais/japonais par des sujets adultes : pourcentage moyen de réponses correctes par phrase, en fonction de sa valeur de* $\%V$.

marquons toutefois que les trois phrases japonaises les plus mal catégorisées sont justement les trois qui ont la plus faible valeur de $\%V$. Ainsi, même si la corrélation ne révèle rien dans ce cas, les valeurs extrêmes restent compatibles avec notre hypothèse.

3.4.4 Conclusion

Au Chapitre 2, nous avons proposé un modèle de la perception du rythme de parole selon lequel les sujets segmentent la parole en intervalles vocaliques et consonantiques, et effectuent des jugements sur la base des valeurs de $\%V$ des énoncés. Tester ce modèle nécessitait de développer une méthodologie particulière, que nous avons exposée dans ce chapitre : nous avons montré comment utiliser la resynthèse de parole pour dégrader sélectivement certaines propriétés de la parole, et notamment pour produire des stimuli qui n'en conservent que le rythme. Après avoir testé cette technique sur la discrimination de l'anglais et du japonais, il nous semble que les résultats obtenus renforcent la plausibilité de notre modèle. Premièrement, les sujets ont réussi

cipalement des variations de la complexité des syllabes, en japonais elles doivent plutôt refléter des variations de la longueur des voyelles. Or ce phénomène n'étant pas contrastif en français, les sujets français n'y sont pas très sensibles (cf. Dupoux, Kakehi, Hirose, Pallier, & Mehler, sous presse).

à discriminer les deux langues dans la condition *sasasa plat*, qui ne dévoile que les intervalles vocaliques et consonantiques. Il semble donc que les sujets soient capables d'analyser ces intervalles, au moins lorsque la segmentation consonne/voyelle est rendue évidente par la resynthèse. Deuxièmement, la concordance entre les scores de l'expérience et les valeurs de $\%V$ suggère que les sujets analysent ces intervalles en termes de la variable $\%V$, ou en tous cas d'une mesure équivalente.

Ces indications en faveur du modèle ne sont bien sûr pas suffisantes, et autorisent d'autres interprétations (voir discussion page 173). Puisque nous disposons maintenant d'un outil pour étudier la discrimination des langues sur la base du rythme, nous allons donc continuer à tester notre modèle en effectuant de nouvelles expériences sur d'autres paires de langues, dont les résultats seront à comparer avec les prédictions des simulations basées sur le modèle (Section 2.3). C'est l'objet du prochain chapitre.

Chapitre 4

Discrimination de langues par des sujets adultes

Les simulations effectuées au Chapitre 2 ont fourni un grand nombre de prédictions concernant les paires de langues susceptibles d'être discriminées ou pas sur la base du rythme. Grâce à la technique de resynthèse de la parole, nous sommes maintenant en mesure de tester ces prédictions par des expériences, que nous effectuerons dans un premier temps sur des sujets adultes. Comme il n'est pas envisageable de réaliser les expériences correspondant aux 28 paires de langues de notre corpus, nous avons sélectionné certaines paires qui nous paraissaient plus intéressantes à tester que d'autres. Nous avons décidé de nous intéresser plus particulièrement au catalan et au polonais, dans la mesure où ces deux langues ont été désignées par Nespor (1990) comme des langues intermédiaires (voir Section 2.1.2.2). Or, nos mesures et simulations suggèrent que le catalan est en fait en tous points similaires aux langues syllabiques, et que le polonais est effectivement une langue à part, mais à condition seulement de considérer la variable ΔV (variabilité des intervalles vocaliques). Il paraît donc important d'obtenir des données expérimentales permettant 1° de départager les arguments phonologiques de Nespor et les résultats de nos mesures dans le cas du catalan, 2° de déterminer si la variable ΔV a effectivement un rôle à jouer dans un modèle de la perception du rythme. Pour ce faire, nous allons comparer le catalan et le polonais à deux langues de référence, l'anglais pour les langues accentuelles et l'espagnol pour les langues syllabiques [1]. Nous utiliserons pour cela la resynthèse de parole, avec la transformation *sasasa* tout d'abord, puis avec la transformation *sasasa plat*. Mais pour commencer, il nous paraît nécessaire

1. Nous ne considérons pas le japonais car aucune étude, pas plus que nos mesures, n'a jamais suggéré que le catalan et le polonais puissent être des langues moraïques.

d'optimiser un peu la tâche de discrimination.

4.1 Exp. 3a–3b : Optimisation de la tâche – anglais-espagnol en *sasasa*

Comme nous l'avons remarqué au chapitre précédent, les scores obtenus par les sujets dans la catégorisation anglais-japonais ne sont pas extrêmement élevés. Pourtant, si l'on en croit la Figure 2.1 et l'ensemble de la littérature, ces deux langues présentent des différences rythmiques maximales. On peut donc s'inquiéter de la sensibilité de la tâche, maintenant que nous allons tester la discrimination de différences rythmiques a priori plus fines.

On peut reprocher à la tâche de catégorisation que nous avons utilisée qu'elle requiert des sujets qu'ils créent des catégories de toutes pièces, une pour chaque langue. On peut penser qu'une tâche n'impliquant que des comparaisons immédiates entre des phrases serait moins exigeante en termes d'attention et de mémoire. Nous allons donc comparer la discrimination des deux langues de référence, l'anglais et l'espagnol, dans deux tâches différentes : la tâche de catégorisation, et une tâche de détection d'intrus, dans laquelle chaque essai consiste en l'écoute de 3 phrases. Les deux premières (phrases de contexte) sont toujours dans la même langue, et la troisième (phrase de test) est dans une langue indéterminée. Le sujet doit répondre « Pareil » s'il estime que la troisième phrase appartient à la même langue que les deux premières, et « Différent » sinon. Ainsi, cette tâche n'exige que de mémoriser trois phrases à la fois, et pas nécessairement d'extraire des catégories à travers les essais [2]. Suivant la terminologie usuelle, nous désignons la tâche de catégorisation par X et la tâche de détection d'intrus par AAX, où X=A ou B dans les deux cas.

4.1.1 Méthode

Bien que les expériences 2b et 2d n'indiquent pas de discrimination supérieure lorsque l'intonation est jointe au rythme, un réflexe de prudence nous a incité à effectuer une première série d'expériences avec les nouvelles langues en utilisant des stimuli de type *sasasa* plutôt que *sasasa plat*. Nous avons utilisé les 20 phrases anglaises dont la segmentation phonétique a été effectuée au Chapitre 2, et 20 phrases espagnoles appartenant au corpus de

2. Cette tâche est en fait dérivée de la tâche dite *odd ball*, dans laquelle le sujet écoute 4 ou 5 exemplaires de la même catégorie, puis un intrus. La tâche *odd ball* est une des plus sensibles qui soit, mais elle nécessite plus de stimuli que nous n'en possédons. La tâche proposée est donc un compromis entre la tâche *odd ball* et le nombre de nos phrases.

Thierry Nazzi[3]. Elles étaient appariées à travers les langues en nombre de syllabes (de 15 à 19, 17 en moyenne pour chaque langue), et avaient des durées comparables (2852 ms ±207 pour l'anglais, 2928 ms ±214 pour l'espagnol, $F(1,38) < 1$) et des fréquences fondamentales significativement différentes (230 Hz ±16 pour l'anglais, 206 Hz ±17 pour l'espagnol, $F(1,38) = 22$, $p < 0.001$). Elles ont été resynthétisées suivant la transformation *sasasa*, selon la méthode décrite au 3.2. Cependant, pour remédier à l'écart des fréquences fondamentales moyennes entre les deux langues, nous avons multiplié par 1.12 tous les points de F_0 des phrases espagnoles, de manière à ce qu'elle aient une fréquence fondamentale moyenne de 230 Hz. Cette transformation, qui constitue une homothétie sur le contour intonatif de chaque phrase, en préserve la mélodie mais la joue à la hauteur tonale voulue. De ce fait il n'est pas possible de distinguer les phrases anglaises des phrases espagnoles grâce à un jugement de hauteur. Ces phrases ont été utilisées dans deux expériences de discrimination, ne différant que par la tâche utilisée.

L'expérience 3a reprend la tâche X dans les mêmes conditions que les expériences 2a–2d (Section 3.3). La seule modification apportée au protocole est la suivante : pendant les sessions d'entraînement uniquement, à chaque essai, après que le sujet a donné une réponse et obtenu le résultat, le programme lui donne la possibilité de réécouter la phrase s'il le désire. Cette possibilité a été ajoutée dans le but de permettre aux sujets de tirer un meilleur enseignement de leurs erreurs. Les instructions données aux sujets sont en Annexe B.4.

L'expérience 3b utilise la tâche AAX qui se déroule comme suit. Les phrases ont été divisées une fois pour toutes en deux ensembles égaux, comprenant chacun 10 phrases par langue et 2 locutrices par langue. Le premier ensemble a fourni les phrases de contexte (AA dans AAX), tandis que le second a fourni les phrases de test (X dans AAX)[4]. L'expérience se déroulait en deux blocs de 20 essais chacun, pendant lesquels la langue de contexte était maintenue constante. Le sujet était prévenu au début de chaque bloc que les phrases de contexte appartenaient toujours à la même langue (Sahatu ou Moltec). L'ordre des blocs était tiré au hasard pour chaque sujet. De plus, à chaque essai, les deux phrases de contexte étaient tirées au sort avec la contrainte de toujours être prononcées par deux locutrices différentes, et la phrase de test était bien sûr prononcée par une troisième locutrice. L'ordre des phrases de test était aléatoire, et une même phrase de test n'apparaissait qu'une fois par bloc. L'Annexe B.6 donne une séquence type de phrases entendues par un sujet au cours de l'expérience 3b. Après chaque essai de trois

3. Voir note 7 p. 91 concernant ces phrases.

4. Ces deux ensembles sont en fait identiques respectivement à l'ensemble des phrases d'entraînement et à l'ensemble des phrases de test de l'expérience 3a.

phrases, le sujet répondait au clavier P (pareil) si les trois phrases lui semblaient appartenir à la même langue, et D (différent) sinon. Les instructions données aux sujets sont en Annexe B.5.

4.1.2 Participants

Seize étudiants ont participé à l'expérience 3a, et seize autres à l'expérience 3b. Il s'agissait de 22 hommes et 10 femmes, âgés en moyenne de 21 ans. Ils ont été recrutés et testés suivant les mêmes modalités que pour les expériences 2a–2d.

4.1.3 Résultats

Comme pour les expériences 2a–2d, les pourcentages de réponses correctes ont été convertis en scores de discrimination A' (voir page 72). Pour l'expérience 3a, seule la phase de test a été considérée, le taux de détection correspondant au pourcentage de phrases espagnoles correctement catégorisées, et le taux de fausses alarmes correspondant au pourcentage de phrases anglaises incorrectement attribuées à l'espagnol. Pour l'expérience 3b, le taux de détection est le pourcentage d'essais « Pareil » corrects, et le taux de fausses alarmes est le pourcentage d'essais « Différent » erronés. Le Tableau 4.1 donne ces valeurs pour les deux expériences.

TAB. 4.1 – *Discrimination anglais-espagnol en* sasasa– *Comparaison des tâches X et AAX.*

	Réponses correctes	Détection	Fausses alarmes	A'	B''_D
Exp. 3a : tâche X	0.58 ±0.14	0.56	0.41	0.62 ±0.19	0.06
Exp. 3b : tâche AAX	0.56 ±0.06	0.55	0.42	0.61 ±0.09	0.04

La distribution des scores de A' peut être considérée comme normale ($p > 0.20$ pour les deux expériences par un test de Lilliefors[5]). Les valeurs de A' sont donc comparées à 0.5 à l'aide d'un test de t. La différence est significative aussi bien pour l'expérience 3a [$t(15) = 2.52$, $p = 0.024$] que pour l'expérience 3b [$t(15) = 4.45$, $p < 0.001$]. La différence importante de significativité entre les deux expériences, alors que les données sont en tous points similaires, suggère que la tâche AAX procure une puissance statistique supérieure à la tâche X. Pourtant, les scores de discrimination sont extrêmement voisins dans

5. Il s'agit d'une variante du test de normalité de Kolmogorov-Smirnov, qui ne suppose pas de connaître à l'avance la moyenne et la variance de la distribution.

les deux expériences. Les écarts-types rapportés au Tableau 4.1 montrent en fait que la différence vient de la variabilité des scores : la tâche AAX produit des scores beaucoup plus homogènes que la tâche X. Cette tendance, bien que non significative, est confirmée par un test de *Levene* d'égalité des variances ($p = 0.12$). Notons enfin que les sujets n'ont pas eu de biais de réponse [$t(15) < 1$ dans les deux expériences].

4.1.4 Conclusion

Ces deux expériences montrent tout d'abord que, conformément à nos prédictions, l'anglais et l'espagnol sont distinguables sur la base de leur rythme et de leur intonation. Il faudra bien sûr réitérer ce résultat avec des stimuli de type *sasasa plat* pour cerner complètement le rôle du rythme seul.

Par ailleurs, le fait que la tâche AAX semble donner des résultats plus homogènes et donc plus significatifs que la tâche X, nous incite à préférer cette première tâche, dans la mesure où les prochaines expériences vont concerner des langues dont les différences rythmiques risquent d'être moindres. Bien évidemment, la différence de variance entre les deux expériences n'étant pas significative, il est possible qu'elle soit due au hasard ou à un facteur non contrôlé. Néanmoins, comme nous n'avons pas d'autre argument qui nous fasse préférer une tâche plutôt que l'autre, celui-ci nous suffit pour décider d'adopter la tâche AAX pour les expériences suivantes.

4.2 Exp. 4a–4b : anglais-polonais et espagnol-polonais en *sasasa*

Maintenant que notre choix s'est porté sur la tâche AAX, nous pouvons aborder les discriminations qui nous intéressent plus particulièrement, à savoir la comparaison entre le polonais, le catalan et les deux langues de référence. Ici, nous commençons par le cas du polonais, en utilisant dans un premier temps la transformation *sasasa*.

4.2.1 Méthode

Nous avons utilisé les mêmes phrases anglaises et espagnoles resynthétisées que pour les expériences 3a–3b. Les 20 phrases polonaises ont été tirées du corpus segmenté au Chapitre 2. Elles étaient appariées en nombre de syllabes avec les phrases anglaises et espagnoles, et avaient une durée moyenne de 3034 ms ±330 et une fréquence fondamentale moyenne de 219 Hz ±22.

Pour ajuster ces valeurs moyennes au niveau des phrases anglaises et espagnoles, nous avons multiplié les durées et les fréquences fondamentales des phrases par certains coefficients. Comme certaines locutrices polonaises étaient très différentes des autres en termes de durée ou de F_0, nous avons appliqué des coefficients différents pour chaque locutrice : nous avons multiplié les durées des phrases des 4 locutrices polonaises respectivement par 1, 0.92, 0.9, et 0.93, et leurs points de F_0 par 1.07, 1.09, 1.17 et 0.91. De cette manière les phrases resynthétisées avaient une durée moyenne de 2839 ms, et une fréquence fondamentale moyenne de 231 Hz, sans différence majeure entre les locutrices ni avec les autres langues. L'expérience 4a a testé la discrimination entre l'anglais et le polonais, et l'expérience 4b la discrimination entre l'espagnol et le polonais. Les expériences se sont déroulées suivant le même protocole et dans les mêmes conditions que l'expérience 3b, en adoptant notamment la tâche AAX.

4.2.2 Participants

Seize étudiants ont participé à l'expérience 4a, et seize autres à l'expérience 4b. Il s'agissait de 23 hommes et 9 femmes, âgés de 22 ans en moyenne. Ils ont été recrutés et testés suivant les mêmes modalités que pour les expériences 2a–2d.

4.2.3 Résultats

Le Tableau 4.2 donne les scores de discrimination et les indices de biais de réponses pour les deux expériences, et reprend les valeurs de l'expérience 3b pour comparaison. Les sujets semblent n'avoir discriminé que les paires anglais/espagnol et espagnol/polonais.

TAB. 4.2 – *Discrimination anglais-espagnol, anglais-polonais et espagnol-polonais en* sasasa.

	A'	B''_D
Exp. 3b : anglais/espagnol	0.61 ±0.09	0.04
Exp. 4a : anglais/polonais	0.55 ±0.18	-0.13
Exp. 4b : espagnol/polonais	0.65 ±0.15	-0.037

La Figure 4.1 donne les histogrammes des scores A' pour chacune des trois expériences. Notons que les intervalles sont du type $]n,n+0.05]$; la ligne verticale continue sépare donc les scores ≤ 0.5 des scores > 0.5. Les lignes

pointillées délimitent la zone des scores qui, considérés individuellement, ne sont pas significativement différents du hasard[6].

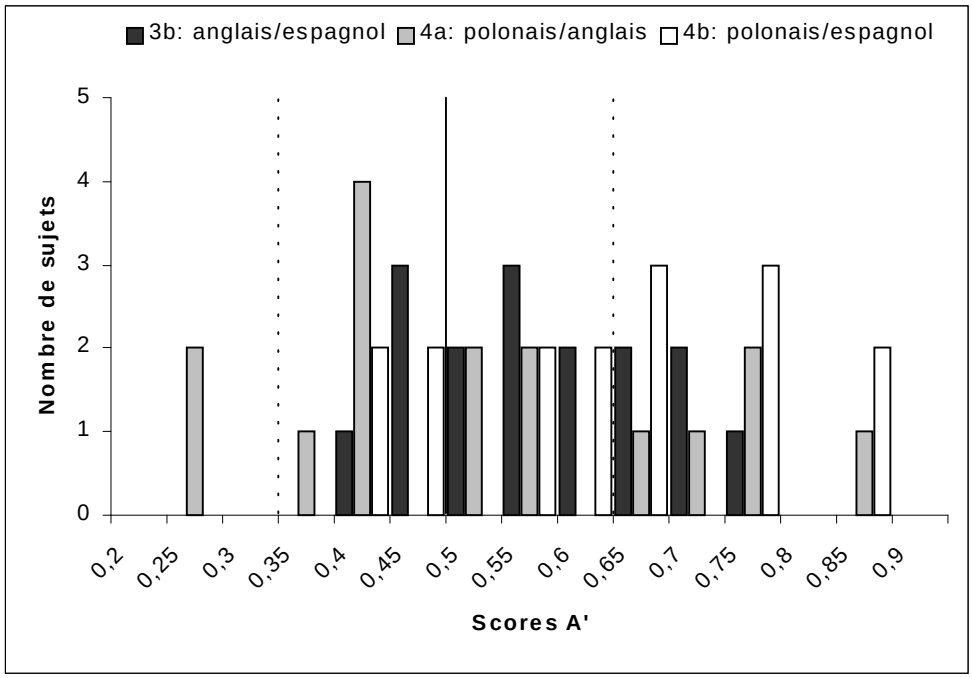

FIG. 4.1 – *Histogramme des scores A' pour les expériences 3b, 4a et 4b.*

La distribution des scores A' peut-être considérée comme normale pour les expériences 4a et 4b ($p > 0.20$ par le test de Lilliefors). Pour l'expérience 4a, A' n'est pas significativement supérieur à 0.5 [$t(15) = 1.183$, $p = 0.25$]. C'est le cas en revanche pour l'expérience 4b [$t(15) = 4.142$, $p = 0.001$]. Dans aucune des deux expériences les sujets n'ont eu de biais de réponse significatif (respectivement $p = 0.19$ et $p = 0.7$). Ainsi, les sujets ont discriminé l'espagnol aussi bien du polonais que de l'anglais. Il n'ont cependant pas discriminé l'anglais du polonais.

En regardant la Figure 4.1 de plus près, on peut observer que les distributions des scores A' sont peut-être bimodales, se composant d'une distribution normale centrée autour de 0.5 et d'une autre distribution centrée sur des scores supérieurs ($\simeq 0.7$). Cette observation est compatible avec la constatation selon laquelle certains sujets comprennent très facilement la nature de la tâche et des indices à trouver, alors que d'autres n'y parviennent jamais. De ce point de vue, l'expérience anglais/polonais n'est pas radicalement dif-

6. La limite est 66.7% (et 33.3%) pour un test binômial bi-directionnel avec 40 observations. Nous avons placé les lignes à 0.35 et 0.65.

férente des autres : les « bons » sujets atteignent des scores tout aussi élevés que dans les deux autres expériences ; ils sont simplement moins nombreux. Ceci suggère que la discrimination anglais/polonais est possible, mais plus difficile, et par conséquent moins de sujets y parviennent.

Néanmoins le nombre de sujets est présentement trop faible pour pouvoir en être sûr : le test de Lilliefors ci-dessus n'a pas rejeté l'hypothèse de normalité. Nous y reviendrons à la Section 4.5.4. En attendant, nos résultats laissent supposer que le polonais s'apparente à une langue accentuelle, mais il est nécessaire de refaire ces expériences avec des stimuli de type *sasasa plat* pour s'assurer que les sujets n'utilisent que des indices rythmiques.

4.3 Exp. 5a–5c : anglais-espagnol, anglais-polonais et espagnol-polonais en *sasasa plat*

Nous allons maintenant à nouveau comparer le polonais à l'anglais et à l'espagnol, en utilisant cette fois la transformation *sasasa plat*, pour s'assurer que seuls les indices rythmiques sont utilisés par les sujets.

4.3.1 Méthode

Les mêmes phrases anglaises, espagnoles et polonaises qu'aux expériences 3b, 4a et 4b ont été utilisées. Elles ont été resynthétisées en utilisant la transformation *sasasa plat* décrite au 3.2. Les mêmes corrections de durée ont été appliquées aux phrases polonaises que pour les expériences 4a et 4b. Les corrections de fréquence fondamentale n'ont pas été appliquées dans la mesure où celle-ci a été rendue constante à 230 Hz pour toutes les phrases. Les expériences se sont déroulées suivant le même protocole et dans les mêmes conditions que les expériences 3b et 4a–4b.

4.3.2 Participants

Quarante-huit sujets, seize par expérience, ont été recrutés et testés suivant les mêmes modalités que lors des expériences précédentes. Il s'agissait de 34 hommes et 14 femmes, âgés de 22 ans en moyenne.

4.3.3 Résultats

Le Tableau 4.3 donne les scores de discrimination et les indices de biais de réponse pour chacune des trois expériences, et la Figure 4.2 les histogrammes

des scores A'. Ici, les sujets ont sans ambiguïté discriminé les trois paires de langues.

TAB. 4.3 – *Discrimination anglais-espagnol, anglais-polonais et espagnol-polonais en* sasasa plat.

	A'	B''_D
Exp. 5a : anglais/espagnol	0.65 ±0.13	0.02
Exp. 5b : anglais/polonais	0.61 ±0.14	-0.08
Exp. 5c : espagnol/polonais	0.74 ±0.08	-0.23

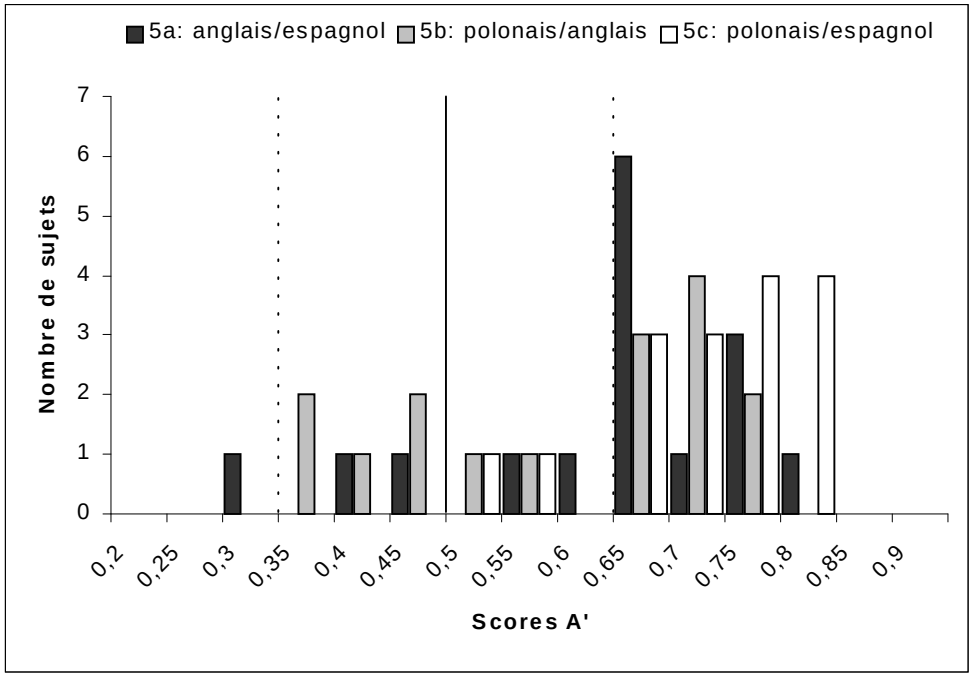

FIG. 4.2 – *Histogramme des scores A' pour les expériences 5a–5c.*

La distribution des scores A' de l'expérience 5a ne peut pas être considérée comme normale, et celle de l'expérience 5b est vraiment à la limite (respectivement $p = 0.01$ et $p = 0.056$), ce qui confirme nos soupçons de bimodalité de la distribution (p. 87). Pour ces deux expériences, nous procédons donc à un test binômial prenant en compte le nombre de sujets dont les scores sont ≤ 50 et ceux dont les scores sont > 50. Celui-ci est significatif pour l'expérience 5a ($p = 0.01$) et pour l'expérience 5b ($p = 0.04$). L'expérience 5c, quant à elle, présente une distribution normale ($p = 0.15$), et des scores de discrimination significativement supérieurs à 0.5 [$t(15) = 11.59$, $p < 0.001$]. Dans les

expériences 5a et 5b, les sujets n'ont pas eu de biais de réponse ($p = 0.85$ et $p = 0.36$). En revanche, dans l'expérience 5c, les sujets ont significativement préféré répondre « Pareil » [$t(15) = 2.41$, $p = 0.03$; cf. Section 4.5.1 pour une discussion]. Ainsi, comme dans les expériences 3b et 4b, les sujets ont été capables de discriminer l'espagnol de l'anglais et du polonais. De plus, on peut cette fois affirmer qu'ils l'ont fait à l'aide du rythme seul. Par ailleurs, les sujets ont également discriminé l'anglais du polonais sur la base du rythme, ce en quoi ils avaient échoué dans l'expérience 4a. On peut donc penser que dans cette expérience, l'intonation des phrases a gêné la perception du rythme. En effet, sur l'ensemble des six expériences employant la tâche AAX (3b, 4a–4b, 5a–5c), les scores sont significativement meilleurs avec les stimuli *sasasa plat* qu'avec les stimuli *sasasa* ($F(1,94) = 5.1$, $p = 0.03$). Cette observation renforce notre remarque de la page 77, selon laquelle la surabondance d'indices n'est pas nécessairement une bonne chose, mais risque au contraire d'augmenter la variabilité des stimuli et de nuire à l'analyse de l'indice principal, le rythme. Ici, la paire de langues anglais/polonais est moins bien discriminée que les deux autres ; lorsque la tâche devient plus difficile (du fait de l'intonation), c'est donc la première à ne plus être discriminée. La discrimination anglais-polonais n'en est pas moins réelle dans l'expérience 5b, et nous concluons donc que l'anglais et le polonais ont également des propriétés rythmiques distinguables.

4.3.4 Discussion

Ainsi, les sujets ont été capables de discriminer le rythme du polonais à la fois du rythme de l'anglais et de celui de l'espagnol. Le polonais semble donc présenter des propriétés rythmiques qui ne sont assimilables ni aux langues accentuelles, ni aux langues syllabiques. Nos résultats accréditent ainsi l'idée de Nespor (1990) selon laquelle le polonais est une langue intermédiaire entre les langues accentuelles et les langues syllabiques. En conséquence, notre modèle de la perception du rythme de la parole reposant sur la variable $\%V$ se révèle inadéquat : en effet, il ne prédit aucune discrimination entre l'anglais et le polonais. Pour rendre compte de cette discrimination, il y aura sans doute lieu d'intégrer la variable ΔV dans le modèle ; nous y reviendrons dans la Section 4.7. En termes phonologiques, nos résultats suggèrent que la complexité syllabique (mesurée par $\%V$) ne suffit pas à rendre compte du rythme. La discrimination entre le polonais et l'anglais doit donc provenir d'une autre propriété phonologique, par exemple la réduction vocalique, présente en anglais et absente en polonais (cf. Section 2.1.2.2). Pour y voir plus clair, nous allons maintenant considérer le cas du catalan, une autre langue supposée être intermédiaire.

4.4 Exp. 6a–6c : catalan-espagnol, anglais-catalan et catalan-polonais en *sasasa plat*

Nous allons maintenant étudier la discrimination du catalan avec les deux langues de référence, l'anglais et l'espagnol. Nous comparerons également le catalan au polonais, afin de savoir comment les deux langues supposées intermédiaires se situent l'une par rapport à l'autre.

4.4.1 Méthode

Nous avons utilisé les phrases anglaises et polonaises resynthétisées selon la transformation *sasasa plat* qui ont déjà servi aux expériences 5a–5c. Les 20 phrases espagnoles[7] et catalanes sont extraites du corpus segmenté décrit au Chapitre 2. Les phrases catalanes ont une durée moyenne de 2856 ms ±285, et les phrases espagnoles une durée de 2928 ms ±214. En tenant compte de la correction de durée appliquée aux phrases polonaises, les quatre langues ont des phrases de durées similaires ($F(3,76) < 1$). Les phrases ont été resynthétisées en *sasasa plat* en portant tous leurs points de fréquence fondamentale à 230 Hz. Dans les trois expériences qui suivent, le catalan a été opposé aux deux langues de référence, l'espagnol (6a)[8] et l'anglais (6b),

7. Les phrases espagnoles précédemment utilisées dans les expériences 3a–3b, 4b, 5a et 5c ont été critiquées par des locuteurs natifs de l'espagnol, qui leur reprochaient d'être issues d'autres dialectes de l'espagnol que le castillan. Bien qu'il n'y ait pas de raison de supposer que cela ait pu influer sur les résultats précédents, nous avons décidé d'effectuer les expériences suivantes avec un corpus plus consensuel. Les phrases espagnoles ont donc été ré-enregistrées par des locuteurs natifs du castillan, sous la supervision de Laura Bosch et Núria Sebastián-Gallés à l'université de Barcelone. Ce sont ces phrases qui sont utilisées dans toutes les expériences ultérieures, ainsi qu'au Chapitre 2. Cf. également la note suivante.

8. **Expérience 6a bis – Discrimination espagnol-catalan.**
Une première version de cette expérience a été réalisée, dans laquelle les stimuli étaient resynthétisés en *mamama plat* (les consonnes étaient remplacées par [m]), ceci afin de rendre les stimuli moins désagréables à écouter. 16 sujets ont été testés, et ont donné le résultat suivant : score moyen de discrimination $A' = 0.57$, significativement supérieur à 0.5 [$t(15) = 2.64$, $p = 0.02$]. Ces sujets ont donc réussi à discriminer les phrases espagnoles des phrases catalanes. Ce résultat nous a paru surprenant, dans la mesure où 1° il était contraire à nos prédictions, 2° l'auteur, malgré sa connaissance des phrases, n'a pas réussi à discriminer les deux langues en passant l'expérience en tant que sujet, 3° deux bilingues catalans/espagnols, qui plus est respectivement phonéticienne et psycholinguiste, ont également échoué dans la tâche de discrimination. Comme il n'y avait pas de raison de suspecter que le changement de type de synthèse de *sasasa plat* à *mamama plat* ait pu être seul responsable de ce résultat, nous avons recherché un artefact dans les phrases qui pourrait expliquer la discrimination par les sujets. Nous avons tout d'abord pensé que les phrases catalanes se terminaient peut-être plus souvent par une consonne que les phrases

et également au polonais (6c). Les expériences se sont déroulées suivant le même protocole et dans les mêmes conditions que les expériences 5a–5c.

Dans le but de savoir si des instructions plus précises pouvaient aider les sujets à mieux effectuer la tâche, la moitié des sujets a reçu les instructions standard comme dans les autres expériences, et l'autre moitié a reçu des instructions attirant l'attention sur le fait qu'il fallait rechercher des différences de rythme entre les langues. Le texte des deux types d'instructions est donné en Annexe B.7. L'effet des instructions sera analysé à la Section 4.5.2.

4.4.2 Participants

48 sujets, 16 par expérience, ont été recrutés dans les grandes écoles parisiennes et à la faculté de psychologie. Il s'agissait de 21 hommes et 27 femmes, âgés de 24 ans en moyenne. Ils ont été testés suivant les mêmes modalités que lors des expériences précédentes.

4.4.3 Résultats

Le Tableau 4.4 donne les scores de discrimination et les indices de biais de réponse pour chacune des trois expériences, et la Figure 4.3 les histogrammes des scores A'.

La distribution des scores des expériences 6a, 6b et 6c peut être considérée comme normale (respectivement $p > 0.20$, $p > 0.20$ et $p = 0.066$). Dans l'expérience 6a (catalan/espagnol), les sujets n'ont pas du tout réussi à discriminer les deux langues [$t(15) = -1.1$, $p = 0.3$], ayant un score A' moyen qui est même inférieur à 0.5[9]. Dans les expériences 6b et 6c, les scores de discrimination sont supérieurs à 0.5, avec une significativité marginale :

espagnoles. Il s'est avéré que c'était faux. Nous avons alors remarqué que certaines phrases comportaient des silences (entre deux groupes intonationnels) : 11 phrases espagnoles sur 20 comportaient au moins un silence, mais seulement 2 phrases catalanes sur 20. Ainsi, une stratégie de réponse qui consisterait à répondre « espagnol » lorsqu'il y a un silence et « catalan » sinon aboutirait à un score maximal de 72.5% de bonnes réponses, et un score A' de 0.83. Pour vérifier que cet artefact était bien la cause de la discrimination observée, nous avons refait l'expérience en supprimant les silences des phrases catalanes et espagnoles au cours de la resynthèse : c'est l'expérience 6a décrite dans le corps du texte (ici, nous sommes revenus au *sasasa plat* pour préserver l'homogénéité de la série d'expériences).

N. B. : nous avons par la suite recherché d'éventuels silences aussi dans les phrases anglaises et polonaises des expériences 3a–3b, 4a–4b et 5a–5c, et nous n'en avons trouvé qu'un seul, dans une phrase polonaise. Ces expériences ne sont donc pas touchées par le même artefact.

9. Ce résultat suggère qu'effectivement la discrimination dans l'expérience 6a bis était due aux silences dans les phrases espagnoles.

TAB. 4.4 – *Discrimination catalan-espagnol, anglais-catalan et catalan-polonais en* sasasa plat.

	A'	B''_D
Exp. 6a : catalan/espagnol	0.46 ±0.13	0.13
Exp. 6b : anglais/catalan	0.57 ±0.14	0.06
Exp. 6c : catalan/polonais	0.57 ±0.16	-0.35

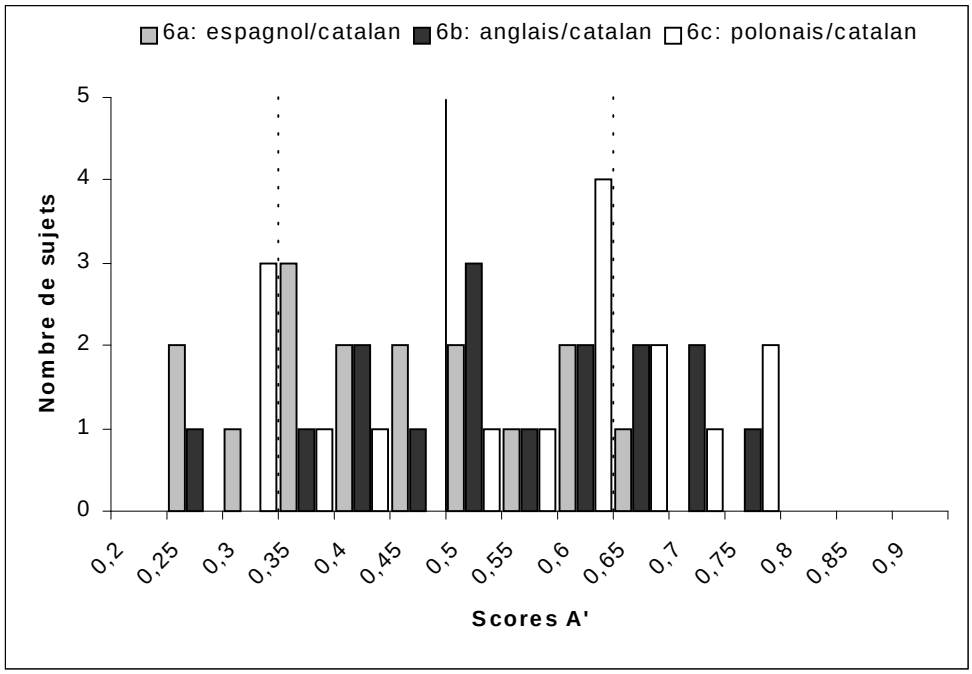

FIG. 4.3 – *Histogramme des scores A' pour les expériences 6a–6c.*

$t(15) = 2$, $p = 0.06$ pour anglais/catalan, $t(15) = 1.81$, $p = 0.09$ pour catalan/polonais. En fait, on peut remarquer que les tests statistiques que nous appliquons depuis le début sont bi-directionnels, c'est-à-dire qu'ils testent l'hypothèse selon laquelle les sujets ont des scores supérieurs *ou* inférieurs à 0.5, alors que nous ne considérons que l'hypothèse selon laquelle ils ont des scores supérieurs à 0.5[10]. Ainsi, bien qu'il ne soit pas d'usage d'utiliser des tests statistiques uni-directionnelles, on peut raisonnablement considérer que les scores sont différents du hasard jusqu'à un seuil de $\alpha = 0.10$. Par ailleurs, il est remarquable que certains sujets des expériences 6b et 6c atteignent des scores supérieurs à 75% (Figure 4.3). Ici encore, on a l'impression que la tâche est possible, mais que le nombre de sujets qui y parviennent est plus réduit que pour d'autres paires de langues. Nous concluons donc que les sujets ont discriminé le catalan à la fois de l'anglais et du polonais, mais pas de l'espagnol.

Remarquons enfin que les sujets n'ont pas eu de biais de réponse dans l'expérience 6a [$t(15) = 1.58$, $p = 0.14$] ni dans l'expérience 6b [$t(15) < 1$]. Dans l'expérience 6c en revanche, les sujets ont répondu significativement plus souvent « Pareil » que différent ($t(15) = -4.6$, $p < 0.001$; cf. Section 4.5.1).

4.4.4 Discussion

Nous avons montré que le rythme du catalan est discriminable de celui de l'anglais, mais pas de celui de l'espagnol. Ainsi, contrairement au polonais, le catalan ne semble pas se comporter comme une langue intermédiaire, mais plutôt comme une langue syllabique. De plus, comme l'espagnol, le catalan est également distinguable du polonais, ce qui confirme que ces deux langues n'ont pas les mêmes propriétés rythmiques.

Ainsi, des deux langues supposées intermédiaires par Nespor (1990), nos résultats suggèrent que l'une (le catalan) est en fait syllabique, et l'autre (le polonais) est véritablement inclassable. Le positionnement du catalan aux côtés de l'espagnol parmi les langues syllabiques était déjà prédit par les mesures et les simulations effectuées au Chapitre 2. Le comportement du polonais, en revanche, n'était pas prédit par les simulations basées sur la variable $\%V$. Il s'agit maintenant de savoir s'il sera possible d'ajuster le

10. Remarquons que dans notre tâche AAX, il n'est pas possible de faire pire que le hasard, car la bonne réponse est révélée après chaque essai, et l'association entre chaque langue et la touche de réponse correspondante est également rappelée à chaque essai. Ainsi, lorsqu'un sujet, voire un groupe de sujets dans le cas de l'expérience 6a, obtient un score moyen inférieur à 0.5, cela ne peut correspondre qu'à l'absence de discrimination, plus l'effet du hasard.

modèle (en incorporant notamment la variable ΔV) pour lui permettre de prédire correctement l'ensemble des discriminations attendues. Ce sera l'objet de la Section 4.7. Avant cela, nous allons revenir sur les données recueillies ci-dessous afin d'effectuer quelques analyses plus globales.

4.5 Analyses post-hoc

4.5.1 Biais de réponse

Dans les expériences 5c et 6c, les sujets ont eu des biais de réponse significatifs, allant dans le même sens ($B''_D < 0$). Au-delà de la significativité, on peut remarquer que B''_D a tendance a être plutôt négatif sur l'ensemble des expériences utilisant la tâche AAX. Cette constatation est confirmée par un test de t comparant B''_D à 0 sur l'ensemble de ces 9 expériences : $t(143) = -2.1$, $p = 0.04$. La négativité de B''_D correspondant au biais de réponse « Pareil », il s'agit là d'un biais bien compréhensible : lorsqu'ils ne savent pas quel indice ou quelle stratégie utiliser, les phrases paraissent toutes semblables aux sujets, d'où la réponse par défaut « Pareil » (certains sujets se sont même plaints de ce que « toutes les phrases étaient pareilles »). Une interprétation alternative pourrait être un biais en faveur de la main droite (commandant la lettre P). N'ayant pas contrebalancé les côtés de réponse à travers les sujets, il ne nous est pas possible de départager les deux hypothèses, mais la première nous paraît plus plausible. Il se pourrait aussi, bien sûr, que de tels effets soient des artefacts ininterprétables.

4.5.2 Instructions

Dans les expériences 6a–6c nous avions introduit une manipulation supplémentaire dans les instructions : la moitié des sujets ont reçu des instructions plus précises leur expliquant que les stimuli préservaient le rythme des phrases et que c'était donc ce paramètre qu'il fallait analyser pour distinguer les deux langues. Le but était d'évaluer l'impact d'une meilleure compréhension de la tâche sur la discrimination, dans le souci d'épargner aux sujets des difficultés non pertinentes pour l'objet de la recherche.

Les 8 premiers sujets testés dans chaque expérience ont donc reçu les instructions habituelles, et les 8 suivants ont reçu les instructions plus précises. Il s'avère que ces derniers n'en n'ont pas tiré avantage. Nous avons comparé les scores des deux groupes de sujets sur les deux expériences donnant lieu à une discrimination (6b et 6c), et nous n'avons trouvé aucune différence significative : $F(1,30) < 1$. La tendance était même contraire à l'effet prédit, avec

un score A' moyen de 0.59 pour les instructions normales et de 0.55 pour les instructions plus précises. Cette tentative pour faciliter la tâche et améliorer les scores des sujets s'avère donc infructueuse.

4.5.3 Recrutement

Comme les scores des expériences 6a–6c sont globalement inférieurs à ceux des expériences 3b, 4a–4b et 5a–5c, nous nous sommes demandés si cela n'était pas dû à un biais de recrutement. En effet, nous avons décelé 3 différences dans le recrutement des deux séries d'expériences : dans les expériences 6a–6c, 1° les sujets sont légèrement plus âgés en moyenne, 2° il y a une plus grande proportion de femmes, 3° la moitié des sujets ont été recrutés à l'université et la moitié en grandes écoles, alors que dans les expériences précédentes ils avaient tous été recrutés en grandes écoles.

Bien que nous n'ayons pas d'hypothèse sur l'influence de ces facteurs sur la discrimination, nous avons étudié leur effet sur les expériences 6b et 6c [11]. Nous n'avons trouvé aucune influence du sexe des sujets [$F(1,31) < 1$] ni du lieu de recrutement [$F(1,31) < 1$]. Par ailleurs, nous n'avons trouvé aucune corrélation entre l'âge des sujets et les scores A' [$R = -0.17$, $p = 0.36$]. Nous en concluons donc que la bas niveau des scores de discrimination dans les scores 6a–6c par rapport aux expériences précédentes n'est pas dû à un biais de recrutement.

4.5.4 Distribution des scores

La Figure 4.4 donne la distribution des scores A' pour l'ensemble des expériences 3b, 4a–4b, 5a–5c et 6a–6c.

Cette distribution est fortement non-normale, ainsi que le confirme le test de Lilliefors ($p = 0.001$). Cette observation suggère que les sujets testés ne forment pas une population homogène : d'après nous, ils se divisent entre ceux qui ont trouvé quel indice écouter (les « bons » sujets), et ceux qui ne l'ont pas trouvé (les « mauvais » sujets). Selon cette interprétation, les scores se répartissent dans une distribution bimodale, dont un mode est centré autour de la valeur 0.5, et l'autre correspond à une valeur plus élevée. La Figure 4.5 montre qu'en effet la distribution des scores peut être vue comme la somme de deux distribution normales, dont l'une est centrée sur la valeur 0.5.

11. Nous n'avons pas effectué l'analyse sur la totalité des expériences parce que le facteur « série d'expériences » aurait dans ce cas été confondu avec les facteurs d'intérêt. Nous avons de plus exclu l'expérience 6a qui, donnant un résultat négatif, n'était pas susceptible de mettre en évidence des différences de performance.

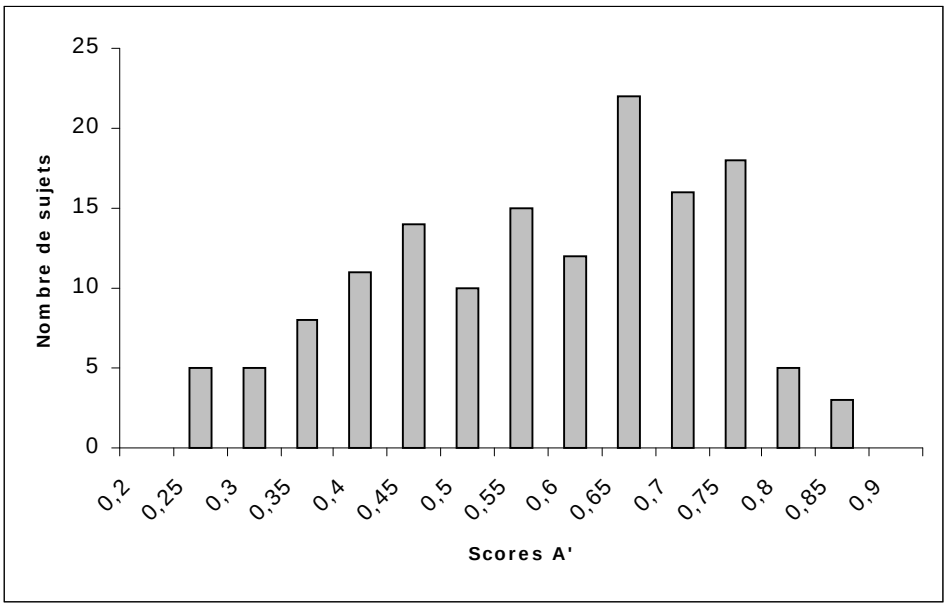

FIG. 4.4 – *Histogramme des scores A' pour l'ensemble des expériences 3b, 4a–4b, 5a–5c et 6a–6c.*

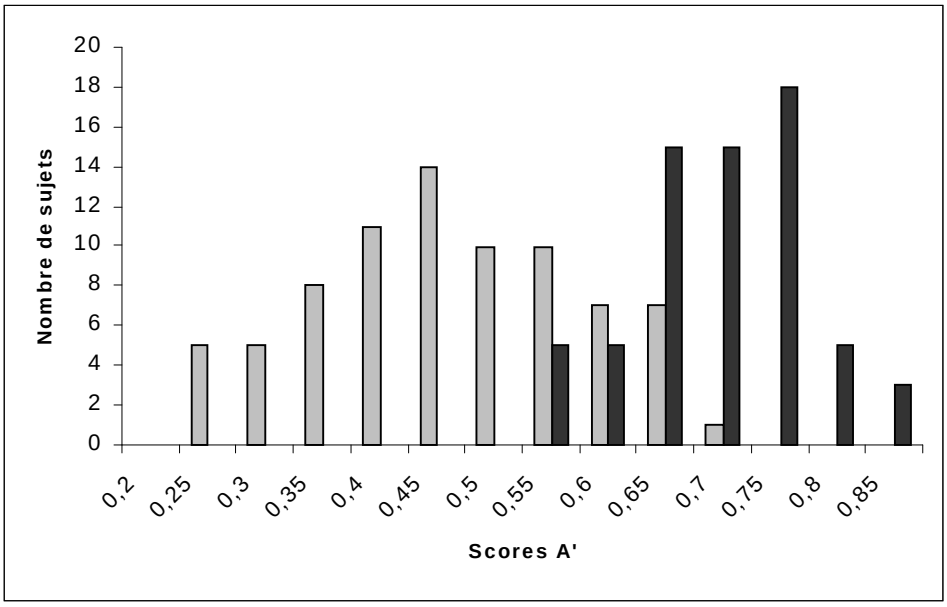

FIG. 4.5 – *Répartition des scores A' en deux distributions normales.*

Ainsi, les expériences précédentes diffèrent entre elles essentiellement par 1° la proportion de sujets appartenant à chacune des deux distributions, et 2° la valeur sur laquelle la distribution des « bons » sujets est centrée. De toutes ces expériences, une seule se distingue en ce qu'elle ne comporte pas de distribution de « bons » sujets : il s'agit de l'expérience 6a (catalan/espagnol), dont on peut dire qu'aucun sujet n'a réussi à extraire des indices fiables pour la discrimination. Dans toutes les autres expériences, la discrimination est possible, à des degrés divers de difficulté qui sont reflétés par les degrés significativité de l'effet de discrimination.

4.6 Résumé

Pour permettre une meilleure vue d'ensemble, le Tableau 4.5 récapitule tous les résultats expérimentaux obtenus dans les Chapitres 3 et 4.

TAB. 4.5 – *Expériences de discrimination de langues chez les sujets adultes.*

Exp.	Paire	Tâche	Stimuli	Sujets	A'	p
2a	Anglais Japonais	X	saltanaj	16	0.72	<0.001
2b	Anglais Japonais	X	sasasa	16	0.68	0.009
2c	Anglais Japonais	X	aaaa	16	0.52	0.67
2d	Anglais Japonais	X	sasasa plat	16	0.72	<0.001
2e	Anglais Japonais	X	aaaa	16 [a]	0.61	0.005
3a	Anglais Espagnol	X	sasasa	16	0.62	0.02
3b	Anglais Espagnol	AAX	sasasa	16	0.61	<0.001
4a	Anglais Polonais	AAX	sasasa	16	0.55	0.25
4b	Espagnol Polonais	AAX	sasasa	16	0.65	0.001
5a	Anglais Espagnol	AAX	sasasa plat	16	0.65	0.01 [b]
5b	Anglais Polonais	AAX	sasasa plat	16	0.61	0.04 [b]
5c	Espagnol Polonais	AAX	sasasa plat	16	0.74	<0.001
6a bis	Espagnol Catalan 1	AAX	mamama plat avec silences	16	0.57	0.02
6a	Espagnol Catalan 2	AAX	sasasa plat sans silence	16	0.46	0.30
6b	Anglais Catalan	AAX	sasasa plat	16	0.57	0.06
6c	Polonais Catalan	AAX	sasasa plat	16	0.59	0.09

[a] Sujets anglophones natifs.

[b] En raison de la non normalité de la distribution, il s'agit ici d'un test binômial, dont la puissance n'est pas comparable aux tests de t utilisés pour les autres expériences.

Pour la discussion qui suit, nous prendrons en compte uniquement les

expériences qui ont employé des stimuli de type *sasasa plat*, et qui sont récapitulées dans le Tableau 4.6.

TAB. 4.6 – *Expériences de discrimination de langues chez les sujets adultes : sasasa plat seulement.*

Paire	Tâche	A'	p
Anglais Japonais	X	0.72	<0.001
Anglais Espagnol	AAX	0.65	0.01 [a]
Anglais Polonais	AAX	0.61	0.04 [a]
Espagnol Polonais	AAX	0.74	<0.001
Anglais Catalan	AAX	0.57	0.06
Polonais Catalan	AAX	0.59	0.09
Espagnol Catalan	AAX	0.46	0.30

[a] En raison de la non normalité de la distribution, il s'agit ici d'un test binômial, dont la puissance n'est pas comparable aux tests de t utilisés pour les autres expériences.

4.7 Ajustement du modèle

Nous disposons donc maintenant d'un certain nombre de données concernant la discrimination de langues sur la base du rythme. Les résultats obtenus sont compatibles avec les simulations effectuées au Chapitre 2, à l'exception de la discrimination anglais/polonais. Il ne s'agit pas là d'un résultat anodin : il remet en cause les traditionnelles classes rythmiques telles qu'elles ont été décrites dans la littérature. Dans la mesure où nous avions bâti notre modèle de la perception du rythme sur la variable $\%V$, de manière à ce qu'il cadre justement avec les classes de rythme, il n'est donc pas étonnant que les simulations qui en découlent n'aient pas prédit la discrimination anglais/polonais.

Néanmoins, nous n'avons pas encore exploité toutes les ressources de notre modèle, et on peut espérer adapter celui-ci de manière à ce qu'il puisse rendre compte des données empiriques obtenues. En effet, nous avions identifié trois variables décrivant nos mesures des intervalles vocaliques et consonantiques : $\%V$, ΔV et ΔC. Le choix de $\%V$ pour effectuer les simulations se basait avant tout sur le souci de concorder avec les données connues à l'époque, c'est-à-dire les classes de rythme. Nous avions cependant déjà remarqué que le choix de la variable ΔV était susceptible de conduire à des prédictions différentes, notamment concernant le polonais. Nos résultats présents donnent raison à cette hypothèse. Il paraît donc maintenant judicieux d'introduire la variable ΔV dans le modèle.

Plusieurs façons d'ajuster le modèle sont envisageables. L'une d'entre elles pourrait être de purement et simplement remplacer la variable $\%V$ par ΔV. On voit pourtant bien que ce choix ne serait pas très heureux. En effet, mise à part la paire anglais/polonais, tous nos résultats restent compatibles avec les classes rythmiques. Or la variable ΔV seule ne rend pas compte des classes de rythme : sur la Figure 2.2 par exemple, l'axe ΔV tend à rapprocher les langues syllabiques et accentuelles et à classer le japonais parmi les langues syllabiques. On a donc tout intérêt à conserver la variable $\%V$ pour rendre compte des classes de rythme, et à ajouter ΔV pour le cas particulier du polonais.

4.7.1 Simulation 3 : Discrimination de langues par les adultes, basée sur $\%V$ et ΔV

La méthode employée est toujours la régression logistique (voir Section 2.3.1.1), dans la mesure où c'est une procédure multivariée, c'est-à-dire qu'elle peut prendre en compte plusieurs variables prédictives. Les données consistent donc maintenant en triplets de valeurs ($\%V$,ΔV,L), où L représente l'une des deux langues de la paire considérée. La procédure trouve alors la valeur seuil d'une combinaison linéaire de $\%V$ et ΔV qui sépare au mieux les deux langues. De même qu'au Chapitre 2, l'analyse est effectuée sur 4 locutrices et évaluée sur les 4 autres. Elle est ensuite refaite en échangeant les locutrices d'entraînement et de test, et le score rapporté est la moyenne de ceux des deux analyses. Les pourcentages de réponses correctes prédits pour les 28 paires de langues sont présentés au Tableau 4.7.

Certaines grandes tendances qui étaient vraies des simulations basées sur $\%V$ le sont encore ici :

- des scores élevés de discrimination entre langues syllabiques et accentuelles ;
- des scores faibles au sein des langues syllabiques ;
- des scores très élevés entre le japonais et toutes les autres langues.

Ces constatations sont compatibles avec les classes de rythme. Néanmoins, les scores de 75% et 70% pour les paires anglais/néerlandais et espagnol/italien sont en contradiction avec les classes de rythme. Enfin, les résultats obtenus pour le polonais sont mitigés : grâce à l'influence de ΔV, celui-ci semble être maintenant distinguable de toutes les langues, sauf pourtant de l'espagnol et du catalan, ce qui est contraire aux résultats de nos expériences, notamment pour ce qui est de l'espagnol qui a été très bien discriminé du polonais par les sujets.

Les résultats de cette simulation ne sont donc pas très satisfaisants :

d'une part il ne rendent pas entièrement compte des données empiriques que nous avons obtenues (polonais/espagnol, polonais/catalan), d'autre part ils s'écartent légèrement des classes de rythme, en suggérant des discriminations entre l'anglais et le néerlandais et entre l'espagnol et l'italien. Il est donc tentant d'entreprendre une nouvelle série de simulations. Dans les présents résultats, il nous semble que nous nous sommes trop éloignés des classes de rythme : des discriminations non prévues sont apparues, et nous en avons perdu deux (polonais/espagnol et catalan) qui étaient pourtant acquises dans la simulation 2. La troisième variable, ΔC, étant corrélée avec $\%V$, devrait pouvoir justement nous rapprocher à nouveau des classes de rythme.

TAB. 4.7 – *Simulations des expériences de discrimination de langues chez l'adulte basées sur $\%V$ et ΔV pour les 28 paires de langues.*

	Ang.	Née.	Fra.	Ita.	Cat.	Esp.	Jap.
Née.	75						
Fra.	80	65					
Ita.	85*	62.5	45				
Cat.	**95***	77.5	52.5	57.5			
Esp.	**77.5**	75	55	70	**57.5**		
Jap.	**	**	87.5	90*	80	95*	
Pol.	**	90*	75	80*	**57.5**	**55**	**

Les scores sont les pourcentages de phrases de test correctement classifiées par les régressions logistiques effectuées sur les phrases d'entraînement. Les variables prédictives sont $\%V$ et ΔV, et le niveau du hasard est 50%. Les valeurs indiquées en gras sont celles pour lesquelles des données empiriques existent.

** Pour ces paires de langues l'une des deux régressions n'a pas convergé, c'est-à-dire que la solution de la régression n'était pas unique. Cela se produit lorsque les variables prédictives ($\%V$ et ΔV)) permettent de complètement séparer les phrases des deux langues (100% de classifications correctes dans la phase d'entraînement). Dans ce cas, c'est le score de la régression qui a convergé qui est indiqué dans le tableau.*

*** Pour ces paires, aucune des deux régressions n'a convergé. On peut alors considérer que les scores correspondants sont très élevés.*

4.7.2 Simulation 4 : Discrimination de langues par les adultes, basée sur $\%V$, ΔV et ΔC

Ici encore, nous utilisons la régression logistique, prenant cette fois en compte des quadruplets de valeurs ($\%V$,ΔV,ΔC,L). Mise à part la variable additionnelle, la procédure est exactement la même qu'à la section précédente. Les pourcentages de réponses correctes prédits figurent dans le Tableau 4.8.

Cette fois, les scores prédits sont compatibles avec les données empiriques obtenues : des scores de 60% ou plus sont prédits pour les paires anglais/catalan, anglais/espagnol, anglais/japonais, anglais/polonais, espagnol/polonais et catalan/polonais, et un score de 47.5% est prédit pour la paire espagnol/catalan.

TAB. 4.8 – *Simulations des expériences de discrimination de langues chez l'adulte basées sur $\%V$, ΔV et ΔC pour les 28 paires de langues.*

	Ang.	Née.	Fra.	Ita.	Cat.	Esp.	Jap.
Née.	57.5						
Fra.	85*	70					
Ita.	75*	60	55				
Cat.	**90***	80*	57.5	57.5			
Esp.	**77.5**	75	40	60	**47.5**		
Jap.	**	**	92.5	90*	80	95*	
Pol.	**	90*	77.5	80*	**60**	**60**	**

Les scores sont les pourcentages de phrases de test correctement classifiées par les régressions logistiques effectuées sur les phrases d'entraînement. Les variables prédictives sont $\%V$, ΔV et ΔC, et le niveau du hasard est 50%. Les valeurs indiquées en gras sont celles pour lesquelles des données empiriques existent.
** Pour ces paires de langues l'une des deux régressions n'a pas convergé, c'est-à-dire que la solution de la régression n'était pas unique. Cela se produit lorsque les variables prédictives ($\%V$, ΔV et ΔC) permettent de complètement séparer les phrases des deux langues (100% de classifications correctes dans la phase d'entraînement). Dans ce cas, c'est le score de la régression qui a convergé qui est indiqué dans le tableau.*
*** Pour ces paires, aucune des deux régressions n'a convergé. On peut alors considérer que les scores correspondants sont très élevés.*

En dehors des paires testées, nous prédisons de plus que le polonais peut être discriminé de toutes les autres langues. Par ailleurs, en dehors du polo-

nais qui semble maintenant former une classe à part, les classes de rythme ne sont pas remises en question : nous obtenons, comme dans la simulation 1, des scores supérieurs ou égaux à 60% entre langues de classes différentes, et des scores inférieurs entre langues appartenant à la même classe, à la seule exception de la paire espagnol/italien, pour laquelle un score de 60% est prédit.

4.8 Conclusion

Au Chapitre 2, nous avons proposé que l'on pouvait rendre compte des propriétés rythmiques des langues par des variables telles que $\%V$, ΔV et ΔC. Ces variables sont globalement compatibles avec la notion de classes rythmiques de langues, et ΔV suggère de plus que le polonais pourrait former une classe à part. Nous avons également proposé un modèle de la discrimination de langues sur la base du rythme, reposant sur la variable $\%V$, et prédisant des discriminations de langues entre classes de rythme, mais pas en leur sein.

Aux Chapitres 3 et 4, nous avons étudié la discrimination d'un certain nombre de paires de langues sur la base d'indices rythmiques, et nos résultats ont confirmé le statut à part du polonais. Nous avons donc été conduits a ré-ajuster le modèle, en lui incorporant notamment la variable ΔV. Les simulations découlant du nouveau modèle basé sur les trois variables $\%V$, ΔV et ΔC prédisent avec succès les résultats de nos expériences. Elles continuent par ailleurs à prédire les discriminations inter-classes mais pas intra-classe.

Ainsi, nous pensons que les classes de rythme sont toujours supportées par nos données. Celles-ci nous obligent néanmoins à postuler l'existence d'une quatrième classe, comportant au moins le polonais. Ce postulat n'a bien sûr rien d'arbitraire, puisque d'un point de vue phonologique, la mise en exergue du polonais est parfaitement interprétable (cf Sections 2.1.2.2 et 2.2.4) et avait été prédite par Nespor (1990).

Il reste néanmoins à savoir ce qu'il en est pour le nouveau-né qui, rappelons-le, est la motivation principale de notre étude. Il serait notamment intéressant de déterminer s'il est capable de discriminer l'anglais du polonais, et ce que cela implique pour l'acquisition du langage. Dans ce but, le Chapitre 5 va tenter d'adapter la technique de resynthèse de parole exposée au Chapitre 3 aux études de discrimination de langues par des nouveau-nés.

Chapitre 5

Discrimination de langues par des nouveau-nés

NOUS AVONS FAIT L'HYPOTHÈSE que le rythme de la parole pourrait être un indice phonologique utile pour initialiser l'apprentissage d'autres propriétés linguistiques (p. 36). Nous avons par ailleurs montré que la perception du rythme de la parole peut être étudiée de manière adéquate par des expériences de discrimination de langues exploitant la technique de la resynthèse de parole (Chapitre 3). Le modèle de la perception du rythme de la parole que nous avons élaboré (Chapitre 2) a fourni un grand nombre de prédictions quant aux langues qui peuvent être discriminées (ou non) sur la base de leur rythme. Ces prédictions ont été largement corroborées par des expériences de discrimination de langues sur des adultes (Chapitre 4). Le moment est donc venu de tester la perception du rythme de la parole directement chez le nouveau-né, par des expériences de discrimination de langues utilisant la resynthèse de parole.

Notons que la discrimination de langues n'est pas le seul moyen d'étudier la perception du rythme chez le nourrisson. Demany, McKenzie, et Vurpillot (1977) ont montré que des enfants de 2-3 mois étaient capables de distinguer des séquences de tons de rythmes différents. Fowler, Smith, et Tassinary (1986) ont de plus montré que des enfants de 3-4 mois étaient capables de distinguer des séquences de syllabes dont les « P-centres »[1] étaient isochrones ou non. Il est donc acquis que le nourrisson humain a une capacité basique à percevoir des différences de rythme. En étudiant la discrimination de langues, notre propos n'est donc pas seulement d'étudier la perception générale du

1. Le P-centre d'une syllabe (ou d'un autre objet acoustique) est son centre rythmique, tel que perçu par le système auditif humain. Notamment, une séquence de syllabes semble régulière lorsque leurs P-centres (plutôt que leurs débuts) sont placés à intervalles réguliers (voir Morton, Marcus, & Frankish, 1976).

rythme. Notre but est d'étudier la perception des propriétés rythmiques particulières qui sont à la base des classes rythmiques de langues, et qui sont susceptibles de jouer un rôle dans l'acquisition du langage.

5.1 Revue des expériences de discrimination de langues par des nourrissons

Nous allons commencer par passer en revue les expériences de discrimination de langues effectuées à ce jour sur des nourrissons.

5.1.1 À quelques jours de vie

Mehler et coll. (1988) ont enregistré des histoires prononcées par une locutrice bilingue français/russe, et les ont présentées à des nouveau-nés français âgés de 4 jours suivant une procédure de succions de haute amplitude non contingente (cf. Annexe A.1). Les nouveau-nés ont effectivement discriminé les phrases russes des phrases françaises, et ont de plus montré une préférence pour les phrases françaises, en tétant plus fort en moyenne à l'écoute de celles-ci. Les mêmes auteurs ont répété l'expérience avec des nouveau-nés nés en France, mais de parents de langue maternelle étrangère. Ces bébés-là n'ont pas réagi à la différence entre les deux langues, ce qui a été interprété comme montrant qu'une certaine familiarité avec l'une des deux langues était nécessaire à la discrimination. Une autre expérience a testé la discrimination entre des histoires anglaises et italiennes, également enregistrées par une locutrice bilingue. Dans l'article d'origine, Mehler et coll. (1988) ont rapporté que les nouveau-nés français n'avaient pas discriminé les deux langues. En fait, en effectuant une nouvelle analyse des mêmes données, Mehler et Christophe (1995) ont montré que les nouveau-nés avaient bien discriminé les deux langues[2]. Ainsi, il semble que la familiarité avec l'une des deux langues ne soit pas une condition nécessaire à la discrimination (ce qui a été confirmé par Nazzi et coll. (1998), voir ci-dessous). L'échec des nouveau-nés de parents étrangers à discriminer le français du russe devient de ce fait difficile à interpréter.

2. Dans l'article de Mehler et coll. (1988), les données des sujets entendant l'anglais dans la première phase de l'expérience étaient analysées séparément de celles des sujets entendant l'italien en premier. Mehler et Christophe (1995) ont fait remarquer qu'aucune différence significative n'étant avérée entre les deux groupes de sujets, on pouvait les fusionner pour effectuer une analyse unique. Celle-ci a révélé une différence significative entre le groupe contrôle et le groupe expérimental (qui était bien sûr déjà présente sous forme de tendance dans les données séparées).

Moon et coll. (1993) ont également étudié l'influence de la langue maternelle sur la discrimination de langues par les nouveau-nés. Ils ont recrutés des bébés de mères dont la langue maternelle était soit l'anglais, soit l'espagnol, et les ont exposés à des histoires anglaises et espagnoles. Ils ont trouvé que les bébés tétaient plus fort pour écouter leur langue maternelle que pour écouter l'autre langue, montrant 1° qu'ils distinguaient les deux langues, 2° qu'ils étaient déjà familiers avec leur langue maternelle et préféraient écouter celle-ci. Ce résultat confirme donc le résultat de Mehler et coll. (1988) avec une nouvelle paire de langue.

Se basant sur leurs premiers résultats, Mehler et coll. (1988) ont tenté de préciser un peu le type de traitement effectué par les bébés sur la parole pour discriminer les langues. Dans une nouvelle expérience, ils ont présenté à des nouveau-nés français les mêmes phrases françaises et russes que précédemment, mais jouées à l'envers[3]. Dans ces conditions, les bébés n'ont pas distingué les deux langues. Les auteurs en ont déduit qu'ils n'analysent pas simplement des propriétés spectrales globales, qui sont parfaitement préservées dans la parole à l'envers, mais qu'ils s'appuient sur les propriétés temporelles caractéristiques du langage parlé qui, elles, sont détériorées par cette manipulation. Dans une autre expérience, ils ont présenté aux bébés les mêmes phrases russes et françaises, jouées à l'endroit, mais filtrées passe-bas à 400 Hz (cf. détails et discussion sur le filtrage à la Section 3.1). Cette fois, les bébés ont bien discriminé les deux langues, répliquant la première expérience dans des conditions où l'information acoustique était considérablement amoindrie. Les auteurs en ont conclu que les nouveau-nés n'utilisaient pas les propriétés phonétiques ni phonotactiques de la parole, qui sont détériorées par le filtrage, mais plutôt ses propriétés prosodiques, qui sont préservées, pour effectuer la discrimination.

En utilisant la procédure de succions de haute amplitude, Nazzi et coll. (1998) ont étendu les résultats précédents à de nouvelles paires de langues : ils ont montré que des nouveau-nés français distinguaient l'anglais du japonais, confirmant que la discrimination était possible sans familiarité préalable avec l'une des langues. Ils ont également montré que les nouveau-nés ne discriminaient pas l'anglais du néerlandais, ce qui les a conduits à proposer que la discrimination n'était possible qu'entre des langues appartenant à des classes rythmiques différentes. D'un point de vue méthodologique, ces deux expériences ont utilisé de la parole filtrée, confirmant le rôle des propriétés prosodiques de la parole. Nazzi et coll. (1998) ont également introduit un nouvel aspect dans leurs expériences : alors que les histoires de Mehler

3. C'est-à-dire comme si la cassette tournait à l'envers, ou encore comme si le temps était inversé.

et coll. (1988) étaient prononcées par une locutrice bilingue, Nazzi et coll. (1998) ont fait appel à plusieurs locutrices (monolingues) de chaque langue. Ainsi, dans la phase d'habituation les bébés étaient habitués avec des phrases prononcées par deux locutrices d'une même langue ; dans la phase de test, les bébés du groupe contrôle passaient à deux autres locutrices de la même langue, tandis que ceux du groupe expérimental écoutaient deux locutrices de l'autre langue. Ce choix expérimental permet 1° d'écarter les critiques liées au fait qu'une locutrice bilingue pourrait ne pas parler les deux langues avec la même compétence, la même voix, le même ton ou le même enthousiasme, 2° d'affirmer qu'un changement de langue est plus significatif pour un nouveau-né qu'un changement de voix.

Pour tester plus en détail l'hypothèse des classes de rythme, Nazzi et coll. (1998) ont élaboré une expérience mettant en jeu quatre langues, toujours en parole filtrée. Les bébés du groupe expérimental ont été habitués à des phrases dans deux langues appartenant à la même classe de rythme (anglais et néerlandais, ou espagnol et italien), et dans la phase de test ils ont été exposés aux deux langues appartenant à l'autre classe de rythme. Dans le groupe contrôle, les bébés étaient habitués à deux langues n'appartenant pas à la même classe de rythme (anglais et espagnol, anglais et italien, néerlandais et espagnol, ou néerlandais et italien), puis exposés aux deux autres langues, qui elles non plus n'appartenaient pas à la même classe. Les bébés du groupe expérimental se sont déshabitués significativement plus que ceux du groupe contrôle, montrant qu'ils étaient plus sensibles à un changement de classe rythmique qu'à un changement de langues non conforme aux classes rythmiques.

Les résultats obtenus par Nazzi et coll. (1998) suggèrent que les nouveau-nés catégorisent les langues selon leur appartenance aux classes rythmiques. Il serait d'ailleurs dans ce cas plus juste de dire qu'ils discriminent les classes rythmiques ou les types de rythme, plutôt que les langues. Néanmoins, on ne peut pas considérer que cette interprétation est définitivement prouvée. Premièrement, parce que l'ensemble des langues étudiées est infime par rapport à l'inventaire des langues du monde. Les données sont trop peu nombreuses pour pouvoir affirmer que ce sont les classes de rythme qui déterminent quelles langues sont discriminées ou pas par le nouveau-né. Une autre hypothèse, entièrement compatible avec les données actuelles, pourrait être que ce sont les familles génétiques (historiques) de langues qui gouvernent la discrimination : en effet, l'anglais et le néerlandais sont des langues germaniques, l'espagnol et l'italien sont des langues latines, le russe est une langue slave, et le japonais appartient à un groupe linguistique extrêmement éloigné de ces langues. Pour que cette hypothèse soit tenable, il faudrait bien sûr préciser par quel moyen le nouveau-né pourrait évaluer la proximité génétique de deux

langues. C'est pourquoi il est crucial de contrôler les paramètres acoustiques et linguistiques mis à disposition de l'enfant dans les expériences. Comme nous l'avons déjà dit à la Section 3.1, le filtrage préserve également l'intonation des phrases. Il est donc parfaitement possible que les discriminations observées soient dues aux différences d'intonation entre les langues testées. C'est pour cette raison que nous avons développé la technique de resynthèse de parole (Chapitre 3), qui va nous permettre de séparer le rythme de l'intonation.

5.1.2 À 2-3 mois

Outre les nouveau-nés français, Mehler et coll. (1988) ont également testé des nourrissons américains âgés de 2-3 mois, en utilisant cette fois une procédure de succion avec présentation audiovisuelle (voir l'article). Ils ont montré que ceux-ci étaient capables de discriminer les histoires anglaises et italiennes, aussi bien sous leur forme naturelle que lorsqu'elles étaient filtrées. En revanche, les bébés n'ont pas réussi à discriminer le français du russe. Ce résultat avait été interprété à l'époque comme confirmant la nécessité d'une familiarité avec l'une des deux langues pour la discrimination. Or, grâce à la ré-analyse de Mehler et Christophe (1995) et aux expériences de Nazzi et coll. (1998), nous savons maintenant que les nouveau-nés français distinguent l'anglais de l'italien et du japonais, et donc que cette familiarité n'est nullement nécessaire. Nous sommes donc conduits à une autre interprétation, qui tient compte de l'âge des enfants : après 2 mois de familiarisation avec l'anglais, il est possible que les langues étrangères éveillent moins l'intérêt des nourrissons américains.

Cette interprétation a été confirmée par les expériences de Hesketh, Christophe, et Dehaene-Lambertz (1997) et Christophe et Morton (1998). Hesketh et coll. (1997) ont testé des nourrissons anglais de 2-3 mois avec la méthode de succions de haute amplitude, et des phrases anglaises et japonaises prononcées par 4 locutrices par langue, non filtrées : ces nourrissons ont bien discriminé l'anglais du japonais. Cependant, Christophe et Morton (1998) ont trouvé qu'ils ne distinguaient pas le français du japonais dans les mêmes conditions, bien que ces deux langues appartiennent à des classes rythmiques différentes et présentent d'autres différences importantes. Ceci accrédite l'interprétation selon laquelle les nourrissons de deux mois ne s'intéressent plus aux langues étrangères, et discriminent seulement leur langue maternelle d'une langue étrangère. Deux expériences supplémentaires ont testé les paires anglais-néerlandais et néerlandais-japonais : les deux ont donné des résultats de discrimination marginalement significatifs. Ceci a été interprété par les auteurs comme montrant qu'entre 2 et 3 mois, les nourrissons deviennent

sensibles à des distinctions plus fines : ils commencent à distinguer le néerlandais de l'anglais. De plus, ils se focalisent sur leur langue maternelle, et commencent à mettre toutes les langues étrangères dans la même catégorie. Par conséquent, au fur et à mesure qu'ils se rendent compte que le néerlandais est une langue étrangère, ils s'intéressent moins à ses différences avec le japonais. Christophe et Morton (1998) prédisent qu'un mois plus tard, ils discrimineront parfaitement le néerlandais de l'anglais (étrangère contre maternelle), et plus du tout le néerlandais du japonais (2 langues étrangères).

En utilisant une technique de latence d'orientation (cf. Annexe A.3), Dehaene-Lambertz et Houston (1998) ont par ailleurs testé des bébés américains et français de 2 mois avec des phrases anglaises et françaises prononcés par une locutrice bilingue. Ils ont montré que ceux-ci orientaient plus vite leur regard vers un haut-parleur diffusant leur langue maternelle que vers un haut-parleur diffusant l'autre langue, exhibant ainsi leur familiarité avec leur langue maternelle et leur aptitude à la distinguer de l'autre. Ce résultat a été obtenu aussi bien avec de la parole naturelle que filtrée. Dans une autre expérience, les stimuli des deux langues ont été réduits soit à des groupes intonationnels extraits des phrases précédentes, soit à des mots extraits des phrases et juxtaposés dans un ordre incohérent. Les bébés se sont bien orientés plus rapidement vers leur langue maternelle, mais seulement dans la condition où les phrases intonationnelles étaient préservées. Outre une nouvelle confirmation de la discrimination des langues et de la préférence pour la langue maternelle, cette étude montre 1° que la discrimination peut s'effectuer sur des stimuli relativement courts (1178 ms en moyenne), mais 2° à condition qu'ils préservent la structure prosodique de la langue. C'est une nouvelle indication du rôle de la prosodie dans la discrimination des langues.

Enfin, Guasti et coll. (sous presse) rapportent une expérience de succions de haute amplitude sur des nourrissons français de 2-3 mois. Les stimuli étaient des phrases françaises et turques appariées syllabe pour syllabe, de manière à ce qu'il n'y ait aucune différence de structure syllabique, ni de nombre de syllabes entre les deux langues. Ces phrases étaient resynthétisées de façon équivalente à la transformation *saltanaj*, de manière à ne préserver que des différences de rythme et d'intonation entre les deux langues, celles-ci étant dues au fait que le français a la proéminence à droite dans les groupes phonologiques, alors que le turc l'a à gauche (cf. Section 1.3.4.2). Ici encore, les nourrissons ont discriminé les deux langues, dont l'une était leur langue maternelle, et ce grâce à leurs propriétés prosodiques. Il est à noter cependant que la discrimination n'a en principe pas pu s'effectuer sur des indices purement rythmiques, dans la mesure où les deux langues sont syllabiques. Ceci suggère que, au moins à l'âge de 2 mois, les classes de rythme ne sont

pas la seule source possible de discrimination.

5.1.3 À 4-5 mois

Bahrick et Pickens (1988) ont présenté à des enfants américains de 5 mois des séquences audiovisuelles d'une locutrice bilingue racontant une histoire soit anglaise, soit espagnole. Il s'agissait d'une procédure d'habituation où la mesure était le temps de regard vers l'écran diffusant la séquence. Les enfants subissaient une phase d'habituation pendant laquelle ils entendaient une histoire dans l'une des deux langues, puis une phase de test pendant laquelle ils entendaient une autre histoire de la même langue (groupe contrôle), ou une autre histoire dans l'autre langue (groupe expérimental). Les enfants se sont déshabitués significativement plus lors du changement de langue que du changement d'histoire, montrant qu'ils distinguaient l'anglais de l'espagnol.

Bosch et Sebastián-Gallés (1997) ont réalisé une série d'expériences sur des enfants de 4 mois qui étaient issus d'environnements monolingues catalans, monolingues espagnols, ou bilingues catalan/espagnol. Ces expériences ont été effectuées avec des phrases catalanes, espagnoles, italiennes et anglaises prononcées par une locutrice quadrilingue, et ont utilisé la procédure de latence d'orientation (Annexe A.3). Une première expérience a montré que les enfants monolingues (catalan ou espagnol) s'orientent plus vite vers leur langue maternelle que vers l'anglais. Une deuxième expérience a montré que les enfants monolingues discriminent également l'espagnol du catalan, et s'orientent plus vite vers celle des deux qui est leur langue maternelle. Une troisième expérience a répliqué ce résultat avec les mêmes phrases filtrées. Une quatrième expérience a montré que les enfants bilingues espagnol/catalan s'orientent plus vite vers l'anglais que vers l'espagnol ou le catalan. Ainsi, ces enfants discriminent leurs deux langues maternelles de l'anglais, mais exhibent l'effet d'orientation inverse par rapport aux enfants monolingues. De plus, une cinquième expérience a montré que ces enfants bilingues ne s'orientent pas plus vite vers le catalan ou vers l'espagnol lorsqu'on leur présente ces deux langues. Cela ne signifie pas nécessairement qu'ils n'arrivent pas à discriminer les deux langues, mais simplement qu'ils ne préfèrent pas particulièrement l'une à l'autre. Enfin, une dernière expérience montre que ces enfants bilingues s'orientent plus vite vers l'italien que vers l'une de leurs langues maternelles, comme c'était le cas pour l'anglais. En résumé, cette étude étend l'inventaire des langues étudiées aux paires catalan/anglais, catalan/espagnol, catalan/italien et espagnol/italien. On peut en conclure qu'à l'âge de 4 mois, les enfants ont suffisamment raffiné leur analyse de la parole et ont suffisamment acquis de connaissances sur leur langue maternelle pour pouvoir discriminer des langues appartenant à la même classe

de rythme comme l'espagnol, le catalan et l'italien. La nature précise des indices utilisés pour ces discriminations reste à explorer, mais l'expérience utilisant la parole filtrée suggère qu'il s'agit une nouvelle fois d'indices prosodiques. Nos mesures suggèrent de plus qu'il ne s'agit pas d'indices purement rythmiques, et qu'il s'agit donc plutôt de l'intonation (Section 2.1). Enfin, le cas des enfants bilingues semble particulièrement complexe et mérite plus d'investigations.

Nazzi et coll. (soumis) se sont également intéressés à la discrimination de langues appartenant à la même classe rythmique. En utilisant la procédure d'orientation préférentielle de la tête (Annexe A.2) et des phrases non filtrées prononcées par 4 locutrices par langue, ils ont tout d'abord montré que des enfants américains de 5 mois discriminaient l'anglais du japonais. Ils ont de plus montré que les bébés étaient capables de discriminer l'anglais du néerlandais, et l'anglais de l'américain. Ce résultat confirme qu'à 5 mois, les enfants ont des aptitudes perceptives suffisantes pour distinguer deux langues appartenant à la même classe rythmique, même si elles sont extrêmement proches comme l'anglais et l'américain. Pour savoir si cette capacité reposent sur les connaissances que ces enfants ont acquis sur l'anglais (ou sur les langues accentuelles), Nazzi et coll. (soumis) ont effectué trois nouvelles expériences, et ont montré que les bébés discriminaient l'italien du japonais, mais pas l'espagnol de l'italien ni l'allemand du néerlandais. Ainsi, contrairement aux enfants de 2-3 mois, ceux de 5 mois peuvent distinguer des langues autres que leur langue maternelle, si elles appartiennent à des classes rythmiques différentes. En revanche, les discriminations plus fines sont cantonnées non seulement à la classe des langues accentuelles, mais aux comparaisons avec la langue maternelle.

5.1.4 Synthèse

L'ensemble des études répertoriées ci-dessus autorisent encore de multiples interprétations. Néanmoins, on peut se risquer à donner une interprétation générale qui fait un minimum d'hypothèses additionnelles. Au cours des cinq premiers mois de la vie, on observe deux phénomènes qui influent sur la capacité à discriminer les langues.

1. **La perception et la représentation de la parole par les bébés tend à s'affiner.** À la naissance, toutes les discriminations observées peuvent s'expliquer en supposant que l'enfant analyse uniquement le rythme de la parole. À deux mois, la discrimination français/turc nécessite de supposer la capacité de distinguer des propriétés prosodiques plus fines, notamment la position de la proéminence. C'est également le cas à 4 mois pour la discrimination espagnol/catalan. Enfin, à cinq

mois, la discrimination anglais/américain suggère que l'enfant utilise des indices phonétiques et/ou prosodiques subtils.

2. **L'enfant acquiert de plus en plus d'informations sur sa langue maternelle, se focalise sur elle et s'intéresse moins aux autres langues.** C'est ce que montre, à partir de l'âge de 2 mois, l'absence de discrimination dès lors que la langue maternelle n'est pas concernée. C'est également le cas à 5 mois pour les discriminations les plus fines.

Une seule expérience ne cadre pas dans cette analyse : il s'agit de l'expérience impliquant des nouveau-nés de familles ne parlant pas le français. D'après toutes les données recueillies depuis, ces enfants auraient dû discriminer le français du russe. Notons cependant que cette expérience ne comporte que 24 bébés. Elle mériterait d'être répliquée.

Soulignons encore une fois qu'il ne s'agit que de la synthèse la plus économique des données, mais que d'autres interprétations sont possibles, notamment, comme nous l'avons déjà suggéré, que les nouveau-nés discriminent les langues en se basant sur l'intonation. Nous allons maintenant présenter une série d'expériences qui vise à tester cette hypothèse.

5.2 Méthode générale

L'objectif de cette série d'expériences est le même que celui de la série 2a–2d, c'est-à-dire d'éliminer progressivement les indices présents dans les stimuli, jusqu'à ne plus préserver que le rythme des phrases, et de déterminer par conséquent quels indices sont nécessaires aux nouveau-nés pour discriminer deux langues. Nous avons choisi une unique paire de langues : le néerlandais et le japonais. Ces deux langues ont été choisies pour maximiser les chances d'observer une discrimination. Le néerlandais étant très proche de l'anglais, on peut attendre que la discrimination néerlandais/japonais soit aussi facile que la discrimination anglais/japonais[4].Dans cette section, nous allons d'abord présenter en détail la méthode expérimentale utilisée, qui sera commune à toutes les expériences.

5.2.1 Stimuli

Nous avons utilisé les 20 phrases néerlandaises et les 20 phrases japonaises décrites au Chapitre 2. Elles sont donc prononcées par 4 locutrices par langue, soit 5 phrases par locutrice. Elles sont appariées en nombre de syllabes et ont

4. Nous avons évité d'utiliser l'anglais car la présente série d'expériences a été en partie planifiée avec celle du Chapitre 6, pour laquelle l'utilisation de l'anglais n'était pas souhaitable (voir discussion page 151).

des durées équivalentes (3120 ms ±186 pour le néerlandais, 3040 ms ±292 pour le japonais, $F(1,39) = 1.1, p = 0.3$). Elles ont en revanche des fréquences fondamentales moyennes différentes (216 Hz ±19 pour le néerlandais, 235 Hz ±15 pour le japonais, $F(1,39) = 11.8$, $p = 0.001$). Cette différence a été compensée à la resynthèse dans les expériences 8 et 9, mais pas dans l'expérience 7[5]. Nous verrons cependant que cela n'a pas eu d'influence sur la discrimination.

Ce sont ces mêmes phrases qui ont été utilisées dans toutes les expériences, qui ne diffèrent donc que par le traitement que les phrases ont subi (resynthèse ou autre). Ces phrases ont été divisées en deux ensembles égaux comprenant deux locutrices par langue, l'un servant à la phase d'habituation et l'autre à la phase de test (sauf dans l'expérience 11, cf. p. 139).

5.2.2 Participants

Les expériences se déroulent à la maternité Port-Royal, à l'hôpital Cochin, avec l'agrément du Comité Consultatif de Protection des Personnes dans la Recherche Bio-médicale de Paris-Cochin. Les bébés sont recrutés auprès de leur mère pendant leur séjour dans le service des suites de couche de la maternité. Le consentement libre et éclairé de la mère ou du père est recueilli avant le début du test. Les bébés sont sélectionnés au vu de leur dossier médical selon les critères suivants :

1. âge compris entre 2 et 5 jours de vie ;
2. terme supérieur ou égal à 38 semaines de gestation ;
3. poids supérieur ou égal à 2800g à la naissance ;
4. pas de souffrance à la naissance (score APGAR = 10 à 5 minutes) ;
5. examens médicaux normaux (notamment l'évaluation neurologique) à la naissance et à deux jours de vie ;
6. pas de séro-conversion à la rubéole ou à la toxoplasmose ;
7. pas de virus grave, ni d'alcoolisme, toxicomanie ou tabagisme poussé chez la mère ;
8. pas d'antécédents de surdité ni de problèmes neurologiques dans la famille ;
9. pas de néerlandais ni de japonais parlé par les parents.

Les bébés sont testés environ 3 heures après leur dernière tétée, à un moment où ils peuvent être maintenus calmes et éveillés, et où leur réflexe de succion est maximal.

5. Pour la simple raison que dans l'expérience 7 les phrases ne sont pas resynthétisées. Par ailleurs le problème ne se pose pas pour les expériences 10 et 11, qui ont recours à une intonation artificielle.

5.2.3 Procédure

La procédure utilisée est celle de succions de haute amplitude, décrite brièvement dans l'Annexe A.1. Son déroulement est contrôlé par un ordinateur PC et le programme `bbexpe`, développé au LSCP. Ce programme a été considérablement remanié avant le début de nos expériences, notamment pour intégrer la possibilité (inexistante dans les études antérieures) d'effectuer les expériences en double aveugle, c'est-à-dire sans que l'expérimentateur ne soit au courant de la condition expérimentale. Nous présentons maintenant en détail les paramètres particuliers qui ont servi dans nos expériences.

5.2.3.1 Groupes de sujets

Chaque expérience (sauf la n° 11, cf. p. 139) se compose de 4 conditions, correspondant aux groupes contrôle et expérimental et au contrebalancement de l'ordre d'apparition des langues (Tableau 5.1). Les bébés sont habitués à deux locutrices d'une langue ; ceux du groupe contrôle sont testés avec les deux autres locutrices de la même langue, alors que ceux du groupe expérimental, sont testés avec deux locutrices de l'autre langue.

TAB. 5.1 – *Conditions expérimentales pour les expériences 7–10.*

Habituation :	Néerlandais	Japonais
Groupe contrôle	née. → née.	jap. → jap.
Groupe expérimental	née. → jap.	jap. → née.

Chaque expérience est planifiée pour un nombre donné de sujets (généralement 32) et l'appariement entre les 4 conditions et les 32 sujets est tiré au sort une fois pour toutes par l'ordinateur[6] et stocké hors de la vue de l'expérimentateur.

Au début de chaque test, l'ordinateur attribue au bébé la condition qui lui revient dans la liste, sans l'afficher à l'écran. Lorsque les résultats d'un bébé sont rejetés (voir ci-dessous Section 5.2.3.3), la condition n'est pas révélée à l'expérimentateur et reste la même pour le bébé suivant.

5.2.3.2 Déroulement du test

Le dernier moyen par lequel l'expérimentateur pourrait prendre connaissance de la condition du sujet est tout simplement l'écoute des phrases pendant le test. Pour cette raison, pendant tout le test, l'expérimentateur porte

6. Par blocs de quatre pour préserver l'équilibre des conditions tout au long de l'expérience.

un casque anti-bruit contenant des écouteurs qui diffusent un bruit de masquage. Le bruit de masquage consiste en un son de parole continu composé des stimuli de l'expérience mis bout à bout et superposés en quatre flux. Par sa composition, ce bruit masque idéalement les stimuli joués pendant le test.

Le test se déroule dans une cabine insonorisée qui isole l'enfant des bruits extérieurs. Seul l'expérimentateur reste dans la cabine pendant le test avec l'enfant. Celui-ci est installé dans une chaise inclinée, et est invité à téter sur une tétine fermée maintenue par un bras flexible. La pression de l'air dans la tétine est mesurée par un transducteur de pression, puis amplifiée et transmise à l'ordinateur via une carte analogique/digital. L'ordinateur détecte les succions de l'enfant et calcule leur amplitude.

Pendant les deux premières minutes du test, l'enfant tète en silence sans recevoir de stimulation. L'ordinateur calcule alors un seuil de haute amplitude, de manière à ce que 75% des succions de ces deux minutes aient une amplitude supérieure au seuil[7]. Dans toute la suite du test, les succions dont l'amplitude dépassera le seuil seront dites des succions de haute amplitude (ou succions HA)[8]. L'ordinateur comptabilise le nombre de succions HA produites par minute.

À partir de la 3e minute commence la phase d'habituation. Chaque succion HA de l'enfant déclenche une phrase. Les succions HA se produisant pendant la diffusion d'une phrase ne déclenchent pas de phrase supplémentaire. Un intervalle minimal de 500 ms est imposé entre deux phrases consécutives. Les phrases sont jouées directement depuis le disque dur de l'ordinateur par une carte son, puis amplifiées et diffusées par deux enceintes situées de part et d'autre de l'enfant. L'ordinateur joue les phrases correspondant à la condition et à la phase dans un ordre aléatoire. Lorsque les 10 phrases correspondantes ont été jouées, l'ordinateur effectue un nouveau tirage aléatoire et rejoue les mêmes phrases, et ainsi de suite jusqu'à la fin de la phase.

La phase d'habituation se poursuit jusqu'à ce que le critère d'habituation soit atteint. Cela se produit lorsque pendant deux minutes consécutives, le nombre de succions HA est de 25% inférieur au nombre maximal de succions observées en une minute depuis le début de la phase de test[9]. On impose également que la phase d'habituation dure au minimum 5 minutes, afin d'être sûr que le bébé a eu une exposition suffisante aux stimuli d'habituation. Dans les expériences 7–11, la phase d'habituation a duré en moyenne 8.7 minutes

7. Sauf dans l'expérience 11, cf. p. 139.

8. Ceci sert à éliminer la partie la plus bruitée du signal : mâchonnements et mouvements des mâchoires involontaires, et à ne conserver que les succions les plus actives.

9. Sans compter la première minute, qui n'est pas considérée comme représentative. Il faut de plus que ce nombre maximal de succions soit au moins égal à 20, pour éviter que le critère soit atteint alors que le bébé tète trop peu.

±3.2.

Une fois le critère d'habituation atteint, commence la phase de test, qui dure 4 minutes. Pendant cette phase, les phrases jouées sont les phrases de test correspondant à la condition du bébé. A la fin de cette phase, le test est terminé.

5.2.3.3 Conditions de délai et de rejet

La plupart du temps, l'expérimentateur se contente de surveiller le bébé sans intervenir, dans la mesure où la tétine est maintenue par le bras flexible. Lorsque le bébé perd la tétine, ou commence à s'agiter ou à pleurer, ou au contraire s'endort, l'expérimentateur intervient pour lui remettre la tétine en place, lui tenir les mains ou le réveiller. Dans ce cas il appuie sur la touche F1 du clavier, qui signale à l'ordinateur l'intervention (ou tout autre événement perturbant le déroulement de l'expérience).

Les données habituellement analysées sont les taux de succion HA pendant les deux minutes précédant et les deux minutes suivant le changement de phase. Pour minimiser le bruit dans les données, il est important de s'assurer que, pendant ces 4 minutes, 1° l'expérimentateur n'est pas intervenu, et 2° le bébé était bien éveillé et a entendu des phrases. Le programme `bbexpe` intègre un certain nombre de conditions de délai et de rejet qui permettent d'en tenir compte.

Conditions de délai.

1. si l'expérimentateur a appuyé sur la touche F1 pendant l'une des deux minutes qui ont précédé ;
2. si le bébé n'a entendu aucune phrase pendant l'une des deux minutes qui ont précédé ;

alors la phase de test ne peut commencer, même si le critère d'habituation est atteint. La phase d'habituation continue, jusqu'à ce qu'à la fois le critère soit atteint et qu'aucune des conditions de délai ne soit vérifiée.

Conditions de rejet.

1. si l'expérimentateur appuie sur la touche F1 pendant l'une des deux premières minutes de la phase de test ;
2. si le bébé n'a pas entendu de phrase pendant l'une des deux premières minutes de la phase de test ;
3. si la phase d'habituation a déjà duré 20 minutes ;

alors l'ordinateur interrompt l'expérience et classe les données du bébé comme « rejetées ». De plus, l'expérimentateur peut lui-même décider d'interrompre le test pour l'une des raisons suivantes :

1. le bébé refuse la tétine ;
2. il ne se maintient pas assez éveillé ou ne tète pas assez ;
3. il pleure ou est trop agité.

Dans ce cas les données du bébé sont également rejetées. Au total, les données d'environ la moitié des bébés sont rejetées.

Nous allons maintenant passer aux expériences proprement dites.

5.3 Exp. 7 : néerlandais-japonais en parole naturelle

Cette première expérience va tester la discrimination du néerlandais et du japonais en parole naturelle, sans aucune modification des phrases[10].

5.3.1 Stimuli

Avant d'utiliser la resynthèse de parole, nous avons par prudence souhaité commencer par effectuer une expérience maximisant les chances d'observer une discrimination, en n'effectuant aucun traitement sur la parole. Nous avons donc utilisé les 40 phrases néerlandaises et japonaises telles quelles.

5.3.2 Participants

Dans cette expériences, 32 bébés ont été testés avec succès, 18 garçons et 14 filles, dont les caractéristiques sont données dans le tableau 5.2. Vingt-neuf d'entre eux étaient issus de familles françaises monolingues, 2 de familles où l'on parle le français plus une ou plusieurs autres langues, 1 d'une famille où l'on ne parle pas le français.

De plus, les résultats de 42 bébés ont été rejetés pour les raisons suivantes : rejet de la tétine (1), pas assez éveillé/succion insuffisante avant le changement (12), pleurs/agitation (9), n'atteint pas le critère d'habituation (9), s'endort/succion insuffisante après le changement (6), perd la tétine après le changement (4), panne d'ordinateur (1).

10. Deux autres expériences liées à l'expérience 7 ont été effectuées. Comme elles n'apportent pas de résultats significatifs dans le cadre de ce chapitre, elles sont décrites en Annexe C.

TAB. 5.2 – *Caractéristiques des nouveau-nés de l'Expérience 7.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	40	154	67	22
Terme (semaines + jours)	38	41+5	40	1+1
Poids (g)	2820	4260	3530	402

5.3.3 Résultats

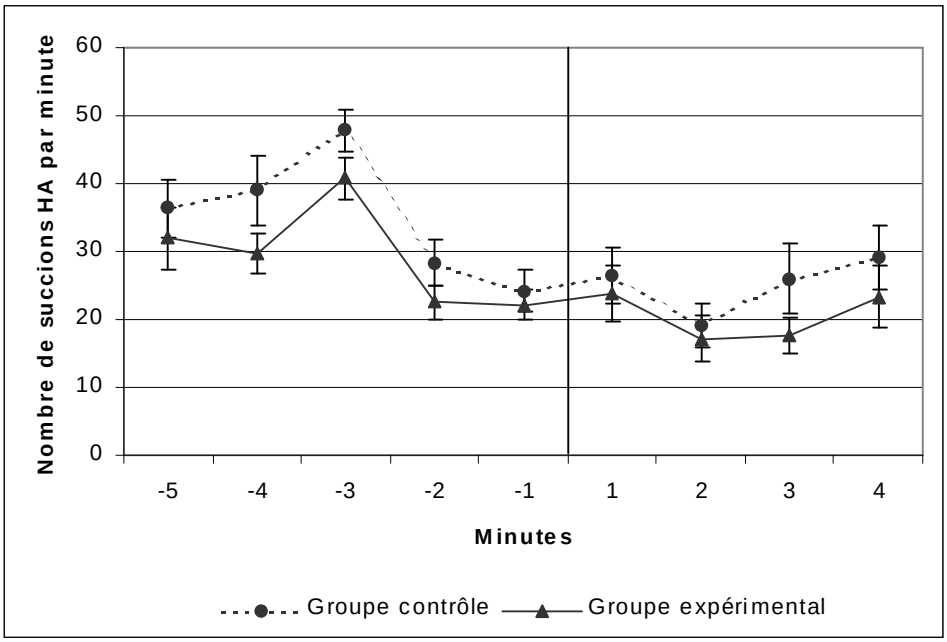

FIG. 5.1 – *Exp. 7 : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle. Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

La Figure 5.1 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet. On observe que les nombres de succions sont similaires pour les deux groupes tout au long du test, avec les bébés du groupe contrôle qui semblent téter légèrement plus que ceux du groupe expérimental. Le changement ne semble pas provoquer des effets différents dans les deux groupes. Si les bébés discriminaient les deux langues, on s'attendrait typiquement à ce que les bébés du groupe expérimental, stimulés par la nouveauté de la deuxième langue, se mettent à téter significativement

plus que ceux du groupe contrôle durant les minutes suivant le changement. Ici, cela ne semble pas être le cas.

Pour les analyses statistiques nous considérons les variables suivantes.

variables indépendantes : *groupe* (expérimental ou contrôle), *habituation* (néerlandais ou japonais en phase d'habituation) ;

variables dépendantes : *ante5* (nombre moyen de succions HA sur les 5 minutes précédant le changement), *ante2* (idem sur les 2 minutes précédant le changement), *post2* (idem sur les 2 minutes suivant le changement).

Ici, il n'y a pas d'effet des facteurs *groupe* ni *habituation* sur la variable *ante5*, c'est-à-dire que les bébés ne tètent pas significativement différemment dans la phase d'habituation selon qu'ils appartiennent au groupe contrôle ou expérimental [$F(1,31) = 2.6$, $p = 0.12$] ou entendent le néerlandais ou le japonais [$F(1,31) < 1$].

Pour tester s'il y a discrimination ou pas, on effectue une analyse de covariance dans laquelle la variable dépendante est *post2*, la covariable *ante2*, et la variable indépendante le *groupe*[11]. Ici, il n'y a pas d'effet significatif de *groupe* [$F(1,29) < 1$].

5.3.4 Discussion

Les nouveau-nés n'ont donc pas discriminé le néerlandais du japonais, alors que les stimuli étaient de la parole naturelle. Ce résultat est contraire à nos attentes, qui se basaient sur la discrimination anglais/japonais par les nouveau-nés (Nazzi et coll., 1998) et la similarité entre néerlandais et anglais. Il nous faut donc chercher quelles différences entre notre expérience et celles-ci pourraient être la cause de ce résultat.

Il se trouve que notre expérience est la première expérience de discrimination de langues à exposer des nouveau-nés à 4 voix différentes. En effet, dans les expériences de Mehler et coll. (1988) et de Moon et coll. (1993), les histoires étaient à chaque fois prononcées par une locutrice bilingue. Les

11. L'analyse standard de ce type de résultats consiste à former une variable *déshabituation* qui ici serait égale à ($post2 - ante2$), et à chercher l'effet du groupe sur cette variable. Cette analyse suppose que $post2 = ante2$ + l'effet de groupe + l'erreur. L'analyse de covariance suppose seulement que $post2 = \alpha \times ante2$ + l'effet de groupe + l'erreur, où α est un coefficient quelconque. Celle-ci fait donc moins d'hypothèses sur le lien entre les minutes précédant et les minutes suivant le changement. Cette analyse statistique est donc plus juste et plus puissante (pour plus de détails, voir Kennedy & Bush, 1985; Christophe, 1993; Christophe et coll., 1994). Dans toutes nos expériences, nous avons également effectué l'analyse standard de déshabituation : elle n'a jamais conduit à des conclusions différentes de l'analyse de covariance et n'est donc pas mentionnée.

bébés n'entendaient donc qu'une seule voix pendant toute l'expérience. Dans les expériences de Nazzi et coll. (1998), 4 locutrices par langue étaient utilisées, mais la parole était filtrée passe-bas, ce qui a pour effet de détériorer les caractéristiques des voix, et d'en atténuer grandement les différences.

Quelques études se sont déjà penchées sur les effets de la variabilité des locuteurs sur la perception de la parole par les nourrissons. Kuhl (1979b, 1983) a trouvé que les bébés de 6 mois peuvent catégoriser des voyelles sans être gênés par une multiplicité de voix. Jusczyk, Pisoni, et Mullenix (1992) ont retrouvé un résultat similaire en testant des bébés de 2 mois sur des discriminations de syllabes. Ils ont cependant montré que, dans une condition plus difficile où les bébés devaient mémoriser les syllabes pendant 2 minutes, ils n'arrivaient plus à discriminer lorsque les syllabes étaient prononcées par plusieurs locuteurs. Il semble donc que pour les nourrissons les plus jeunes, la variabilité des locuteurs puisse poser un problème.

La resynthèse de parole nous offre le moyen idéal pour tester cette hypothèse. En effet, quel que soit le nombre de locutrices dans les phrases d'origine, une fois resynthétisées, nos phrases ne comportent plus qu'une seule voix : celle de la personne qui a prononcé la base de diphones utilisée par le synthétiseur vocal (voir Section 3.2.2). Si la multiplicité des voix est effectivement la cause de l'absence de discrimination dans notre expérience, on doit donc prédire que la discrimination redevient possible dès lors que les phrases sont resynthétisées (sous réserve qu'elles préservent les indices nécessaires à la discrimination, bien sûr). Nous allons donc tester cette hypothèse dans la prochaine expérience.

5.4 Exp. 8 : néerlandais-japonais en *saltanaj*

5.4.1 Stimuli

Les phrases utilisées dans l'expérience 7 ont été resynthétisées selon la transformation *saltanaj* (cf. Section 3.2.3), en utilisant la base de diphones du français fournie avec le logiciel MBROLA. Comme cette base de diphones a été réalisée à partir d'une voix d'homme, nous avons utilisé un paramètre de MBROLA (longueur du tract vocal) pour « féminiser » la voix[12]. Cette manipulation a eu pour effet annexe de sur-échantillonner les phrases à 17500 Hz,

12. Dans toutes les expériences de discrimination de langues effectuées à ce jour (et dans quasiment toutes les expériences effectuées sur le nourrisson), seules des voix de femmes ont été utilisées. Bien que cet usage frôle la superstition, nous avons préféré nous y conformer pour ne pas introduire une différence de plus entre nos expériences et celles des autres. De plus, les phrases d'origine étant prononcées par des femmes, il paraissait naturel de préserver cette caractéristique.

qui ont donc été jouées à cette fréquence. Par ailleurs, pour remédier à l'écart de fréquence fondamentale moyenne entre les deux langues, nous avons multiplié tous les points de F_0 du néerlandais par 1.04, et ceux du japonais par 0.96 [13].

Nous avons conservé le principe d'habituation à deux locutrices et de test sur de autres locutrices. Bien que cela puisse semble inutile dans la mesure où la resynthèse supprime les différences de voix, il reste possible que les locutrices aient des prosodies légèrement différentes ; ce facteur doit donc être contrôlé.

5.4.2 Participants

Trente-deux bébés ont été testés avec succès, 16 garçons et 16 filles, dont les caractéristiques sont données dans le tableau 5.3. Vingt-cinq d'entre eux étaient issus de familles françaises monolingues, 3 de familles où l'on parle le français plus une ou plusieurs autres langues, 4 de familles où l'on ne parle pas le français.

TAB. 5.3 – *Caractéristiques des nouveau-nés de l'Expérience 8.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	38	136	80	25
Terme (semaines + jours)	38	41+5	39+6	1+1
Poids (g)	2830	4750	3508	477

De plus, les résultats de 20 bébés ont été rejetés pour les raisons suivantes : pas assez éveillé/succion insuffisante avant le changement (6), pleurs/agitation (4), n'atteint pas le critère d'habituation (1), s'endort/succion insuffisante après le changement (3), perd la tétine après le changement (6).

5.4.3 Procédure

À titre exploratoire, nous avons voulu introduire un deuxième changement pendant le test, comme cela se fait avec les nourrissons de 2-3 mois (Hesketh et coll., 1997; Christophe & Morton, 1998). Plus précisément, après le 1er changement et les 4 minutes qui suivent, nous avons autorisé un éventuel deuxième changement, de nature différente du premier (contrôle ou expérimental), permettant ainsi d'obtenir une mesure intra-sujet de la discrimi-

13. En réalité, l'absence de discrimination dans l'expérience 7, où la fréquence fondamentale n'était pas corrigée, prouve bien qu'un écart de F_0 moyenne d'environ 10% n'est pas exploité par les nouveau-nés pour discriminer les deux langues.

nation. Ainsi, les bébés qui ont subi un premier changement de locutrices (contrôle) ont eu la possibilité de subir un deuxième changement de langue (expérimental), et vice-versa, afin que chaque bébé puisse être son propre contrôle.

Le deuxième changement a été soumis au même critère d'habituation que le premier (deux minutes consécutives avec 25% de succions en moins par rapport à la minute maximale), et aussi aux mêmes critères de délai et de rejet. Un délai minimal de 4 minutes a été imposé entre les deux changements. Vis-à-vis du premier changement, l'introduction du deuxième n'a donc absolument rien modifié. Tout s'est déroulé comme lors d'une expérience à un seul changement, plus la possibilité, pour les bébés qui le pouvaient et qui vérifiaient toutes les conditions, de subir le deuxième changement. Nous avons arrêté l'expérience lorsque 32 bébés, 16 par groupe, ont passé le premier changement avec succès.

5.4.4 Résultats

5.4.4.1 Premier changement

La Figure 5.2 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet. Ici, on observe un effet de groupe assez important : pendant les 2 minutes suivant le changement, les succions du groupe expérimental augmente, alors que celles du groupe contrôle diminue, suggérant que les bébés du groupe expérimental ont bien réagi au changement de langue.

Les bébés n'ont pas tété différemment pendant l'habituation selon leur groupe ni selon leur langue d'habituation [respectivement $F(1,31) < 1$ et $F(1,31) = 1.8$, $p = 0.19$]. L'analyse de covariance donne un effet significatif du groupe [$F(1,29) = 6.3$, $p = 0.018$], montrant que les bébés du groupe expérimental ont augmenté leurs succions pendant les 2 minutes suivant le changement significativement plus que les bébés du groupe contrôle. On peut donc en conclure que les bébés du groupe expérimental ont distingué les deux langues.

5.4.4.2 Deuxième changement

Sur 32 bébés ayant passé correctement le premier changement, seuls 11 ont réussi à passer le deuxième. Les 21 autres ont été rejetés pour les raisons suivantes : pas assez éveillé/succion insuffisante avant le 2e changement (7), pleurs/agitation (2), n'atteint pas le critère d'habituation au bout de 20

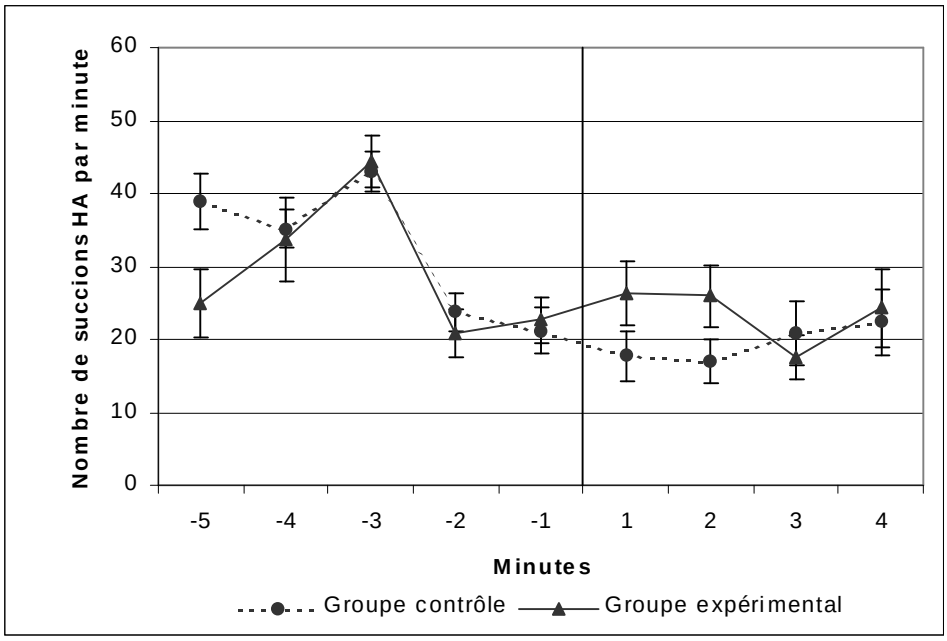

FIG. 5.2 – *Exp. 8 : Discrimination néerlandais-japonais par des nouveau-nés –* saltanaj. *Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

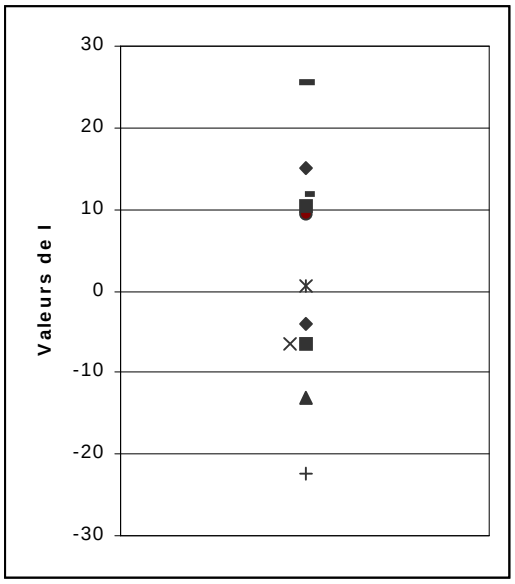

FIG. 5.3 – *Exp 8 : Valeurs de l'indice I pour les 11 bébés ayant effectué deux changements.*

minutes de test (6), s'endort/succion insuffisante après le 2e changement (4), perd la tétine après le 2e changement (2). La faible proportion de bébés ayant réussi à passer les deux changements montre déjà que cette procédure ne peut s'imposer comme la procédure standard pour tester les nouveau-nés, car elle conduirait à rejeter beaucoup trop de données (ici, 41 bébés sur 52 en comptant ceux qui ont été rejetés dès le premier changement). Pour ces 11 bébés, un indice I de discrimination intra-sujet a été calculé selon la formule suivante (Hesketh et coll., 1997) :

$$I = (post2_{expe} - ante2_{expe}) - (post2_{cont} - ante2_{cont})$$

où $post2$ et $ante2$ réfèrent à la moyenne du nombre de succions HA sur les deux minutes respectivement suivant et précédant un changement, et où les indices *cont* et *expe* réfèrent au type de changement (contrôle ou expérimental). Ainsi, un bébé qui aura plus augmenté ses succions lors du changement de langue que lors du changement de locutrices aura un I strictement positif. La Figure 5.3 donne les valeurs de I pour les 11 bébés ayant effectué les deux changements. Les scores de déshabituation $post2 - ante2$ se répartissent comme indiqué dans le Tableau 5.4.

TAB. 5.4 – *Exp. 8: Répartition des scores de déshabituation* $post2 - ante2$ *en fonction de l'ordre et de la nature du changement.*

	1er changement	2e changement
contrôle	0 (4)	+4.5 (7)
expérimental	+6.6 (7)	+0.8 (4)

Entre parenthèses le nombre d'observations par case.

On observe une très légère tendance dans le bon sens : les bébés ayant subi le changement contrôle en premier augmentent leurs succions de 0.8 au changement expérimental, alors que ceux ayant subi le changement expérimental en premier perdent 2.1 succions au changement contrôle. Néanmoins, la valeur moyenne de I n'est que de 1.9 ± 14, et n'est pas significativement différente de 0 [$t(10)<1$].

Ainsi, d'après l'indice de discrimination I, ces bébés n'ont pas plus réagi au changement de langue qu'au changement de locutrices. Pourtant, en ce qui concerne le 1er changement, ces bébés ont réagi comme l'ensemble des 32 bébés testés, et conformément à l'effet de discrimination observé : ceux du groupe expérimental ont augmenté leurs succions (+6.6), mais pas ceux du groupe contrôle. C'est donc lors du deuxième changement que leur comportement est inattendu: ce sont alors les bébés qui subissent un changement

de locutrices qui augmentent le plus leurs succions. On peut analyser ce résultat en disant que ces bébés se sont comportés normalement au premier changement et ont exhibé un effet de persévérance au deuxième changement : notamment, les bébés ayant subi un premier changement contrôle se sont définitivement lassés, et n'ont pu ultérieurement augmenter leurs succions (ils sont d'ailleurs moins nombreux à avoir pu patienter jusqu'au deuxième changement).

Ces analyses suggèrent qu'il n'y a pas grand-chose à tirer d'un deuxième changement avec les nouveau-nés. En ajoutant à cela que trop peu de bébés réussissent à passer les deux changements, on peut conclure qu'une telle procédure n'est pas adaptée aux nouveau-nés.

5.4.5 Discussion

Néanmoins, l'analyse des données des 32 bébés qui ont subi le 1er changement montre sans ambiguïté leur réaction au changement de langues. On peut donc en déduire que :

1. comme on s'y attendait, les nouveau-nés français sont capables de discriminer le néerlandais du japonais ;
2. ils sont capables d'effectuer cette discrimination alors que les phrases sont resynthétisées. La resynthèse étant de type *saltanaj*, on peut affirmer que la discrimination n'est pas due aux propriétés lexicales ni phonétiques des deux langues, mais plutôt à leurs propriétés prosodiques au sens large.
3. comme nous l'avions supposé, l'absence de discrimination dans l'expérience 7 était probablement due à la multiplicité des voix[14]. Ce résultat confirme l'observation de Jusczyk et coll. (1992), selon laquelle la variabilité des voix induit un « coût » dans le traitement de la parole par le nourrisson.

Ayant obtenu un premier résultat positif, nous allons maintenant essayer de mieux cerner les limites de la capacité des nouveau-nés à discriminer les langues.

14. L'interaction *groupe* × *expérience* n'est que marginale [$F(1,59) = 2.6$, $p = 0.11$]. Remarquons néanmoins que ce type d'interaction est très rarement significatif dans les expériences sur les nouveau-nés, dû à leur faible puissance statistique. Notamment aucune interaction significative n'a été rapportée pour des expériences directement comparables (Mehler et coll., 1988; Nazzi et coll., 1998).

5.5 Exp. 9 : néerlandais-japonais en *saltanaj* à l'envers

Mehler et coll. (1988) ont montré que les nouveau-nés français ne distinguaient plus le français du russe lorsque les histoires étaient jouées à l'envers. Ce résultat nous semble intéressant à répliquer pour plusieurs raisons. Un certain nombre d'études ont suggéré que dès les premières étapes du traitement de la parole, la parole à l'envers est traitée différemment par le cerveau de la parole normale. Par exemple, la parole jouée à l'envers n'active qu'un sous-ensemble des zones du cerveau activées par l'écoute de la parole normale, même lorsque celle-ci est dans une langue inconnue, et bien que les sujets prétendent n'avoir pas remarqué la différence (Perani et coll., 1996). De plus, même des enfants de 6 mois ont des potentiels évoqués différents à l'écoute de mots joués à l'envers par rapport à des mots normaux (Neville & Mills, 1997). Ces résultats suggèrent que la perception de la parole repose sur certaines propriétés spécifiques du signal, qui ne sont pas vérifiées par la parole à l'envers. Il est donc important de savoir si la sensibilité à ces propriétés est présente dès la naissance. De plus, l'utilisation de la resynthèse de parole peut nous renseigner sur la nature même de ces propriétés.

5.5.1 Stimuli

Nous avons utilisé les mêmes phrases néerlandaises et japonaises que pour l'expérience 8, resynthétisées selon la transformation *saltanaj*, mais jouées à l'envers (l'ordre des échantillons a été inversé pour chaque phrase).

5.5.2 Participants

Trente-deux bébés ont été testés avec succès, 16 garçons et 16 filles, dont les caractéristiques sont données dans le tableau 5.5. Vingt-deux d'entre eux étaient issus de familles françaises monolingues, 9 de familles où l'on parle le français plus une ou plusieurs autres langues, 1 d'une famille où l'on ne parle pas le français.

TAB. 5.5 – *Caractéristiques des nouveau-nés de l'Expérience 9.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	40	122	63.9	22.4
Terme (semaines + jours)	38	41+5	39+6	1
Poids (g)	2750	3940	3360	343

De plus, les résultats de 7 bébés ont été rejetés pour les raisons suivantes : pleurs/agitation (1), s'endort/succion insuffisante après le changement (4), perd la tétine après le changement (2)[15].

5.5.3 Procédure

Nous sommes revenus à la procédure à un seul changement utilisé pour l'expérience 7.

5.5.4 Résultats

La Figure 5.4 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet.

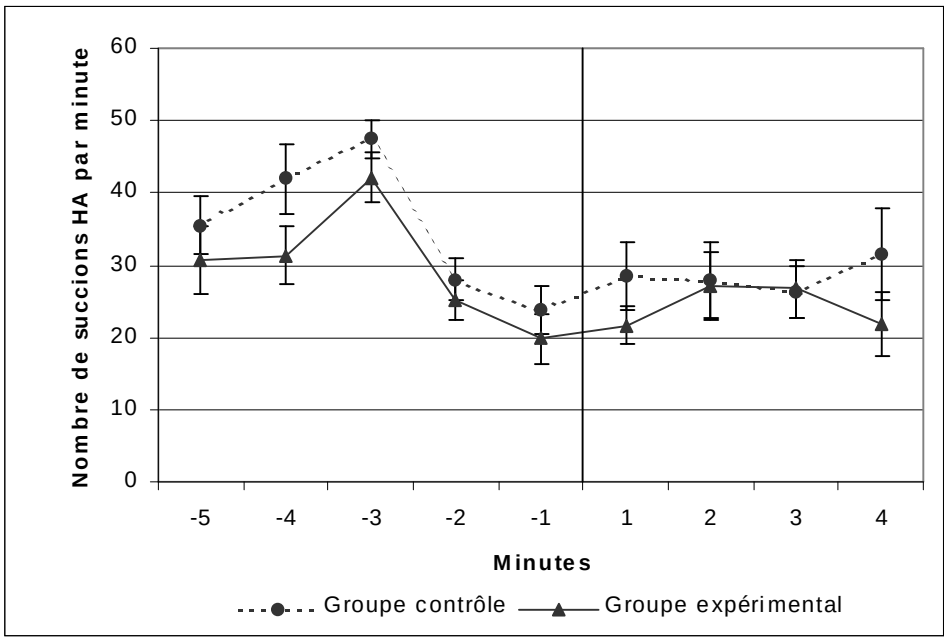

FIG. 5.4 – *Exp. 9 : Discrimination néerlandais-japonais par des nouveau-nés* – saltanaj *à l'envers. Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

Les bébés n'ont pas tété différemment pendant l'habituation selon leur groupe ni selon leur langue d'habituation [$F(1,31) = 2.7$, $p = 0.11$ et

15. Le lecteur attentif aura remarqué que le nombre de bébé rejetés varie grandement d'une expérience à l'autre (42 dans l'expérience 7, 7 dans l'expérience 9). Nous n'avons a priori pas d'hypothèse autre que le hasard pour expliquer ces variations.

$F(1,31) < 1$ respectivement]. L'analyse de covariance ne donne aucun effet significatif du groupe [$F(1,29) < 1$], montrant que les bébés du groupe expérimental n'ont pas augmenté leurs succions consécutivement au changement de langue. Par ailleurs, l'interaction *groupe* × *expérience* entre l'expérience 8 et celle-ci est marginalement significative [$F(1,59) = 3.6$, $p = 0.06$].

5.5.5 Discussion

Cette expérience confirme donc, avec une autre paire de langues et un autre type de stimuli, que la parole à l'envers ne permet pas au nouveau-né de discriminer les langues, alors que la discrimination est possible lorsque les mêmes phrases sont jouées à l'endroit. Comment cela se fait-il ? Si l'on reprend à la lettre les hypothèses de notre modèle de la perception du rythme de parole (Section 2.3), on devrait prédire que la discrimination est aussi facile à l'envers qu'à l'endroit, dans la mesure où les intervalles vocaliques et consonantiques, et les statistiques qui en découlent sont les mêmes, que l'on prenne la parole dans un sens ou dans l'autre.

En fait, on peut tout de même voir au moins deux explications pour lesquelles notre modèle serait inopérant sur la parole à l'envers.

- La première, c'est que la segmentation en intervalles consonantiques et vocaliques elle-même pourrait échouer sur la parole à l'envers. Plus précisément, le mécanisme qui reconnaît les consonnes et les voyelles dans le cerveau humain pourrait reposer de manière cruciale sur certaines propriétés des phonèmes, qui ne seraient pas conservées dans les mêmes phonèmes à l'envers. De fait, certains phonèmes sont notoirement asymétriques par rapport au temps : il s'agit des occlusives. À l'écoute de la parole à l'envers, celles-ci ne sonnent plus comme des phonèmes, mais plutôt comme des bruits étrangers à la langue. Dans le spectrogramme, les occlusives voisées ont généralement des transitions de formant qui ne sont pas symétriques ; quant aux occlusives non voisées, leur explosion est toujours précédée de leur fermeture, ces deux événements ne pouvant intervenir dans l'ordre inverse. Il est donc plausible que ces propriétés typiques des occlusives, qui sont altérées dans la parole à l'envers, provoquent l'échec de la segmentation consonne/voyelle.
- La deuxième, c'est que la parole à l'envers pourrait ne pas être dirigée vers le module de segmentation consonne/voyelle et d'extraction du rythme. Il pourrait en effet exister un « filtre » dans les entrées auditives qui séparerait la parole des autres sons et bruits de l'environnement, qui eux ne nécessitent pas un traitement spécifiquement linguistique. Comment un tel filtre pourrait-il fonctionner ? Par exemple en essayant

de détecter certaines propriétés typiques de la parole, comme celles que nous avons mentionnées ci-dessus : les formants et le sens de leurs transitions, les fermetures et les explosions, et l'ordre dans lequel elles se produisent. Une autre propriété générale de la parole qui est inversée dans la parole à l'envers est la déclinaison : dans toutes les langues du monde, la fréquence fondamentale a tendance à descendre au cours de la phrase. Dans la parole à l'envers, elle monte. Les stimuli qui ne vérifient pas cet universel linguistique pourraient donc être « filtrés ».

Nous sommes donc amenés à supposer que, sous une forme ou sous une autre, le cerveau des nouveau-nés « connaît » (sans doute de manière innée) certaines propriétés caractéristiques de la parole. Un tel dispositif serait sans aucun doute un atout pour l'acquisition du langage, car si l'enfant devait considérer tous les bruits de la nature comme des sources potentielles d'information linguistique, ses analyses et ses statistiques sur les propriétés de sa langue maternelle risqueraient d'en être fort altérées. Néanmoins, quelques questions restent encore en suspens : quelles sont précisément ces propriétés de la parole que l'enfant connaîtrait de manière innée ? Et sont-elles vraiment spécifiques à la parole ? Nous ne répondrons pas ici à la première question, mais nous proposerons à la Section 5.9 quelques pistes pour poursuivre les investigations. Quant à la seconde, nous tenterons d'y répondre au Chapitre 6. Nous allons maintenant reprendre notre quête du rythme de la parole.

5.6 Exp. 10 : néerlandais-japonais en *sasasa* avec intonation artificielle

Les stimuli de type *saltanaj* que nous avons utilisés dans l'expérience 8 préservent au moins autant d'indices que la parole filtrée : la prosodie, et peut-être quelques indications concernant la structure des syllabes. Ces stimuli tombent donc sous le coup des critiques que nous avons déjà faites à la parole filtrée (p. 65). Pour déterminer si le rythme est suffisant pour discriminer les langues, il nous faut maintenant restreindre encore les indices disponibles pour la discrimination.

Dans les expériences sur les adultes, nous avons utilisé à cet effet la transformation de type *sasasa plat*. Ces stimuli extrêmement monotones avaient déjà demandé de la part de nos sujets un effort d'attention et de concentration non négligeable. On a donc toutes les raisons de craindre l'effet qu'ils pourraient provoquer sur les nouveau-nés (l'endormissement, ou bien l'absence de réaction...). Nous allons donc essayer réduire les indices disponibles sans pour autant rendre les stimuli trop monotones.

5.6.1 Stimuli

Nous avons choisi d'utiliser des stimuli de type *sasasa* assortis d'une intonation artificielle ne pouvant servir à la discrimination entre les deux langues.

Pour cela nous avons « composé » cinq contours de $F0$ qui, sans être directement inspirés de phrases réelles, paraissent à l'oreille être des contours d'intonation possibles, au moins dans la langue française. Ces contours sont illustrés dans la Figure 5.5. Ils comportent tous une déclinaison régulière sur la fin, pour pouvoir s'adapter à des phrases de différentes longueurs.

Chaque contour a été appliqué à 4 phrases néerlandaises et 4 phrases japonaises. Le numéro du contour dans la figure réfère au nombre de syllabes que comportaient les phrases auxquelles il a été appliqué. Chaque phrase a alors été resynthétisée suivant la transformation *sasasa*, avec le contour de $F0$ qui lui a été attribué. Ainsi les phrases comportent une intonation qui n'est pas monotone, et qui est partiellement variée à travers les phrases, mais sans qu'elle puisse pour autant servir à distinguer les deux langues.

On peut craindre que l'application à une phrase d'un contour d'intonation qui n'est pas le sien engendre des phénomènes illégaux du point de vue linguistique. L'un des risques, dans ce cas, serait que les nouveau-nés se désintéressent des stimuli, ou ne parviennent pas à les analyser : on n'observerait alors aucune discrimination. Ce qui serait plus grave encore, c'est que de tels phénomènes n'apparaissent dans les phrases que de l'une des deux langues. Cela pourrait se passer si par exemple les intonations artificielles étaient plus adaptées à une langue qu'à l'autre. On pourrait craindre alors d'observer un effet de discrimination dû à un artefact : si le nouveau-né distinguait une langue à l'intonation normale d'une langue à l'intonation illégale, l'effet observé n'aurait rien à voir avec l'hypothèse du rythme. Pour en savoir un peu plus sur la manière dont les stimuli de l'expérience 10 sont perçus, nous avons effectué une expérience préliminaire sur des sujets adultes.

Expérience 10 bis : jugement par des sujets adultes des stimuli avec intonation artificielle.
Nous avons choisi de présenter les stimuli de l'expérience à des sujets adultes, dans la tâche suivante : les sujets écoutaient les phrases une par une, dans un ordre aléatoire, et devaient noter le caractère naturel de l'intonation sur une échelle de 0 à 5 (0 : très étrange – 5 : parfaitement naturelle). Les instructions sont données en Annexe B.8. Si les intonations artificielles étaient également adaptées (ou inadaptées) aux deux langues, on s'attendrait à obtenir le même score moyen pour les phrases des deux langues.

En fait, nous avons pressenti que les phrases néerlandaises seraient plus mal notées que les phrases japonaises, pour une raison n'ayant rien à voir

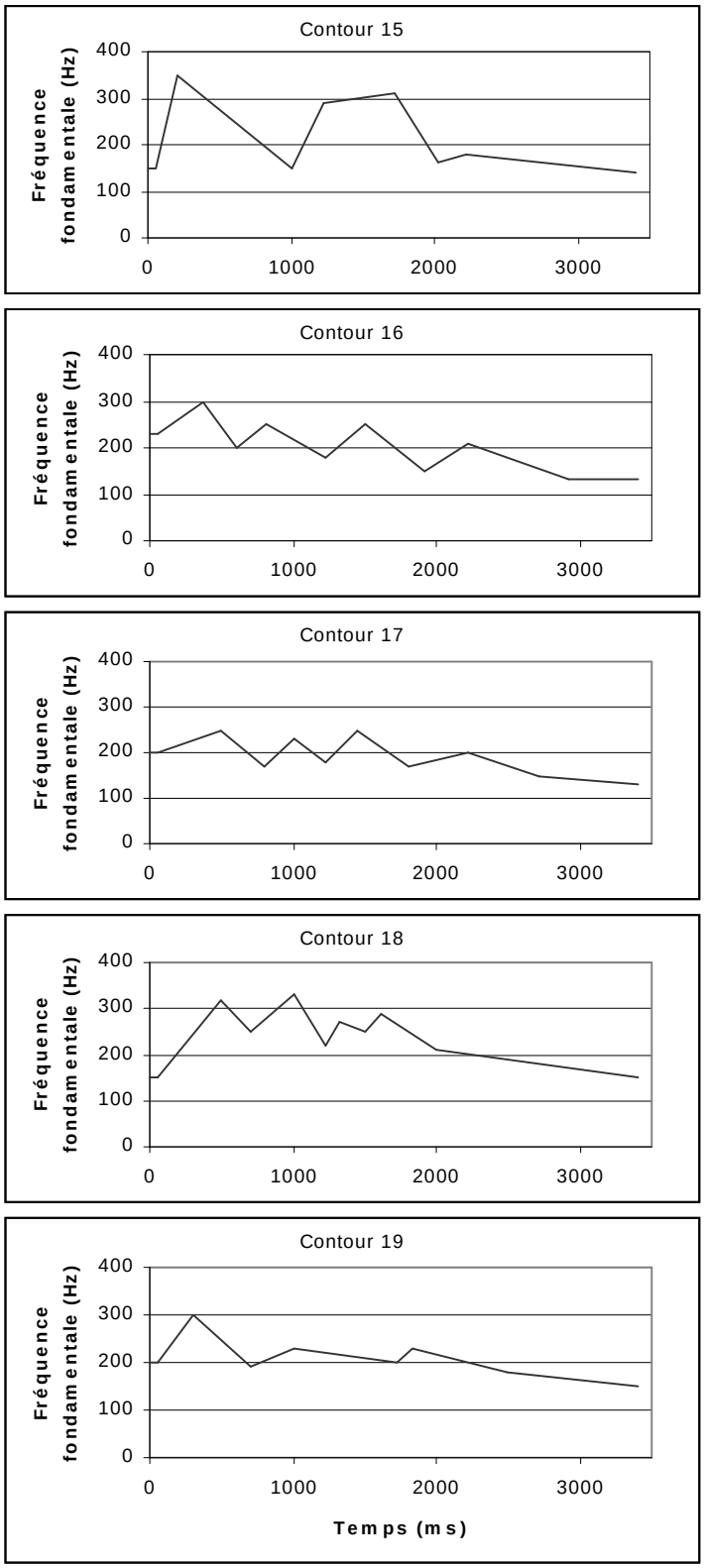

FIG. 5.5 – *Contours d'intonation utilisés dans l'expérience 10.*

avec l'intonation : en effet les phrases néerlandaises sont plus désagréables à écouter que les phrases japonaises, car possédant de nombreux groupes de consonnes, leurs versions *sasasa* comportent des [s] plus longs qui rendent la phrase plus chuintante. Bien sûr, les sujets avaient pour instruction de ne prêter attention qu'à l'intonation, mais il était prévisible qu'ils se laisseraient influencer par leur appréciation générale de la phrase. Pour contrôler cet effet, nous avons d'emblée prévu de tester aussi les sujets sur un deuxième bloc de phrases, dans leur version *saltanaj* avec l'intonation normale. L'ordre d'apparition des blocs a été contre-balancé à travers les sujets. Le dessin expérimental résulte donc du croisement de deux facteurs : le type de resynthèse (*saltanaj* ou *sasasa* avec intonation artificielle) et la langue (néerlandais ou japonais).

Nous avons recruté 12 sujets de langues maternelles diverses (français : 4, roumain : 1, espagnol : 3, allemand : 1, anglais : 2, néerlandais : 1). Ils s'agissait de 4 hommes et 8 femmes, âgés en moyenne de 34 ans. Ils ont participé gratuitement et ont été testés au LSCP.

La Figure 5.6 donne les scores moyens attribués aux phrases en fonction de la langue et du type de resynthèse.

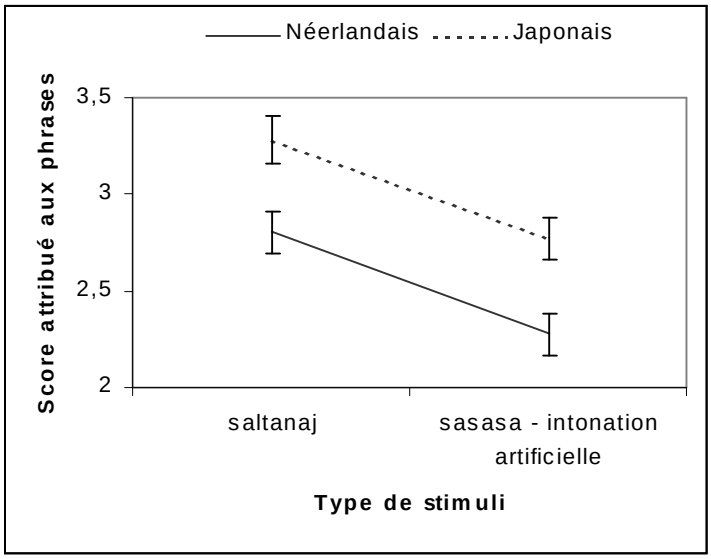

FIG. 5.6 – *Exp. 10bis : Jugements des sujets adultes sur l'intonation des phrases* sasasa *avec intonation artificielle et* saltanaj. *Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne.*

Les deux effets principaux sont significatifs : les phrases néerlandaises sont jugées avoir une intonation moins naturelle que les phrases japonaises

[$F(1,11) = 21.8, p = 0.001$], et les stimuli *saltanaj* sont également mieux jugés que les stimuli *sasasa* avec intonation artificielle [$F(1,11) = 8.39$, $p = 0.015$]. Comme nous l'avions prévu, la tâche joue donc en défaveur du néerlandais, et ce aussi bien avec les stimuli *saltanaj* que *sasasa*. Il semble donc que les sujets préfèrent les syllabes simples au syllabes complexes, indépendamment de la manière dont elles sont resynthétisées. Par ailleurs, les sujets préfèrent les stimuli de type *saltanaj* à ceux de type *sasasa*, ce qui est bien compréhensible étant donnée la monotonie de ces derniers. De toute évidence, ces effets ne reflètent pas la tâche demandée, qui ne concernait que l'intonation, mais ils sont sans doute inévitables. Le plus important, c'est qu'il n'y a aucune interaction entre la langue et le type de stimuli [$F(1,11) < 1$]. Ceci montre que les intonations artificielles ne sont pas moins adaptées à une langue qu'à l'autre, ce qui était ce dont nous voulions nous assurer. Nous pouvons donc maintenant utiliser ces stimuli sur les nouveau-nés.

5.6.2 Participants

Quarante bébés[16] ont été testés avec succès, 19 garçons et 21 filles, dont les caractéristiques sont données dans le Tableau 5.6. Vingt-six d'entre eux étaient issus de familles françaises monolingues, 11 de familles où l'on parle le français plus une ou plusieurs autres langues, 3 de familles où l'on ne parle pas le français.

TAB. 5.6 – *Caractéristiques des nouveau-nés de l'Expérience 10.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	40	132	65.9	23.1
Terme (semaines + jours)	38	41+6	40+1	1+1
Poids (g)	2810	4290	3428	424

De plus, les résultats de 20 bébés ont été rejetés pour les raisons suivantes : rejet de la tétine (2), pas assez éveillé/succion insuffisante avant le changement (7), pleurs/agitation (3), n'atteint pas le critère d'habituation (1), s'endort/succion insuffisante après le changement (4), perd la tétine après le changement (3).

16. Une analyse effectuée sur les 32 premiers bébés ayant montré des résultats un peu ambigus, nous avons testé huit bébés supplémentaires pour être certains des résultats.

5.6.3 Résultats

La Figure 5.7 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet. Cette fois, ce sont les bébés du groupe expérimental qui semblent téter légèrement plus que ceux du groupe contrôle. De ce fait, la différence apparente entre les deux groupes à la deuxième minute après le changement ne traduit pas une déshabituation du groupe expérimental : il faut prendre en compte la différence déjà présente sur les minutes précédentes. Les bébés du groupe expérimental ne semblent donc pas avoir réagi au changement de langues.

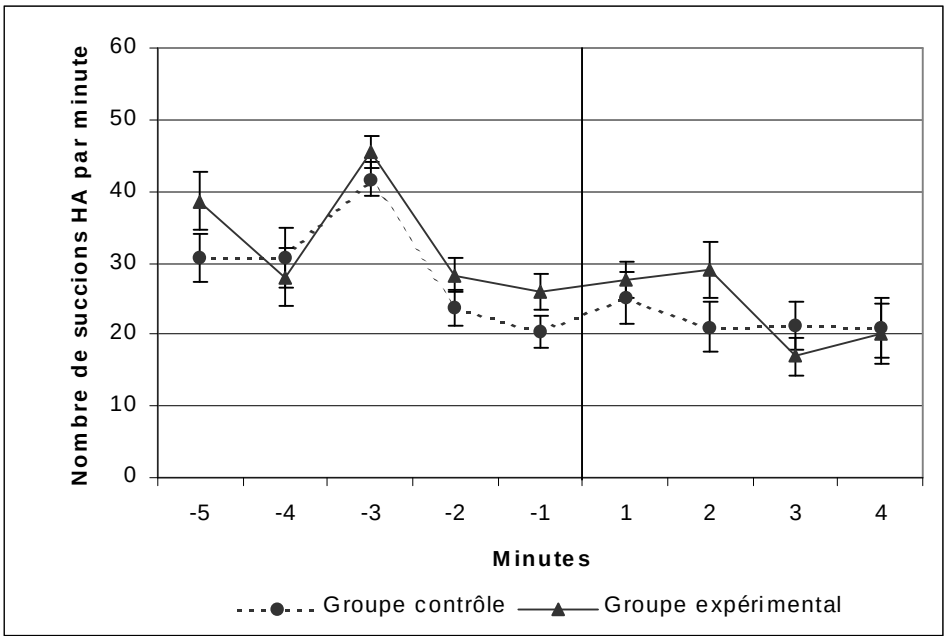

FIG. 5.7 – *Exp. 10 : Discrimination néerlandais-japonais par des nouveau-nés* – sasasa *avec intonation artificielle. Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

Les bébés n'ont pas tété différemment pendant l'habituation selon leur groupe [$F(1,39) = 1.5$, $p = 0.23$]. En revanche, les bébés qui ont entendu le néerlandais pendant la phase d'habituation ont tété significativement plus que ceux qui ont entendu le japonais [35.6 ± 8.3 succions HA/minute en moyenne sur les 5 minutes précédant le changement pour le néerlandais, 25.8 ± 9.1 pour le japonais, $F(1,39) = 12.8$, $p = 0.001$]. Pour prendre en compte cet effet, nous incluons le facteur *habituation* dans l'analyse de co-

variance. Nous trouvons que celui-ci n'a pas d'effet sur la variable *post2* [$F(1,35) < 1$] et, surtout, qu'il n'interagit pas avec le groupe [$F(1,35) < 1$]. Nous n'observons par ailleurs aucun effet de groupe [$F(1,35) < 1$], montrant que les bébés du groupe expérimental n'ont pas discriminé les deux langues.

Pour information, nous montrons en Figures 5.8 et 5.9 les nombres de succions HA par minute pour les bébés ayant écouté respectivement le néerlandais et le japonais en habituation.

On peut se demander quelle est la cause de la différence de succion entre les bébés habitués au néerlandais et ceux habitués au japonais. Dans la première expérience de Mehler et coll. (1988), un tel effet avait été interprété comme une préférence pour le stimulus qui avait engendré le plus de succions (les histoires françaises). Ici, on pourrait donc en déduire que les bébés ont préféré écouter le néerlandais au japonais. Ce résultat peut paraître étonnant dans la mesure où les sujets adultes ont eu la préférence opposée. Plutôt que de parler de préférence, on a peut-être plutôt intérêt à parler de stimulation : les bébés tètent d'autant plus qu'ils sont stimulés par leur environnement. On peut alors imaginer que les phrases néerlandaises, plus chuintantes, les aient maintenus plus éveillés et stimulés... Il reste que cet effet pourrait aussi tout simplement être un effet d'échantillonnage, c'est-à-dire que par pur hasard les bébés qui avaient la plus forte succion aient été assignés à l'écoute du néerlandais en habituation.

5.6.4 Discussion

Une interprétation possible de l'absence de discrimination est que les bébés sont incapables d'analyser le rythme de la parole, et que c'est en fait l'intonation qui est la cause des discriminations observées. Une interprétation alternative pourrait être que le rythme et l'intonation forment un tout indissociable, et que si les nouveau-nés peuvent effectuer des discriminations sur la base du rythme de parole, il faut néanmoins que les deux éléments soient réunis. Enfin, une troisième interprétation serait que les nouveau-nés sont peut-être capables d'analyser le rythme de parole, mais que les stimuli présentés n'ont pas éveillé leur intérêt, ou tout au moins n'ont pas suscité un traitement linguistique de leur part. Par exemple, on pourrait imaginer que les stimuli de type *sasasa*, comme la parole à l'envers, ne sont pas conformes aux propriétés typiques de la parole. De fait, aucune langue du monde n'a un inventaire phonétique réduit à une consonne et une voyelle. Pour reprendre l'hypothèse du filtre, les stimuli *sasasa* auraient pu être filtrés par le système auditif des nouveau-nés comme étant des bruits non linguistiques. Notre prochaine expérience va donc s'efforcer de tester cette hypothèse.

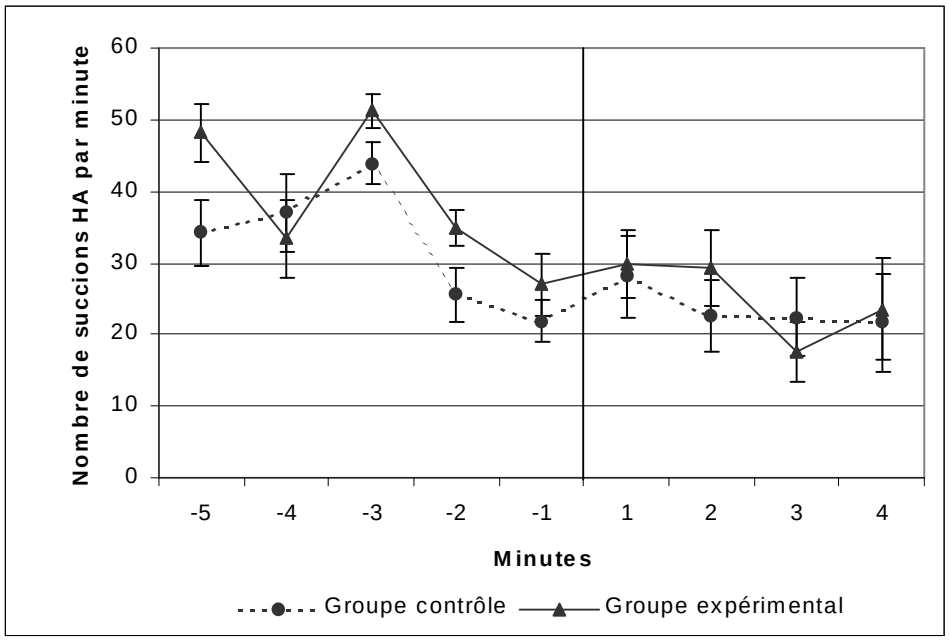

FIG. 5.8 – *Données des bébés de l'expérience 10 qui ont entendu le néerlandais en habituation (10 bébés par groupe).*

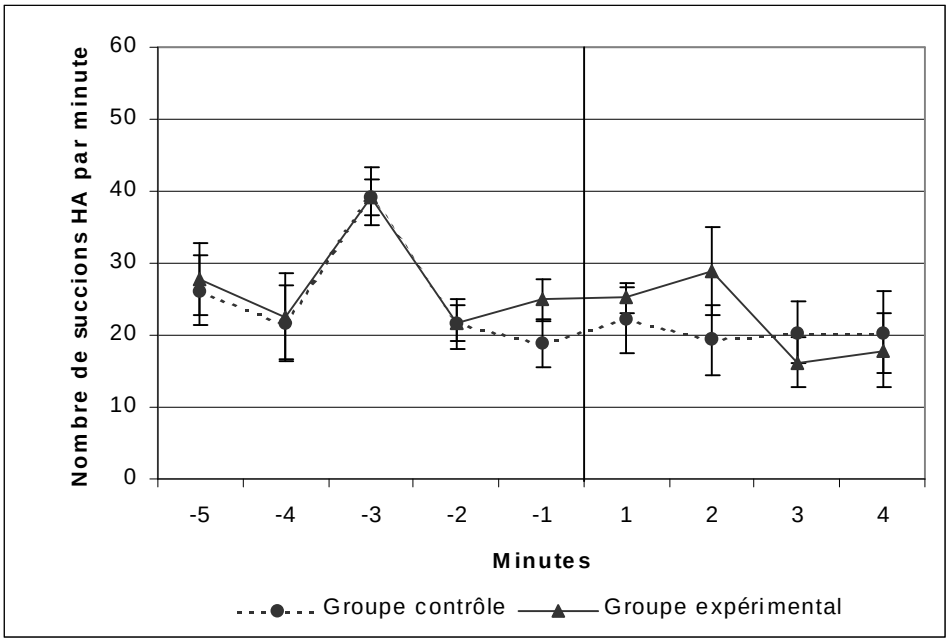

FIG. 5.9 – *Données des bébés de l'expérience 10 qui ont entendu le japonais en habituation (10 bébés par groupe).*

5.7 Exp. 11 : néerlandais-japonais en *saltanaj* avec intonation artificielle

Entre l'expérience 8 et l'expérience 10, deux changements sont intervenus : 1° la transformation des phonèmes est passée de *saltanaj* à *sasasa*, 2° l'intonation est passée de naturelle à artificielle. Visiblement, au moins l'un de ces deux changements provoque l'échec de la discrimination par les nouveau-nés. Nous allons donc faire marche arrière et n'effectuer qu'un seul des deux changements, celui concernant l'intonation.

5.7.1 Stimuli

Partant toujours des mêmes phrases néerlandaises et japonaises, nous les avons cette fois resynthétisées selon la transformation *saltanaj*, mais en leur appliquant les contours d'intonation artificielle, de la même manière que pour l'expérience 10.

5.7.2 Participants

Quarante bébés ont été testés avec succès, 21 garçons et 19 filles, dont les caractéristiques sont données dans le Tableau 5.7. Vingt-sept d'entre eux étaient issus de familles françaises monolingues, 12 de familles où l'on parle le français plus une ou plusieurs autres langues, 1 d'une famille où l'on ne parle pas le français.

TAB. 5.7 – *Caractéristiques des nouveau-nés de l'Expérience 11.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	41	129	67.6	21.1
Terme (semaines + jours)	38	41+6	40+1	1
Poids (g)	2800	4090	3512	341

De plus, les résultats de 44 bébés ont été rejetés pour les raisons suivantes : rejet de la tétine (5), pas assez éveillé/succion insuffisante avant le changement (22), pleurs/agitation (8), n'atteint pas le critère d'habituation (2), s'endort/succion insuffisante après le changement (4), perd la tétine après le changement (3).

5.7.3 Procédure

Conditions. Dans le but de toujours améliorer le dessin de nos expériences, nous avons affiné la répartition des bébés par condition. Alors que dans les expériences précédentes quatre locutrices étaient assignées de manière définitive à la phase d'habituation, et les quatre autres à la phase de test, nous avons souhaité faire varier ce facteur. Chaque condition de l'ancien dessin expérimental a donc été divisée en deux sous-conditions, pour contrebalancer l'attribution des locutrices à chaque phase. Il en résulte huit conditions au lieu de quatre, que l'ordinateur a assigné par avance aux 40 bébés de l'expérience, comme auparavant.

Seuil de haute amplitude. Par ailleurs, les résultats d'une expérience qui n'est pas rapportée ici nous ont suggéré que la procédure d'ajustement automatique du seuil de haute amplitude (cf. 5.2.3.2) pendant les deux premières minutes de silence présentait le risque de fixer ce seuil trop bas, et d'autoriser ainsi trop de succions HA. L'inconvénient est apparu surtout pour les bébés qui tètent le plus, qui peuvent facilement dépasser 60 succions HA par minute, augmentant alors de manière déraisonnable la variabilité des résultats. Alors que la procédure d'ajustement automatique du seuil avait été instaurée pour éviter l'ajustement manuel tel qu'il était effectué dans les études antérieures, il est donc apparu préférable de revenir à l'ajustement manuel.

Le seuil HA est donc maintenant fixé de manière définitive pour toute l'expérience. La période de silence au début de l'expérience a été réduite à une minute, dans la mesure où elle ne sert plus à calculer le seuil. Pendant cette période, et pendant les premières minutes de la phase d'habituation, l'expérimentateur ajuste manuellement le gain de l'amplificateur du signal de pression, de manière à obtenir plus ou moins de succions qui dépassent le seuil HA fixé. Le but est d'obtenir pour chaque bébé un nombre moyen de succions HA par minute compris entre 20 et 40, et ne dépassant en aucun cas 60. Pendant les minutes pendant lesquelles le gain est ajusté, l'expérimentateur appuie sur la touche F1 pour signaler à l'ordinateur de ne pas les prendre en compte pour le calcul du critère d'habituation.

5.7.4 Résultats

La Figure 5.10 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet. Ici, les bébés du groupe expérimental semblent s'être déshabitués significativement plus que ceux du groupe contrôle, mais sur une seule minute après le changement.

Les bébés n'ont pas tété différemment pendant l'habituation selon leur groupe ni selon leur langue d'habituation [respectivement $F(1,39) < 1$ et $F(1,39) = 2.1$, $p = 0.16$]. L'analyse de covariance habituelle prenant comme variable dépendante la moyenne des deux minutes suivant le changement ne donne pas d'effet de groupe [$F(1,37) = 1.46$, $p = 0.24$]. Néanmoins, l'analyse des courbes de succion suggère qu'il y a un effet de groupe, mais qu'il est localisé seulement sur la première minute suivant le changement. En prenant cette minute comme variable dépendante, l'analyse de covariance trouve effectivement un effet de groupe significatif [$F(1,37) = 4.48$, $p = 0.04$].

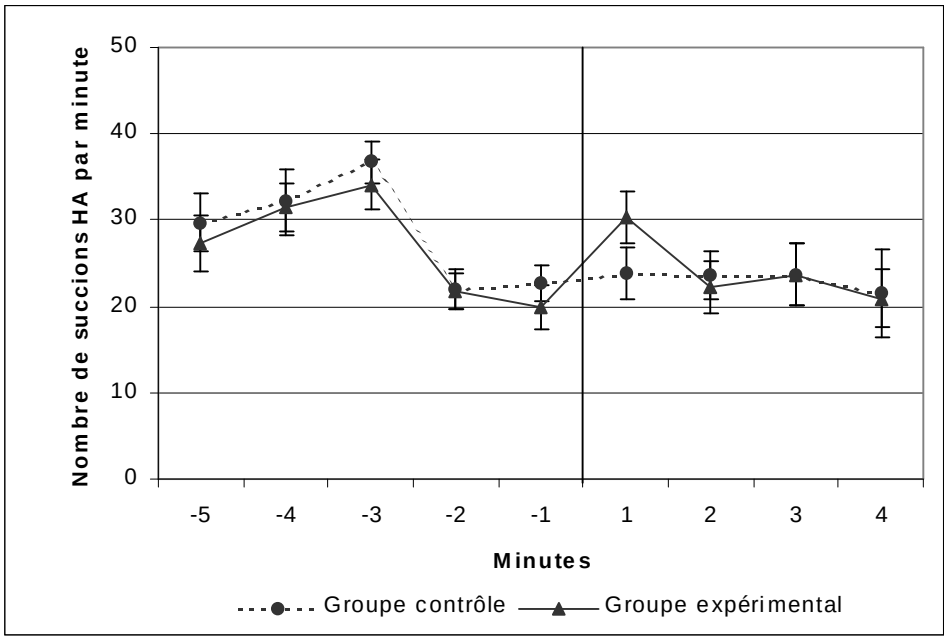

FIG. 5.10 – *Exp. 11 : Discrimination néerlandais-japonais par des nouveau-nés* – saltanaj *avec intonation artificielle. Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

5.7.5 Discussion

On peut donc penser que les bébés ont discriminé les deux langues, mais cet effet est visiblement plus fugitif que dans l'expérience 8, et que dans les études antérieures de manière générale : les effets de discrimination s'étalent habituellement sur les deux, trois ou quatre minutes suivant le changement. Devant la modestie de l'effet que nous avons obtenu ici, on ne peut considérer

ce résultat que comme provisoire. Il serait souhaitable de répliquer cette expérience pour vérifier la reproductibilité de l'effet et de sa localisation sur la première minute après le changement. En attendant, nous poursuivons cette discussion en supposant que les bébés du groupe expérimental ont bien discriminé les deux langues.

Étant donnée la nature des stimuli, il ne peuvent y être parvenus grâce à l'intonation. Nous avons donc là une bonne indication que le rythme de la parole peut suffire aux nouveau-nés pour discriminer les langues. Néanmoins, il y a tout de même dans les stimuli de type *saltanaj* un petit peu d'information supplémentaire par rapport aux stimuli *sasasa* : il s'agit d'indications partielles sur la structure des syllabes et les types de groupes de consonnes autorisés dans la langue. Se pourrait-il que cette phonotactique primitive soit la base de la discrimination ? Il y a de bonnes raisons d'en douter. En effet, dans une expérience de Bijeljac-Babic et coll. (1993), des nouveau-nés n'ont pas discriminé des listes de mots bi-syllabiques de structures syllabiques complexes (comportant 4 consonnes, comme CCVCCV) et de structures syllabiques simples (comportant 2 consonnes, comme CVCV), alors que par ailleurs ils étaient capables de distinguer des mots bi-syllabiques CVCV de mots trisyllabiques CVCVCV. De même, Bertoncini et coll. (1995) ont montré que les nouveau-nés ne discriminaient pas des listes de mots de structure CVCCV (tri-moraïque) et de structure CVCV (bi-moraïque), alors que dans une expérience similaire, ils avaient distingué des mots bi-syllabiques (C)VCV de mots tri-syllabiques (C)VCVCV. Ainsi, il semble que les nouveau-nés perçoivent les syllabes comme des unités, sans s'attacher à leur structure, même lorsque c'est la seule différence entre les deux listes de mots[17]. Par ailleurs, rappelons que les nouveau-nés discriminent les langues lorsque la parole est filtrée (Mehler et coll., 1988; Nazzi et coll., 1998), or la structure des syllabes n'est clairement pas préservée dans ces stimuli. Les données disponibles suggèrent donc que 1° les nouveau-nés sont indifférents à la structure des syllabes, 2° ils peuvent discriminer les langues sans cela. Le résultat de l'expérience 11 s'interprète donc par la sensibilité des bébés au rythme de la parole.

5.8 Préférence pour le néerlandais

Suite à la découverte de succions significativement supérieures à l'écoute du néerlandais qu'à l'écoute du japonais dans l'expérience 10, nous avons eu

17. S'ils ne s'intéressent pas à la structure phonotactique des syllabes, ils sont en revanche très sensibles à leurs propriétés rythmiques, notamment tonales et accentuelles (van Ooyen et coll., 1997; Sansavini, Bertoncini, & Giovanelli, 1997; Nazzi, Floccia, & Bertoncini, 1998)

l'idée de regarder s'il y avait une tendance identique dans les autres expériences. Le Tableau 5.8 donne pour chaque expérience le nombre moyen de succions suscitées par chacune des langues sur les cinq minutes précédant le changement.

TAB. 5.8 – *Nombre moyen de succions HA par minute pendant les cinq dernières minutes d'habituation, selon la langue écoutée et l'expérience.*

	Exp. 7	Exp. 8	Exp. 9	Exp. 10	Exp. 11
Née.	33 ± 11	32.2 ± 11.2	32.4 ± 12.7	35.6 ± 8.3	29.4 ± 9.6
Jap.	30.9 ± 11.1	27 ± 10.4	31.6 ± 10.3	25.8 ± 9.1	25.3 ± 8.3

Il se trouve que dans toutes les expériences, le néerlandais a suscité plus de succions que le japonais. Dans certains cas la différence est infime, mais dans d'autres la tendance est plus prononcée. En faisant la moyenne sur toutes les expériences, on obtient 32.5 ± 10.5 succions HA/minute pour le néerlandais et 27.9 ± 9.9 pour le japonais. La différence est significative : $F(1,175) = 9.1$, $p = 0.003$. La Figure 5.11 illustre l'effet sur les cinq minutes précédant le changement.

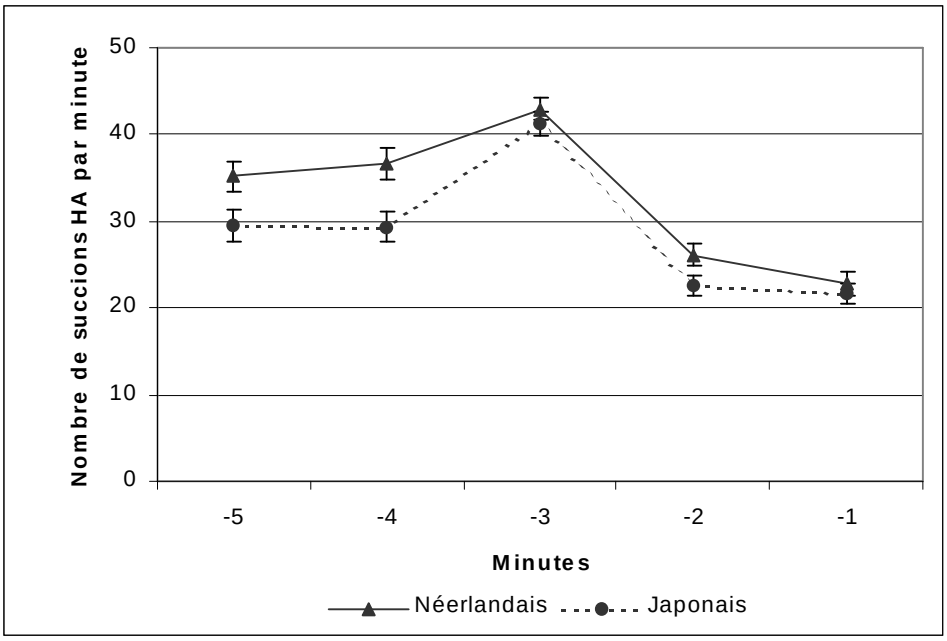

FIG. 5.11 – *Nombre de succions HA par minute pendant l'habituation, en fonction de la langue écoutée.*

La généralité de l'effet sur les cinq expériences suggèrent que celui-ci n'est

pas dû à un effet d'échantillonnage des bébés. On peut de plus remarquer que Nazzi et coll. (1998) avaient observé un effet similaire, les bébés tétant plus pour écouter l'anglais que le japonais pendant les deux dernières minutes d'habituation. Comme il n'y a toujours pas de raison de supposer que les nouveau-nés français aient une véritable préférence pour le néerlandais ni pour l'anglais, nous concluons que ces langues doivent plus les stimuler, pour des raisons liées à leurs sonorités, et potentiellement amplifiées par la manière dont elles sont rendues dans nos stimuli resynthétisés. En s'inspirant du résultat obtenu sur les adultes, on pourrait même arguer que les stimuli néerlandais excitent plus les nouveau-nés parce qu'ils sont plus désagréables à écouter. Cela peut sembler étonnant dans la mesure où plus les bébés tètent, plus ils entendent les stimuli en question, et ils auraient donc tout intérêt à téter moins pour moins entendre des stimuli désagréables. En l'absence de véritable modèle des variations de la succion, l'interprétation reste ouverte.

5.9 Conclusion et perspectives

Au cours de ce chapitre, nous avons montré :

- que les nouveau-nés français discriminent le néerlandais du japonais ;
- que cette discrimination repose au moins sur une propriété de la parole bien identifiée, le rythme ;
- que néanmoins la faculté de distinguer les langues et/ou de percevoir le rythme peut être altérée, notamment en présence de nombreuses voix, et aussi lorsque la parole est jouée à l'envers ;

De nombreuses questions sont bien sûr encore en suspens, et on peut imaginer des expériences supplémentaires pour y répondre.

Premièrement, bien sûr, il serait bon de répliquer l'expérience 11 pour s'assurer que l'effet de discrimination obtenu sur une seule minute est bien réplicable. On pourrait ensuite essayer d'éliminer les derniers indices phonotactiques qui sont préservés dans le *saltanaj* avec intonation artificielle. Sur la base d'études antérieures, nous avons argué à la Section 5.7.5 que les nouveau-nés sont incapables de discriminer ce type de différences. Néanmoins, il serait encore plus convaincant de montrer qu'ils sont encore capables de discriminer les langues lorsqu'elles sont éliminées. C'était bien le but de l'expérience 10 utilisant des stimuli de type *sasasa*. Nous pensons que l'absence de discrimination dans cette expérience est due au fait que ces stimuli, ne comportant que deux phonèmes, sont trop monotones pour les nouveau-nés. Pour remédier à ce problème, on pourrait réintroduire de la diversité phonétique sans pour autant préserver les indices phonotactiques. On pourrait pour cela resynthétiser les stimuli en *salatanaj* : comme dans le *sasasa*, un seul phonème

remplacerait tout un intervalle consonantique (ou vocalique), mais la nature même du phonème serait tirée au sort dans un certain inventaire, par exemple [saltnj] qui a bien réussi aux expériences 8 et 11. Ainsi, seules les durées des intervalles consonantiques et vocaliques seraient préservées, pas la forme des groupes de consonnes. On appliquerait bien sûr la même intonation artificielle qu'à l'expérience 11. Une telle expérience permettrait d'écarter de manière définitive l'hypothèse de la discrimination basée sur la phonotactique.

Deuxièmement, on aimerait bien sûr étudier de nouvelles paires de langues, pour pouvoir en confronter les résultats aux prédictions de notre modèle (Section 2.3.2.1). Notamment, il serait intéressant de tester la discrimination anglais/polonais, afin de savoir si, à l'instar de l'adulte, le nouveau-né est capable de percevoir cette différence, qui a priori a trait à la présence ou non de réduction vocalique. On en déduirait alors s'il est nécessaire d'inclure la variable ΔV également dans le modèle de la perception du rythme par le nouveau-né.

Enfin, le résultat de l'expérience 9 sur la parole à l'envers nous interpelle, car il nous pose la question de savoir exactement ce qui provoque cette différence de traitement entre la parole à l'envers et la parole naturelle. Nous avons proposé à la Section 5.5.5 deux indices possibles : le premier est la forme acoustique des occlusives à l'envers, et le second est la direction de l'intonation (bas-haut ou haut-bas). On peut imaginer deux nouvelles expériences, employant deux manières différentes de resynthétiser les stimuli, qui permettraient de tester séparément chaque indice.

- La première manière consisterait à resynthétiser les phrases en *saltanaj*, à l'endroit, mais en inversant le contour de $F0$ dans le fichier de resynthèse. L'intonation serait alors jouée à l'envers, mais les phonèmes à l'endroit. Si la discrimination échouait dans cette condition, cela ne pourrait être dû qu'à la direction de l'intonation.
- La deuxième manière consisterait à jouer à l'envers ces mêmes phrases. Les phonèmes seraient alors joués à l'envers, mais cela rétablirait en même temps la direction de l'intonation. Dans cette condition, seuls les phonèmes à l'envers pourraient être responsables de l'absence de discrimination.
- Dans ce deuxième cas de figure, pour bien s'assurer que ce sont les occlusives à l'envers qui sont l'indice crucial, on pourrait resynthétiser les phrases en *salmanaj*, c'est-à-dire en remplaçant les occlusives par une autre catégorie de consonnes, comme la nasale [m]. En jouant ces phrases à l'envers, on devrait alors observer une discrimination, puisqu'il n'y a pas d'occlusives à l'envers.

Indépendamment de la nature exacte des indices acoustiques utilisés par le nouveau-né pour caractériser la parole, ce résultat est également remarquable parce qu'il suggère que le cerveau du nouveau-né possède déjà des connaissances concernant les caractéristiques acoustiques de la parole. Nous sommes donc amenés à supposer que le nouveau-né possède une capacité innée à reconnaître la parole d'après certaines de ses propriétés, et à en extraire des propriétés rythmiques. Avant de pouvoir affirmer que nous sommes en présence d'une capacité innée spécifique au langage, il est néanmoins nécessaire de réellement tester cette spécificité. C'est ce que nous allons faire au Chapitre 6.

Chapitre 6

Discrimination de langues par des singes

Les résultats du chapitre précédent montrent que l'enfant, dès la naissance, possède un mécanisme de traitement du rythme de la parole. Il s'en sert notamment pour discriminer des langues, mais nous faisons l'hypothèse supplémentaire que ce mécanisme joue un rôle d'initialisation phonologique dans l'acquisition du langage (cf. Chapitre 1). Nos résultats obtenus avec la parole jouée à l'envers suggèrent que l'enfant possède en outre un mécanisme qui sélectionne, dans les entrées auditives, les sons qui nécessitent un traitement linguistique spécifique, parmi tous les bruits de l'environnement.

Une question importante, et récurrente, dans l'étude du langage, est celle de la spécificité des mécanismes que l'on postule dans les modèles de traitement de la parole. Cette question est d'autant plus cruciale lorsque l'on met en évidence certaines capacités chez le nourrisson, à un âge où il est peu probable qu'elles soient dues à l'expérience. Faut-il supposer que les mécanismes de traitement que nous découvrons chez l'enfant sont des mécanismes dus à l'évolution de l'espèce humaine, qui auraient été sélectionnés spécialement pour analyser la parole ? Ou doit-on au contraire supposer qu'il s'agit de capacités générales du cerveau, et notamment du système auditif, qui auraient été recrutées entre autres dans un domaine particulier, la parole ?

Pour répondre à cette question, Patricia Kuhl a proposé une démarche originale, consistant à tester ces mêmes capacités de perception de la parole sur des animaux :

> « Lorsque les humains perçoivent la parole, nous pensons qu'ils utilisent les niveaux de traitement à la fois auditif et phonétique. Cependant, il est difficile de distinguer les deux niveaux, à la fois

> sur le plan théorique et sur le plan expérimental. Nous ne savons pas quels effets attribuer au niveau auditif et lesquels attribuer au niveau phonétique. Un modèle animal est utile parce qu'il résout ce problème. Si l'espèce animale est choisie convenablement (Kuhl, 1979a), elle peut fournir un bon modèle du niveau de traitement auditif de l'homme, en l'absence de tout traitement de haut niveau (phonétique). L'avantage est assez évident : l'animal reflète ce qui est naturel pour le système de traitement auditif, lorsque les influences du niveau phonétique sont éliminées et que seules restent les influences du niveau auditif. » (Kuhl, 1987, p. 360, traduction libre)

Nous allons donc maintenant étudier la discrimination des langues par des singes. Mais nous allons tout d'abord passer en revue les principaux résultats de perception de la parole obtenus chez l'animal.

6.1 Revue des expériences de perception de la parole chez l'animal

Dans les tous débuts de la recherche sur la perception de la parole, a été mis en évidence un phénomène remarquable : la perception catégorielle (Liberman, 1957; Liberman, Harris, Hoffman, & Griffith, 1957). On observe typiquement ce phénomène dans le protocole suivant. Deux syllabes sont choisies, ne différant que par un trait phonétique (par exemple [da] et [ta], qui ne diffèrent que par le temps entre l'explosion de la consonne et le début du voisement [1]), et une série de syllabes intermédiaires sont synthétisées en faisant varier régulièrement les paramètres physiques responsables de la différence (ici, le VOT). À l'écoute des stimuli, au lieu d'entendre des syllabes intermédiaires entre [da] et [ta], les sujets n'entendent jamais que l'une ou l'autre, comme s'il existait une frontière phonétique au passage de laquelle la perception change de manière radicale et non linéaire. De plus, dans les tâches de discrimination, les sujets n'arrivent à discriminer que les syllabes situées de part et d'autre de la frontière, pas celles situées du même côté, à distance acoustique égale. Ainsi, contrairement aux autres domaines de la perception auditives, la perception phonétique n'est pas linéaire et continue, mais « catégorielle ». Cette découverte a conduit à supposer que la parole était « spéciale » (Liberman, 1982), et par conséquent nécessitait des mécanismes de traitement spécifiques. Cette supposition a été renforcée par la découverte de la perception catégorielle chez des nourrissons d'un mois, et a

1. Cet intervalle est appelé VOT, pour voice-onset time.

de plus suggéré que cette capacité spécifique à la parole était innée (Eimas, Siqueland, Jusczyk, & Vigorito, 1971).

Pour tester si la perception catégorielle requiert réellement des mécanismes de traitement spécifiques à la parole, Kuhl et Miller (1975) ont testé des chinchillas[2]. Ils ont entraîné les chinchillas à reconnaître deux exemplaires extrêmes de [da] et [ta] (de VOT respectivement 0 et 80 ms), en utilisant une procédure de conditionnement : lorsque les animaux entendaient l'une des deux syllabes, ils devaient changer de côté dans la cage, sinon ils recevaient un léger choc électrique. Lorsqu'ils entendaient l'autre syllabe, ils devaient rester sur place, étaient alors récompensés par un peu d'eau à boire. Après plusieurs mois d'entraînement, ils sont arrivés à des performances de discrimination supérieures à 96%. Ils ont alors été testés sur les syllabes de VOT intermédiaires entre [da] et [ta]. Il est apparu que leurs scores de discrimination sur ces syllabes intermédiaires étaient les mêmes que ceux de sujets humains. Ainsi, non seulement les chinchillas ont exhibé de la perception catégorielle sur le continuum [da]–[ta], mais celle-ci est organisée autour de la même frontière phonétique que les humains (un VOT d'environ 35 ms).

D'autres expériences ont généralisé ces résultats à d'autres contrastes phonétiques, et ont pris en compte d'autres phénomènes. Notamment, la frontière phonétique correspondant au voisement n'est pas toujours au même endroit selon la place d'articulation : ainsi, la frontière entre [ba] et [pa] correspond à un VOT de 25 ms, alors que celle entre [ga] et [ka] correspond plutôt à 45 ms. En employant la même méthode que ci-dessus, Kuhl et Miller (1978) ont montré que les chinchillas percevaient les continuums [ba]–[pa] et [ga]–[ka] de la même manière que les humains, avec la frontière phonétique au même endroit. Ainsi, la frontière phonétique correspondant au voisement bouge selon le contexte, de la même manière que chez l'humain.

Kuhl (1981) s'est également intéressée à la discrimination par les chinchillas de paires de syllabes sur le continuum [da]–[ta]. Elle a trouvé que comme chez l'humain, la discrimination était maximale de part et d'autre de la frontière phonétique, et beaucoup plus difficile au sein des catégories phonétiques. De tels résultats ont été répliqués chez le macaque, à la fois pour des contrastes de voisement (Kuhl & Padden, 1982) et pour des contrastes de place d'articulation (Kuhl & Padden, 1983).

Ces expériences ont engendré un domaine de recherches devenu trop vaste pour être traité ici exhaustivement. Notons simplement que des résultats tous aussi intéressants ont également été obtenus sur des animaux plus éloignés sur le plan phylogénétique, comme la caille (Kluender, Diehl, & Killeen, 1987) ou la perruche (Dent, Brittan-Powell, Dooling, & Pierce, 1997).

2. Le chinchilla est un rongeur vivant en Amérique du Sud.

De manière générale, ces expériences ont montré qu'il n'est pas nécessaire de postuler des mécanismes qui auraient évolué spécifiquement pour le langage pour expliquer les principaux aspects de la perception phonétique. Il s'agit bien sûr de la toute première étape du traitement de la parole, et il reste probable qu'à un certain niveau de traitement plus élaboré, des mécanismes spécifiques sont bel et bien nécessaires. Le tout est de savoir à quel niveau précisément les mécanismes spécifiques entrent en jeu. Dans ce chapitre, en étudiant la discrimination de langues chez le singe, nous allons justement nous intéresser à un niveau a priori supérieur[3] au traitement phonétique, la perception du rythme de la parole.

6.2 Démarche générale

Pourquoi supposer que des singes pourraient discriminer des langues ? Nous avons proposé au Chapitre 2 un algorithme simple, basé sur la segmentation de la parole en consonnes et voyelles et le calcul de statistiques de durées, permettant de rendre compte de la discrimination des langues. Compte tenu des capacités de perception phonétique des singes, il est possible qu'ils puissent disposer d'un tel algorithme. Celui-ci ne serait alors pas spécifique au langage. L'absence de discrimination de langues lorsque la parole est jouée à l'envers (Expérience 9) suggère un deuxième volet à notre programme. En effet, nous avons émis l'hypothèse que le nouveau-né pourrait posséder une connaissance innée de certaines propriétés de la parole, qui lui permettrait de filtrer les stimuli et de diriger la parole, et seulement elle, vers un traitement spécifiquement linguistique. Suivant cette hypothèse, et dans l'éventualité où les singes seraient capables de discriminer des langues, on doit alors prédire qu'ils devraient aussi être capables de les discriminer à l'envers. En effet, ils n'ont pas, eux, de raison de posséder un mécanisme de filtrage spécifique à la parole, et de traiter préférentiellement la parole à l'endroit de la parole à l'envers.

Les expériences sur les singes ont été planifiées simultanément avec les expériences correspondantes chez le nouveau-né, l'idée étant de comparer directement le comportement des nouveau-nés et des singes dans des expériences similaires et sur des stimuli identiques. L'expérience 12 fait le parallèle avec les expériences 7 et 7 bis, et l'expérience 13 avec les 8 et 9 sur les nouveau-nés. Les expériences décrites dans ce chapitre ont été réalisées avec la collaboration de Marc Hauser et Cory Miller, au Primate Cognitive Neuroscience Laboratory, Harvard University.

3. Mais ce pourrait être une étape de traitement indépendante de la perception phonétique.

6.3 Exp. 12 : Discrimination néerlandais-japonais par des tamarins – parole naturelle et parole à l'envers

Comme pour le nouveau-né, nous préférons, avant d'utiliser la parole resynthétisée, faire une première série d'expériences avec de la parole naturelle, préservant ainsi le maximum d'indices pour la discrimination. Nous allons donc tester la discrimination japonais-néerlandais sur des tamarins, en parole naturelle, et nous testerons de plus la discrimination sur ces mêmes phrases jouées à l'envers. Cette expérience fait donc le parallèle des expériences 7 et 7 bis (voir l'Annexe C pour cette dernière) effectuées sur le nouveau-né.

6.3.1 Stimuli

Les stimuli utilisés sont les mêmes phrases néerlandaises et japonaises que pour l'expérience 7 et leurs versions jouées à l'envers utilisées dans l'expérience 7 bis. La répartition des locutrices entre phases d'habituation et de test est la même. Le choix du néerlandais plutôt que l'anglais pour cette série d'expériences est dû à la crainte que les singes, ayant passé toute leur vie dans un laboratoire américain, soient déjà sensibilisés à la langue anglaise. Le néerlandais, en revanche, est une langue également inconnue des singes et des nouveau-nés testés.

6.3.2 Participants

Le Primate Cognitive Neuroscience Laboratory dispose d'une colonie de 13 tamarins (« cotton-top tamarins », ou *saguinus œdipus œdipus*), une espèce de singe originaire d'Amérique du sud (voir Figure 6.1). Ces tamarins sont nés en captivité. Au cours de la série d'expériences 12 et 13, deux d'entre eux sont morts. Deux tamarins supplémentaires ont alors été acquis et ont servi à compléter les groupes.

6.3.3 Procédure

Nous avons utilisé une procédure d'habituation/déshabituation similaire dans son principe à celle utilisée sur les nouveau-nés, mais en prenant comme mesure expérimentale l'orientation du regard plutôt que la succion.

FIG. 6.1 – *Photographie d'un tamarin « cotton-top » (à panache blanc).* ©*Roger Williams Park Zoo.*

6.3.3.1 Groupes de sujets

Dû au faible nombre de sujets, chacun a participé à 4 conditions différentes. Huit conditions ont été obtenues par le croisement des facteurs groupe (contrôle/expérimental), stimuli (endroit/envers) et langue d'habituation (néerlandais/japonais). Les tamarins ont donc été répartis en deux groupes selon les conditions effectuées (Tableau 6.1). L'ordre de passage des conditions a été contrebalancé à travers les sujets. Un intervalle de temps minimal de 4 jours a été observé entre deux passages d'un même individu dans une condition expérimentale. La médiane de ces intervalles a été de 7 jours.

TAB. 6.1 – *Exp. 12 : Répartition des tamarins dans les huit conditions.*

Groupe :	Expérimental		Contrôle	
Stimuli :	Endroit	Envers	Endroit	Envers
Groupe A	née. → jap.	jap. → née.	née. → née.	jap. → jap.
Groupe B	jap. → née.	née. → jap.	jap. → jap.	née. → née.

Le groupe A comprend 7 tamarins, 3 mâles et 4 femelles, et le groupe B 6 autres tamarins, 3 mâles et 3 femelles.

6.3.3.2 Déroulement du test

Le test se déroule dans une cabine isolée visuellement et auditivement du reste de la colonie. Le tamarin est installé dans une cage au-dessus de laquelle, dans un coin, se trouve un haut-parleur. Le test est filmé par une caméra vidéo qui englobe toute la cage dans son champ. Deux expérimentateurs suivent le déroulement du test sur un moniteur situé à l'extérieur de la cabine.

L'idée est qu'à la diffusion d'une phrase par le haut-parleur, le tamarin va se retourner et regarder dans cette direction. Puis, rien ne se passant, il reprend son activité (qui consiste à se déplacer dans tous les sens dans la cage). À force de diffuser des phrases, il arrive un moment où le tamarin s'habitue, et ne se retourne plus vers le haut-parleur. Si à ce moment-là on diffuse une phrase dans une nouvelle langue, va-t-il se déshabituer et se retourner à nouveau ? Si c'est le cas, on aura une bonne indication que le tamarin a remarqué une nouveauté dans la nouvelle langue.

Un essai est lancé par l'expérimentateur à un moment où le tamarin regarde vers le bas et n'est pas face au haut-parleur. L'essai consiste en la diffusion d'une phrase. Une réponse positive de la part du singe consiste à s'orienter vers le haut-parleur avant la fin de la diffusion de la phrase. Le début et la fin de l'orientation vers le haut-parleur sont codés en temps réel par l'expérimentateur. Si le tamarin ne s'oriente pas vers le haut-parleur, ou s'oriente après la fin de la phrase, la réponse est codée comme négative. Lorsqu'une orientation est ambiguë, il n'est pas tenu compte de l'essai pour le critère d'habituation. L'intervalle minimum entre deux essais est de 15 secondes, et l'intervalle maximum est de 60 secondes[4].

Pendant la phase d'habituation, les phrases correspondant à la phase et à la condition sont diffusées dans un ordre aléatoire, comme pour les bébés. Le critère d'habituation est atteint lorsque le singe ne s'oriente pas vers le haut-parleur lors de deux essais consécutifs. Il est alors « habitué » aux stimuli. Un minimum de 2 phrases d'habituation doivent également avoir été entendues (et suivies d'une orientation vers le haut-parleur) pour que le critère soit atteint.

La phase de test consiste en deux essais, diffusant chacun une phrase de test. Seul le premier essai de test est pris en compte pour déterminer si le singe s'est déshabitué ou pas. Dans l'éventualité où le singe ne réagit pas au cours du premier essai de test, un essai de post-test est lancé : il consiste en la diffusion d'un cri long de tamarin[5]. L'idée derrière le post-test est de

4. Théoriquement, si l'intervalle maximum est atteint sans avoir pu lancer un nouvel essai, l'expérience est interrompue et les données sont rejetées. En pratique, ce cas de figure ne s'est jamais produit.

5. Il s'agit d'un cri servant à établir le contact entre les individus.

vérifier si par hasard le tamarin ne s'est pas habitué, non seulement aux stimuli d'habituation, mais à toute la session expérimentale. Si le tamarin ne réagit pas au post-test, on considère donc qu'il n'était pas en état de réagir aux stimuli de test, et les résultats de cette session sont rejetés. Lorsque l'orientation vers le haut-parleur est ambiguë pour l'une des deux phrases de test ou pour le post-test, les résultats sont également rejetés.

Après la session, les essais de test sont recodés par deux observateurs d'après la bande vidéo, sans connaître la condition expérimentale. L'accord mesuré entre les deux observateurs est de 0.92.

Dû au faible nombre de sujets, les tamarins dont les résultats ont été rejetés pour une condition ont repassé celle-ci ultérieurement, jusqu'à ce que toutes les conditions soient complétées.

6.3.4 Résultats

Sur les 52 sessions (13 tamarins $\times$ 4 conditions), 6 ont été rejetées pour réponse ambiguë au test (2) ou absence de réaction au post-test (4). Ces sessions ont été refaites pour compléter toutes les conditions à 13 sujets.

Le nombre moyen d'essais pour atteindre le critère d'habituation a été de 16.9 ± 9.2 (intervalle : 4 à 52), sans différence significative selon le type de stimuli (endroit ou envers) ou la langue d'habituation ($F(1,51) < 1$ pour les deux). La Figure 6.2 donne le nombre de sujets ayant répondu positivement et négativement au premier essai de test, en fonction du type de stimuli (à l'endroit ou à l'envers), et du type de changement (de langue ou de locutrices).

Dans la condition changement de langue à l'endroit, 10 tamarins sur 13 se sont déshabitués, ce qui est significativement différent du hasard ($p<0.05$ par un test binômial). Mais comme pour les nouveau-nés, cela ne signifie rien si l'on ne prend pas en compte le groupe contrôle, qui ne changeait que de locutrices. Dans ce groupe, une majorité de tamarins ne s'est pas déshabituée. La différence entre les deux groupes est significative ($\chi^2 = 3.94$, $p < 0.05$)[6], montrant que les tamarins se sont plus déshabitués au changement de langues qu'au changement de locutrices. Les tamarins ont donc discriminé le néerlandais du japonais, là où les nouveau-nés ont échoué (Expérience 7) à cause de la variabilité des voix.

En revanche, dans les conditions où les phrases étaient jouées à l'envers,

6. En théorie un test statistique intra-sujet aurait été plus approprié que le test de χ^2, car ce dernier requiert que les deux séries de données soient indépendantes, or ici ce sont les mêmes sujets qui sont testés dans les deux conditions. En fait, ce ne sont pas tout à fait les mêmes sujets à cause des deux tamarins qui ont été remplacés en cours de route. Éliminer leurs données pour effectuer le test intra-sujet réduirait le nombre de sujets (déjà faible) à 11, ce qui diminuerait la puissance statistique de manière déraisonnable.

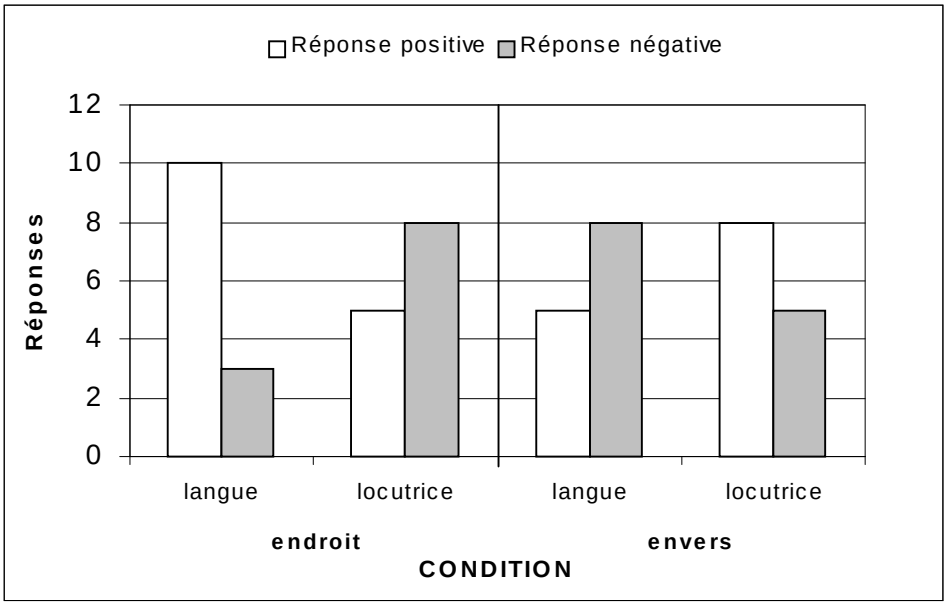

FIG. 6.2 – *Exp. 12 : Discrimination néerlandais-japonais par les tamarins, en parole naturelle à l'endroit et à l'envers.*

la situation est tout à fait différente : une majorité de tamarins ne s'est pas déshabituée au changement de langue, mais s'est déshabituée au changement de voix, sans que la tendance soit significative ($p = 0.29$ pour les deux). De plus il n'y a pas de différence entre le changement de langue et le changement de voix ($\chi^2 = 1.38$, $p > 0.25$). Les tamarins n'ont donc pas discriminé les deux langues lorsque les phrases étaient jouées à l'envers.

6.3.5 Discussion

L'expérience ci-dessus nous apporté deux résultats remarquables. Le premier, c'est que des animaux ayant un système auditif comparable au nôtre, les tamarins, peuvent discriminer deux langues comme le néerlandais et le japonais. Ceci suggère que l'aptitude du nouveau-né à distinguer les langues pourrait reposer sur des capacités auditives générales, non spécifiques au langage. Plus étonnant encore, les tamarins ont discriminé les langues dans une condition où les bébés ont échoué, comportant de multiples voix. Faut-il en déduire que les singes discriminent mieux les langues que les humains ? Peut-être pas. Après tout, les tamarins, eux, sont adultes, et ont donc un système auditif arrivé à maturité. De plus, ce sont des êtres sociables, qui vivent en groupe et communiquent à l'aide d'un certain nombre de cris. Ils

sont donc parfaitement familiers avec les situations comportant plusieurs locuteurs, même si ceux-ci n'ont pas les mêmes propriétés que les locuteurs humains. Par conséquent, dès lors qu'on accepte qu'ils puissent discriminer des langues, c'est une moindre surprise que de constater qu'ils puissent le faire à travers plusieurs locuteurs.

Le deuxième résultat remarquable, c'est que cette capacité des tamarins à discriminer les langues ne se maintient pas lorsque les phrases sont jouées à l'envers. Ceci suggère que pour expliquer le même comportement chez le nouveau-né, il va falloir trouver une autre explication que celle d'un filtre spécifique à la parole. Avant de discuter plus en détail toutes les implications théoriques de nos résultats, il est important d'essayer de préciser un peu *comment* le tamarin est parvenu à distinguer le japonais du néerlandais. Peut-être y parvient-il par des moyens très différents de ceux du nouveau-né, auquel cas il y aurait peu de conclusions à en tirer en ce qui concerne la capacité du nouveau-né. L'expérience 13 va donc s'efforcer de préciser la nature de la discrimination par le tamarin.

6.4 Exp. 13 : Discrimination néerlandais-japonais par des tamarins – *saltanaj* à l'endroit et à l'envers

Nous allons donc maintenant tester la discrimination japonais-néerlandais, en utilisant cette fois les stimuli synthétisés suivant la transformation *saltanaj*, joués à l'endroit, puis à l'envers. Cette expérience est donc l'équivalent des expériences 8 et 9 effectuées sur le nouveau-né.

6.4.1 Stimuli

Il s'agit des phrases néerlandaises et japonaises resynthétisées selon la transformation *saltanaj*, déjà utilisées dans l'expérience 8, et de leurs versions jouées à l'envers, utilisées dans l'expérience 9.

6.4.2 Participants

Les tamarins testés sont les mêmes que dans l'expérience précédente. Ils ont été testés dans cette expérience environ un an après leur passage dans l'expérience 12.

6.4.3 Procédure

C'est la même que pour l'expérience 12, y compris la répartition des sujets dans les différentes conditions.

6.4.4 Résultats

Sur les 52 sessions de test, 7 ont été rejetées pour réponse ambiguë au test (2) ou absence de réaction au post-test (5). Ces sessions ont été refaites pour compléter toutes les conditions à 13 sujets. Le nombre moyen d'essais pour atteindre le critère d'habituation a été de 32.5 ± 33.3 (intervalle : 4 à 211), sans différence significative selon le type de stimuli ($F(1,51) < 1$) ou la langue d'habituation ($F(1,51) = 1.1$, $p = 0.3$). La Figure 6.3 donne le nombre de sujets ayant répondu positivement et négativement, en fonction du type de stimuli (à l'endroit ou à l'envers), et du type de changement (de langue ou de locutrices).

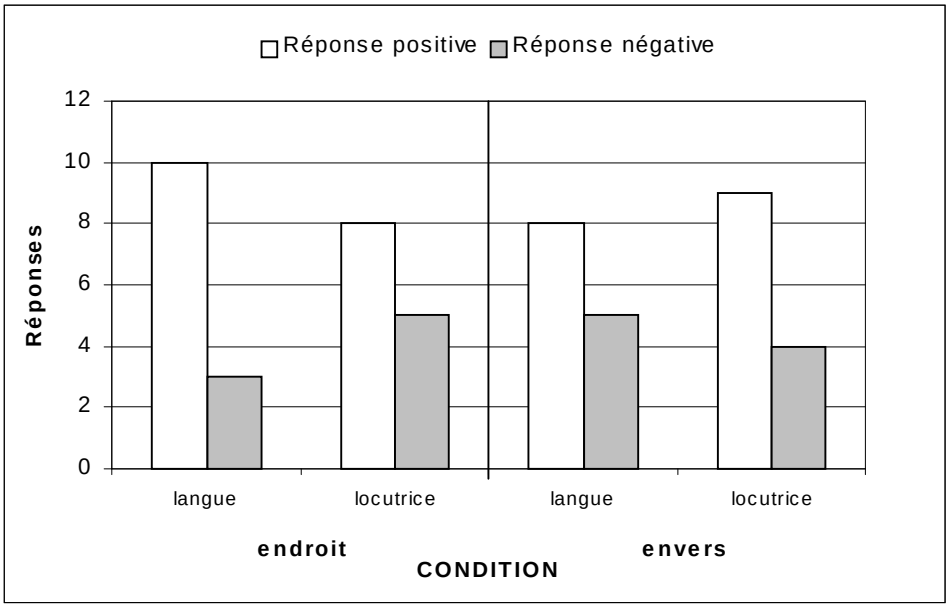

FIG. 6.3 – *Exp. 13 : Discrimination néerlandais-japonais par les tamarins, en parole resynthétisée* saltanaj *à l'endroit et à l'envers.*

Comme dans l'expérience 12, 10 tamarins sur 13 se sont déshabitués au changement de langue à l'endroit ($p<0.05$ par un test binômial). Ici cependant, dans le groupe contrôle exposé au changement de locutrices, 8 tamarins sur 13 se sont également déshabitués ($p=0.29$). De ce fait, la différence entre le groupe contrôle et le groupe expérimental n'est pas significative ($\chi^2 = 0.72$,

$p > 0.3$). On ne peut donc pas affirmer que les tamarins discriminent le néerlandais du japonais lorsque les phrases sont resynthétisées en *saltanaj*.

Dans les conditions où les phrases étaient jouées à l'envers, un nombre important mais non significatif[7] ($p = 0.29$ et $p = 0.13$) de tamarins s'est déshabitué au moment du changement. Il n'y a pas de différence entre le groupe subissant le changement de langue et celui subissant le changement de locutrices ($\chi^2 = 0.17$, $p > 0.5$).

Il semble donc que les stimuli de type *saltanaj* n'aient pas suscité de discrimination significative chez les tamarins.

6.5 Comparaison des expériences 12 et 13

Les expériences 12 et 13 présentent un certain nombre de différences frappantes. Tout d'abord, les tamarins ont mis un nombre d'essais pour s'habituer bien supérieur avec les stimuli *saltanaj* qu'avec les stimuli naturels ($F(1,102) = 11.2$, $p = 0.001$), avec des valeurs extrêmes étonnantes : si avec les stimuli naturels toutes les habituations ont pris au plus 52 essais, avec les stimuli *saltanaj* elles ont souvent dépassé ce nombre, avec un maximum de 211 essais !

Par ailleurs, les tamarins ont eu tendance à se déshabituer plus fréquemment aux stimuli *saltanaj* (35 fois sur 52) qu'aux stimuli naturels (28 fois sur 52) ($\chi^2 = 1.97$, $p > 0.15$, tendance non significative).

Nous n'avons pas d'explication valable pour ces différences. On aurait pu prédire que les stimuli *saltanaj*, étant moins variés et plus monotones, auraient suscité moins de réactions de la part des tamarins. On aurait alors prédit qu'ils s'y habitueraient plus facilement et se déshabitueraient plus difficilement, or c'est exactement le contraire qui s'est produit. Un autre argument aurait pu être que les tamarins, ayant déjà subi une première expérience similaire, se seraient lassés dans la deuxième expérience. Mais dans ce cas on aurait également prédit le contraire de l'effet observé.

À la fois le nombre d'essais jusqu'à l'habituation et le nombre total de déshabituations concordent pour suggérer que les stimuli de type *saltanaj* sont hautement plus attrayants, stimulants ou excitants, que les stimuli naturels. Pourrait-il s'agir de propriétés spectrales, ou d'un son présent dans la voix synthétique auquel les tamarins auraient été particulièrement sensibles ? D'autres expériences seraient nécessaires pour le savoir.

7. On voit bien ici que le nombre limité de sujets est problématique : 10 sur 13 (condition langue à l'endroit) représente un proportion significative, mais pas 9 sur 13 (condition locutrice à l'envers).

De ce fait, il est difficile de tirer des conclusions de l'expérience 13. Peut-être que les tamarins sont bien capables de discriminer le japonais du néerlandais dans les conditions propres au *saltanaj* (10 sur 13 se sont tout de même déshabitués au changement de langue à l'endroit), mais que ce résultat a été noyé dans un bruit d'origine à déterminer.

La Figure 6.4 donne le nombre de réponses positives et négatives par condition, pour l'ensemble des deux expériences 12 et 13. Elle montre que globalement, les singes se déshabituent significativement plus dans le changement de langue à l'endroit que dans le changement de locutrices ($\chi^2 = 4.06$, $p < 0.05$). En revanche, aucune différence n'est visible lorsque les phrases sont jouées à l'envers.

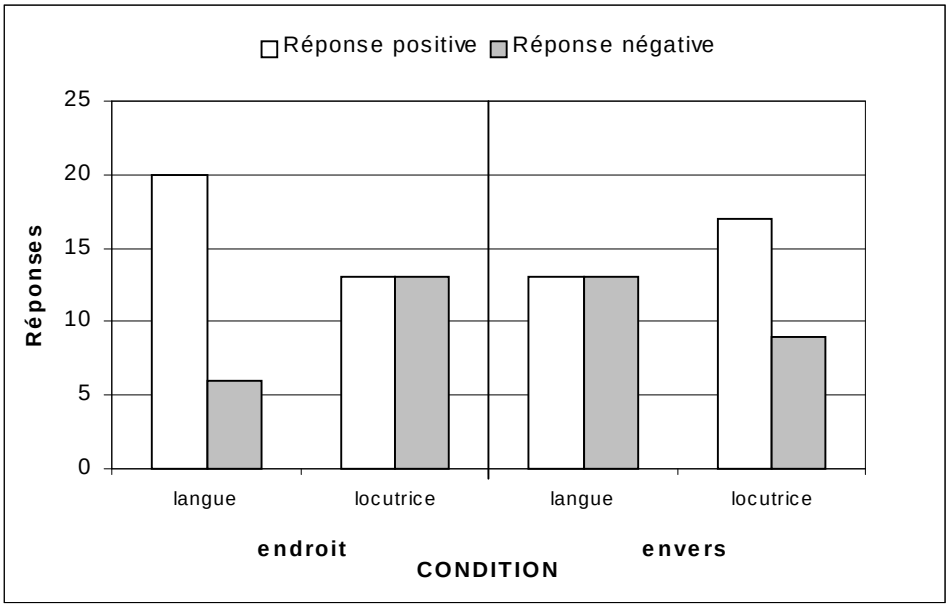

FIG. 6.4 – *Discrimination néerlandais-japonais par les tamarins – expériences 12–13 confondues.*

Cette analyse globale renforce notre conclusion principale de l'expérience 12, qui est que les tamarins discriminent le néerlandais du japonais à l'endroit mais pas à l'envers. En revanche, l'expérience 13 ne permet pas de conclure quant à la nature des indices qu'utilisent les tamarins pour la discrimination : s'agit-il comme pour les nouveau-nés, d'indices prosodiques ? S'agit-il aux contraires d'indices plutôt phonétiques, ou d'une autre nature ? D'autres expériences seront nécessaires pour répondre à cette question.

6.6 Discussion générale

Ainsi, nous avons trouvé que les tamarins sont, comme les nouveau-nés humains, capables de discriminer le néerlandais et le japonais. Cette découverte suggère, au premier abord, que l'aptitude des nouveau-nés à discriminer des langues ne repose pas nécessairement sur des capacités de traitement spécifiques au langage, mais sur des capacités auditives générales que nous partageons au moins avec d'autres primates. Cependant, il convient de rester prudent. En effet, du fait que nous n'avons pas pu préciser si les tamarins discriminent les langues avec des stimuli de type *saltanaj*, il reste possible qu'ils effectuent la discrimination en utilisant des indices totalement différents de ceux des nouveau-nés. Par exemple, il reste possible que les tamarins aient repéré des différences phonétiques entre les deux langues, ce qui serait finalement plausible compte tenu de leurs capacités de catégorisation phonétique passées en revues à la Section 6.1. Dans ce cas, cela n'impliquerait pas que la capacité de discrimination du nouveau-né soit une capacité auditive générale : la capacité à discriminer des langues *d'après leurs propriétés rythmiques*, tant qu'elle n'aura pas été démontrée chez le singe, peut très bien être spécifique au langage et à l'être humain.

Notre deuxième découverte est que comme les nouveau-nés, les tamarins ne discriminent plus les deux langues lorsqu'elles sont jouées à l'envers. Une fois de plus, cela suggère que l'absence de discrimination par le nouveau-né des langues jouées à l'envers n'est pas due à un dispositif de filtrage du son spécifique au langage. Cependant, nous voyons plusieurs interprétations possibles de ce résultat.

- La première interprétation est qu'effectivement la discrimination du rythme des langues repose sur des capacités auditives générales, communes à l'humain et au singe. Le modèle de perception du rythme de la parole que nous avons proposé pourrait rentrer dans ce cadre. En effet, si l'on admet qu'au moins certains aspects de la catégorisation phonétique ne sont pas spécifiques au langage (Section 6.1), ceux-ci pourraient suffire pour effectuer la segmentation en intervalles consonantiques et vocaliques requise par notre modèle. Comme nous avons de plus émis l'hypothèse que cette segmentation pourrait être altérée par des sons comme les occlusives à l'envers (Section 5.5.4), cela pourrait expliquer pourquoi la discrimination des types de rythme n'est pas possible sur la parole à l'envers, aussi bien chez le nouveau-né humain que chez le singe.
- Une autre possibilité est que le nouveau-né possède bien un système de filtrage des entrées auditives, mais que celui-ci ne soit pas à proprement parler spécifique au langage. Il pourrait s'agir d'un dispositif plus

général servant à sélectionner les stimuli pertinents pour l'organisme, et ce dispositif pourrait également être présent chez le singe. On doit alors se demander sur quelles propriétés du signal acoustique se base ce dispositif de filtrage, propriétés non spécifiques à la parole, mais qui permettraient néanmoins de distinguer la parole à l'envers de la parole à l'endroit ? Nous voyons au moins deux possibilités.

  - La première est que certaines propriétés de la parole à l'envers, et notamment des occlusives à l'envers, pourraient ne pas être « naturelles » dans un sens assez général. Par exemple, il se pourrait que les sons émis par des objets entrant en vibration respectent toujours un certain profil d'émission d'énergie acoustique dans le temps. Dans ce cas, les mêmes sons joués à l'envers pourraient avoir un profil d'énergie non naturel, et être détectés (et rejetés) comme tels par le système auditif, à la fois chez l'humain et chez le singe.
  - Nous avons également pensé à un autre indice qui pourrait « démasquer » la parole à l'envers. Toutes nos phrases (et aussi la base de diphones du logiciel de synthèse MBROLA) ont été enregistrées dans une cabine insonorisée, mais pas anéchoïque. Ainsi, tous nos stimuli comportent un peu de réverbération, due à la réflexion du son sur les parois de la cabine. Bien entendu, dans les enregistrements, la réverbération de chaque son arrive *après* le son source. Lorsque la parole est jouée à l'envers, la réverbération arrive alors *avant* sa source, ce qui est parfaitement anti-naturel. Il se pourrait donc que cette propriété de la parole à l'envers suscite le rejet de ces sons par le système auditif, et ce aussi bien chez le singe que chez l'humain. Pour tester cette hypothèse, on pourrait refaire les expériences de discrimination de langues à l'envers avec des phrases enregistrées en chambre anéchoïque : on prédirait alors une discrimination correcte.

- Enfin, la dernière possibilité est que la réaction négative des nouveau-nés et des singes à la parole à l'envers n'est pas due aux mêmes raisons. Il se pourrait par exemple que les nouveau-nés discriminent les langues sur la base de leurs propriétés rythmiques, et possèdent un dispositif de filtrage spécifique à la parole, qui rejette la parole à l'envers parce qu'il ne reconnaît pas les occlusives à l'envers. Dans le même temps, les tamarins pourraient discriminer les langues sur la base de leurs propriétés phonétiques, et leur absence de discrimination avec la parole à l'envers pourrait être due à son caractère non naturel, démasqué par la forme acoustique des occlusives à l'envers, ou par la réverbération.

Ainsi, dans l'état actuel des choses, la similarité de comportement entre le nouveau-né et le tamarin ne doit pas nécessairement être interprétée comme prouvant la similarité de leurs mécanismes de perception de la parole.

Ainsi, les résultats de nos expériences autorisent encore de nombreuses interprétations. Ceci est dû au fait qu'il ne suffit pas de montrer que les singes et les nouveau-nés ont les mêmes performances dans une tâche donnée pour en déduire qu'ils résolvent la tâche de la même manière. Encore faut-il explorer précisément la manière dont chacune des espèces résout la tâche. En ce qui concerne le nouveau-né, nous sommes arrivés au chapitre précédent à la conclusion que c'est le rythme de la parole qui lui permet de discriminer les langues. De nouvelles expériences seraient nécessaires pour répondre à la même question chez le tamarin. Pour savoir si les tamarins discriminent les langues sur la base de leur rythme ou de leurs propriétés phonétiques, on pourrait les tester sur d'autres paires de langues. Par exemple, deux langues ayant des propriétés phonétiques différentes mais des propriétés rythmiques similaires, comme l'anglais et le néerlandais. La resynthèse de parole peut bien sûr, comme nous l'avons fait avec les nouveau-nés, servir à départager les différents indices phonologiques. Quant à l'absence de discrimination avec la parole à l'envers, nous avons déjà proposé à la Section 5.9 une série d'expériences visant à déterminer précisément quels sont les indices acoustiques qui « démasquent » la parole à l'envers. Il va de soi que ces mêmes expériences seraient toutes aussi intéressantes à mener chez le singe, car il est tout à fait concevable que le rejet de la parole à l'envers ait des causes différentes chez le singe et chez l'humain.

Même si nos résultats ne nous permettent pas de conclure sur tous ces points précis, nous pensons qu'ils montrent (une fois de plus) que les hypothèses de spécificité au langage et à l'être humain méritent souvent d'être testées. Ces deux premières expériences, et la découverte de la discrimination de langues par le tamarin, suscitent de nombreuses questions passionnantes et ouvrent la voie à toute une série d'expériences visant à mieux comprendre la spécificité des différentes composantes du traitement de la parole. De plus, par rapport aux expériences de Kuhl et ses collègues, nous pensons que nos expériences apportent un certain nombre d'avancées significatives :

- D'un point de vue méthodologique, l'utilisation de la procédure d'habituation/déshabituation est un réel progrès. Rappelons en effet que la procédure de conditionnement utilisée par Kuhl demandait aux chinchillas plus de 6 mois d'entraînement à la tâche sur les exemplaires extrêmes du continuum phonétique. Sans parler du coût et de la lourdeur de telles expériences, on peut se demander à quel point cette exposition

permanente et prolongée aux phonèmes n'a pas sensibilisé le système auditif de ces animaux. La procédure d'habituation/déshabituation, elle, ne requiert aucune exposition préalable, et teste directement les réactions spontanées des singes face aux stimuli. Les conclusions que l'on peut en tirer en ont donc d'autant plus de portée. Le fait que cette procédure soit la même (sur le principe) que celle utilisée sur les nouveau-nés facilite également la comparaison directe entre les deux populations.

- Les expériences de Kuhl visaient à explorer l'une des toutes premières étapes de la perception de la parole : la catégorisation phonétique. Il était sous-entendu que si des étapes du traitement de la parole pouvaient être dues à un traitement auditif général, ce ne pouvait être que les toutes premières : celles concernant le codage du signal acoustique en représentations phonétiques. Le débouché naturel de ce programme de recherche était donc l'étude des autres étapes du traitement de la parole. En nous intéressant au rythme, nous avons commencé l'exploration de l'une de ces autres étapes, le traitement phonologique de la parole.
- Enfin, Kuhl (1987) a suggéré que l'une des caractéristiques du traitement de la parole par l'humain qui était susceptible de ne pas être retrouvée chez le singe était la « perception de classes d'équivalence auditives ». Elle faisait référence à la capacité du nourrisson (et de l'adulte) à considérer comme équivalentes des formes acoustiques différentes du même phonème, à savoir le même phonème prononcé par plusieurs locuteurs (Kuhl, 1979b, 1983). Ici, nous avons montré que les singes discriminaient le néerlandais du japonais à travers les voix de plusieurs locutrices, et réagissaient plus au changement de langue qu'au changement de voix. Il s'agit donc d'une toute première indication de la perception de classes d'équivalence auditives par le singe.

Chapitre 7

Conclusion : le rôle du rythme dans l'acquisition du langage

Apprendre le langage nécessite de la part de l'enfant l'apprentissage de tout un ensemble de connaissances abstraites, à partir d'un medium unique : le signal de parole. Au Chapitre 1, nous avons vu que les théories classiques de l'acquisition du langage avaient des difficultés à expliquer un tel apprentissage en l'espace de quelques années. Ces théories sont confrontées à des paradoxes, postulant par exemple que la syntaxe ne peut s'acquérir sans le lexique, mais que réciproquement l'apprentissage du lexique nécessite la connaissance préalable de quelques éléments de syntaxe. L'hypothèse de *l'initialisation phonologique* tente de pallier à cette circularité en proposant que des régularités présentes dans le signal de parole peuvent apporter de l'information utile sur le lexique et la syntaxe de la langue. Cette source supplémentaire d'information peut alors permettre d'initialiser l'apprentissage de certaines propriétés du lexique et de la syntaxe.

L'initialisation phonologique constitue donc un programme de recherches visant 1° à étudier les propriétés et régularités de la parole, 2° à étudier leurs liens avec des propriétés plus abstraites de la langue dont l'apprentissage peut sembler problématique, et 3° à explorer la perception de ces régularités par des nourrissons, et l'utilisation qu'ils en font au cours de l'acquisition du langage. Nous avons passé en revue au Chapitre 1 un grand nombre d'études dont les résultats montrent que cette démarche permet de faire progresser notre compréhension de l'acquisition du langage, au moins dans certains domaines qui ont été bien explorés comme l'apprentissage du lexique.

Dans le cadre de l'initialisation phonologique, nous avons proposé d'étudier les propriétés rythmiques de la parole, comme possible source d'information sur des propriétés plus abstraites de la langue. Cette hypothèse dérive de la classification des langues selon leurs propriétés rythmiques en un

nombre restreint de classes, et d'expériences suggérant que des nouveau-nés sont sensibles à ces propriétés. Néanmoins, pour confirmer une telle hypothèse, plusieurs éléments de démonstration étaient encore nécessaires :

- Montrer que le rythme est à la fois une propriété détectable dans le signal de parole, et une entité qui a un sens du point de vue linguistique.
- Montrer que le rythme de la parole est perçu par les nourrissons.
- Montrer les bébés se servent effectivement de l'information fournie par le rythme pour apprendre d'autres propriétés de la langue.

Nous allons maintenant discuter notre contribution à chacun de ces trois points.

7.1 Le rythme: une entité linguistique détectable dans le signal de parole

Comme nous l'avons vu au Chapitre 2, le rythme de la parole est une notion avant tout linguistique, selon laquelle il existe différents types de rythme à travers les langues du monde, les langues se répartissant en trois « classes rythmiques » : les langues accentuelles, syllabiques et moraïques. Cette notion linguistique a trouvé son équivalent acoustique dans la notion d'isochronie, selon laquelle les différents types de rythme sont dus à l'espacement régulier de différentes unités de la parole, à savoir le pied, la syllabe ou la more. Cette double assise, à la fois linguistique et acoustique, est typiquement un bon départ pour imaginer un scénario d'initialisation phonologique. Cela n'avait pas échappé à Mehler et coll. (1996) et Nazzi et coll. (1998), qui avaient déjà proposé que le rythme de la parole puisse jouer un rôle dans l'acquisition du langage.

Néanmoins, la notion de rythme des langues a été doublement remise en question. Du côté linguistique, par la mise en évidence de langues qui ne semblaient correspondre à aucun type de rythme déjà défini. Du côté acoustique, par l'impossibilité de vérifier empiriquement l'hypothèse de l'isochronie. Dès lors, toute spéculation sur l'initialisation phonologique était vaine.

L'alternative est venue d'une théorie plus phonologique du rythme, selon laquelle celui-ci résulte de la structure des syllabes qui composent la parole, et des phénomènes phonologiques qui influent sur cette structure. Ainsi, plus une langue a des syllabes complexes, plus elle serait du type « accentuel », et plus elle a des syllabes simples, plus elle serait du type « syllabique » ou « moraïque ». Du point de vue de l'initialisation phonologique, deux questions surgissent alors : 1° cette nouvelle conception est-elle compatible avec les classes de rythme, ou bien implique-t-elle que les propriétés rythmiques

sont graduelles ? 2° cette conception possède-t-elle un pendant acoustique, permettant de retrouver le rythme depuis le signal de parole ?

Au Chapitre 2, nous avons tenté de répondre à ces deux questions en cherchant dans le signal de parole des indices des propriétés phonologiques donnant naissance au rythme. En analysant huit langues, nous avons montré que ces propriétés pouvaient se déduire d'une segmentation en consonnes/voyelles de la parole, et de statistiques sur les durées des intervalles vocaliques et consonantiques. Plus précisément, nous avons identifié trois variables qui reflètent la complexité des groupes de consonnes et des noyaux vocaliques dans la langue. De plus, nous avons trouvé que les langues ne se répartissent pas uniformément suivant ces dimensions, mais au contraire se regroupent en trois classes : les classes rythmiques. Nous avons donc établi un lien entre le signal de parole, la théorie phonologique du rythme, et les classes rythmiques de langues. Il redevient ainsi possible d'envisager l'initialisation phonologique.

Il existe donc une corrélation entre certaines propriétés du signal de parole et les propriétés rythmiques des langues. Mais plus précisément, par quel mécanisme le cerveau humain extrait-il les propriétés rythmiques d'une langue à partir du signal ? Nous proposons un modèle de la perception du rythme, selon lequel le cerveau segmente la parole en intervalles vocaliques et consonantiques, et déduit le type de rythme des statistiques effectuées sur leurs durées. Peut-on tester un tel modèle de manière empirique ? Sur la base d'étude antérieures, nous avons adopté le paradigme de la discrimination de langues, comme indicateur de la perception de différences rythmiques. À partir de notre modèle de la perception du rythme, nous avons alors simulé des expériences de discrimination de langues, chez le nouveau-né et chez le sujet adulte. Nous avons obtenu ainsi des prédictions concernant les paires de langues qui, selon notre modèle, peuvent (ou pas) être discriminées sur la base de leur rythme par des nouveau-nés et par des adultes.

Pour évaluer la cohérence de notre modèle, nous en avons testé un certain nombre de ces prédictions sur des sujets adultes. Nous avons pour cela mis au point une technique, la resynthèse de parole, permettant de contrôler les propriétés prosodiques de la parole, afin de produire des stimuli dont toutes les propriétés sauf le rythme seraient éliminées (Chapitre 3). Les résultats de nos expériences de discrimination sur 7 paires de langues (Chapitre 4) montrent que :

1. les adultes sont capables de distinguer des langues sur la base de leurs seules propriétés rythmiques ;
2. ils discriminent globalement les paires prédites par les simulations ;
3. outre les trois classes rythmiques traditionnelles, le polonais semble former une classe à part. Cet atypie du polonais avait néanmoins été

prédite sur la base de ses propriétés phonologiques (Nespor, 1990), et est compatible avec nos simulations.

Ainsi, les données empiriques que nous avons recueillies renforcent la crédibilité de notre modèle de la perception du rythme.

Le lien que nous avons établi entre le signal et le rythme de la parole, et le modèle que nous en avons fourni, rendent d'autant plus plausible que le nouveau-né puisse extraire directement le type de rythme de sa langue depuis le signal. Par ailleurs, le fait que les propriétés rythmiques des langues se regroupent en un nombre limité de classes, plutôt que de se distribuer de manière uniforme, laisse espérer que des paramètres linguistiques puissent être reliés au rythme. Grâce à ces résultats, le rythme paraît donc être un bon candidat pour un scénario d'initialisation phonologique.

7.2 Perception du rythme de la parole par le nouveau-né

Il nous faut donc maintenant savoir si le rythme est effectivement perçu par le nouveau-né. Cette hypothèse a déjà été émise d'après des études de discrimination de langues antérieures, montrant que des nouveau-nés distinguent des langues de rythmes différents, mais pas des langues partageant le même type de rythme. Néanmoins, l'échantillon des langues étudiées était trop restreint, et les indices présents dans les stimuli trop nombreux, pour pouvoir être sûr que les discriminations de langues ont été effectuées grâce au rythme.

Plutôt que de multiplier les paires de langues pour augmenter la généralité des études antérieures, notre stratégie a plutôt été de nous concentrer sur une seule paire de langues, le néerlandais et le japonais, et d'éliminer un maximum d'indices présents dans la parole, afin de ne plus en préserver que le rythme. Nous avons utilisé pour cela la technique de resynthèse du Chapitre 3. Dans une condition où le seul indice préservé était le rythme, nous avons finalement trouvé que les nouveau-nés réagissaient significativement plus au changement de langue qu'à un simple changement de locutrices (Chapitre 5). Cet effet, localisé sur une seule minute suivant le changement, est plus faible que dans une condition similaire où l'intonation était également préservée. Il montre cependant que les nouveau-nés, dans des conditions où le signal est considérablement appauvri, sont capables de discriminer le néerlandais et le japonais sur la base de leurs seules propriétés rythmiques. Par conséquent, ils sont bien sûr capables de percevoir ces propriétés rythmiques.

7.3 Le rôle du rythme dans l'acquisition du langage

Nous avons jusqu'ici montré qu'il était plausible d'envisager un scénario d'initialisation phonologique à partir du rythme de la parole, et que les nouveau-nés semblaient effectivement avoir les capacités perceptives requises. Il nous reste maintenant à aborder la question la plus difficile, qui est de comprendre ce que font les nouveau-nés une fois qu'ils ont analysé le rythme, et qu'ils ont donc déterminé à quelle classe rythmique appartient leur langue.

Certains auteurs ont proposé des liens entre le type de rythme d'une langue et certains paramètres syntaxiques. Notamment, Donegan et Stampe (1983) ont proposé que le rythme accentuel se retrouvait dans les langues ayant la tête à gauche (voir Section 1.3.4.2) et utilisant la suffixation inflexionnelle, alors que le rythme syllabique serait plutôt associé aux langues ayant la tête à droite et utilisant la préfixation. Il est bien évident que de telles corrélations seraient d'une aide significative dans l'acquisition de la syntaxe, mais malheureusement il semble qu'elles ne se généralisent pas au-delà des langues étudiées par ces auteurs[1].

En l'absence d'un paramètre syntaxique qui soit lié de manière fiable au rythme, nous proposons de revenir sur la corrélation que nous avons observée depuis le début entre le rythme et la structure des syllabes de la langue. Celle-ci est véritablement fiable sur l'ensemble des langues qui ont été étudiées, les langues accentuelles possédant des syllabes complexes notamment au niveau des groupes de consonnes, les langues syllabiques possédant des syllabes moins complexes, et les langues moraïques possédant les syllabes les plus simples. Nous faisons donc l'hypothèse que l'enfant, ayant détecté le type de rythme de sa langue, pourrait en déduire des informations sur la structure de ses syllabes. Il s'agirait d'un exemple « d'initialisation phonologique de la phonologie », dans le sens où une propriété phonologique directement observable dans le signal de parole, le rythme, faciliterait l'apprentissage d'une propriété phonologique plus abstraite, la structure des syllabes.

7.3.1 Apprendre la structure des syllabes de la langue

Pour décrire l'ensemble des phénomènes phonologiques observées dans la parole, les théories génératives postulent des représentations phonologiques (stockées dans le cerveau du locuteur), et une grammaire qui permet de passer de ces représentations phonologiques aux représentations phonétiques,

1. Voir Auer (1993) pour une critique détaillée, et Nespor et coll. (1996) et la Section 1.3.4.2 pour d'autres indices concernant l'ordre des mots.

qui sont articulées dans la parole. Une partie de cette grammaire concerne la structure des syllabes.

Dans la théorie des « Principes et Paramètres » (Chomsky, 1981), la structure syllabique est décrite par des paramètres binaires tels que :

+/− **ObligatoryOnset** Toute syllabe possède nécessairement (+) ou non (−) un onset[2].

+/− **Coda** Une syllabe peut posséder (+) ou non (−) un coda.

+/− **ComplexCoda** Une syllabe peut posséder (+) ou non (−) un coda complexe.

+/− **ComplexOnset** Une syllabe peut posséder (+) ou non (−) un onset complexe.

Par exemple, la structure syllabique de l'anglais, qui est particulièrement complexe, est (partiellement) décrite par les valeurs [− ObligatoryOnset], [+ Coda], [+ ComplexOnset] et [+ ComplexCoda].

La théorie de l'Optimalité (Prince & Smolensky, 1993) décrit la structure syllabique dans un autre formalisme. Dans cette théorie, la grammaire syllabique est définie par l'ordre dans lequel sont ordonnées un certain nombre de contraintes. Les contraintes structurelles, qui tirent les syllabes vers une structure simple par défaut, interagissent avec des contraintes de fidélité, qui poussent au contraire la forme phonétique à être identique à la forme phonologique. Les contraintes sont éventuellement contradictoires, chaque contrainte pouvant être enfreinte si cela permet d'en satisfaire une autre qui la domine. Ce sont donc les relations de domination entre ces différents types de contraintes qui forment la grammaire d'une langue. Voici les principales contraintes utilisées pour décrire la structure syllabique :

Contraintes structurelles :

Onset Toute syllabe possède un onset.

NoCoda Une syllabe ne doit pas posséder de coda.

NoComplexOnset Une syllabe ne doit pas posséder d'onset complexe.

NoComplexCoda Une syllabe ne doit pas posséder de coda complexe.

Contraintes de fidélité :

Fill Tous les segments de la forme phonétique doivent provenir de la forme phonologique.

2. Dans le cas général, une syllabe est composée d'un noyau (sa voyelle, ou son maximum de sonorité), éventuellement précédé de consonnes (l'onset) et/ou suivi de consonnes (le coda). Un onset ou un coda est dit complexe s'il comporte plus d'une consonne.

Parse Tous les segments de la forme phonologique doivent se retrouver dans la forme phonétique.

Par exemple, en anglais, les contraintes Fill et Parse dominent toutes les contraintes structurelles. Elles imposent donc l'identité entre la forme phonologique et la forme phonétique, sans aucune contrainte sur la structure des syllabes.

Que l'on utilise un formalisme ou l'autre, la grammaire syllabique doit être apprise par l'enfant. En effet, elle ne se contente pas de décrire la structure de surface des syllabes dans la langue. La grammaire syllabique est réellement utilisée dans la production de la parole, et interagit avec les autres règles phonologiques. On la voit particulièrement à l'œuvre dans les emprunts à d'autres langues. En japonais, par exemple, où la structure syllabique est très simple, les mots empruntés à d'autres langues sont adaptés par les locuteurs ; typiquement, des voyelles sont insérées dans les groupes de consonnes pour que les mots se conforment à la grammaire syllabique du japonais : « festival » devient [fesutibaru], sphinx devient [sufinkusu], etc. (Itô & Mester, 1995).

Pour apprendre la grammaire syllabique de sa langue, l'enfant doit donc faire le travail du phonologue : en observant les syllabes au niveau phonétique, reconstituer la forme phonologique et les règles qui ont permis de passer de l'une à l'autre. Par ailleurs, notons que les syllabes ne sont pas livrées telles quelles à l'enfant : la parole est continue ! Pour pouvoir observer les syllabes de sa langue, il faudrait donc qu'il ait auparavant segmenté la parole en syllabes : c'est un problème équivalent à la segmentation de la parole en mots, que nous avons longuement détaillé au Chapitre 1, à ceci près qu'il faut pousser la segmentation un cran plus loin puisque les syllabes sont des unités inférieures aux mots. En revanche, on peut supposer que l'enfant dispose d'un algorithme universel de segmentation de la parole en syllabes. Mais celui-ci ne pourra avoir qu'un succès limité, à cause même de son caractère universel : en effet une même chaîne de phonèmes peut être syllabifiée de différentes manières selon la langue. Notre problème se reformule alors en : comment l'enfant rend-il son algorithme de segmentation spécifique à sa langue maternelle.

7.3.2 L'initialisation par le rythme

Nous proposons donc que la détermination du type de rythme de la langue permet de réduire la complexité de l'apprentissage de la grammaire syllabique, en apportant directement une partie de la solution. Nous allons maintenant expliciter l'apport potentiel du rythme pour l'apprentissage de la structure syllabique, dans chacune des deux théories phonologiques que nous avons présentées.

Principes et Paramètres. En se basant sur les huit langues présentes dans notre corpus, on peut proposer la généralisation suivante:

- les langues moraïques ont [− ComplexOnset] et [− ComplexCoda][3] ;
- les langues syllabiques ont [+ Coda] et [+ ComplexOnset][4] ;
- les langues accentuelles ont [+ Coda], [+ ComplexOnset] et [+ ComplexCoda].

Il est bien entendu indispensable d'étudier plus de langues pour déterminer si ces implications sont très générales. Il est également possible qu'une étude plus fine de la structure syllabique de ces langues permette d'en trouver d'autres. En attendant, nous postulons que la connaissance du type de rythme de sa langue peut permettre à l'enfant de fixer directement deux ou trois paramètres de sa grammaire syllabique.

Théorie de l'Optimalité. Il est bien entendu possible de traduire les trois implications ci-dessus dans le formalisme de l'optimalité, le type de rythme permettant de positionner les contraintes Fill et Parse par rapport aux contraintes structurelles correspondant aux paramètres sus-cités. On peut par ailleurs envisager une formulation plus générale de notre hypothèse. Levelt et van de Vijver (1998) proposent pour simplifier de considérer les contraintes de fidélité comme une contrainte unique, et de ne considérer que le positionnement de cette contrainte par rapport aux contraintes structurelles. Comme il y a 5 manières possibles de la placer par rapport aux 4 contraintes structurelles[5], Levelt et van de Vijver (1998) avancent que cela correspond à 5 classes de complexité syllabique[6]. Comme nous l'avons déjà remarqué p. 62, celles-ci incluent les trois classes de rythme. Sur les deux classes additionnelles, l'une aurait une complexité syllabique intermédiaire entre les langues syllabiques et les langues moraïques, l'autre aurait une complexité inférieure aux langues moraïques. On peut donc imaginer deux possibilités : soit il existe deux classes rythmiques supplémentaires que nous n'avons pas mises en évidence avec le corpus dont nous disposons, soit certaines classes de complexité syllabique correspondraient à la même classe rythmique. Dans les deux cas, connaître la classe rythmique de sa langue permettrait à l'enfant de connaître sa classe de complexité syllabique, et par

3. Plus, éventuellement, une contrainte sur la nature des codas, ce qu'il faudrait vérifier sur d'autres langues moraïques.

4. Plus, éventuellement, [− ComplexCoda], selon l'analyse que l'on fait du français.

5. Et ces 5 manières sont effectivement illustrées par des langues du monde ; on suppose en outre que la position relative des 4 contraintes structurelles n'entre pas en compte.

6. En réalité, la contrainte Onset ne participe pas vraiment à la complexité syllabique. On pourrait donc la négliger, et on aboutirait à seulement 4 classes.

conséquent le positionnement global des contraintes de fidélité par rapport aux contraintes structurelles.

On voit donc que notre proposition n'est pas dépendante d'un formalisme particulier. Dans les deux cas, la connaissance du type de rythme peut permettre de faire une inférence sur la nature des syllabes. Bien entendu, nous devons en outre supposer que la correspondance entre type de rythme et complexité syllabique est donnée par la grammaire universelle, qui est une hypothèse commune aux deux théories.

7.4 Discussion et perspectives

Nous allons maintenant aborder un certain nombre de questions que soulève notre proposition.

7.4.1 Est-ce vraiment le rythme ?

Nous avons attribué deux propriétés remarquables à la structure syllabique : d'une part elle se manifeste dans le signal de parole, par exemple par le biais des intervalles vocaliques et consonantiques ; d'autre part elle induit différents types de rythmes. Nous postulons alors que grâce aux indices présents dans le signal, l'enfant perçoit le rythme de la parole, et que grâce au rythme, il a accès à la structure syllabique (Figure 7.1). Pourquoi ce détour par le rythme ? Pourquoi ne pas postuler que grâce aux indices présents dans le signal, l'enfant accède directement à la structure syllabique ?

Pour aller plus loin, ne peut-on pas penser que dans les expériences de discrimination de langues, le nouveau-né discrimine les structures syllabiques, plutôt que les types de rythme ? C'est certainement une hypothèse alternative, mais il semble que ce ne soit pas la bonne[7]. En effet, comme nous l'avons déjà remarqué à la Section 5.7.5, les nouveau-nés ne semblent pas être capables de discriminer des mots de différentes structures syllabiques (Bertoncini et coll., 1995; Bijeljac-Babic et coll., 1993). Ils sont en revanche capables de discriminer des langues sans information sur la structure syllabique, lorsque la parole est filtrée (Mehler et coll., 1988; Nazzi et coll., 1998). Nous

7. Au moins en ce qui concerne le nouveau-né. Les sujets adultes, eux, sont susceptibles d'avoir utilisé toute stratégie efficace à leur disposition pour résoudre la tâche, notamment des jugements sur la structure ou le poids des syllabes. Mais quelle que soit leur stratégie, l'intérêt pour nous des expériences effectuées sur l'adulte est qu'elles révèlent s'il existe ou non des indices présents dans le signal et perceptibles par l'humain.

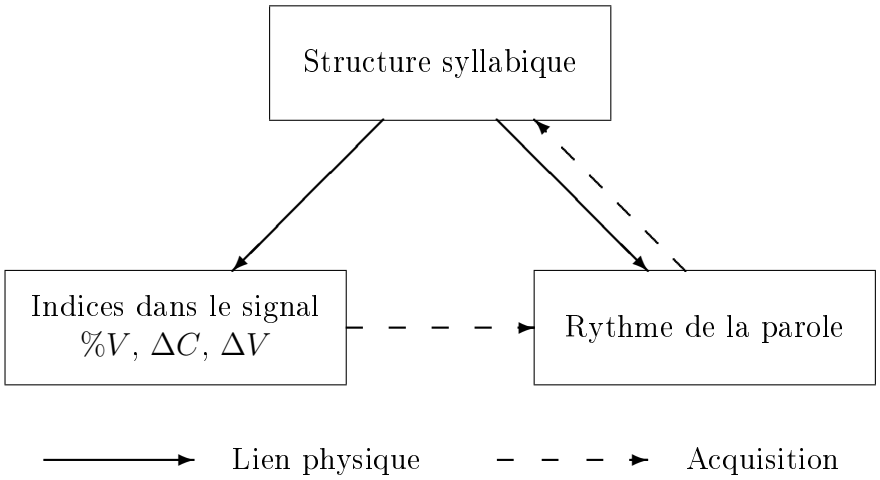

FIG. 7.1 – *Illustration des relations entre le signal, la structure syllabique, et le rythme, et du parcours postulé pour l'acquisition des syllabes.*

avons donc toutes les raisons de croire qu'à la naissance, la structure syllabique n'est pas perçue alors que le rythme l'est. C'est sur la base de ces données empiriques que nous sommes amenés à proposer le rythme comme un détour nécessaire pour accéder à la structure syllabique, au moins à cet âge.

7.4.2 Le cas du polonais

Au Chapitre 2, nous avons mis en évidence que suivant certaines mesures sur le signal (la variable ΔV), le polonais semblait avoir des propriétés différentes des langues accentuelles auxquelles il est habituellement assimilé. Cette différence peut être attribuée au fait que contrairement aux autres langues accentuelles, le polonais ne pratique pas la réduction vocalique. Nous avons de plus montré au Chapitre 4 que les sujets adultes sont sensibles à cette différence rythmique. Le nouveau-né doit-il donc supposer que le polonais a une structure syllabique différente des langues accentuelles ? Et alors laquelle ?

En fait, la structure syllabique du polonais est très complexe, comme c'est le cas pour les autres langues accentuelles. De ce point de vue, le nouveau-né n'a donc pas intérêt à le considérer comme appartenant à une autre classe rythmique. En fait, il faudrait avant tout tester la discrimination anglais/polonais par le nouveau-né : peut-être sa perception n'est-elle

pas aussi fine que celle de l'adulte, et ne lui permet pas de distinguer le polonais des autres langues accentuelles. Dans le cas où il discrimine effectivement les deux langues, cela n'implique pas non plus qu'il doive postuler une structure syllabique différente pour le polonais. Il suffit qu'il sache que l'information apportée par ΔV n'est pas pertinente du point de vue de la structure des syllabes (de la même manière qu'il doit savoir que la phonétique et l'intonation ne sont pas non plus pertinentes pour cet apprentissage). En revanche, ΔV pourrait apporter de l'information concernant les voyelles de la langue, les phénomènes de réduction et d'allongement vocalique. Nous n'avons cependant pas d'hypothèse détaillée dans ce domaine.

7.4.3 Et les singes ?

Au Chapitre 6, nous avons montré que des tamarins étaient également capables de discriminer le néerlandais du japonais, et échouaient de même lorsque la parole était jouée à l'envers. Ceci suggère que la capacité du nouveau-né à discriminer des langues et à percevoir leurs propriétés rythmiques n'est pas *nécessairement* spécifique au langage. Cela remet-il en question l'utilisation de cette capacité pour un apprentissage linguistique ?

En fait, comme nous l'avons déjà expliqué à la Section 6.6, si nos résultats montrent qu'effectivement les tamarins ont le même comportement que les nouveau-nés dans la tâche de discrimination de langues, ils ne permettent pas d'en déduire qu'ils utilisent les mêmes mécanismes de perception de la parole. Il reste possible, par exemple, que les tamarins utilisent des indices phonétiques plutôt que rythmiques pour la discrimination de langues, auquel cas cela ne remettrait pas en cause l'hypothèse selon laquelle la sensibilité aux propriétés rythmiques des langues est spécifique au langage.

En tout état de cause, même s'il s'avérait que la perception du rythme de la parole n'est pas spécifique au langage, cela n'impliquerait en rien qu'elle ne peut pas servir dans l'apprentissage du langage. Il est certain que la perception de la parole repose en partie sur des capacités auditives générales, qui sont pourtant recrutées par le cerveau humain pour accomplir des tâches spécifiques au langage. Par exemple, de nombreux travaux ont suggéré que la perception phonétique universelle était en grande partie une capacité auditive générale également présente chez d'autres mammifères (Kuhl, 1987), et néanmoins il est incontestable qu'elle est à la source d'apprentissages spécifiquement linguistiques, qui ne sont jamais observés chez d'autres animaux. Il n'y a donc aucune incompatibilité entre les résultats que nous avons obtenus sur les singes et notre hypothèse d'initialisation phonologique.

7.4.4 Les enfants bilingues

La majorité des enfants du monde naissent et grandissent dans un environnement multilingue (Hakuta, 1985; Harris & Nelson, 1992). Ils n'en deviennent pas dysphasiques pour autant, et aucune étude n'a montré de trouble ou de retard de l'acquisition du langage chez ces enfants, par rapport aux enfants monolingues. Ils ne semblent pas non plus avoir tendance à confondre les deux langues. Comment font-ils pour se rendre compte que plusieurs langues sont parlées dans leur milieu, et pour apprendre deux grammaires sans les mélanger ?

Il leur est sans aucun doute d'un grand secours de distinguer des langues le plus précocement possible. D'après nos résultats et ceux de Mehler et coll. (1988) et Nazzi et coll. (1998), un enfant en présence de 2 langues de classes rythmiques différentes peut les distinguer dès la naissance. À 2-3 mois, il pourra distinguer des langues ayant des contours d'intonation différents (Guasti et coll., sous presse ; voir Section 5.1.2). Et à 4-5 mois, il pourra distinguer des langues aussi proches que l'espagnol et le catalan ou l'anglais et le néerlandais, qui ne possèdent que des différences beaucoup plus subtiles (Bosch & Sebastián-Gallés, 1997; Nazzi et coll., soumis). On voit donc qu'il n'y aucune raison de redouter la confusion.

Mehler et coll. (1996) prédisaient cependant des difficultés, ou tout au moins un léger retard dans l'acquisition du langage, aux enfants bilingues dont les deux langues appartiennent à la même classe rythmique. On pourrait effectivement penser que les enfants confrontés à des langues présentant des différences rythmiques pourraient prendre un peu d'avance dans l'acquisition des deux langues, par rapport à des enfants confrontés à des langues dont les différences seraient seulement phonétiques, qui devraient attendre 4 mois de plus pour les distinguer. Il nous semble cependant que cette prédiction est liée à une conception selon laquelle la discrimination de langues serait un but en soi de l'enfant, pour se prémunir contre la confusion.

Du fait de notre engagement en faveur de l'initialisation phonologique, nous proposons une vision légèrement différente de la discrimination de langues. Selon nous, la capacité du nourrisson à distinguer des langues n'a pas évolué spécifiquement pour répondre aux problèmes du bilinguisme[8]. Le nourrisson n'a pas une capacité à discriminer les langues en soi : il a une capacité à percevoir et à distinguer des propriétés phonologiques, et cela parce qu'il doit apprendre ces propriétés, et parce qu'elles vont lui servir pour initialiser l'apprentissage d'autres propriétés de la langue.

8. Bien qu'il s'agisse de pure spéculation, il peut sembler logique que le langage n'ait évolué qu'une fois dans un seul groupe humain, et donc *avant* que la diversité linguistique apparaisse (Ruhlen, 1996).

Pour l'enfant confronté à l'anglais et au néerlandais, le fait de ne pas distinguer ces deux langues à la naissance est informatif : il signifie que quel que soit le nombre de langues que l'enfant est en train d'apprendre, elles ont toutes le même rythme, c'est-à-dire, selon notre hypothèse, la même complexité syllabique. Cet enfant acquiert donc une information sur les syllabes de sa/ses langue(s) aussi précocement que l'enfant confronté à l'anglais et au japonais, pour qui la discrimination entre les deux langues signifierait qu'il y a deux grammaires syllabiques à apprendre.

Cela signifie de manière plus générale que l'intérêt de l'enfant n'est pas nécessairement de discriminer les langues le plus tôt possible. Lorsque l'enfant discrimine deux langues à un âge donné, cela lui indique que les propriétés sur la base desquelles il a effectué la discrimination diffèrent dans les deux langues, et donc qu'il y a lieu d'en apprendre les deux versions différentes. En revanche, lorsqu'il ne discrimine pas deux langues sur la base de ces propriétés, il suffit de les apprendre une seule fois pour que cela vaille pour les deux langues. Dans le cas où ces propriétés phonologiques servent à initialiser d'autres propriétés de la langue, la même démarche convient. Ainsi, plus les langues sont proches, moins il est dommageable de ne pas les distinguer précocement.

Nous prédisons donc que l'apprentissage de la phonologie se fait de manière parallèle chez l'enfant monolingue et chez l'enfant bilingue. Et compte tenu du calendrier des acquisitions phonologiques que nous avons dressé au Chapitre 1, on peut estimer qu'à la fin de la première année de vie, lorsque l'enfant apprend ses premiers mots, il possède l'essentiel de la phonologie de sa ou ses langues. On ne s'attend donc pas à observer de différence développementale dans l'acquisition du langage entre l'enfant monolingue et l'enfant multilingue[9].

9. À cette réserve près que l'acquisition de la phonologie nécessite certainement une quantité d'exposition minimale à chaque langue, notamment pour que les analyses distributionnelles et statistiques convergent vers des résultats fiables. Pour un bébé exposé à 10 langues, l'exposition à chacune d'entre elles serait divisée en moyenne par 10 par rapport à un bébé monolingue, à exposition totale constante. Il n'est pas certain que cela suffise, mais nous n'avons pas de données concernant des enfants décalingues (s'il y en a !). Nous n'avons de manière générale aucune donnée sur l'exposition minimale nécessaire à l'acquisition du langage (en dehors des cas normaux et des cas d'isolement extrême). Comme le trilinguisme est assez répandu sans que cela ait l'air de présenter des difficultés, on peut estimer qu'un tiers de l'exposition totale standard est suffisant.

7.4.5 L'hypothèse de l'initialisation par le rythme est-elle testable ?

Nous supposons que l'enfant utilise sa perception des propriétés rythmiques de la parole pour acquérir des informations sur la structure des syllabes de sa langue. Cette hypothèse ne repose actuellement que sur la corrélation entre les propriétés rythmiques et la structure syllabique, et la capacité du nouveau-né à percevoir les propriétés rythmiques. Est-il possible de prouver que l'enfant utilise réellement les indices rythmiques comme nous le supposons ?

Cela peut s'avérer difficile. En effet, nous postulons que le rythme renseigne l'enfant sur la structure des syllabes très précocement. Or l'enfant ne commence à babiller qu'à partir d'environ 7 mois. Et à cet âge, il ne prononce qu'un sous-ensemble des syllabes autorisées dans sa langue, la plupart du temps les syllabes de type CV. On pourrait voir cela comme indiquant que l'enfant de 7 mois n'a, malgré sa sensibilité au rythme, toujours pas appris les syllabes de sa langue, ce qui infirmerait notre hypothèse. Cependant il est probable que des contraintes articulatoires l'empêchent de prononcer correctement les syllabes plus complexes (Macken, 1995). Dans ces conditions, il est bien difficile d'évaluer sa grammaire syllabique.

Ce que l'on souhaiterait, finalement, c'est observer l'acquisition des syllabes sans la contribution du rythme. Serait-elle impossible, altérée, retardée, ou normale ? De telles observations peuvent être éventuellement apportées par la neuropsychologie. Nous prédisons qu'un enfant ayant un déficit congénital de la perception du rythme devrait avoir des difficultés avec les syllabes de sa langue. Un tel déficit existe-t-il ? Peut-être... À titre de spéculation, nous proposons que les enfants dyslexiques sont peut-être dans cette situation.

La dyslexie est un trouble neurologique héréditaire, qui se manifeste par des difficultés d'apprentissage de la lecture, chez des enfants qui n'ont pas de problèmes familiaux, et qui ont par ailleurs des compétences intellectuelles normales (Morgan, 1896; Miles & Haslum, 1986; Olson, Wise, Conners, Rack, & Fulker, 1989). De nombreuses recherches ont montré que ce déficit n'est pas directement lié à la lecture, mais semble être un trouble spécifiquement phonologique (Brady & Shankweiler, 1991; Frith, 1985; Wagner & Torgesen, 1987) qui n'est généralement pas dépisté avant l'apprentissage de la lecture[10]. Ce déficit phonologique se manifeste dans un grand nombre de

10. D'autres hypothèses existent, notamment concernant des troubles plus généraux d'attention visuelle ou auditive (voir par exemple Galaburda & Livingstone, 1993, et Tallal, 1980). Toutes ces hypothèses ne sont pas mutuellement exclusives, car la population dyslexique est assez inhomogène, et de nombreuses causes peuvent être responsables de troubles de la lecture. Il est donc probable que l'on regroupe sous le terme de dyslexie des

tâches impliquant des jugements et des manipulations sur les sons de la parole[11]. De manière générale, les dyslexiques ont du mal à prononcer, répéter, distinguer, les mots longs et les syllabes complexes. Quelques études ont par ailleurs montré que les dyslexiques avaient certains troubles de la perception du rythme (McGivern, Berka, Languis, & Chapman, 1991; Kujala et coll., soumis). Il semble donc que les dyslexiques aient des difficultés assez variées dans tous les domaines impliquant la phonologie. Jusqu'à ce jour, il n'a pas été possible d'identifier un déficit fonctionnel unique, qui déclencherait en cascade l'ensemble des troubles observés. Puisque cette hypothèse découle directement de notre travail, nous proposons que le déficit de la perception du rythme est la cause de la dyslexie. En effet, selon nous, un tel déficit entraînerait des difficultés dans l'apprentissage de la grammaire syllabique de la langue. Si ces difficultés ne peuvent être surmontées à l'aide d'autres sources d'information ou d'autres mécanismes psychologiques, elles pourraient donc entraîner le large éventail de troubles phonologiques observés chez les dyslexiques.

Une manière d'évaluer cette hypothèse serait de tester la discrimination de langues chez des nouveau-nés de parents dyslexiques. En effet, ceux-ci ont une chance sur deux d'être eux-mêmes dyslexiques. Nous prédisons donc que la moitié des bébés testés devraient être insensibles aux différences rythmiques entre les langues. Étant donnée la puissance statistique typique des expériences sur les nouveau-nés, cela suffirait certainement à observer une absence de discrimination sur l'ensemble des bébés testés. On pourrait également suivre ces enfants pendant quelques années, en testant régulièrement leur perception et leur production des syllabes. Une fois la dyslexie (ou son absence) diagnostiquée de manière certaine (vers 6 ans), on pourrait alors évaluer la corrélation entre la perception du rythme à la naissance, l'apprentissage des syllabes et la dyslexie. Si la perception du rythme à la naissance est un bon prédicteur des troubles ultérieurs, ce serait alors une bonne indication en faveur de notre hypothèse.

---

déficits assez différents. Ici nous nous intéressons donc à la sous-population des dyslexiques qui ont des troubles spécifiquement phonologiques.

11. Deux exemples : 1° la tâche de rime, dans laquelle il faut juger si deux mots riment l'un avec l'autre, 2° la tâche dite de « spoonerisme », dans laquelle il faut échanger par exemple le premier son de deux mots (**J**ohn **L**ennon → **L**ohn **J**ennon), comme dans les contrepèteries.

Annexe A

Méthodes pour l'étude des nourrissons

La plupart des méthodes décrites ci-dessous admettent un certain nombre de variantes rapportées dans la littérature, correspondant aux améliorations successives de la méthode, ou aux préférences de certains expérimentateurs. Dans chaque cas, nous décrivons la version la plus récente ou la plus utilisée, et nous renvoyons le lecteur intéressé aux articles décrivant les autres variantes.

A.1 Succions de haute amplitude

La méthode des succions de haute amplitude permet de tester si des bébés âgés de 0 à 2 mois peuvent discriminer deux catégories de stimuli (par exemple des phrases prononcées en deux langues, ou bien deux syllabes, ou encore des mots bi- de mots tri-syllabiques). Elle repose sur le principe de l'habituation/déshabituation : pendant une première phase l'enfant est familiarisé avec un type de stimuli, et pendant une deuxième phase il est testé sur un autre type de stimuli. On essaye alors d'évaluer la réaction de l'enfant au changement de catégorie.

L'activité de l'enfant est mesurée à travers ses succions sur une tétine fermée, dont on mesure la pression interne. Un seuil de haute amplitude est défini a priori, et les succions dont l'intensité dépassent ce seuil (les succions dites de haute amplitude) sont renforcées par la diffusion d'un stimulus. Ainsi, l'enfant détecte un lien de cause à effet entre ses succions et l'apparition des stimuli.

Le passage des stimuli d'habituation aux stimuli de test s'effectue lorsqu'un critère d'habituation est atteint, à savoir une baisse sensible des suc-

cions pendant deux minutes consécutives. On dit alors que le bébé est « habitué » aux stimuli. Si l'enfant détecte un changement entre les stimuli d'habituation et ceux de test, il le manifeste généralement par une augmentation de ses succions. Cette augmentation est bien sûr à comparer avec le profil de succions d'un groupe contrôle, dont les stimuli de test sont de la même catégorie que ceux d'habituation. Ainsi, lorsque les tests statistiques montrent que les bébés du groupe expérimental (changement de catégorie) augmentent leurs succions après le changement significativement plus que ceux du groupe contrôle (pas de changement, ou changement au sein de la même catégorie), on en déduit que les bébés sont capables de discriminer les deux types de stimuli.

Références

La plus ancienne description de la méthode est de Eimas et coll. (1971). La version la plus récente est décrite par Nazzi et coll. (1998) et Shi et coll. (1999) (voir également la Section 5.2.3). Mehler et coll. (1988) ont utilisé une variante dite « non contingente », dans laquelle il n'y avait pas de lien entre la succion de l'enfant et l'apparition des stimuli. Hesketh et coll. (1997) ont introduit une variante comportant deux changements par sujet, pour tester des bébés de 2 mois (voir également la Section 5.4 et l'Annexe C.2). Enfin, Floccia (1996 ; Floccia, Christophe, & Bertoncini, 1997) a étudié en grand détail les procédures de renforcement des succions chez le nouveau-né, afin de mieux comprendre les mécanismes sous-tendant la méthode.

A.2 Orientation préférentielle de la tête

La méthode d'orientation préférentielle de la tête (*head-turn preference procedure*) est une technique relativement souple qui permet de poser des questions assez diverses, et qui peut s'appliquer à des enfants d'âge compris entre 4 et 12 mois (voire plus). L'aspect principal de la méthode est de mesurer la préférence de l'enfant lorsqu'il lui est donné la possibilité d'entendre deux types de stimuli. On suppose que la préférence de l'enfant révèle le type de stimuli auquel il est le plus sensible, ou qui se conforme le mieux à sa connaissance ou à ses attentes concernant sa langue maternelle ou le langage en général. Les développements récents de la méthode introduisent de plus une période de familiarisation. On suppose alors que la préférence de l'enfant est guidée par le lien qu'il établit entre les stimuli de test et ceux de familiarisation.

Le test se déroule dans une cabine, au centre de laquelle l'enfant est assis

sur les genoux d'un parent. Sur le mur faisant face à l'enfant se trouve une lumière verte, et sur les murs latéraux se trouvent deux lumières rouges. Derrière chaque lumière rouge se trouve un haut-parleur. L'expérimentateur suit l'expérience depuis l'extérieur de la cabine (par un orifice dans le mur ou par une caméra vidéo). Pour démarrer un essai, l'expérimentateur fait clignoter la lumière centrale. Lorsque l'enfant regarde dans cette direction, l'essai proprement dit commence : l'une des lumières latérales clignote. Lorsque l'enfant se tourne vers cette lumière latérale, le haut-parleur correspondant diffuse un stimulus. La diffusion du stimulus continue jusqu'à ce que l'enfant arrête de regarder dans cette direction pendant plus de 2 secondes. C'est donc la durée d'orientation de l'enfant dans la direction du haut-parleur qui détermine la durée pendant laquelle il entend les stimuli. La mesure est le temps pendant lequel l'enfant s'est orienté vers chaque type de stimuli. Si le temps d'orientation est significativement supérieur pour une catégorie de stimuli que pour l'autre, on en déduit que l'enfant a une préférence ou une sensibilité accrue pour cette catégorie. Par exemple, il a été ainsi montré que des enfants de 7 mois préfèrent écouter des portions de phrases coupées aux frontières de groupes intonationnels, plutôt qu'au milieu, révélant ainsi leur sensibilité à ces unités prosodiques (Hirsh-Pasek et coll., 1987).

La phase de test peut éventuellement être précédée d'une phase d'entraînement, qui se déroule dans les mêmes conditions[1]. La préférence que l'enfant exhibe éventuellement pendant la phase de test peut alors être influencée par les stimuli entendus pendant l'entraînement. Par exemple les stimuli d'entraînement peuvent être des phrases contenant toutes un même mot, et les stimuli de test se composent 1° de ce mot prononcé en isolation, et 2° d'un autre mot n'apparaissant pas dans les phrases de familiarisation. Si les enfants ont segmenté et mémorisé le mot test dans la première phase, et qu'ils le reconnaissent dans la deuxième, on s'attend à ce qu'ils réagissent différemment à ce mot reconnu qu'au mot nouveau. Dans ce cas, on observe effectivement que l'enfant s'oriente plus longtemps vers le mot familier que vers le mot nouveau (voir Jusczyk & Aslin, 1995 ; Jusczyk et coll., sous presse ; Houston et coll., sous presse). Pour prendre un autre exemple, on peut familiariser l'enfant avec des phrases dans une langue, et pendant la phase de test lui donner à écouter d'autres phrases dans la même langue, et dans une langue différente. Lorsque l'enfant distingue les deux langues, on observe qu'il s'oriente plus longtemps vers la nouvelle langue (Nazzi et coll., soumis). On peut remarquer que dans certaines expériences, l'enfant mani-

1. Dans certains cas, la phase d'entraînement peut également être passive, c'est-à-dire que les stimuli sont diffusés sans conditions particulières sur l'orientation de l'enfant. C'est le cas notamment lorsque les stimuli d'entraînement consistent en un long flux continu de parole (Saffran et coll., 1996; Aslin et coll., 1998).

feste une préférence pour le stimulus nouveau, et dans d'autres il préfère le stimulus familier. Les causes exactes de la préférence de l'enfant ne sont à l'heure actuelle pas bien comprises[2].

Références

La version originale de la méthode (sans entraînement) est due à Fernald (1985), qui l'a elle-même adaptée à partir d'une méthode de préférence visuelle (Colombo & Bundy, 1981). La version la plus récente de la méthode (avec phase d'entraînement) est décrite dans deux articles méthodologiques : Kemler Nelson et coll. (1995), Jusczyk (sous presse).

A.3 Latence d'orientation

La méthode de latence d'orientation permet également de mesurer des différences de réaction face à différents types de stimuli auditifs. Elle a été testée sur des enfants de 2 à 4 mois, et est en principe utilisable à des âges bien supérieurs.

L'enfant est installé face à un écran, et deux haut-parleurs sont situés de part et d'autre, à un angle d'environ 35° par rapport à l'enfant. Au début de chaque essai, l'écran montre des images destinées à attirer l'attention de l'enfant. Une fois que celui-ci regarde dans cette direction, l'un des deux haut-parleurs joue un stimulus. La mesure est le temps entre le début du stimulus et le début de l'orientation des yeux de l'enfant vers le haut-parleur. Elle est obtenue en codant ultérieurement les enregistrements vidéo image par image. Lorsque l'enfant ne s'oriente pas, ou s'oriente vers le haut-parleur ne diffusant pas de stimulus, la réponse n'est pas prise en compte. Les stimuli des différentes catégories sont présentés dans un ordre aléatoire. Si les enfants s'orientent plus vite vers un type de stimuli que vers un autre, on en déduit qu'ils perçoivent les différences entre les deux types, et éventuellement qu'ils préfèrent ou qu'ils sont plus familiers avec le type vers lequel ils s'orientent le plus vite.

Références

Cette méthode est une adaptation récente des méthodes d'orientation vers des stimuli visuels (Johnson, Posner, & Rothbart, 1991; de Schonen, Deruelle,

2. Cela n'empêche pas d'interpréter les expériences : une préférence significative pour un type de stimuli reflète dans tous les cas la sensibilité de l'enfant aux différences entre les deux types de stimuli testés.

Mancini, & Pascalis, 1993), et a été utilisée pour montrer que des bébés s'orientent plus vite vers des phrases dans leur langue maternelle que dans une autre langue (Dehaene-Lambertz & Houston, 1998; Bosch & Sebastián-Gallés, 1997).

A.4 Conditionnement

La technique de conditionnement consiste à entraîner des enfants à tourner la tête vers un haut-parleur latéral lorsqu'ils détectent un certain changement ou un certain son. Une fois l'entraînement effectué on peut tester la manière dont ils généralisent ce comportement à des stimuli nouveaux. Cette méthode est utilisée sur des enfants âgés de 6 à 12 mois.

L'enfant est assis sur les genoux d'un parent, face à un expérimentateur. À la gauche (ou à la droite) de l'enfant, se trouve un haut-parleur masqué par un dispositif de renforcement. Ce dernier consiste en une vitrine opaque à l'intérieur de laquelle est placée un jouet. Lorsque l'expérimentateur déclenche un renforcement, la vitrine s'allume et le jouet devient visible et s'anime.

Au début d'un essai, l'expérimentateur attire l'attention de l'enfant vers lui en utilisant différents jouets ou objets. Une fois que l'enfant a son attention centrée dans sa direction, le haut-parleur latéral diffuse des suites de sons, par exemples des syllabes, dont l'une est différente des autres. Si l'enfant tourne sa tête vers le haut-parleur lorsqu'il entend la syllabe différente, l'expérimentateur déclenche le renforcement. En revanche, il n'y a pas de renforcement si l'enfant tourne la tête alors que la syllabe n'a pas changé. La mesure expérimentale est la présence ou l'absence d'une orientation lors du changement, comparé à une situation contrôle où il n'y a pas de changement. Une première phase d'entraînement sert uniquement à ce que l'enfant apprenne la tâche, par conditionnement dû au renforcement. Dans la phase de test, on peut alors tester la réaction de l'enfant sur de nouvelles syllabes, ou de nouveaux locuteurs, ou en présentant la syllabe dans des contextes différents (par exemple dans des mots). On en déduit des informations sur la nature des différences perçues et mémorisées par l'enfant.

Une variante récente de cette méthode consiste à entraîner l'enfant à s'orienter vers le haut-parleur lorsqu'il entend un certain bruit, qui est superposé aux stimuli. La mesure est la latence d'orientation de l'enfant vers le haut-parleur. Une fois que l'enfant est conditionné, on peut faire varier la position du bruit au sein des stimuli. L'hypothèse est que le bruit devrait être plus difficile à détecter lorsqu'il est superposé à des stimuli qui forment une unité perceptive pour l'enfant, que lorsqu'il intervient entre des unités

perceptives. Cette difficulté se traduit alors par une latence d'orientation plus élevée. Par exemple, Morgan (1994) a montré que les latences d'orientation était supérieures lorsque le bruit était superposé à des bi-syllabes accentués systématiquement en première syllabe, que lorsqu'il était superposé à des bi-syllabes qui n'étaient pas toujours accentués de la même manière. Il en déduit que pour un enfant américain de 7-8 mois, un bi-syllabe accentué en première syllabe a une cohésion plus forte, car c'est la forme prosodique typique des mots en anglais.

Références

La technique originale de conditionnement est décrite dans Kuhl (1983, 1985). La variante utilisant la détection de bruit et la latence d'orientation est due à Morgan (1994).

Annexe B

Instructions et fichiers d'exemple

B.1 Fichier utilisé pour la resynthèse

Prenons en exemple la phrase anglaise « A hurricane was announced this afternoon on the TV » transcrite phonétiquement en SAMPA[1] :
[9h9rik@nwOz@naUnstD@saft@nunOnD@tivi]

Le fichier listé page suivante est celui pris en entrée par MBROLA si l'on veut resynthétiser la phrase anglaise telle quelle. La première colonne donne le code de chaque phonème dans la base de diphones de l'anglais. La deuxième colonne donne la durée du phonème en millisecondes. Les colonnes suivantes sont optionnelles, et donnent des couples de nombres définissant des points de fréquence fondamentale (F_0) : le premier situe le point en pourcentage par rapport à la durée du phonème, et le deuxième spécifie la fréquence fondamentale voulue en ce point. Par exemple, le premier phonème de la phrase est [9] ([Œ] en IPA), sa durée est 47 ms, et son premier point de F_0 est situé à 28% en durée du début du phonème (soit à environ 13 ms du début), et sa valeur est 202 Hz. La première et la dernière ligne du fichier spécifient de plus des silences initiaux et finaux de 50 ms, afin que le premier et le dernier phonème soient correctement synthétisés. Enfin, précisons que MBROLA interpole linéairement la courbe de F_0 entre les points spécifiés.

1. Speech Assessment Methods Phonetic Alphabet : un équivalent de l'alphabet phonétique international en caractères ASCII.

```
%%%%%%%%%%%%%%%%%% début du fichier %%%%%%%%%%%%%%%%%%%%%%%%
_ 50
9 47 28 202 38 201 49 200 60 197 70 192 81 189 91 187
h 88 7 172 13 168 18 168 30 169 35 172 47 176 52 176 81 193 86 203
9 53 15 230 25 238 34 244 43 249 53 252 62 253 72 254 81 255 91 256
r 72 7 257 14 258 21 259 35 263 42 266 56 268 63 269 76 271 83 272
i 87 9 271 15 270 21 269 26 269 32 268 38 267 44 266 49 265 55 264
k 67
@ 61 15 216 23 215 31 213 39 209 48 203 56 198 64 194 72 192 80 192
n 88 9 196 15 195 26 193 38 191 49 189 55 189 66 191 77 193 89 193
w 44 11 192 23 192 34 192 45 192 57 192 68 193 80 193 91 193
O 37 16 193 30 194 43 195 57 196 70 195 84 194 97 191
z 59 15 179 24 172 92 208 100 208
@ 42 12 207 24 206 36 205 48 203 60 202 71 199 83 196 95 194
n 60 13 194 22 196 30 197 38 197 47 197 55 196 63 196 72 196 80 196
aU 152 5 196 15 191 25 188 35 185 45 185 55 185 64 185 74 184 84 184
n 33 94 65
s 59 14 154 22 184 31 184
t 87
D 26
@ 36 17 209 31 209 44 206 58 203 72 199 86 196 100 191
s 83 6 188 96 222
a 123 6 221 14 216 26 210 34 206 46 201 54 200 67 200 75 200 87 201
f 84
t 58 95 214
@ 44 16 213 27 212 39 212 50 212 61 212 73 213 84 213 95 213
n 29 28 211 45 210 62 209 79 208 97 208
u 102 9 207 19 205 28 203 38 201 48 199 58 197 68 196 77 194 87 191
n 59 12 185 20 186 29 187 37 188 46 188 54 187 63 186 71 185 80 185
O 76 11 188 17 187 24 186 37 185 43 185 57 184 63 184 76 184 83 183
n 71 10 181 17 182 24 183 38 185 45 185 52 185 66 185 73 185 80 185
D 33 18 184 33 184 48 184 64 185 79 185 94 186
@ 88 9 189 15 189 20 189 26 189 32 189 38 189 43 186 49 182
t 80 100 212
i 132 4 212 11 210 23 205 34 202 45 200 53 198 64 194 76 193 87 192
v 102 8 183 18 175 27 173 37 173 47 175 57 176 67 182 76 189 86 195
i 308 2 201 12 195 21 182 31 176 41 172 51 168 60 166 70 168 80 166
_ 50
%%%%%%%%%%%%%%%%%% fin du fichier %%%%%%%%%%%%%%%%%%%%%%%%
```

Après transformation en mode *sasasa plat*, ce même fichier devient :

```
%%%%%%%%%%%%%%%%%% début du fichier %%%%%%%%%%%%%%%%%%%%%%%%
_ 50 0 230
a 47 0 230
s 88 0 230
a 53 0 230
s 72 0 230
a 87 0 230
s 67 0 230
a 61 0 230
s 88 0 230
s 44 0 230
a 37 0 230
s 59 0 230
a 42 0 230
s 60 0 230
a 152 0 230
s 33 0 230
s 59 0 230
s 87 0 230
s 26 0 230
a 36 0 230
s 83 0 230
a 123 0 230
s 84 0 230
s 58 0 230
a 44 0 230
s 29 0 230
a 102 0 230
s 59 0 230
a 76 0 230
s 71 0 230
s 33 0 230
a 88 0 230
s 80 0 230
a 132 0 230
s 102 0 230
a 308 0 230
_ 50 0 230
%%%%%%%%%%%%%%%%%% fin du fichier %%%%%%%%%%%%%%%%%%%%%%%%
```

B.2 Instructions pour les expériences 2a–2d

Vous allez entendre des phrases en Sahatu et en Moltec, ayant subi une transformation acoustique de manière à les rendre méconnaissables. Votre tâche sera de déterminer, pour chaque phrase, s'il s'agit d'une phrase prononcée en Sahatu ou en Moltec.
Vous allez entendre dans un premier temps une phrase sahatu et une phrase moltèque pour vous familiariser. Puis vous entendrez les phrases une par une et, après écoute de chaque phrase, il vous sera demandé d'entrer votre réponse : "s" si vous pensez que la phrase était en Sahatu, "m" si vous pensez qu'elle était en Moltec.

Ne vous pressez pas pour répondre, vous avez tout votre temps. Après chaque réponse, le résultat vous sera donné. Vous pourrez ainsi en tirer un enseignement et vous améliorer au fur et à mesure.
Si tout ne vous paraît pas clair, n'hésitez pas à poser des questions avant de commencer.
Lorsque vous êtes prêt, appuyez sur une touche pour commencer.

B.3 Instructions pour l'expérience 2e

You are about to listen to sentences in Sahatu and in English. These sentences have been acoustically modified in order to preserve only their intonation. Your task will be to decide for each sentence whether it was uttered in Sahatu or in English. You will first hear a sahatu and an english sentence as examples. Then you will hear more sentences one by one, and at the end of each sentence, you will be asked whether it was sahatu or english. Type "s" for sahatu and "e" for english. Take your time to answer, try to be accurrate.
After each answer, you will be told whether you are right or wrong, in order for you to understand your errors and try to improve your scores.

If there is anything you think you may have misunderstood, feel free to ask questions to the experimenter.
Don't forget to wear the headphones!

When you are ready, press any key to start.

B.4 Instructions pour l'expérience 3a

Vous allez entendre des phrases en Sahatu et en Moltec, ayant subi une transformation acoustique de manière à les rendre méconnaissables. Votre tâche sera de déterminer, pour chaque phrase, s'il s'agit d'une phrase prononcée en Sahatu ou en Moltec.
Vous allez entendre dans un premier temps une phrase sahatu et une phrase moltèque pour vous familiariser. Puis vous entendrez les phrases une par une et, après écoute de chaque phrase, il vous sera demandé d'entrer votre réponse : "s" si vous pensez que la phrase était en Sahatu, "m" si vous pensez qu'elle était en Moltec.
Ne vous pressez pas pour répondre, vous avez tout votre temps. Après chaque réponse, le résultat vous sera donné. Vous pourrez réécouter la phrase lorsque vous le désirerez.
Vous pourrez ainsi en tirer un enseignement et vous améliorer au fur et à mesure.
Si tout ne vous paraît pas clair, n'hésitez pas à poser des questions avant de commencer.
Lorsque vous êtes prêt, appuyez sur une touche pour commencer.

B.5 Instructions pour les expériences 3b, 4a–4b, et 5a–5c

Les phrases que vous allez entendre sont des phrases de deux langues différentes: le Sahatu et le Moltec. Ces phrases ont subi une transformation acoustique de manière à les rendre méconnaissables.
L'expérience se déroulera de la façon suivante: Vous entendrez des groupes de trois phrases

phrase 1 phrase 2 phrase X

Les deux premières seront toujours dans la même langue. Votre tâche consistera à déterminer, pour

```
chaque groupe, si la phrase X est ou non dans la même langue
que les deux premières.

Ainsi, un choix de deux réponses vous
sera proposé: (P)areil ou (D)ifférent Vous devrez alors tapez
P si vous pensez que la troisième phrase est dans la même
langue que les deux premières, D sinon.
Ne vous pressez pas pour répondre, vous avez tout votre
temps. Après chaque réponse, le résultat vous sera donné.
Vous pourrez ainsi en tirer un enseignement et vous
améliorer au fur et à mesure.
Si tout ne vous paraît pas clair, n'hésitez pas à poser des
questions avant de commencer.
Lorsque vous êtes prêt, appuyez
sur une touche pour commencer.
```

B.6 Séquence de phrases entendues par un sujet dans l'expérience 3b

Nous reproduisons ci-dessous le fichier de résultats du premier sujet de l'expérience 3b. Les indications « English » et « Spanish » indiquent le début d'un nouveau bloc. Chaque ligne représente un essai. Les colonnes 2 et 3 indiquent les deux phrases de contexte de chaque essai, et la colonne 4 la phrase de test. La colonne 5 donne la réponse du sujet (P pour pareil, D pour différent), et la colonne 6 indique si la réponse est bonne ou pas.

Le nom de fichier de chaque phrase indique la langue (« enl » pour anglais et « spl » pour espagnol, et la locutrice (numéro de 1 à 4 dans chaque langue, premier chiffre en partant de la gauche). Les deux chiffres suivants indiquent le nombre de syllabes (de 15 à 19), et le dernier est le numéro de la phrase pour cette locutrice et ce nombre de syllabes.

```
Sujet   Phrase 1    Phrase 2    Phrase 3 Réponse Résultat
sujet1  %%%  English
sujet1  enl1155.raw enl2185.raw enl3164.raw P True
sujet1  enl1182.raw enl2176.raw spl5185.raw D True
sujet1  enl2191.raw enl1194.raw spl5161.raw P False
sujet1  enl1166.raw enl2176.raw enl4156.raw D False
sujet1  enl2152.raw enl1155.raw enl4184.raw D False
sujet1  enl2162.raw enl1182.raw spl4183.raw D True
sujet1  enl1194.raw enl2176.raw enl4195.raw D False
```

```
sujet1  enl2152.raw enl1155.raw enl4174.raw D False
sujet1  enl2162.raw enl1194.raw spl5191.raw P False
sujet1  enl1182.raw enl2191.raw enl3153.raw P True
sujet1  enl2162.raw enl1171.raw spl4172.raw D True
sujet1  enl1166.raw enl2185.raw enl3175.raw D False
sujet1  enl2176.raw enl1171.raw spl5176.raw D True
sujet1  enl2152.raw enl1166.raw spl4162.raw P False
sujet1  enl1155.raw enl2185.raw spl4197.raw D True
sujet1  enl1194.raw enl2191.raw enl3192.raw P True
sujet1  enl2185.raw enl1171.raw enl3181.raw P True
sujet1  enl1166.raw enl2152.raw spl4154.raw D True
sujet1  enl2162.raw enl1182.raw enl4161.raw D False
sujet1  enl1171.raw enl2191.raw spl5153.raw D True
sujet1  %%%  Spanish
sujet1  spl2182.raw spl3184.raw spl5191.raw P True
sujet1  spl3155.raw spl2194.raw spl5176.raw D False
sujet1  spl3196.raw spl2151.raw enl3164.raw D True
sujet1  spl2182.raw spl3174.raw enl4161.raw P False
sujet1  spl3155.raw spl2164.raw enl4184.raw D True
sujet1  spl3196.raw spl2182.raw spl4197.raw D False
sujet1  spl2151.raw spl3184.raw enl3153.raw D True
sujet1  spl2171.raw spl3163.raw enl4195.raw D True
sujet1  spl3184.raw spl2164.raw enl4156.raw P False
sujet1  spl2194.raw spl3196.raw spl4162.raw P True
sujet1  spl2171.raw spl3174.raw enl3181.raw D True
sujet1  spl3155.raw spl2182.raw spl4172.raw D False
sujet1  spl3163.raw spl2164.raw spl5161.raw P True
sujet1  spl2151.raw spl3155.raw spl5185.raw P True
sujet1  spl3163.raw spl2194.raw enl3192.raw P False
sujet1  spl2171.raw spl3174.raw enl3175.raw D True
sujet1  spl3184.raw spl2194.raw spl5153.raw P True
sujet1  spl3163.raw spl2171.raw spl4154.raw D False
sujet1  spl2164.raw spl3174.raw spl4183.raw P True
sujet1  spl2151.raw spl3196.raw enl4174.raw D True
```

B.7 Instructions pour les expériences 6a–6c

B.7.1 Première moitié des sujets

```
Les phrases que vous allez entendre sont des phrases de
```

```
deux langues différentes: le Sahatu et le Moltec.
Ces phrases ont subi une transformation
acoustique de manière à les rendre méconnaissables.

L'expérience se déroulera de la façon suivante:
Vous entendrez des groupes de trois phrases
        phrase 1     phrase 2     phrase X
Les deux premières seront toujours dans la même langue.

Votre tâche consistera à déterminer, pour chaque groupe, si la
phrase X est ou non dans la même langue que les deux premières.
Ainsi, un choix de deux réponses vous sera proposé:
        (P)areil ou (D)ifférent
Vous devrez alors tapez P si vous pensez que la troisième phrase
est dans la même langue que les deux premières, D sinon.

Ne vous pressez pas pour répondre, vous avez tout votre temps.
Après chaque réponse, le résultat vous sera donné. Vous
pourrez ainsi en tirer un enseignement et vous améliorer
au fur et à mesure.
Si tout ne vous paraît pas clair, n'hésitez pas à poser des
questions avant de commencer.
La tâche n'est pas facile, mais elle est possible.
Accrochez-vous!

Lorsque vous êtes prêt, appuyez sur une touche pour commencer.
```

B.7.2 Deuxième moitié des sujets

```
Les phrases que vous allez entendre sont des phrases de
deux langues différentes: le Sahatu et le Moltec.
Ces phrases ont subi une transformation
acoustique de manière à les rendre méconnaissables.

L'expérience se déroulera de la façon suivante:
Vous entendrez des groupes de trois phrases
        phrase 1     phrase 2     phrase X
Les deux premières seront toujours dans la même langue.

Votre tâche consistera à déterminer, pour chaque groupe, si la
phrase X est ou non dans la même langue que les deux premières.
```

Ainsi, un choix de deux réponses vous sera proposé:
(P)areil ou (D)ifférent
Vous devrez alors tapez P si vous pensez que la troisième phrase est dans la même langue que les deux premières, D sinon.
Les deux langues ne diffèrent que par leur rythme. Il est donc important de vous concentrer sur le rythme des phrases, afin d'arriver à distinguer les deux langues.

Ne vous pressez pas pour répondre, vous avez tout votre temps.
Après chaque réponse, le résultat vous sera donné. Vous pourrez ainsi en tirer un enseignement et vous améliorer au fur et à mesure.
Si tout ne vous paraît pas clair, n'hésitez pas à poser des questions avant de commencer.
La tâche n'est pas facile, mais elle est possible.
Accrochez-vous!

Lorsque vous êtes prêt, appuyez sur une touche pour commencer.

B.8 Instructions pour l'expérience 10 bis

You are going to listen to 2 series of bizarre synthesized sentences uttered in undetermined languages.
Your task will be to judge the intonation of each of these sentences, that is, does the global intonation (melody) of this sentence sound right, does it sound like a possible sentence in some human language, or does the intonation sound somewhat unnatural, is there some local or global feature of this intonation that does not sound natural to your ears?
After asking yourself such questions, please rate each sentence on a 0-to-5 scale, 0= very unnatural, 5= perfectly natural.

If the instructions are not yet clear to you, feel free to ask questions to the experimenter.

Press any key to start.

Annexe C

Expériences supplémentaires de discrimination de langues par des nouveau-nés

C.1 Exp. 7 bis : néerlandais-japonais en parole naturelle à l'envers

Les expériences 7, 7 bis, 12 et 13 ont été planifiées ensemble. L'idée était de tester simultanément la discrimination néerlandais-japonais en parole naturelle à l'endroit et à l'envers, sur les nouveau-nés et sur les singes. L'expérience 7 bis est celle qui testait la parole naturelle à l'envers sur le nouveau-né, et elle a été effectuée simultanément avec l'expérience 7, testant la parole à l'endroit. Nous avons effectué une analyse partielle de ces deux expériences alors qu'elles comportaient chacune 20 sujets, et nous avons constaté que ni l'une ni l'autre n'indiquaient une tendance à la discrimination. Nous avons alors décidé de compléter l'expérience 7 à 32 sujets pour obtenir un résultat définitif, mais il est bien évident que l'expérience 7 bis perdait dès lors tout son intérêt. Nous avons donc arrêté cette expérience après les 20 premiers sujets.

C.1.1 Stimuli

Nous avons utilisé les mêmes phrases néerlandaises et japonaises que dans l'expérience 7, mais jouées à l'envers (l'ordre des échantillons a été inversé).

C.1.2 Participants

Vingt bébés ont été testés avec succès, 13 filles et 7 garçons, dont les caractéristiques sont données dans le Tableau C.1. Dix-sept d'entre eux étaient issus de familles françaises monolingues, 2 de familles où l'on parle le français plus une ou plusieurs autres langues, 1 d'une famille où l'on ne parle pas le français.

TAB. C.1 – *Caractéristiques des nouveau-nés de l'Expérience 7 bis.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	41	98	61	16
Terme (semaines + jours)	38	41+4	39+6	1+1
Poids (g)	2950	4000	3341	248

De plus, les résultats de 24 bébés ont été rejetés pour les raisons suivantes : rejet de la tétine (1), pas assez éveillé/succion insuffisante avant le changement (4), pleurs/agitation (5), n'atteint pas le critère d'habituation (6), s'endort/succion insuffisante après le changement (6), perd la tétine après le changement (2).

C.1.3 Résultats

La Figure C.1 présente le nombre de succions HA par minute pour les deux groupes de sujets. Elle reprend les 5 minutes précédant et les 4 minutes suivant le changement pour chaque sujet.

L'aspect général des courbes de succions est plus bruité que pour les autres expériences, ce qui n'est pas étonnant compte tenu du fait qu'il y a seulement 10 bébés par groupe. Les bébés du groupe expérimental semblent avoir tété plus fort que ceux du groupe contrôle tout au long du test. Sur les 5 dernières minutes de la phase d'habituation, cet effet est marginalement significatif [$F(1,19) = 4.13$, $p = 0.057$]. Il ne l'est véritablement que si l'on considère les 2 dernières minutes de l'habituation [$F(1,19) = 8.8$, $p = 0.008$]. Il s'agit vraisemblablement d'un effet d'échantillonnage des bébés, ceux qui tétaient le plus fort ayant été attribués par hasard au groupe expérimental. Il n'y a pas d'effet de la langue d'habituation sur le nombre de succions pendant cette phase [$F(1,19) < 1$].

Après le changement, les bébés n'ont pas plus augmenté leurs succions dans le groupe expérimental que dans le groupe contrôle [$F(1,17) = 1.43$, $p = 0.25$ sur *post2*, la tendance allant dans le sens inverse, c'est-à-dire plus d'augmentation pour le groupe contrôle]. Sur la figure, il semble que l'augmentation des succions par le groupe expérimental soit significative lors de

la 3e minute après le changement, mais rapportée à la différence de succions déjà présente entre les deux groupes deux minutes avant le changement, l'analyse de covariance sur cette 3e minute en prenant *ante2* comme covariable ne donne aucun effet de groupe [$F(1,17) < 1$].

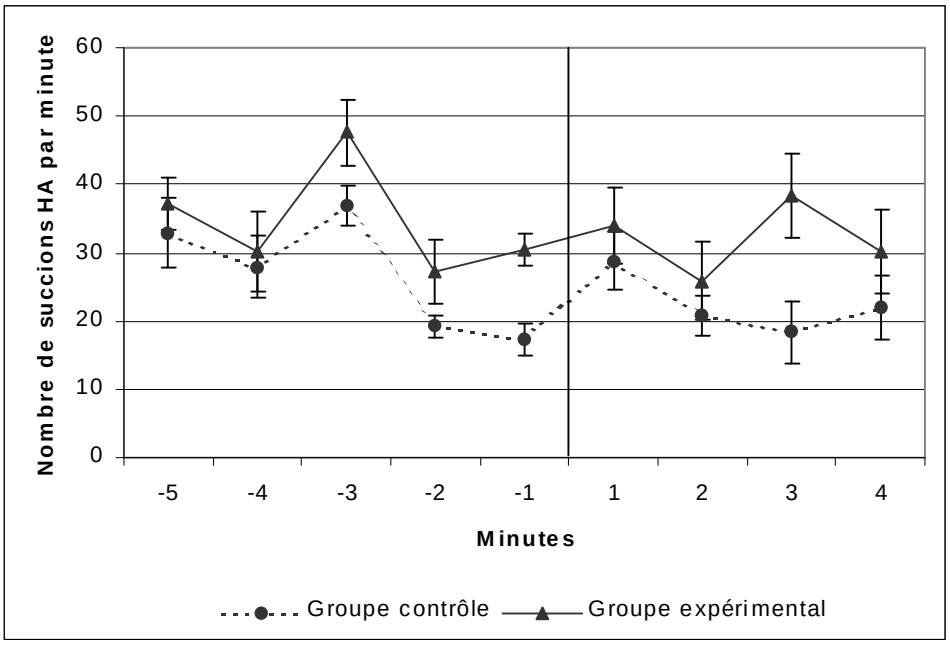

FIG. C.1 – *Exp. 7bis : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle à l'envers (10 bébés par groupe). Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne. La ligne verticale indique le moment du changement.*

Ainsi, les bébés n'ont pas réagi plus au changement de langue qu'au changement de locutrices. Même si avec 20 bébés on ne s'attend pas à observer un effet de discrimination significatif, le fait qu'il n'y ait aucune tendance dans cette direction suggère fortement que le résultat aurait été le même avec 12 ou 20 bébés de plus. A posteriori, le résultat n'est de toute façon pas étonnant, puisque l'absence de discrimination est prédite à la fois par la présence de 4 voix (comme dans l'expérience 7) et par la parole à l'envers (comme dans l'expérience 9). Il est donc parfaitement justifié d'avoir interrompu cette expérience dans l'état.

C.2 Exp. 7 ter : néerlandais-japonais en parole naturelle, plusieurs changements par sujet

À l'issue de l'expérience 7, nous n'avons pas immédiatement pensé que la variabilité des locutrices puisse être la cause de l'absence de discrimination. Nous avons au contraire pensé que le problème principal était la faible sensibilité de la méthode de succion de haute amplitude. Nous avons donc choisi de refaire l'expérience en utilisant une nouvelle méthode de test potentiellement plus puissante.

C.2.1 Stimuli

Nous avons utilisé les mêmes phrases néerlandaises et japonaises que dans l'expérience 7, c'est-à-dire les phrases originales jouées à l'endroit.

C.2.2 Participants

Quarante bébés ont été testés avec succès, 17 filles et 23 garçons, dont les caractéristiques sont données dans le Tableau C.2. Trente-six d'entre eux étaient issus de familles françaises monolingues, 1 d'une famille où l'on parle le français et une autre langue, 3 de familles où l'on ne parle pas le français.

TAB. C.2 – *Caractéristiques des nouveau-nés de l'Expérience 7 ter.*

	Min.	Max.	Moyenne	Écart-type
Âge (heures)	41	110	68	17
Terme (semaines + jours)	38+1	41+4	39+5	0+6
Poids (g)	2880	4740	3482	414

De plus, les résultats de 19 bébés ont été rejetés pour les raisons suivantes : pas assez éveillé/succion insuffisante (3), pleurs/agitation (9), rejet a posteriori (5) (voir ci-dessous), test interrompu par un parent (1), fichier perdu (1).

C.2.3 Procédure

Le principe de la nouvelle procédure est de multiplier le nombre de changements subi par chaque sujet, afin d'obtenir plusieurs mesures par sujet, et notamment une mesure intra-sujet de la discrimination. C'est le même souci qui a conduit à l'emploi de 2 changements par sujet lors des expériences sur les enfants de 2 mois (Hesketh et coll., 1997; Christophe & Morton, 1998),

et qui nous a incité à faire de même dans l'expérience 8. Néanmoins, comme nous l'avons vu à la Section 5.4.4.2, trop peu de nouveau-nés tiennent jusqu'au deuxième changement pour que la technique présentent un intérêt pour nous.

Ici, nous proposons d'augmenter radicalement le nombre de changements : jusqu'à 8 par sujet. Comme il n'est pas possible d'augmenter la durée du test, l'intervalle entre deux changements consécutifs est réduit à 2 minutes, soit juste le temps nécessaire à la manifestation d'une déshabituation consécutivement à un changement. Les changements sont alternativement du type contrôle (changement de locutrices) et du type expérimental (changement de langue et de locutrices). Du fait que les changements sont programmés toutes les deux minutes, il n'y a plus de critère d'habituation. On peut donc s'attendre à une augmentation du bruit dans les données, dans la mesure où il n'y a plus de garantie qu'un bébé aura la possibilité d'augmenter ses succions consécutivement à un changement de langue. Néanmoins, nous supposons que la perte de sensibilité due à l'absence de critère d'habituation peut être compensée par la multiplication des mesures : jusqu'à 8 par bébé. Le Tableau C.3 décrit l'ordre d'écoute des langues et des locutrices selon les différentes conditions.

TAB. C.3 – *Conditions de l'expérience 7ter.*

Minutes:	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18
Cond. J1J2	J1	J2	D1	D2	J2	J1	D2	D1	J1
Cond. J1D2	J1	D2	D1	J1	J2	D1	D2	J2	J1
Cond. D2D1	D2	D1	J2	J1	D1	D2	J1	J2	D2
Conc. D2J1	D2	J1	J2	D2	D1	J2	J1	D1	D2

Les conditions sont nommées d'après le premier changement.
J1: 10 phrases prononcées par deux locutrices japonaises.
J2: 10 phrases prononcées par deux autres locutrices japonaises.
D1: 10 phrases prononcées par deux locutrices néerlandaises.
D2: 10 phrases prononcées par deux autres locutrices néerlandaises.

Les conditions de rejet sont similaires à celles décrites à la Section 5.2.3.3, à ceci près qu'elles s'appliquent changement par changement. Ainsi, si le bébé n'entend aucune phrase pendant une minute, ou si l'expérimentateur est amené à appuyer sur la touche F1, seul le changement correspondant à cette minute est rejeté (soit 4 minutes de données en tout), ce qui laisse à l'enfant la possibilité de passer d'autres changements avec succès. Pour analyser les données d'un bébé, nous imposons qu'il ait effectué avec succès au moins un changement contrôle et un changement expérimental, ce qui permet d'avoir

pour chaque bébé une mesure intra-sujet de la discrimination. Les données des bébés qui ne vérifient pas cette condition à l'issue du test sont rejetées a posteriori. Les données analysées comportent donc entre 2 et 8 changements par sujet.

C.2.4 Résultats

Chaque bébé a effectué en moyenne 5.8 changements avec succès. Nous considérons deux groupes de sujets, recouvrant les quatre conditions ci-dessus : le Groupe 1 rassemble les sujets dont le premier changement est du type contrôle (J1J2 ou D1D2), et le Groupe 2 ceux dont le premier changement est du type expérimental (J1D2 ou D2J1). Du fait de l'alternance des changements contrôles et expérimentaux, ces deux groupes de sujets sont en permanence dans la situation opposée l'un de l'autre. Nous allons donc rechercher un effet significatif du facteur groupe sur les succions des bébés. La Figure C.2 présente le nombre de succions HA par minute pour les deux groupes de sujets.

On observe tout d'abord que les bébés du Groupe 1 ont tendance à téter moins que ceux du Groupe 2 sur l'ensemble du test. Cet différence n'est cependant pas significative lorsque l'on considère le nombre moyen de succions par minute [$F(1,39) = 1.94$, $p = 0.17$]. Pour chaque changement, et pour chaque bébé ayant effectué ce changement avec succès, nous calculons un score de déshabituation, égal au nombre de succions pendant les 2 minutes suivant le changement, moins le nombre de succions pendant les 2 minutes précédant le changement. Ce score reflète l'augmentation des succions consécutivement à un changement. Si les bébés discriminent les deux langues, on s'attend à ce que les scores de déshabituation soient supérieurs lors des changements expérimentaux que lors des changements contrôles. La Figure C.3 donne les valeurs de ces scores pour les huit changements successifs, en fonction du type de changement.

Dans 6 changements sur 8, les scores de déshabituation sont effectivement supérieurs pour les changements expérimentaux que contrôles, mais les différences entre les deux conditions sont très faibles. Nous avons conduit une analyse de variance prenant comme variable dépendante les scores de déshabituation, comme facteurs le changement (de 1 à 8) et son type (contrôle ou expérimental). Celle-ci confirme qu'il n'y a pas d'effet significatif du type de changement [$F(1,217) = 1.65$, $p = 0.20$]. Il n'y a pas non plus d'effet de l'ordre dans lequel interviennent les changements [$F(7,217)<1$], ni d'interaction entre les deux facteurs [$F(7,217)<1$].

Nous avons effectué une analyse supplémentaire, en calculant pour chaque bébé, la différence entre la moyenne de ses scores lors de changements ex-

périmentaux, et la moyenne de ses scores lors de changements contrôles. Ce nouvel indice donne une mesure intra-sujet de la différence de réaction aux changements de langues et aux changements de voix. La moyenne de cet indice est 2.5 ± 28.4, et n'est pas significativement supérieure à 0 [$t(39) < 1$; $p = 0.87$ par un test binômial].

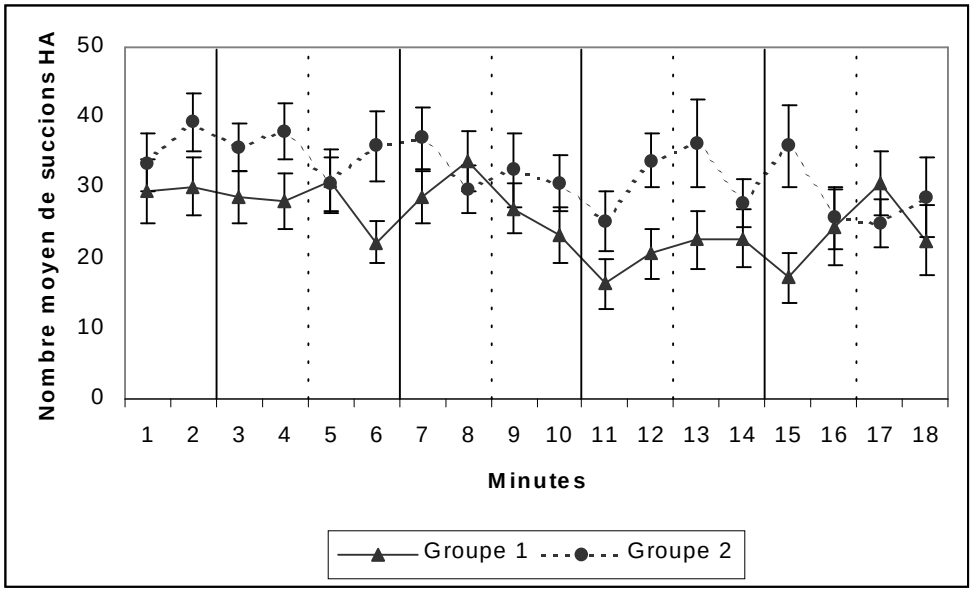

FIG. C.2 – *Exp. 7ter : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle à l'endroit – plusieurs changements par sujets. Les lignes verticales continues représentent les changements contrôles pour le Groupe 1, et expérimentaux pour le Groupe 2. Les lignes verticales pointillées représentent les changements expérimentaux pour le Groupe 1, et contrôles pour le Groupe 2. Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne.*

C.2.5 Discussion

Il est donc clair que les bébés n'ont pas réagi plus aux changements de langues qu'aux changements de voix, et n'ont donc pas discriminé les langues. Ce résultat est identique à celui obtenu à l'expérience 7. Néanmoins, il n'y a pas de conséquences à en tirer, dans la mesure où c'est la première fois que cette procédure est utilisée : nous ne savons pas si elle serait susceptible de révéler une discrimination le cas échéant.

Cette expérience était donc mal choisie pour tester une nouvelle procédure expérimentale. Il faudrait réutiliser celle-ci dans des conditions où un

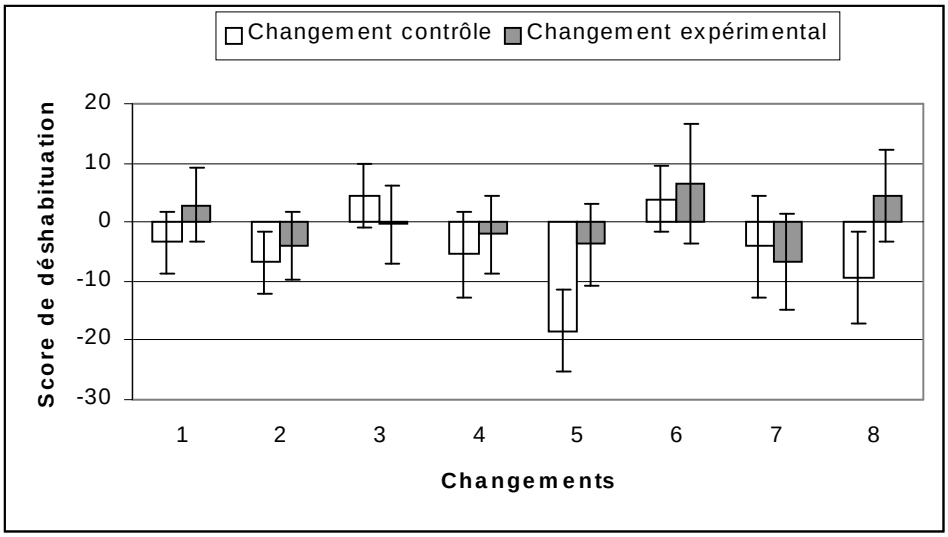

FIG. C.3 – *Scores de déshabituation de l'expérience 7 ter, changement par changement, en fonction du type de changement (expérimental ou contrôle). Les barres d'erreur représentent* ± 1 *erreur standard de la moyenne.*

résultat positif est attendu (comme l'expérience 8 avec les phrases resynthétisées), pour savoir si elle permet effectivement de mettre en évidence des discriminations, et éventuellement si elle est plus sensible que la méthode d'habituation/déshabituation traditionnelle. Alors seulement, le résultat négatif obtenu ici pourrait devenir interprétable.

Bibliographie

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine.

Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, *22*, 239-268.

Aslin, R., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 117-134). Mahwah, NJ: Lawrence Erlbaum Associates.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321-324.

Auer, P. (1993). *Is a rhythm-based typology possible? A study of the role of prosody in phonological typology* (KontRI Working Paper No. 21). Universität Hamburg.

Bahrick, L. E., & Pickens, J. N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant Behavior and Development*, *11*, 277-296.

Bertinetto, P. M. (1981). *Strutture prosodiche dell' italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici.* Firenze: Accademia della Crusca.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21-33.

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech*, *38*, 311-329.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, *29*, 711-721.

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (p. 206-244). Cambridge: Blackwell.

Bolinger, D. (1965). Pitch accent and sentence rhythm. In *Forms of English: Accent, morpheme, order.* Cambridge, MA: Harvard University Press.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of Human Language* (p. 471-524). Stanford, CA: Stanford University Press.

Bolinger, D. L., & Gerstman, L. J. (1957). Disjuncture as a cue to constraints. *Word*, *13*, 246-255.

Bond, Z. S., & Fokes, J. (1991). Identifying foreign languages. In *XIIth International Congress of Phonetic Sciences* (p. 198-201).

Borzone de Manrique, A. M., & Signorini, A. (1983). Segmental durations and the rhythm in Spanish. *Journal of Phonetics*, *11*, 117-128.

Bosch, L., & Sebastián-Gallés, N. (1997). Native language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*, 33-69.

Bourgeois, T. C. (1991). *Instantiative phonology.* Thèse de doctorat non publiée, University of Arizona.

Brady, S., & Shankweiler, D. (1991). *Phonological processes in literacy.* Hillsdale, NJ: Lawrence Erlbaum.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93-125.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111-153.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, *30*, 348-369.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, *35*, 26-58.

Chomsky, N. (1976). *Reflections on language.* London: Temple Smith.

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht: Foris.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221-268.

Christophe, A. (1993). *Rôle de la prosodie dans la segmentation en mots.* Thèse de doctorat non publiée, EHESS.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*(3), 1570-1580.

Christophe, A., Guasti, T., Nespor, M., Dupoux, E., & van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, *12*(5/6), 585-612.

Christophe, A., & Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-month-olds. *Developmental Science*, *1*(2), 215-219.

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*(1-2), 53-69.

Colombo, J., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, *4*, 219-223.

Comrie, B. (1989). *Language Universals and Linguistic Typology.* Chicago: The University of Chicago Press.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (p. 105-121). Cambridge Mass: MIT Press.

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 87-99). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*, 385–400.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113-121.

Cutler, A., & Otake, T. (1994). Mora or phoneme - further evidence for language-specific listening. *Journal of Memory and Language*, *33*, 824-844.

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, *12*, 58-69.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *XIth International Congress of Phonetic Sciences* (Vol. 5, p. 447-450). Tallinn.

de Pijper, J. R. (1983). *Modelling British English intonation.* Dordrecht - Holland: Foris.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, *96*, 2037-2047.

de Schonen, S., Deruelle, C., Mancini, J., & Pascalis, O. (1993). Hemispheric differences in face processing and brain maturation. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (p. 149-163). The Netherlands: Kluwer Academic Publishers.

Dehaene-Lambertz, G. (1995). *Capacités linguistiques précoces et leurs bases cérébrales.* Thèse de doctorat non publiée, EHESS.

Dehaene-Lambertz, G., & Houston, D. (1998). Faster orientation latencies toward native language in two-month old infants. *Language and Speech*, *41*(1), 21-43.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*(5604), 718-9.

den Os, E. (1988). *Rhythm and tempo of Dutch and Italian.* Thèse de doctorat non publiée, Rijksuniversiteit.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (Melopsittacus undulatus). *Journal of the Acoustical Society of America*, *102*(3), 1891-1897.

Dominey, P., & Ramus, F. (sous presse). Neural network processing of natural language: I. Sensitivity to Serial, Temporal and Abstract Structure of language in the Infant. *Language and Cognitive Processes.*

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*(3), 275-277.

Donegan, P. J., & Stampe, D. (1983). Rhythm and the holistic organization of language structure. In F. Richardson (Ed.), *CLS Parasession on The Interplay of Phonology, Morphology and Syntax* (Vol. 19, p. 337-353).

Duanmu, S. (1994). Syllabic weight and syllabic duration : a correlation between phonology and phonetics. *Phonology*, *11*, 1-24.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (sous presse). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance.*

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *ICSLP'96.* Philadelphia.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, *36*, 202-225.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303-306.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181-195.

Fernald, A., & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 365-388). Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, C., & Tokura, H. (1996). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 343-363). Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, C. L., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, *67*, 3192-3218.

Floccia, C. (1996). *Perception de la parole et apprentissage chez le nouveau-né: Etude méthodologique de la procédure de renforcement des succions de haute amplitude.* Thèse de doctorat non publiée, EHESS.

Floccia, C., Christophe, A., & Bertoncini, J. (1997). High-amplitude sucking and newborns: The quest for underlying mechanisms. *Journal of Experimental Child Psychology*, *64*, 175-198.

Fodor, J., & Pylyshyn, Z. (1981). How direct is visual perception?: Some reflections on Gibson's "Ecological Approach". *Cognition*, *9*(2), 139-196.

Fowler, C. A., Smith, M. R., & Tassinary, L. G. (1986). Perception of syllable timing by prebabbling infants. *Journal of the Acoustical Society of America*, *79*(3), 814-825.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287-295.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (p. 301-330). London: Lawrence Erlbaum.

Galaburda, A., & Livingstone, M. (1993). Evidence for a magnocellular defect in developmental dyslexia. *Annals Of The New York Academy Of Sciences*, *682*, 70-82.

Gerken, L. (1996). Phonological and distributional information in syntax acquisition. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 411-425). Mahwah, NJ: Lawrence Erlbaum Associates.

Gerken, L., Jusczyk, P., & Mandel, D. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237-65.

Gibson, J. J. (1950). *Perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3-55.

Gleitman, L., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (p. 3-48). Cambridge UK: Cambridge University Press.

Goodsitt, J., Morgan, J., & Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*, 229-252.

Guasti, M. T., Nespor, M., Christophe, A., & van Ooyen, B. (sous presse). Pre-lexical setting of the head-complement parameter through prosody. In J. Weissenborn & B. Höhle (Eds.), *Signal to Syntax II.*

Hakuta, K. (1985). *Mirror of language: The debate on bilingualism.* New York: Basic Books.

Haraguchi, S., & Tanaka, S. (1992). A systematic study of the role and characteristics of mora. In S. Haraguchi (Ed.), *A systematic study of Japanese mora and syllable structure. Progress report.*

Harris, R. J., & Nelson, E. M. M. (1992). Bilingualism: Not the exception anymore. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (p. 3-14). Amsterdam: North Holland.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146-162.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*, 190-222.

Hayes, B., & Puppel, S. (1985). On the rhythm rule in Polish. In H. van der Hulst & N. Smith (Eds.), *Advances in nonlinear phonology* (p. 59-81). Dordrecht: Foris.

Hayes, J., & Clark, H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. Hayes (Ed.), *Cognition and the development of language* (p. 221-234). Wiley New York.

Hesketh, S., Christophe, A., & Dehaene-Lambertz, G. (1997). Non-nutritive sucking and sentence processing. *Infant Behavior and Development*, *20*, 263-269.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension.* Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*, 269-286.

Hockett, C. F. (1958). *A course in modern linguistics.* New York: Macmillan.

Hoequist, C. J. (1983). Durational correlates of linguistic rhythm categories. *Phonetica*, *40*, 19-31.

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, *56*(6), 613-623.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley and Sons.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (sous presse). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review.*

Hung, F.-S., & Peters, A. M. (1997). The role of prosody in the acquisition of grammatical morphemes: evidence from two Chinese languages. *Journal of Child Language*, *24*, 627-650.

Itô, J., & Mester, A. (1995). Japanese phonology. In J. Goldsmith (Ed.), *The handbook of phonological theory* (p. 817-838). Oxford: Blackwell.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*, 335-344.

Jusczyk, P. (sous presse). Using the headturn preference procedure to study language acquisition. In E. Bavin & D. Burnham (Eds.), *Advances in Infancy Research* (Vol. 13).

Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.

Jusczyk, P., Kemler-Nelson, D., Hirsh-Pasek, K., Kennedy, L., Woodward, A., Piwoz, J., et coll.. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252-293.

Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630-645.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675-687.

Jusczyk, P. W., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, *32*, 402-420.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (sous presse). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (sous presse). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*.

Jusczyk, P. W., & Kemler-Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 389-408). Mahwah, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W., Pisoni, D. B., & Mullenix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*, 253-291.

Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language*, *27*, 343-358.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349-364.

Kelly, M. H. (1996). The role of phonology in grammatical category assignment. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 249-262). Mahwah, NJ: Lawrence Erlbaum Associates.

Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 389-403.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright-Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*, 55-68.

Kemler Nelson, D. G., Jusczyk, P., Mandel, D., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111-116.

Kennedy, & Bush. (1985). *An introduction to the design and analysis of experiments in behavioral research.* Lanham: University Press of America.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*, 1195-1197.

Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*, 163-175.

Kucera, H., & Francis, N. (1967). *A computational analysis of present day English.* Providence, RI: Brown University Press.

Kuhl, P. (1987). The special-mechanisms debate in speech research: Categorization tests on animal and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (p. 355-386). Cambridge University Press.

Kuhl, P. K. (1979a). Models and mechanisms in speech perception: Species comparison provide further contributions. *Brain, behavior and evolution*, *16*, 374-408.

Kuhl, P. K. (1979b). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*, 1668-1679.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, *70*, 340-349.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263-285.

Kuhl, P. K. (1985). Methods in the study of infant speech perception. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview.* Norwood, NJ: Ablex.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69-72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*, 905-917.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, *32*, 542-550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, *73*, 1003-1010.

Kujala, T., Myllyviita, K., Tervaniemi, M., Alho, K., Kallio, J., & Näätänen, R. (soumis). Basic auditory dysfunction in dyslexia as demonstrated by brain-activity measurements.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Landau, B., & Gleitman, L. (1985). *Language and Experience - Evidence from the Blind Child.* Cambridge, MA: Harvard University Press.

Lea, W. A. (1974). *Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding* (Univac Report No. PX10791). Sperry Univac.

Lehiste, I., & Wang, W. S.-Y. (1977). Perception of sentence and paragraph boundaries with and without semantic information. In Dressler & Pfeiffer (Eds.), *Phonologica 1976* (p. 277-283).

Levelt, C., & van de Vijver, R. (1998). Syllable types in cross-linguistic and developmental grammars. In R. Kager & W. Zonneveld (Eds.), *Third Biannual Utrecht Phonology Workshop* (Vol. ROA 265). Utrecht.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, *29*, 117-123.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*, 148-167.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Lloyd James, A. (1940). *Speech signals in telephony.* London.

Locke, J. (1690). *An essay concerning human understanding.*

Luangthongkum, T. (1977). *Rhythm in standard Thai.* Thèse de doctorat non publiée, University of Edinburgh.

Luce, P. A., & Charles-Luce, J. (1983). The role of fundamental frequency and duration in the perception of clause boundaries: Evidence from a

speeded verification task. In *the 105th meeting of the Acoustical Society of America.* Cincinatti.

Macken, M. A. (1995). Phonological acquisition. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (p. 671-696). Cambridge: Blackwell.

Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and hearing: Work in progress, University College London*, 74-93.

Maidment, J. A. (1983). Language recognition and prosody: further evidence. *Speech, hearing and language: Work in progress, University College London*, *1*, 133-141.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*, 315-318.

Marcus, G. F. (1998). Can connectionism save constructivism? *Cognition*, *66*, 153-182.

Mascaró, J. (1976). *Catalan Phonology and the phonological cycle.* Thèse de doctorat non publiée, MIT.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465-494.

Mazuka, R. (1996). Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 313-330). Mahwah, NJ: Lawrence Erlbaum Associates.

McGivern, R., Berka, C., Languis, M., & Chapman, S. (1991). Detection of deficits in temporal pattern discrimination using the seashore rhythm test in young children with reading impairments. *J. Learn. Disabil.*, *24*(1), 58-62.

Mehler, J., & Christophe, A. (1995). Maturation and learning of language during the first year of life. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (p. 943-954). Bradford Books / MIT Press.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 101-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-

Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Mehler, J., Sebastián-Gallés, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: the role of rhythm and meaning. *Annals of the New York Academy of Science*, *682*, 272-282.

Miles, T., & Haslum, M. (1986). Dyslexia: anomaly or random variation. *Dyslexia*, *36*, 103-117.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*, 495-500.

Morgan, J. (1986). *From simple input to complex grammar.* Cambridge Mass: MIT Press.

Morgan, J. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development*, *17*, 389-403.

Morgan, J., & Demuth, K. (1996a). *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah NJ: Lawrence Erlbaum Associates.

Morgan, J., & Demuth, K. (1996b). Signal to syntax: An overview. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 1-22). Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J., & Saffran, J. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911-36.

Morgan, J., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 263-283). Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*, 666-688.

Morgan, W. (1896). A case of congenital word blindness. *B. M. J.*, *2*, 1378.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405-408.

Nazzi, T. (1997). *Du rythme dans l'acquisition et le traitement de la parole.* Thèse de doctorat non publiée, Ecole des Hautes Etudes en Sciences Sociales.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of*

Experimental Psychology: Human Perception and Performance, *24*(3), 756-766.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates: Brief report. *Infant Behavior and Development*, *21*(4), 779-784.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (soumis). Discriminating languages from the same rhythmic class: Evidence from English-learning 5-month-olds.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (p. 157-175). Dordrecht: Foris.

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: The rhythmic activation principle. In U. Kleinhenz (Ed.), *Interfaces in Phonology* (p. 1-26). Berlin: Akademie Verlag.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology.* Dordrecht: Foris.

Neville, H. J., & Mills, D. L. (1997). Epigenesis of language. *Mental retardation and developmental disabilities research reviews*, *3*, 282-292.

O'Connor, J. D. (1965). *The perception of time intervals* (Progress Report No. 2). Phonetics Laboratory, University College London.

O'Connor, J. D. (1973). *Phonetics.* Harmondsworth, Middlesex: Penguin Books.

Ohala, J. J., & Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody. In P. Léon & M. Rossi (Eds.), *Problèmes de prosodie* (Vol. II, p. 123-131). Ottawa: Didier.

Olson, R., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: genetic and environmental influences. *J Learn Disabil*, *22*, 339-348.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258-278.

Pagel, V., Carbonell, N., & Laprie, Y. (1996). A new method for speech delexicalization, and its application to the perception of French prosody. In *ICSLP'96.* Philadelphia.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments, & Computers*, *29*(3), 322-327.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: a cross-linguistic study. *Memory & Cognition*, *26*, 844-851.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S., Dupoux, E., Fazzio, F., & Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport*, *7*, 2439-2444.

Peters, A. M., & Strömqvist, S. (1996). The role of prosody in the acquisition of grammatical morphemes. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 215-232). Mahwah, NJ: Lawrence Erlbaum Associates.

Pike, K. L. (1945). *The intonation of American English.* Ann Arbor, Michigan: University of Michigan Press.

Pinker, S. (1984). *Language learnability and language development.* Cambridge: Harvard University Press.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar* (Tech. Rep. No. TR-2). Rutgers University.

Quine, W. V. (1960). *Word and Object.* Cambridge, MA: MIT Press.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax.* Oxford: Blackwell.

Ratner, N. B. (1996). From "Signal to Syntax": But what is the nature of the signal? In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 135-150). Mahwah, NJ: Lawrence Erlbaum Associates.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies.* London: Edward Arnold.

Rubach, J., & Booij, G. E. (1985). A grid theory of stress in Polish. *Lingua*, *66*, 281-319.

Ruhlen, M. (1996). *The Origin of Language: Tracing the Evolution of the Mother Tongue.* New York: John Wiley.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101-105.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*, 3-11.

Schaffer, D. (1984). The role of intonation as a cue to topic management in conversation. *Journal of Phonetics*, *12*, 327-344.

Seidenberg, M. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science*, *275*(5306), 1599-603.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E. O. (1996). The prosodic structure of function words. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (p. 187-213). Mahwah, NJ: Lawrence Erlbaum Associates.

Shady, M., Jusczyk, P. W., & Gerken, L. A. (1998). Infants' sensitivity to function morphemes. In *23rd Annual Boston University Conference on Language Development.* Boston, MA.

Shafer, V., Gerken, L. A., Shucard, J., & Shucard, D. (1992). "The" and the brain: An electrophysiological study of infant's sensitivity to English function morphemes. In *the Boston University Conference on Language Development.* Boston, MA.

Shafer, V., Shucard, D., Shucard, J., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal Of Speech, Language, And Hearing Research*, *41*(4), 874-86. (Hofstra University, Hempstead, NY 11550, USA.)

Shen, Y., & Peterson, G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics - Occasional papers*, *9*, 1-36.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, *25*, 169-201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*, B11-B21.

Skinner, B. F. (1957). *Verbal Behavior.* New York: Appleton-Century-Crofts.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34-50.

Steedman, M. (1996). Phrasal intonation and the acquisition of syntax. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 331-342). Mahwah, NJ: Lawrence Erlbaum Associates.

Steever, S. B. (1987). Tamil and the Dravidian languages. In D. Comrie (Ed.), *The world's major languages* (p. 725-746). Oxford: Oxford University Press.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*, 1582-92.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain And Language*, *9*(2), 182-98.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word recognition in 6-month-olds. *Psychological Science*, *10*, 172-175.

Umeda, N., & Coker, C. H. (1974). Allophonic variation in American English. *Journal of Phonetics*, *2*, 1-5.

van de Weijer, J. (1998). *Language input for word discovery.* Nijmegen: MPI Series in Psycholinguistics.

van Ooyen, B., Bertoncini, J., Sansavini, A., & Mehler, J. (1997). Do weak syllables count for newborns? *Journal of the Acoustical Society of America*, *102*(6), 3735-3741.

Venditti, J. J., Jun, S., & Beckman, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 287-311). Mahwah, NJ: Lawrence Erlbaum Associates.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192-212.

Wenk, B., & Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, *10*, 193-216.

Wexler, K., & Culicover, P. W. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Wheeler, M. (1979). *Phonology of Catalan.* Oxford: Blackwell.

Willems, N. (1982). *English intonation from a Dutch point of view.* Dordrecht - Holland: Foris.

Table des matières

1 Introduction : l'initialisation phonologique 1
1.1 Quelques problèmes classiques de l'acquisition du langage . . . 2
1.1.1 Apprendre le lexique 2
1.1.2 Analyser la syntaxe des phrases 4
1.1.3 Lexique et syntaxe : la poule et l'œuf? 5
1.2 L'initialisation phonologique 6
1.2.1 Présentation de l'hypothèse 6
1.2.2 Le rôle de l'environnement dans l'acquisition du langage 8
1.3 Quelques exemples d'initialisation phonologique 10
1.3.1 Restreindre le domaine des analyses de la parole 11
1.3.1.1 La hiérarchie prosodique 11
1.3.1.2 Réalité acoustique et psychologique des frontières prosodiques 12
1.3.2 L'extraction des mots du flux de parole 14
1.3.2.1 Régularités distributionnelles 14
1.3.2.2 Régularités lexicales 17
1.3.2.3 Exploitation de l'ensemble des régularités . . 23
1.3.2.4 Conclusion 25
1.3.3 L'apprentissage des catégories grammaticales 26
1.3.3.1 Mots de fonction et mots de contenu 26
1.3.3.2 Noms et verbes 31
1.3.4 L'organisation syntaxique de la phrase 32
1.3.4.1 Les constituants syntaxiques 32
1.3.4.2 L'ordre des mots 34
1.4 À la recherche de nouveaux indices phonologiques 36

2 Le rythme de la parole 39
2.1 Revue des théories 39
2.1.1 La théorie de l'isochronie 39
2.1.1.1 Historique 39
2.1.1.2 L'isochronie en échec 40

2.1.2 Alternatives 42
2.1.2.1 Une nouvelle approche du rythme de parole . 42
2.1.2.2 Les langues intermédiaires 43
2.1.2.3 Perspectives 44
2.2 Expérience 1 : Mesures acoustico-phonétiques en huit langues . 44
2.2.1 Justification de la démarche 44
2.2.2 Méthode 45
2.2.3 Résultats 47
2.2.4 Discussion 49
2.3 Un modèle de la perception du rythme, simulations et prédictions 51
2.3.1 Simulation 1 : Discrimination de langues par les adultes, basée sur $\%V$ 52
2.3.1.1 Modèle de la tâche 52
2.3.1.2 Résultats 53
2.3.2 Simulation 2 : Discrimination de langues par les nouveau-nés, basée sur $\%V$ 54
2.3.2.1 Expériences passées 54
2.3.2.2 Modèle de la tâche 55
2.3.2.3 Résultats 58
2.3.2.4 Discrimination entre groupes de langues . . . 61
2.4 Conclusions et perspectives 62

3 La resynthèse de parole 65
3.1 Revue des techniques décrites dans la littérature 65
3.2 La resynthèse 67
3.2.1 Phrases d'origine 68
3.2.2 Traitement général 68
3.2.3 Transformations 69
3.3 Expériences 2a–2d : Discrimination anglais-japonais à l'aide des phrases resynthétisées 70
3.3.1 Méthode 70
3.3.2 Participants 71
3.3.3 Résultats 72
3.4 Discussion 74
3.4.1 Rôles respectifs du rythme et de l'intonation – Expérience 2e 74
3.4.2 Considérations méthodologiques 76
3.4.3 Comparaison de l'expérience et de la simulation 77
3.4.4 Conclusion 79

4 Discrimination de langues par des sujets adultes 81
4.1 Exp. 3a–3b : Optimisation de la tâche – anglais-espagnol en *sasasa* . . . 82
4.1.1 Méthode . . . 82
4.1.2 Participants . . . 84
4.1.3 Résultats . . . 84
4.1.4 Conclusion . . . 85
4.2 Exp. 4a–4b : anglais-polonais et espagnol-polonais en *sasasa* . 85
4.2.1 Méthode . . . 85
4.2.2 Participants . . . 86
4.2.3 Résultats . . . 86
4.3 Exp. 5a–5c : anglais-espagnol, anglais-polonais et espagnol-polonais en *sasasa plat* . . . 88
4.3.1 Méthode . . . 88
4.3.2 Participants . . . 88
4.3.3 Résultats . . . 88
4.3.4 Discussion . . . 90
4.4 Exp. 6a–6c : catalan-espagnol, anglais-catalan et catalan-polonais en *sasasa plat* . . . 91
4.4.1 Méthode . . . 91
4.4.2 Participants . . . 92
4.4.3 Résultats . . . 92
4.4.4 Discussion . . . 94
4.5 Analyses post-hoc . . . 95
4.5.1 Biais de réponse . . . 95
4.5.2 Instructions . . . 95
4.5.3 Recrutement . . . 96
4.5.4 Distribution des scores . . . 96
4.6 Résumé . . . 98
4.7 Ajustement du modèle . . . 99
4.7.1 Simulation 3 : Discrimination de langues par les adultes, basée sur $\%V$ et ΔV . . . 100
4.7.2 Simulation 4 : Discrimination de langues par les adultes, basée sur $\%V$, ΔV et ΔC . . . 102
4.8 Conclusion . . . 103

5 Discrimination de langues par des nouveau-nés 105
5.1 Revue des expériences de discrimination de langues par des nourrissons . . . 106
5.1.1 À quelques jours de vie . . . 106
5.1.2 À 2-3 mois . . . 109

5.1.3 À 4-5 mois . . . 111
5.1.4 Synthèse . . . 112
5.2 Méthode générale . . . 113
5.2.1 Stimuli . . . 113
5.2.2 Participants . . . 114
5.2.3 Procédure . . . 115
5.2.3.1 Groupes de sujets . . . 115
5.2.3.2 Déroulement du test . . . 115
5.2.3.3 Conditions de délai et de rejet . . . 117
5.3 Exp. 7 : néerlandais-japonais en parole naturelle . . . 118
5.3.1 Stimuli . . . 118
5.3.2 Participants . . . 118
5.3.3 Résultats . . . 119
5.3.4 Discussion . . . 120
5.4 Exp. 8 : néerlandais-japonais en *saltanaj* . . . 121
5.4.1 Stimuli . . . 121
5.4.2 Participants . . . 122
5.4.3 Procédure . . . 122
5.4.4 Résultats . . . 123
5.4.4.1 Premier changement . . . 123
5.4.4.2 Deuxième changement . . . 123
5.4.5 Discussion . . . 126
5.5 Exp. 9 : néerlandais-japonais en *saltanaj* à l'envers . . . 127
5.5.1 Stimuli . . . 127
5.5.2 Participants . . . 127
5.5.3 Procédure . . . 128
5.5.4 Résultats . . . 128
5.5.5 Discussion . . . 129
5.6 Exp. 10 : néerlandais-japonais en *sasasa* avec intonation artificielle . . . 130
5.6.1 Stimuli . . . 131
5.6.2 Participants . . . 134
5.6.3 Résultats . . . 135
5.6.4 Discussion . . . 136
5.7 Exp. 11 : néerlandais-japonais en *saltanaj* avec intonation artificielle . . . 138
5.7.1 Stimuli . . . 138
5.7.2 Participants . . . 138
5.7.3 Procédure . . . 139
5.7.4 Résultats . . . 139
5.7.5 Discussion . . . 140

5.8 Préférence pour le néerlandais 141
5.9 Conclusion et perspectives 143

6 Discrimination de langues par des singes 147
6.1 Revue des expériences de perception de la parole chez l'animal 148
6.2 Démarche générale 150
6.3 Exp. 12 : Discrimination néerlandais-japonais par des tamarins – parole naturelle et parole à l'envers 151
6.3.1 Stimuli 151
6.3.2 Participants 151
6.3.3 Procédure 151
6.3.3.1 Groupes de sujets 152
6.3.3.2 Déroulement du test 153
6.3.4 Résultats 154
6.3.5 Discussion 155
6.4 Exp. 13 : Discrimination néerlandais-japonais par des tamarins – *saltanaj* à l'endroit et à l'envers 156
6.4.1 Stimuli 156
6.4.2 Participants 156
6.4.3 Procédure 157
6.4.4 Résultats 157
6.5 Comparaison des expériences 12 et 13 158
6.6 Discussion générale 160

7 Conclusion : le rôle du rythme dans l'acquisition du langage 165
7.1 Le rythme: une entité linguistique détectable dans le signal de parole 166
7.2 Perception du rythme de la parole par le nouveau-né 168
7.3 Le rôle du rythme dans l'acquisition du langage 169
7.3.1 Apprendre la structure des syllabes de la langue 169
7.3.2 L'initialisation par le rythme 171
7.4 Discussion et perspectives 173
7.4.1 Est-ce vraiment le rythme ? 173
7.4.2 Le cas du polonais 174
7.4.3 Et les singes ? 175
7.4.4 Les enfants bilingues 176
7.4.5 L'hypothèse de l'initialisation par le rythme est-elle testable ? 178

Annexes 181

A Méthodes pour l'étude des nourrissons **181**
A.1 Succions de haute amplitude 181
A.2 Orientation préférentielle de la tête 182
A.3 Latence d'orientation 184
A.4 Conditionnement 185

B Instructions et fichiers d'exemple **187**
B.1 Fichier utilisé pour la resynthèse 187
B.2 Instructions pour les expériences 2a–2d 190
B.3 Instructions pour l'expérience 2e 190
B.4 Instructions pour l'expérience 3a 191
B.5 Instructions pour les expériences 3b, 4a–4b, et 5a–5c 191
B.6 Séquence de phrases entendues par un sujet dans l'expérience 3b 192
B.7 Instructions pour les expériences 6a–6c 193
B.7.1 Première moitié des sujets 193
B.7.2 Deuxième moitié des sujets 194
B.8 Instructions pour l'expérience 10 bis 195

C Expériences supplémentaires de discrimination de langues par des nouveau-nés **197**
C.1 Exp. 7 bis : néerlandais-japonais en parole naturelle à l'envers 197
C.1.1 Stimuli 197
C.1.2 Participants 198
C.1.3 Résultats 198
C.2 Exp. 7 ter : néerlandais-japonais en parole naturelle, plusieurs changements par sujet 200
C.2.1 Stimuli 200
C.2.2 Participants 200
C.2.3 Procédure 200
C.2.4 Résultats 202
C.2.5 Discussion 203

Bibliographie **205**

Table des matières **221**

Table des figures **227**

Liste des tableaux **229**

Table des figures

2.1 Projection des huit langues sur le plan (%V,ΔC). 48
2.2 Projection des huit langues sur le plan (%V,ΔV). 48
2.3 Projection des huit langues sur le plan (ΔV,ΔC). 49
2.4 Simulation de la discrimination anglais/japonais par des nouveau-nés. 60
2.5 Simulation de la discrimination anglais/espagnol par des nouveau-nés. 60
2.6 Simulation de la discrimination espagnol/italien par des nouveau-nés. 61

3.1 Exp 2a–2d : Scores A' moyens par condition. 74
3.2 Discrimination anglais/japonais par des sujets adultes : pourcentage moyen de réponses correctes par phrase, en fonction de sa valeur de %V. 79

4.1 Histogramme des scores A' pour les expériences 3b, 4a et 4b. . 87
4.2 Histogramme des scores A' pour les expériences 5a–5c. 89
4.3 Histogramme des scores A' pour les expériences 6a–6c. 93
4.4 Histogramme des scores A' pour l'ensemble des expériences 3b, 4a–4b, 5a–5c et 6a–6c. 97
4.5 Répartition des scores A' en deux distributions normales. . . . 97

5.1 Exp. 7 : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle. 119
5.2 Exp. 8 : Discrimination néerlandais-japonais par des nouveau-nés – *saltanaj*. 124
5.3 Exp 8 : Valeurs de l'indice I pour les 11 bébés ayant effectué deux changements. 124
5.4 Exp. 9 : Discrimination néerlandais-japonais par des nouveau-nés – *saltanaj* à l'envers. 128
5.5 Contours d'intonation utilisés dans l'expérience 10. 132

5.6 Exp. 10bis : Jugements des sujets adultes sur l'intonation des phrases *sasasa* avec intonation artificielle et *saltanaj*. 133
5.7 Exp. 10 : Discrimination néerlandais-japonais par des nouveau-nés – *sasasa* avec intonation artificielle. 135
5.8 Données des bébés de l'expérience 10 qui ont entendu le néerlandais en habituation (10 bébés par groupe). 137
5.9 Données des bébés de l'expérience 10 qui ont entendu le japonais en habituation (10 bébés par groupe). 137
5.10 Exp. 11 : Discrimination néerlandais-japonais par des nouveau-nés – *saltanaj* avec intonation artificielle. 140
5.11 Nombre de succions HA par minute pendant l'habituation, en fonction de la langue écoutée. 142

6.1 Photographie d'un tamarin « cotton-top » (à panache blanc). 152
6.2 Exp. 12 : Discrimination néerlandais-japonais par les tamarins, en parole naturelle à l'endroit et à l'envers. 155
6.3 Exp. 13 : Discrimination néerlandais-japonais par les tamarins, en parole resynthétisée *saltanaj* à l'endroit et à l'envers. 157
6.4 Discrimination néerlandais-japonais par les tamarins – expériences 12–13 confondues. 159

7.1 Illustration des relations entre le signal, la structure syllabique, et le rythme, et du parcours postulé pour l'acquisition des syllabes. 174

C.1 Exp. 7bis : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle à l'envers. 199
C.2 Exp. 7ter : Discrimination néerlandais-japonais par des nouveau-nés – parole naturelle à l'endroit – plusieurs changements par sujets. 203
C.3 Scores de déshabituation de l'expérience 7 ter. 204

Liste des tableaux

2.1 Répartition des langues dans les classes de rythme. 41
2.2 Nombre de mesures et valeurs moyennes des trois variables par langue. 47
2.3 Simulations des expériences de discrimination de langues chez l'adulte basées sur $\%V$ pour les 28 paires de langues. 53
2.4 Revue des expériences de discrimination de langues chez le nouveau-né. 54
2.5 Simulations des expériences de discrimination de langues chez le nouveau-né sur 28 paires de langues. 58

3.1 Exp. 2a–2d : Pourcentages moyens de réponses correctes dans les différentes sessions des 4 conditions. 72
3.2 Exp. 2a–2d : Statistiques par condition. 73

4.1 Discrimination anglais-espagnol en *sasasa*– Comparaison des tâches X et AAX. 84
4.2 Discrimination anglais-espagnol, anglais-polonais et espagnol-polonais en *sasasa*. 86
4.3 Discrimination anglais-espagnol, anglais-polonais et espagnol-polonais en *sasasa plat*. 89
4.4 Discrimination catalan-espagnol, anglais-catalan et catalan-polonais en *sasasa plat*. 93
4.5 Expériences de discrimination de langues chez les sujets adultes. 98
4.6 Expériences de discrimination de langues chez les sujets adultes : sasasa plat seulement. 99
4.7 Simulations des expériences de discrimination de langues chez l'adulte basées sur $\%V$ et ΔV pour les 28 paires de langues. . 101
4.8 Simulations des expériences de discrimination de langues chez l'adulte basées sur $\%V$, ΔV et ΔC pour les 28 paires de langues.102

5.1 Conditions expérimentales pour les expériences 7–10. 115
5.2 Caractéristiques des nouveau-nés de l'Expérience 7. 119

5.3 Caractéristiques des nouveau-nés de l'Expérience 8. 122
5.4 Exp. 8 : Répartition des scores de déshabituation $post2-ante2$ en fonction de l'ordre et de la nature du changement. 125
5.5 Caractéristiques des nouveau-nés de l'Expérience 9. 127
5.6 Caractéristiques des nouveau-nés de l'Expérience 10. 134
5.7 Caractéristiques des nouveau-nés de l'Expérience 11. 138
5.8 Nombre moyen de succions HA par minute pendant les cinq dernières minutes d'habituation, selon la langue écoutée et l'expérience. 142

6.1 Exp. 12 : Répartition des tamarins dans les huit conditions. . . 152

C.1 Caractéristiques des nouveau-nés de l'Expérience 7 bis. 198
C.2 Caractéristiques des nouveau-nés de l'Expérience 7 ter. 200
C.3 Conditions de l'expérience 7ter. 201

Abstract

Within the phonological bootstrapping framework, we have investigated the hypothesis that from birth, infants are sensitive to speech rhythm, and that this sensitivity is one of the very first steps of language acquisition. We made acoustic-phonetic measurements in eight languages, showing that a consonant/vowel segmentation of speech can allow determining certain rhythmic properties of languages. This led us to a computational model of rhythm perception, and to simulations predicting which languages can be distinguished on the basis of their rhythmic properties. In order to test these predictions, we ran a series of language discrimination experiments on adult subjects, which showed consistency between the model and the empirical data. To ensure that subjects could not use any other cue than rhythm, we designed a technique based on speech resynthesis allowing to degrade stimuli and preserve their rhythmic properties exclusively. This technique allowed us to show that newborns can discriminate between Dutch and Japanese utterances, on the sole basis of their rhythmic differences. Finally, we questioned the speech-specificity of this sensitivity to rhythm, by running pilot experiments showing language discrimination by monkeys. We conclude that newborn infants have a capacity to process speech rhythm, and we hypothesize that this capacity may help them learning the syllable structure of their native language.

Keywords: Language acquisition, speech perception, phonological bootstrapping, prosody, rhythm, language discrimination.

Discipline : Cognitive Science

Laboratoire de Sciences Cognitives et Psycholinguistique (UMR 8554)
54 boulevard Raspail
75006 Paris
`http://www.ehess.fr/centres/lscp/`

Résumé

Dans le cadre de la théorie de l'initialisation phonologique, nous avons étudié l'hypothèse selon laquelle le nourrisson est sensible dès la naissance au rythme de la parole, cette sensibilité constituant l'une des toutes premières étapes de l'acquisition du langage. Pour ce faire, nous avons effectué des mesures acoustico-phonétiques sur huit langues, montrant qu'une segmentation de la parole en consonnes/voyelles permet de déterminer certaines propriétés rythmiques des langues. Nous en avons déduit un modèle computationnel de la perception du rythme, et nous avons effectué des simulations prédisant quelles langues peuvent être distinguées sur la base de leurs propriétés rythmiques. Afin de tester ces prédictions, nous avons effectué une série d'expériences de discrimination de langues sur des sujets adultes, qui ont montré l'adéquation du modèle avec les données empiriques. Pour s'assurer que les sujets n'utilisaient pas d'autres indices que le rythme, nous avons mis au point une technique basée sur la resynthèse de parole, permettant de dégrader les stimuli pour n'en préserver que les propriétés rythmiques. Grâce à cette technique, nous avons pu montrer que des nouveau-nés sont capables de distinguer le néerlandais du japonais, sur la base de leurs seules différences rythmiques. Enfin, nous nous sommes demandés si cette sensibilité pour le rythme était spécifique au langage, et donc à l'espèce humaine. Dans cette perspective, nous présentons quelques expériences-pilotes de discrimination de langues par des singes. Nous concluons que le nourrisson possède dès la naissance une capacité à analyser le rythme de la parole, et nous faisons de plus l'hypothèse que cette capacité peut lui être utile dans l'apprentissage de la structure des syllabes de sa langue maternelle.

Mots-clé : Acquisition du langage, perception de la parole, initialisation phonologique, prosodie, rythme, discrimination de langues.

Discipline : Sciences cognitives

Laboratoire de Sciences Cognitives et Psycholinguistique (UMR 8554)
54 boulevard Raspail
75006 Paris
`http://www.ehess.fr/centres/lscp/`