

**UNIVERSITÉ GRENOBLE 3 – STENDHAL**  
**U.F.R. DES SCIENCES DU LANGAGE**

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--	--

**THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ STENDHAL**

**Discipline : « Sciences du Langage »**

préparée à l'Institut de la Communication Parlée CNRS UMR 5009  
dans le cadre de l'École Doctorale « **Langues, littératures et sciences humaines** »

présentée et soutenue publiquement

par

**Virginie DUCEY-KAUFMANN**

le 26 janvier 2007

**LE CADRE DE LA PAROLE ET LE CADRE DU SIGNE :  
UN RENDEZ-VOUS DÉVELOPPEMENTAL**

---

**Sous la Direction de**

**Christian ABRY**

---

**JURY**

M. Jean-François BONNOT Professeur  
M. Jacques VAUCLAIR Professeur  
M. Jean-Marc COLLETTA MCF HDR  
M. Christian ABRY Professeur

Rapporteur  
Rapporteur  
Examineur  
Directeur



**UNIVERSITÉ GRENOBLE 3 – STENDHAL  
U.F.R. DES SCIENCES DU LANGAGE**

**THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ STENDHAL**

**Discipline : « Sciences du Langage »**

préparée à l'Institut de la Communication Parlée CNRS UMR 5009  
dans le cadre de l'École Doctorale « **Langues, littératures et sciences humaines** »

présentée et soutenue publiquement

par

**Virginie DUCEY-KAUFMANN**

le 26 janvier 2007

**LE CADRE DE LA PAROLE ET LE CADRE DU SIGNE :  
UN RENDEZ-VOUS DÉVELOPPEMENTAL**

---

**Sous la Direction de**

**Christian ABRY**

---

**JURY**

M. Jean-François BONNOT Professeur  
M. Jacques VAUCLAIR Professeur  
M. Jean-Marc COLLETTA MCF HDR  
M. Christian ABRY Professeur

Rapporteur  
Rapporteur  
Examinateur  
Directeur

## Remerciements

*Je tiens tout d'abord à remercier mon directeur de thèse, Christian Abry, pour m'avoir guidée tout au long de ces 5 années, une aventure qui a commencé dès la maîtrise et qui s'est poursuivie jusqu'à la fin de cette thèse. Je le remercie pour la qualité de son encadrement, pour sa présence continue. Il a toujours su me guider dans mon travail de recherche et m'a apporté une aide inestimable dans l'élaboration de mes mémoires et de ma thèse. Sa vivacité scientifique hors du commun est un modèle pour tous.*

*Je remercie Jacques Vauclair pour avoir accepté d'être l'un des rapporteurs de ma thèse. Ses recherches en psychologie du développement et en primatologie ont été une stimulation dans le cadre des projets français et européens sur l'Origine de l'Homme, du Langage et des Langues, projets dirigés par Jean-Luc Schwartz, directeur de l'ICP.*

*Je remercie Jean-François Bonnot pour avoir également accepté d'être l'un de mes rapporteurs de thèse. Son apport dans le domaine de la phonétique et tout particulièrement de la motricité est indiscutable, et ses recherches sur le langage de l'enfant très enrichissantes.*

*Je remercie Jean-Marc Colletta, qui me fait l'honneur d'être examinateur dans mon jury. Merci pour nos conversations sur son domaine de prédilection, le geste, notamment lors des différents colloques où nous nous sommes retrouvés. Merci également Jean-Marc pour ta gentillesse et ton intérêt pour mon travail.*

*Merci à tous ceux qui m'ont transmis leurs travaux, des références, des suggestions : Nick Campbell, Denis Creissels, Ann Cutler, Jan Peter De Ruiter, Holger Diessel, Susan Goldin-Meadow, Julie Grèzes, Jana Iverson, Georges Kleiber, Iva Novakova-Nokolova, Mireille Piot, Elena Pizzuto, Robert Remez, Wendy Sandler, Gillian Sankoff, Marc Sato, Marie Savelli, Carmen Scarlat, Brian Scassellati, Stefan Schaal, Jechil Sieratzki, Jean-Emmanuel Tyvaert, Edy Veneziano, Virginia Volterra, Bencie Woll.*

*Je remercie Coriandre Vilain pour son aide tout au long de notre expérience sur la synergie de la voix et du doigt, notamment pour sa patience afin de m'expliquer de nombreuses notions jusqu'alors méconnues pour ma part. Ce fut un réel plaisir de travailler avec toi. Je remercie par ailleurs tous les sujets qui ont accepté de se prêter à cette expérience, souvent harassante : Aude, Claire, les deux Anne, les deux Nico, Fanny, Tiphaine, Matthias, Luca, Muriel, Virginie, Frederic, Christophe, Albert et Emilie.*

*Je tiens à remercier Marie-Agnès Cathiard pour son aide précieuse et indirecte dans mon travail. Ses remarques et son œil de psychologue avertie m'ont été vraiment utiles. Je rajoute à cela un merci tout particulier pour sa qualité de tutrice d'enseignement. Elle m'a accueilli au sein de l'équipe pédagogique de phonétique-phonologie avec beaucoup de gentillesse, de patience, en prenant le temps de m'expliquer ce que je ne comprenais pas. Ce fut un véritable plaisir de travailler avec toi. Avec Virginie Attina que tu as dirigée pour sa thèse sur la production et la perception du Langage Parlé Complété, j'ai pu bénéficier de vos connaissances acquises sur le contrôle de la main coordonnée avec la face.*

*Pour leur aide dans l'élaboration des différents enseignements donnés au cours de mon monitorat et de mon ATER, pour leur sympathie, je remercie également Anne, Solange, Claire, Cécile, Isabelle, les deux Fanny, Marinette, Anna, Elisabetta, Cathy.*

*Un merci tout particulier à Stefanie Brosda, sans qui ce travail ne serait pas. Ses nombreuses heures passées à filmer ces 6 petits bouts constituent un corpus précieux pour nous de l'ICP qui avons voulu étudier le développement vocal et gestuel des enfants.*

*Pour leur aide technique et informatique, leur gentillesse et leur disponibilité, je remercie chaleureusement Alain Arnal et Christophe Savariaux.*

*Je remercie vivement Dominique, Marie-Thé, et plus récemment Aline et Marie-Ange pour toute l'aide qu'elles ont pu m'apporter, tant pour les problèmes administratifs que pour tous les conseils liés au fonctionnement du laboratoire.*

*Merci à tous les membres de l'ICP, campus ou gare, qui ont été là, qui sont là ou qui viennent d'arriver, pour leur gentillesse, et puis pour leurs nombreux témoignages de gentillesse à l'arrivée de Corentin, mon fils, cela m'a beaucoup touché.*

*J'adresse un remerciement tout spécial, empreint d'une affection très profonde, à mes partenaires de « grooming » matinal, Aude, Claire, Fanny et Emilie. Merci pour nos discussions (scientifiques ou non), pour votre compréhension et votre soutien sans faille tout au long de ces années icpéennes. Merci pour nos délires, nos pétages de plomb mutuels, nos pauses bonbons, nos regards complices. Que cette amitié, largement cultivée en dehors du labo, reste toujours aussi belle lorsque chacune de nous aura pris un autre envol.*

*Un autre remerciement particulier va à Isabelle. Isa, tu es vraiment un cœur sur pattes, merci pour ton soutien, pour nos longues discussions qui m'ont toujours reboostée. Et puis spécialement pour tous les petits soucis de fin de thèse que j'ai rencontré, pour ton aide dans ma reconversion, pour ton coup de fil du dernier moment qui j'espère va me porter chance !*

*Je remercie également tous les doctorants de l'ICP campus pour leur sympathie, contribuant à travailler dans une atmosphère appréciable.*

*Un merci aux doctorants de l'ICP gare, rencontrés au gré des séminaires, réunions, pots de Noël. Merci plus particulièrement à Marion, pour son soutien, ses conseils, ses encouragements au cours de ma thèse.*

*Je n'oublie pas ici de remercier les filles d'« en-haut », particulièrement Céline, Carmen et Fanny, pour leur soutien et leur présence à différents moments de ma thèse.*

*Je remercie profondément Fab & Lo, que serais-je sans ton amitié Fab, tu sais trouver les mots qui me réconfortent et qui me boostent, tu as toujours été présente dans les bons et les mauvais moments, je suis fière et tellement chanceuse de t'avoir comme meilleure amie, merci pour tout, merci d'être toi.*

*Pour leurs nombreux encouragements, pour leur intérêt dans mon travail, pour leur soutien, un grand merci à Gé, Tichef, Fred, Sève, Jenny, Pierre & Carine, Anne-Marie & Seb, les mamours, Céline & Nico.*

*Sandrine, Yann, Romane et Téo, merci de m'avoir toujours fait confiance, de m'avoir toujours soutenue et encouragée.*

*Merci à Nadine, Christian, Doudou, Fanny & Mickaël pour ce bol d'air bisontin qui m'a fait énormément de bien après de nombreux coups durs. « La vivi » vous*

*remercie et n'oublie pas les « trois jours de rab » dont nous profiterons dès que cette histoire sera achevée.*

*Je voudrais remercier spécialement Angeline, Gérard et Nelly. Ce petit mot ne sera jamais assez long pour vous dire combien votre amour a toujours été un trésor inestimable, un havre de sécurité et un repli pour faire face à toutes les tempêtes, toutes les incertitudes. Merci à vous, papa et maman pour avoir toujours cru en moi, pour m'avoir encouragé, pour tous les sacrifices que vous avez fait pour que ma vie soit belle.*

*And the last but not the least... mon jp les mots me manquent pour t'exprimer tous les sentiments qui émergent à l'écriture de ton nom. Personne ne me connaît aussi bien que toi et toi seul sait prononcer les mots qu'il faut pour m'apaiser, me rassurer, me calmer, me motiver ou même me faire rire. Tu as toujours cru en moi dès le début, tes remarques objectives me font évoluer, même si tu n'en as pas toujours l'impression. Ton soutien est inébranlable et je n'aurais jamais pu y arriver sans toi. Ton altruisme est tel que tu tais tes soucis pour soulager les miens, tu es une perle et je mesure la chance que j'ai de t'avoir auprès de moi. Merci enfin et surtout pour ce magnifique petit Corentin que tu m'as offert, quel plus beau cadeau pouvais-je rêver de toi... Je vous aime tous deux passionnément.*

*Et une dernière petite pensée, nostalgique mais douce, à mon arrière-grand-mère, qui ne cessait jamais de me dire, avec son plus grand humour, à chacune de mes visites : « Des mercis, j'en ai plein les poches... ».*

---

## **RESUME en français**

Notre hypothèse de travail est qu'il existerait un rendez-vous développemental entre ce que nous nommons le cadre de la parole et le cadre du signe. Tandis que le cadre de la parole (Speech Frame) s'établit sous la forme du babillage canonique, vers 7 mois, le cadre du signe (Sign Frame) se manifeste tout d'abord sous la forme du pointage dit impératif vers 9 mois, avant de donner lieu au pointage dit déclaratif. Ce dernier apparaît avec les premiers mots, tandis que le cadre de la parole permet à ce moment-là de coproduire (coarticuler) voyelle et consonne (Sussman et al. 1999). Les places respectives des ingrédients de ce rendez-vous développemental autour du premier mot restent encore à explorer.

Dans la présente contribution, nous avons voulu tester l'existence d'un rapport harmonique entre cadre de la parole et cadre du signe. Pour cela, il nous a fallu tout d'abord obtenir la distribution des fréquences de babillage, puis celle des durées des pointers. Nos résultats sur 6 sujets, suivis sur 12 mois, montrent qu'avec un mode de babillage à 3Hz et des strokes de pointers de 600-700 ms (1.5Hz), nous pouvons rendre compte du gabarit (*template*) des premiers mots. En effet, ces mots «prosodiques» pouvant varier d'une à deux «syllabes», il est nécessaire de faire appel à la notion de pied (*foot*) comme une unité de contrôle métrique ancrée dans le pointer. Ceci rendra compte des observations courantes dans la littérature à condition qu'au lieu de compter seulement des syllabes/mot, on mesure le pas des cycles mandibulaires entrant dans le stroke des pointers.

---

## **TITRE en anglais**

The Speech Frame and the Sign Frame: A Developmental "Rendez-Vous"

---

## **RESUME en anglais**

Our working hypothesis is based on the idea that there is a "developmental rendez-vous" between what we call the "Sign Frame" and what we call the "Speech Frame". While the Speech Frame is established in the form of the canonical babbling around the age of 7 months, the Sign Frame appears first of all in the form of imperative pointing around the age of 9 months, before giving place to the so-called declarative pointing. Declarative pointing appears along with the first words, while the Speech Frame allows the child at that stage to coproduce (coarticulate) a vowel and a consonant (Sussman et al., 1999). The relative importance of the elements of this developmental "rendez-vous" at the time of the emergence of the first words remains to be explored. In the present contribution, we would like to study the existence of a harmonic relationship between the Speech Frame and the Sign Frame. To this end, we studied the distribution of the babbling frequencies, and of the durations of the pointing or "stroke" gestures. Our results for these six children, followed during 12 months, show that with a babbling mode at 3 Hz and "strokes" gestures at 600-700 ms (1.5 Hz), we can account for the first words template. Thus, with these "prosodic words" that can vary from one to two syllables, it is necessary to call upon the *foot* as a metric control unit rooted in pointing. This will account for the current observations in the literature provided that instead of counting only syllables/words, one measures the mandible cycles control embedded in pointing "strokes".

---

## **DISCIPLINE – SPECIALITE DOCTORALE**

Sciences du Langage

---

## **MOTS-CLES**

Cadre de la parole, Cadre du signe, babillage, syllabe, pointer de l'index, stroke, pied

---

## **INTITULE ET ADRESSE DU LABORATOIRE**

Institut de la Communication Parlée, UMR CNRS 5009 - Université Stendhal  
DU, 1180 Av Centrale  
BP25 - 38040 GRENOBLE Cedex 9 FRANCE  
Phone: +33 4 76 82 4128 Fax: +33 4 76 82 43 35

---

# TABLE DES MATIERES

EN GUISE D'INTRODUCTION :	13
Piaget et Chomsky auraient-ils tous deux tort ?	13
PREMIERE PARTIE : LE CADRE DE LA PAROLE	25
I.1. Le contrôle du « cadre » ou contrôle du cycle syllabique	27
I.1.1. Un chemin évolutif pour expliquer le développement de la parole	27
I.1.1.1. L'explication phylogénétique	27
I.1.1.2. Une explication ontogénétique	29
I.1.1.3. La dominance du cadre et la notion de « cadre pur »	30
I.1.1.4. Différents « cadres purs » idiosyncrasiques	32
I.1.1.5. Le contrôle neural du « cadre »	34
I.2. Le contrôle du « contenu » ou contrôle segmental	36
I.2.1. Vers une lèvre de moins en moins compliante...	37
I.2.2. Un développement séquentiel pour le contrôle de la parole ?	38
I.2.3. Quel développement du contrôle du velum pour la nasalité ?	40
I.2.4. Quel développement du contrôle des articulateurs dans la coproduction des consonnes et des voyelles ?	43
I.3. Conclusion	47
SECONDE PARTIE: LE CADRE DU SIGNE	49
II.1. Le système attentionnel :	50
orientation et alerte	50
II.1.1. Vision et attention	50
II.1.2. Notre première orientation d'alerte	51
II.1.3. Un handicap avantageux	51
II.1.4. Le principe de la coordination en perception active	53
II.2. L'indexation des objets dans le monde	54
II.2.1. La notion d' « objectitude »	54
II.2.2. L'indexation chez l'adulte et l'enfant	55
II.2.3. Indexation dans le cadre robotique	57
II.3. Les grands systèmes sémantiques cérébraux de suivi d'agents/objets/événements dans le monde	60
II.3.1. Le système Where	62
II.3.2. Le système What	71
II.3.3. Le système How	72
II.3.4. L'intégration du système What et du système Where	75
II.3.5. Le système When prioritaire sur le Where et le What	76
II.3.6. Le système That	79
II.3.7. Le système Then	83
II.3.8. Un cerveau pour 7 systèmes sémantiques	85
II.4. Quatre modules dans la voie développementale vers une théorie de l'esprit pour la communication, l'attente et l'attention	88
II.4.1. Le détecteur d'intentionnalité (ID)	88
II.4.2. Le détecteur de direction du regard (EDD)	89
II.4.3. Le mécanisme d'attention partagée (SAM)	92
II.4.4. Le mécanisme de la théorie de l'esprit (ToMM)	97



II.4.5. Pourquoi les agents sont-ils les mieux reconnus ? .....	99
II.4.5.1. Interactions entre l'enfant et autrui.....	101
II.4.5.2. Observations passives entre enfant et autrui.....	101
II.4.5.3. Les interactions et observations entre les enfants et les agents non-humains .....	103
II.4.6. Une étude des modèles de développement social via une robotique humanoïde	105
II.4.6.1. L'œil comme premier outil d'indexation dans le monde .....	107
II.4.6.1.a) Implémentation de l'attention conjointe.....	109
II.4.6.1.b) La maintenance du contact de l'oeil.....	109
II.4.6.1.c) Le suivi du regard.....	109
II.4.6.2. Le doigt comme second outil d'indexation dans le monde.....	114
II.4.6.2.a) Codage du pointage impératif .....	114
II.4.6.2.b) Codage du pointage déclaratif.....	115
II.5. La compréhension de l'intention à partir des mécanismes d'attention et d'imitation	116
II.5.1. La notion d'intention via l'observation des actions.....	116
II.5.2. L'imitation motrice au service de la compréhension de l'intention .....	117
II.5.2.1 Un système miroir chez le singe macaque.....	117
II.5.2.2 Un système miroir chez l'humain.....	120
II.5.3. Le développement du SIM ( <i>Shared Intention Mechanism</i> ) chez l'enfant ou l'intentionnalité selon Tomasello et al. (2004).....	122
II.5.3.1. L'action intentionnelle humaine .....	123
II.5.3.2. Comprendre cette intention .....	124
II.5.3.3. Intentionnalité partagée .....	128
II.5.4. Quelles fonctions cérébrales pour la compréhension de l'action et de l'intention ? .....	132
II.5.5. Comment passe-t-on de SAM à SIM ?.....	140
II.5.5.1 <i>Comprendre les intentions à 3 mois</i> .....	143
II.5.5.2. Compréhension de mouvements violant la biologie à 8 mois .....	144
II.5.5.3. Compréhension de la structure intentionnelle et compréhension du pointer à 10 mois .....	145
II.6. La mise en place de la production du geste de pointage chez l'enfant dans le développement du cadre du signe .....	150
II.6.1. Les origines ontogénétiques de la production de pointage.....	151
II.6.2. Le pointage déictique comme une action communicative de base chez l'enfant	155
II.6.2.1 Pointage impératif, pointage déclaratif.....	155
II.6.2.2. ... ou plutôt pointage coopératif ? .....	157
II.7 Le geste de pointer : « en route vers la syntaxe ! » .....	162
II.7.1. Naissance de la syntaxe dans le bras... ..	162
II.7.1.1. Syntaxe par le geste ou la parole : même combat dans l'hémisphère gauche .....	164
II.7.1.2. Syntactisation, grammaticalisation .....	166
II.8. La monstration et l'interrogation aux origines de la parole.....	167
II.8.1. Trois cris d'alerte pour les singes vervets .....	168
II.8.2. Quel avantage évolutif à un système de communication avec monstration et interrogation chez l'animal ?.....	169
II.8.3. Peut-on parler d'un véritable langage chez les singes vervets ?.....	170
II.8.4. L'apport des suricates .....	171
II.8.4.1. <i>Les cris d'alerte</i> .....	172
II.8.4.2. Réponses aux cris d'alerte .....	173
II.8.5. Les bonobos linguistes de 2 ans ½ .....	174

II.8.6. Conclusions .....	175
II.9. Quels avantages d'un système de monstration/interrogation dans la naissance du langage ?.....	176
II.9.1. La lexicalisation de la monstration.....	177
II.9.2. Aux racines d'un système d'interrogation.....	178
II.10. Du pointer à la morphosyntaxe via la prosodie .....	181
II.10.1. « Les origines de la grammaire » : <i>bootstrapping</i> chez l'enfant .....	181
II.10.2. Le système omniprédicatif selon Launey .....	186
II.10.3. La défektivité spéciale du verbe .....	187
II.10.4. Quelle solution pour le verbe ?.....	189
II.11. Sociogenèse et ontogenèse du processus de grammaticalisation .....	191
II.11.1. Sociogenèse de « that » (ça) à « that » (que).....	191
II.11.1.1. La puissance des déictiques dans le processus de grammaticalisation.....	192
II.11.1.2. La récursivité : une clé pour expliquer la différence entre communication animale et humaine ?.....	196
II.11.1.2.a) La faculté de langage au sens large (FLB) .....	197
II.11.1.2.b) La faculté de langage au sens étroit (FLN) .....	198
II.11.2. Ontogenèse de « that » (ça) à « that » (que).....	201
II.11.3. Conclusion.....	205
II.12. La relation geste-parole chez l'enfant dans l'acquisition du langage.....	208
II.12.1. Les combinaisons geste-parole chez l'enfant .....	210
II.12.2. Quand la non-redondance du geste et de la parole annonce l'énoncé à 2 mots. 214	
II.12.3. Concordance/discordance geste-parole comme index de transition.....	221
II.12.3.1. Les gestes dans l'acquisition de la conservation de quantités .....	221
II.12.3.2. Le geste dans l'acquisition de l'équivalence mathématique .....	223
II.12.4. Relation entre gestes déictiques et mots représentationnels.....	224
II.12.5. Le geste comme outil d'apprentissage pour tous ?.....	226
II.12.5.1. Sourds vs. entendants .....	226
II.12.5.2. Voyants vs. non-voyants.....	229
II.12.5.3. Geste et parole chez les enfants à développement atypique.....	233
II.12.6. Conclusion.....	234
II.13. De l'intégration du langage à travers les gestes dans la cognition humaine .....	235
II.13.1. Plaidoyers pour une théorie gestualiste de l'origine du langage .....	237
II.13.1.1. La dominance de la main droite dans l'évolution du geste vers la parole..	241
II.13.1.2. La théorie gestualiste selon Corballis.....	245
II.14. La coordination geste-parole chez l'adulte.....	248
II.14.1. Quels modèles théoriques pour expliquer la relation geste-parole ?.....	250
II.14.1.1. Modèle de Krauss, Chen et Gottesman (2000) : lien préconceptualiseur .	250
II.14.1.2. Modèle de De Ruiter : lien conceptualiseur .....	254
II.14.1.3. Modèle de Cassel et Prevost : lien post-conceptualiseur.....	257
II.14.1.4. Modèle de McNeill : la théorie Growth Point (GP) .....	258
II.14.1.5. Modèle de Kita et Özyürek: l'hypothèse d'interface.....	261
II.14.1.5.a) Hypothèse d'imagerie libre (HIL) .....	262
II.14.1.5.b) Hypothèse lexico-sémantique (HLS) .....	262
II.14.1.5.c) Hypothèse d'interface (HI).....	263
II.14.1.6. Geste + parole : un bénéfice pour le locuteur mais également pour l'interlocuteur .....	266
II.14. La coordination temporelle geste-parole .....	267
II.14.1. Levelt... ou la voix attend le doigt.....	267
II.14.2. Holender .....	269

II.14.3. Feyereisen.....	270
II.14.4. Quand la main pointe la voyelle sur le visage.....	272
II.14.5. Coordination oui, mais synchronisation... à voir.....	274
II.14.6. Pour une synergie de la voix et du doigt : considérations théoriques pour une synchronisation non obligatoire.....	276
II.15. En guise de transition... ..	279

TROISIEME PARTIE : QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE..... 281

III.1. Une hypothèse de travail pour trois questions princeps: le mot, le pied et la syllabe	282
III.1.1. Quels critères pour les premiers mots ? .....	282
III.1.2. Quand peut-on parler de permanence du mot chez l'enfant ?.....	283
III.1.3. Des premiers mots similaires pour tous ?.....	285
III.1.3.1. Première stratégie : a minima.....	285
III.1.3.2. Deuxième stratégie : « les charmes de la conversation ».....	286
III.1.3.3. Troisième stratégie : une alternative .....	287
III.1.4. Définir le mot ou en percer le contrôle ? Trois questions pour.....	289
III.2. Pourquoi un geste <i>discret</i> si voyant pour le mot ? .....	295
III.3. Retour sur notre hypothèse « étonnante »... ..	297
III.4. Analyses du corpus.....	298
III.4.1. Le mode de distribution des durées de détente du pointer : tendance globale ...	298
III.4.2. Le mode de distribution des durées de détente du pointer : analyses individuelles .....	303
III.4.2.1. Anatole .....	303
III.4.2.2. Célia .....	304
III.4.2.3. Jules.....	305
III.4.2.4. Nicolas.....	307
III.4.2.5. Lise .....	308
III.4.2.6. Tom .....	310
III.4.2.7. Conclusion.....	312
III.4.3. Le mode de distribution du cycle du babillage : apports théoriques.....	312
III.4.3.1. <i>Le babillage canonique comme une activité rythmique</i> .....	317
III.4.3.2. <i>Quelle fréquence pour le rythme du babillage ?</i> .....	318
III.4.3.3. Adaptation de l'enfant au rythme propre à la langue maternelle .....	325
III.4.3.4. <i>Le cas de la naissance de la métrique française de Célia : pointers, syllabes et mots</i> .....	326
III.4.4. Le mode de distribution des durées de cycles de la syllabe : tendance globale.	328
III.4.5. Le mode de distribution des durées de cycles de la syllabe : analyses individuelles .....	331
III.4.5.1. Anatole .....	331
III.4.5.2. Célia .....	332
III.4.5.3. Jules.....	333
III.4.5.4. Nicolas.....	334
III.4.5.5. Lise .....	335
III.4.5.6. Tom .....	336
III.4.5.7. Conclusion.....	337
III.4.6. Le ratio pointer/babillage pour les enfants du corpus .....	337
III.4.6.1. Anatole .....	337
III.4.6.2. Célia .....	338

III.4.6.3. Jules .....	339
III.4.6.4. Nicolas.....	340
III.4.6.5. Lise .....	341
III.4.6.6. Tom .....	342
III.4.7. Le ratio pointer/babillage : notre « hypothèse étonnante » .....	342
CONCLUSION : PROBLEMES ET PERSPECTIVES .....	349
Davantage de questions pour une réponse ?.....	350
Problèmes d'atypicité ? Les Williams et encore les SMA .....	351
BIBLIOGRAPHIE .....	354

**EN GUISE D'INTRODUCTION :**

**Piaget et Chomsky auraient-ils tous deux tort ?**

## A- Développement du langage et développement moteur

Notre travail repose sur une « hypothèse étonnante » (pour ne pas dire présomptueusement « stupéfiante », traduction française du Nobel Francis Crick, 1994, *The astonishing hypothesis...*), l'idée que « le pied serait dans le bras » : le pied de la métrique linguistique s'entend, contenu dans le geste de pointage du doigt (nous dirons par la suite indifféremment pointage ou pointer). Comment démontrer dans ce sens, au cours du développement de l'enfant et particulièrement lors de l'émergence du langage, que les vocalisations autant que les gestes ont une importance fondamentale en interaction ? Ces deux facettes d'un même acte de communication sont-elles liées de façon indiscutable, la nature et la fonction de ce lien restant encore à être déterminées. C'est dans cette perspective d'un liage (ou *binding*) de la phonologie de la syllabe dans une phonologie du mot, qu'est né notre projet de recherche.

Afin de mieux appréhender quels ont été les précurseurs de cette vision impliquant le corps plus largement, et pas seulement la zone oro-faciale, dans la naissance de la parole, replaçons-nous plus de 30 ans avant notre présent, en 1975, dans le débat opposant Jean Piaget et Noam Chomsky à l'abbaye de Royaumont. Lors de ce débat, organisé et recueilli par Massimo Piattelli-Palmarini (*Théorie du langage théories de l'apprentissage*, Seuil, Piattelli-Palmarini, 1979), une proposition d'expérience fut avancée par un autre Nobel, Jacques Monod, lors d'une discussion portant sur les schémas cognitifs et l'acquisition du langage.

Rappelons que Piaget et Chomsky se confrontent sur la relation qui peut exister entre développement moteur et langage. Le langage possède-t-il son propre réseau indépendant (encapsulé), ou alors est-il connecté et interagit-il de manière sensible avec les réseaux qui traitent de fonctions sensori-motrices ? Les conceptions *minimalistes* bien plus récentes (postérieures à 1995) de Chomsky sur un « organe » du langage si parfait (au sens quasi-mathématique), qu'il serait malgré tout immédiatement lisible par les systèmes sensori-moteurs, si imparfaits (bricolages de l'évolution biologiques), ne nous ont vraiment pas dépayés de ce débat.

Bärbel Inhelder, qui participe également au débat, commence celui-ci en posant l'idée que certaines propriétés d'emploi et de structures du langage trouvent leur origine dans la construction sensori-motrice. Chomsky ne conteste pas cette vision mais ne s'accorde pas à dire que toutes ces propriétés proviennent de la construction sensori-motrice, contrairement à ce que pensent Piaget et Inhelder. Selon Chomsky, il est impossible d'affirmer que les principes de la construction sensori-motrice ou de toute autre méthode générale sont en soi des explications exhaustives aux propriétés du noyau fixe du langage, soit la grammaire ou syntaxe. D'autre part on pourrait trouver des aspects de la structure sémantique qui ne seraient pas explicables par l'intelligence sensori-motrice.

Cette réflexion amène Jacques Monod à poser le défi suivant qui s'adresse aux Piagétiens :

Je pense à une expérience théoriquement très simple : si le développement du langage chez l'enfant est étroitement associé à l'expérience sensori-motrice, on peut supposer qu'un enfant né quadriplégique, par exemple, aurait les plus grandes difficultés à développer son langage. (Piattelli-Palmarini, 1979:211)

En d'autres termes, et si on se place d'un point de vue piagétien, le petit enfant qui serait atteint d'un handicap l'empêchant de développer son système de locomotion et par là-même de marcher, serait alors confronté à des problèmes pour parler, son langage ne se développant pas ou en étant appauvri. En effet, la théorie de Piaget part du principe que les actions motrices constituent des précurseurs fondamentaux au développement des représentations cognitives internes de l'enfant. Les abstractions obtenues à partir de schémas sensori-moteurs permettraient selon lui de construire le langage.

Si l'on suit la perspective de Chomsky et sa vision que toutes les propriétés ne relèvent pas de la construction sensori-motrice, dans le cas cité précédemment, le fait de ne pas marcher ne devrait pas interférer sur le développement linguistique de l'enfant, ni en bien, ni en mal, puisque les propriétés du langage ne seraient pas reliées aux autres propriétés motrices. Son postulat repose donc sur le fait que le langage n'est pas en connexion directe avec d'autres aspects moteurs du développement. Inhelder suppose elle aussi que même dans ce type de cas extrêmes, le langage devrait tout de même être assez riche, bien qu'aucune expérience sur des enfants autres que « normaux » n'ait été réalisée au sein du groupe de Genève, répond-elle.

L'intervention de Monod est-elle tout à fait pertinente ? Ce qu'il livre comme une expérience de pensée serait semble-t-il un test décisif, dont les résultats pourraient infirmer la théorie de Piaget et, nous allons le voir, bien qu'il ne semble pas y penser, celle de Chomsky. Cependant, aucune étude ne s'est véritablement attaquée à cette proposition par la suite, et il faudra attendre l'étude de Sieratzki et Woll (2002), près de trente ans après, pour obtenir une première réponse à cette question.

Sieratzki et Woll (2002) se sont donc intéressés à la relation entre langage et développement moteur, chez une population d'enfants âgés de 18 à 35 mois atteints de SMA (*Spinal Muscular Atrophy*, atrophie musculaire spinale ou amyotrophie spinale infantile [ASI] de type II). Il s'agit d'une maladie récessive autosomale qui affecte 1 naissance sur 3000. Cette maladie a pour caractéristique une apoptose, ou involution des neurones moteurs spinaux, sans implication des aires corticales cérébrales. On distingue plusieurs niveaux de gravité dans SMA, d'une paralysie pratiquement générale des membres et du tronc avec une faiblesse respiratoire sévère, à des enfants qui marchent indépendamment –capacité qui peut être perdue plus tard. Les auteurs ont étudié des enfants atteints de SMA de niveau II, le niveau intermédiaire. Ces enfants ont un développement normal durant les 6 premiers mois, ils peuvent s'asseoir sans aide mais ne peuvent pas rester debout ou marcher.

Sieratzki et Woll (2002) ont utilisé l'inventaire du développement communicatif MacArthur (CDI) afin d'évaluer le développement du langage chez les enfants. Les résultats ont fait apparaître que les scores de vocabulaire des enfants SMA sont proches des scores des enfants en développement normal. Mais, plus intéressant encore, les scores pour les hyper-régularisations atteignent en moyenne 78 %, avec trois enfants obtenant un score supérieur à 90%, c'est-à-dire des scores qui représentent jusqu'à 10 fois ceux des enfants normaux.

Le phénomène d'hyper-régularisation est très important dans l'étude du développement du langage chez l'enfant. En effet, ces hyper-régularisations sont des erreurs dans la forme des mots, comme par exemple en anglais le mot *foot* qui au pluriel se retrouve sous la forme *foots* au lieu de *feet*. Comme tels ce sont des témoins de l'émergence précoce de la grammaire. Lorsqu'un enfant entre dans cette étape fondamentale, cela signifie qu'il applique de lui-même des patrons réguliers à



des items qu'il entend comme irréguliers, utilisant ainsi ses connaissances implicites en matière de règles flexionnelles.

Une telle précocité est frappante chez ces enfants SMA, le phénomène d'hyper-régularisation ne survenant généralement pas avant l'âge de 2 ans. Notons que les enfants SMA suivent le modèle de développement typique par la suite pour les hyper-régularisations.

Les résultats montrent ainsi, qu'en dépit d'une altération motrice sévère dans le cas de l'enfant SMA, la capacité de langage reste intacte. Qui plus est, elle se trouve même supérieure à celle des enfants en développement normal.

Sieratzki et Woll (2002) apportent ainsi des éléments nouveaux pour la compréhension de la relation entre langage et développement moteur, suite au débat Piaget vs. Chomsky.

Selon les auteurs, ces résultats ne sont pas compatibles avec la vision de Piaget, dans la mesure où un enfant privé de locomotion peut tout à fait développer son langage. Et de même ces résultats ne s'accordent pas avec la vision chomskyenne, rejetant une séparation nette entre langage et autres aspects du développement, donc sans influence prédictible du comportement moteur sur le développement de l'« organe » du langage. L'existence d'un module de grammaire indépendant ne peut donc pas rendre compte théoriquement du phénomène de précocité en hyper-régularisation observé ici.

Sieratzki et Woll (2002) avancent une explication tentative pour rendre compte de cet avantage observé chez les enfants SMA, qui serait rendu possible par un remapping cortical :

Il pourrait être suggéré que l'apprentissage de la grammaire par les enfants avec SMA peut être augmenté par une ré-assignation au traitement grammatical des aires pré-frontales ordinairement recrutées par le système moteur. Un re-mapping cortical survient en réponse à une dé-afférentation sensorielle, et a été observé pour des régions somatiques [...] et des modalités perceptives [...]. Les enfants avec SMA ont un contrôle moteur fin adéquat et sont capables de planifier des mouvements dans l'espace, mais ils n'ont pas une puissance suffisante pour les exécuter. (Sieratzki et Woll, 2002:429)

Les auteurs suggèrent que l'enfant SMA, privé de marche, ce qui l'empêche d'explorer le monde physique extérieur environnant comme le ferait un enfant de son

âge, va explorer en compensation son « monde intérieur » et, entre autres, utiliser pour cela ses capacités langagières, ce qui pourra entraîner cette avance certaine dans la construction de sa grammaire. Cette avance se réduisant plus tard, les enfants normaux rattraperont les enfants SMA.

Cette étude de Sieratzki et Woll (2002) apporte des pistes de recherche pour comprendre la relation qui s'établit entre développement du langage et développement moteur. Il est clair que l'altération d'un de ces deux « modules » n'empêche pas l'autre de se développer, puisqu'à un déficit moteur correspond une précocité certaine dans le langage. Cette conclusion n'est pas piagétienne. Et l'avantage langagier observé en présence du handicap moteur n'est certes pas non plus une prédiction chomskyenne.

Bien entendu avant de conclure que nos deux grands psychologue et linguiste ont tous deux tort, nous avons demandé par mail à Sieratzki et Woll si leurs sujets SMA II avaient bien la maîtrise de leurs membres supérieurs, un point important pour notre travail sur la deixis visuo-manuelle. La réponse est oui.

La question de Monod ne serait donc pas si décisive qu'elle en a l'air, même dans un cas de quadriplégie, comme l'avait déjà radicalement posé en conclusion Jerry Fodor à Inhelder:

Cela ne répond pas à la question qui est : que feriez-vous en cas de capacité extrêmement réduite ? Je pensais que la réponse était qu'il suffit qu'on fasse très peu de chose, qu'on bouge simplement les yeux [cf. à la limite le syndrome de *locked-in*]. Mais alors pourquoi ne pas dire que tout déclencheur éthologique, toute liaison innée d'une action quelconque à un schème d'intelligence hautement développé pourrait avoir valeur d'apprentissage ? Une telle affirmation rend tout simplement triviale la doctrine selon laquelle l'intelligence naît de l'activité sensori-motrice, si bien qu'en réalité il n'en reste pour ainsi dire rien (Piattelli-Palmarini, 1979:212)

Reste que le résultat d'un avantage linguistique dû au handicap moteur SMA II n'était prévu par aucun de nos deux théoriciens de l'apprentissage et du langage (ni par Fodor). Nous reviendrons rapidement plus loin sur une posture récente de Chomsky, encadré par deux spécialistes de la communication animale (Hauser, Chomsky, Fitch, 2002, dans *Science*, avec ses suites en débats-réponses par Pinker et Jackendoff dans *Cognition*, Pinker et Jackendoff, 2005, Jackendoff et Pinker, 2005). Mais nous serons forcée de constater que les données évoquées, et même

celles apportées depuis, n'ont pas ce caractère stimulant de l'article défiant de Sieratzki et Woll (2002), finalement accepté dans une modeste revue comme *Lingua*.

## **B- A la recherche d'un cadre moteur du signe intégrant le cadre de la parole pour la phonologie du mot**

Si l'on accepte même a minima ces conséquences d'une avance dans l'émergence du langage due au handicap moteur, il faut compter avec la place laissée à l'organisation neurale des propriétés motrices du corps humain. Dans ce sens, notre étude part tout simplement du problème général (dit classiquement bernsteinien) des degrés de liberté, un problème que l'enfant est censé rencontrer dans ses différents domaines de développement : c'est plus généralement le problème de développer les habiletés du contrôle des segments du corps. Et il en va bien entendu de même pour la parole, en ce qui concerne les degrés de liberté de ses articulateurs.

Notre hypothèse de travail repose sur l'idée qu'il existe un « rendez-vous développemental » entre ce que nous avons nommé le « cadre du signe » et le « cadre de la parole ». Cette hypothèse part d'une découverte relativement récente, à savoir la co-émergence du phénomène de coarticulation –qui constitue le contrôle de base de la parole– avec le(s) premier(s) mot(s), juste à la fin de la première année de vie (Sussmann et al, 1999). Pour notre étude, nous serons amenée à nous poser plusieurs questions empiriques, qui pourront éclairer précisément la nature de cette relation entre geste et parole chez l'enfant : (i) De quels mécanismes l'enfant a-t-il besoin pour arriver à l'étape de la coarticulation (ou coproduction) pour le contrôle de la parole ? (ii) De quels mécanismes a-t-il besoin pour arriver au mot, dans le contrôle du signe ? Et enfin (iii) quelle est la métrique qui peut permettre le dialogue (*cross-talk*) et l'intégration entre le signe et la parole ?

Notre cadre de travail (*framework*) sur le signe-parole prend avantage de l'émergence des cadres du babillage et du pointage pour proposer un processus d'intégration de la parole (le babillage canonique à 7 mois) dans le signe (le pointage de l'index vers 9 mois), soit dans le cadre de l'unité prosodique des premiers mots, le pied métrique (ou *foot*) de l'enfant.

Notre travail consistera donc à mettre en évidence toutes les étapes fondamentales nécessaires à la naissance de ces contrôles, pour la mise en place,

d'une part, de la phonologie du langage à travers les différents contrôles pour la parole et, d'autre part, pour l'apparition de la sémantique à travers les mécanismes d'attention et l'indexation des objets du monde par les gestes, pour arriver aux premiers mots. Nous discuterons alors de la nature quantifiable de cette relation, soit plus précisément du rapport harmonique qui s'établit entre geste et parole.

Voici donc ce cadre conceptuel ou dispositif de travail (*framework*) complet (ci-dessous Figure B.1), tel que nous l'avons établi au fil des ans de notre thèse. Il récapitule toutes les étapes importantes qui vont être détaillées par la suite. Nous retrouverons les deux flux essentiels pour arriver au mot, le flux du « cadre de la parole » et le flux du « cadre du signe », déclinés tous deux dans les types de contrôles moteurs acquis par l'enfant au cours de son développement.

Il pourra paraître étonnant de voir autant de domaines divers traités ici dans un univers linguistique. Mais il est nécessaire, à chaque étape, à l'intérieur de chaque flux, de bien comprendre les notions cruciales pour la mise en place de ce système minimal de communication qui se révélera dans la suite du développement et de l'histoire du langage et des langues si puissant. Parmi les domaines qui doivent, selon nous, être absolument abordés pour situer notre travail, celui de la psychologie cognitive est immense et on décèlera aisément à plus d'une reprise notre incompetence par formation dans ce champ de recherche. On s'apercevra notamment que nous nous sommes beaucoup servis du filtre, selon nous efficace, qu'un roboticien a pu utiliser pour faire le point sur les connaissances de la psychologie du développement en vue de construire un robot pointant : il s'agit de la thèse de Brian Scassellati (2001) dans laquelle nous avons beaucoup puisé pour comprendre ce qu'il nous fallait retenir de la théorie de l'esprit (et de bien d'autres points) pour la deixis.

Ce travail étant « jumelé » avec la thèse en cours de Claire Lalevée sur le babillage, nous ne traiterons qu'en parent pauvre ce « cadre de la parole », en empruntant beaucoup à ce qui a été emprunté à MacNeilage (mais on notera aussi notre hommage appuyé aux travaux pionniers sur la production d'une Française non macneilagienne, Bénédicte de Boysson-Bardies) dans les équipes successivement dirigées par Christian Abry, qui ont démarré ces études développementales depuis 7-8 ans, incluant ce qui a pu être originalement proposé à Grenoble jusqu'à cette thèse jumelle. Si nous avons pu traiter exhaustivement pour les pointages l'inestimable

corpus de 6 bébés suivis tous les 15 jours entre 6 et 18 mois, que nous a laissé, à Claire et à nous-mêmes, Stefanie Brosda, nous avons dû mesurer par sondage le babillage sur 5 des enfants en plus des données obligeamment fournies pour Célia que Claire parvenait à décrire exhaustivement.

Le gros de notre réflexion est bien entendu consacré au « cadre du signe ». Nous n'avons pas hésité sur un domaine aussi difficile que celui de cette simple sémiotique, qui a déjà fait couler tant d'encre, à donner plus d'éléments théoriques que nous pourrions en appliquer dans notre partie empirique. L'avenir nous dira si nous pouvions mettre en œuvre expérimentale plus profondément d'autres zones de cette grosse partie théorique immergée.

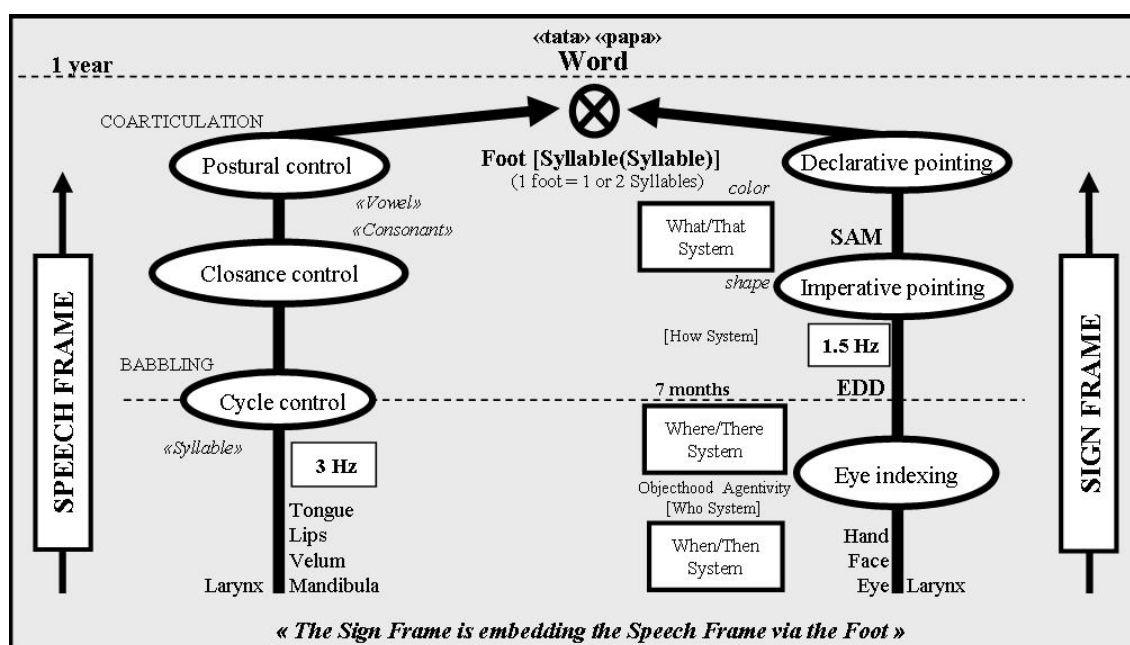


FIGURE B.1 : **Un cadre pour deux cadres.** A environ 1 an, le cadre de la parole va être intégré dans le cadre du signe : une ou deux syllabes dans un gabarit fourni par le pied pour les premiers mots. Pour le cadre de la parole, après l'émergence de la « syllabe » du babillage canonique, il reste deux étapes : le contrôle de la closance pour la « consonne », et la coarticulation (coproduction) pour le contrôle postural de la « voyelle » à l'intérieur de la « consonne ». Pour le cadre du signe, trois flux cérébraux sont en maturation : la voie de détection des événements occipito-pariétale (*When*), la voie dorsale (*Where*) et la voie ventrale (*What*). Leurs produits sont l'objectitude (*Objecthood*) et l'agentivité (*Agentivity*, système *Who*), alors que le système *How* ventro-pariétal fournit l'affordance de la forme, avant la couleur de l'objet via le système *What*. Parmi les « réponses » correspondantes (*Then/There/That*) à ces *Wh*-systèmes, la plus pertinente pour notre posture déictique est la « *That-Path* » fronto-pariétale (mise en évidence par Løevenbruck et al., 2005). Le mécanisme d'attention-intention partagée (SAM-SIM) se développe plus tard que la détection de direction de l'œil (EDD). Avec des cycles de babillage à 3Hz, la prédiction de ce cadre est un ratio Babillage/Pointage de 2:1. (Abry, Ducey-Kaufmann, Vilain, à paraître).

La partie fondamentale de cette étude réside dans ce rendez-vous développemental, offert par les éléments-clés qui permettent cette rencontre. Comme la phonologie de la syllabe est devenue, depuis MacNeilage, plus concrète en neuro-motricité par le contrôle rythmique de la mandibule, la phonologie du mot sera ici viabilisée par la phonologie du pied (*foot*) concrétisée en cognition neuro-motrice par le contrôle du pointage de l'index.

Comme nous nous intéressons aux primitives de la communication face-à-face à travers ses comportements en geste et parole, nous nous devons d'avertir le lecteur que le geste n'est pas envisagé ici d'un point de vue général interactionniste, comme il l'a naturellement souvent été dans la littérature sur la gestualité, en plein essor sur la communication non-verbale et co-verbale (un bon exemple en français est le livre récent de Colletta, 2004). Nous ne cherchons pas à déterrer la fleur avec toute sa terre, tout ce qui pourrait venir autour... Non, nous dépouillons toute cette verdure afin de ne garder que les racines du phénomène, les primitives explicatives. Ainsi le geste de pointage de l'index, au sens où nous l'entendons ici, n'est pas étudié dans un contexte général ou particulier (sur la complexité du contexte, cf. le numéro spécial de *Journal of Pragmatics*, 35, 2003), où nous détaillerions tous les éléments de la scène dans laquelle il survient avec toutes les interactions en jeu, tous les gestes significatifs (« représentationnels ») produits par l'enfant vers l'adulte. Il n'est même pas compris simplement comme un des nombreux dispositifs déictiques. La deixis est encore un phénomène, pour nous ici, beaucoup trop général : en témoignent quelques colloques comme celui précisément sur *La Deixis* (PUF, 1990) ; ou « Ici et maintenant » (qui s'est tenu à Nice les 25-26 nov. 2005). Le livre semble-t-il « pointu », *Pointing: Where Language, Culture and Cognition meet*, enfin édité en 2003, par Kita, à partir des travaux du groupe du Max Planck de Nimègue, manque de psycholinguistique expérimentale (« chez » Levelt !), et totalement de la moindre approche neurale. Pour notre part, nous envisagerons chaque fois que cela nous sera possible le pointer de l'index sous l'angle du contrôle cérébral qu'il implique, c'est-à-dire dans sa fonction fondamentale qui consiste à diriger, avec son propre cerveau, le cerveau d'un congénère (ou du plus vieux compagnon de l'homme, le chien, étudié avec l'enfant et le singe, par le groupe de Tomasello au Max Planck de Leipzig), vers un objet d'intérêt privé, dans un système déictico-attentionnel partageable entre l'enfant et l'adulte (sinon toujours le

chimpanzé ou le chien). Le geste de pointer (et pourquoi pas le museau) est bien un outil pour diriger l'attention, qui possède la fonction d'apporter de l'information monstrativement structurée. En aucun cas ce geste n'est à considérer comme un geste fondamentalement porteur en lui-même d'information sémantique concrète, comme on dit souvent « représentationnelle » (sinon de l'émotion, l'interrogation, etc. ; se discute encore plus la question de l'information de localisation vs. celle de démonstration, cf. Hurford, 2003). Car selon le proverbe bien connu —« Quand le sage montre la lune, le sot regarde le doigt »—, on ne regarde bien entendu pas la main ou l'index ou la bague, mais ce qui est indiqué par un vecteur comme celui de l'index. Une petite histoire rapportée par Paul Radin<sup>1</sup>, le grand spécialiste de la mythologie des Winnebago, celle de leur « fripon divin », dupeur dupé, ou *Trickster* mondialement le plus célèbre (sans doute à cause d'une « liaison dangereuse » de Radin avec Jung) nommé Wadjukaga, résumera parfaitement à elle seule la fonction princeps intentionnelle d'un tel geste dans la communication... Pourvu que l'on en ait sensément la plus petite théorie de l'esprit:

**The Pointing.** [Trickster] went to a particular place. There, as he went along, he came in sight of a lake [...]. So he came towards the edge of the lake. He came nearer. To his surprise, near the edge of the lake someone was standing. [...]. He stood there wearing a black shirt. There he went. To his surprise, he was pointing across the lake. He went to him there and he said, "Kodé! my younger brother, what are you pointing at?" After he said it, he was not answered. Again he did not [answer]. "My younger brother, what are you pointing at?" Useless. Again he asked the third time: "My younger brother, what is it that you are pointing at?" It was useless. There no person said a thing like he thought he would. Again the fourth time he asked him, but he did not now answer him. Thus he continued pointing across. Thus it was. "Well, *hahó*. We'll do that. What? Me too. Likewise, I can point for a long time once I have put on the black shirt," he said. He put on a black shirt. Thus he did and he quickly stepped along side of him. Also where he was pointing, there Trickster pointed. Thus he stood indefinitely. In the course of time, Trickster's arm got tired. He said, "My younger brother, so let's be done with this," said he, but he was still not answering. Again a second time he said, when he could not hold on any longer, "My younger brother, so let's be done with this. My arm is tired," he said, but he was not answered. Thus he said repeatedly, but he was never answered. "My younger brother, I am hungry. Let's eat, then we'll begin again. I will kill a very fine animal. Whatever one you like, I will kill one of that kind. So let's quit," he had said, but he received no answer. "*Howa*, what am I saying? This one, one whose heart has slipped through him, the thing he is doing, I am doing too" he said. He walked away from him. When he looked back at him, to his astonishment, he was a stump [souche]. When he had said that it was pointing, he had been referring to one whose branch had extended out. There said, "*Hohó!* it is because of this the people have called me "foolish" [Trickster]. It is even as they have said," he said. And away he walked.

Typiquement, la littérature a établi abondamment que l'unité cadre (ou *frame*) de la parole était la syllabe et que celle-ci témoigne du processus de phonologisation

---

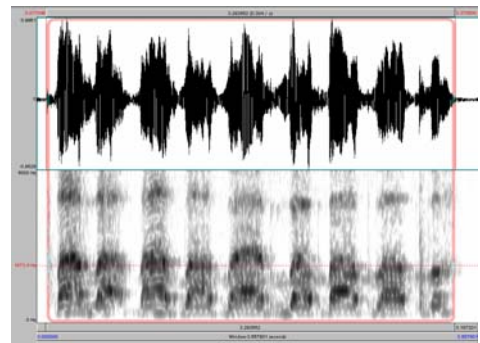
<sup>1</sup> RADIN P., *The Trickster : A study in American Indian Mythology* (New York: Schocken books, 1956) 13-14. The telling by Felix White, Sr. is almost identical. See Kathleen Ann Danker, *The Winnebago Narratives of Felix White, Sr.: Style, Structure and Function* (Ph.D. Thesis, University of Nebraska, Lincoln, May, 1985, 157-161).

chez l'enfant. En revanche très peu (ou pas) d'études à notre connaissance ont tenté d'apporter une unité sémantique et phonologique au signe, et encore moins d'essayer de lier ces deux unités afin d'en dégager une mise en gabarit pour le mot, dans un cadre (*frame*) du signe. Plusieurs études en revanche ont tenté de comprendre comment cohabitent le geste et la parole lorsqu'ils sont combinés dans une même tâche, souvent déictique. Mais chez l'enfant, nous pensons être la première à apporter des éléments de réponse quant à l'harmonisation motrice entre ces deux modules, et surtout à questionner quelle répercussion cruciale elle peut avoir pour la réalisation des premiers mots.

Le dispositif de travail dont nous disposons nous permettra en arrière-plan de mettre en évidence le rôle déterminant des fonctions monstrative et interrogative dans le sens où elles enracinent profondément les bases du langage. Ces deux systèmes montrent quotidiennement leur robustesse dans une interaction humaine face-à-face, et la nécessité de les implémenter dans des outils de communication est une évidence. C'est donc tout naturellement que nous nous pencherons ensuite sur la sociogenèse et l'ontogenèse du processus de grammaticalisation, en soulignant la puissance des déictiques pour ce processus, des déictiques qui trouvent leur origine dans ces mécanismes fondamentaux dans toutes les langues, ces primitives pour le langage, de monstration et d'interrogation.



# PREMIERE PARTIE : LE CADRE DE LA PAROLE



baßababababababaBa...

Le premier cadre étudié ici est le *cadre de la parole*, le flux concernant la partie gauche sur la figure B.1 de notre Introduction. Nous allons nous attacher à décrire les différentes étapes requises dans le développement de l'enfant, dans la mise en place de son système phonologique.

De nombreuses études se sont intéressées à l'acquisition de la parole chez l'enfant, tant en perception qu'en production, et se sont inscrites dans des courants de pensée différents. Avant les années soixante, Jakobson<sup>2</sup> pensait que l'enfant avait la capacité de produire n'importe quel son de parole, dans la période dite « prélinguistique », et qu'ensuite on assistait à une réduction de cette faculté, pour finalement ne conserver que les sons distinctifs et pertinents de la langue maternelle. Sa vision l'a poussé à se désintéresser complètement des productions du babillage, qui n'auraient eu aucun rapport avec le répertoire des premiers mots. De manière assez étonnante, il avance que le babillage, non linguistique, serait séparé des premières productions linguistiques vers 12 mois, par une période silencieuse...

Cette première ligne de pensée assez radicale a largement contribué à ce que cette période prélinguistique ne soit pas étudiée, ne présentant soi-disant que peu d'intérêt, idée confortée par une personnalité comme Chomsky qui, à l'époque, voit le babillage comme une simple étape de maturation. Cette opinion, bien que non justifiée empiriquement, a prévalu pendant longtemps. Elle a présenté pour seul avantage de motiver de nouvelles approches. Parmi celles-ci nous allons privilégier celle qui fait appel aux contraintes biomécaniques du conduit vocal, pour démontrer que le système phonétique de l'enfant est le produit de l'adaptation à des contraintes articulatoires bien particulières dans la production du langage.

Dans cette partie, nous allons donc nous attacher tout spécialement à décrire les étapes essentielles de la mise en place de la parole chez l'enfant, tout en mettant l'accent sur les contraintes biomécaniques existantes et sur les différents contrôles moteurs que l'enfant doit acquérir pour atteindre la norme adulte.

---

<sup>2</sup> JAKOBSON R. (1969), *Langage enfantin et aphasie*, Paris, Editions de Minuit (Traduction de : *Kindersprache, Aphasie und allgemeine Lautgesetze*, Uppsala, 1941)

## I.1. Le contrôle du « cadre » ou contrôle du cycle syllabique

### I.1.1. Un chemin évolutif pour expliquer le développement de la parole

Pour notre étude, nous partons d'une théorie qui vise à expliquer le développement ontogénétique et phylogénétique de la parole par les contrôles articulatoires, particulièrement par la mise en place graduelle du contrôle des différents articulateurs glottiques et supraglottiques. Cette théorie proposée par Peter MacNeilage (MacNeilage, 1990, 1998) est plus connue sous le nom « Cadre, puis Contenu » (*Frame, then Content*). Pour rendre compte du contrôle de la parole à la fois dans son développement et dans son évolution, MacNeilage part d'un questionnement essentiellement évolutionnaire sur notre capacité à organiser et produire les mouvements qui constituent la parole.

MacNeilage argumente que la parole chez l'homme diffère de la communication vocale d'autres mammifères, dans le sens où elle est tout à fait spécifique. Similairement à d'autres espèces animales, l'être humain utilise des séquences d'ouverture et de fermeture du conduit vocal dans le but de communiquer. Cette alternance rythmique constitue ce que MacNeilage appelle le « cadre ». Néanmoins, nous nous distinguons des autres espèces par le fait que nous ajoutons à ce cadre une modulation articulatoire du mouvement, correspondant à la notion de « contenu ». Suite à ce constat, MacNeilage propose de s'interroger sur l'ontogenèse de la parole, si l'on admet qu'elle peut récapituler, au moins en partie, la phylogenèse, suivant en cela Haeckel (1896)<sup>3</sup>.

#### I.1.1.1. L'explication phylogénétique

L'auteur propose de tracer le chemin évolutif de la parole humaine de façon darwinienne, en partant des cyclicités ingestives vers la parole. Il suggère cette idée en se reposant sur l'existence d'une forme intermédiaire, qui serait présente chez beaucoup d'autres grands singes. Ces cyclicités communicatives visuo-faciales se

---

<sup>3</sup> Haeckel E., 1896, *Systematische Phylogenie. Zweiter Theil: Systematische Phylogenie der wirbellosen Thiere (Invertebrata)*. Berlin. Verlag von Georg Reimer. 720 p.

trouvent sous la forme de claquements de lèvres (*lipsmacks*), claquements de langue (*tonguesmacks*) ou encore de dents (*teeth chatters*). Ainsi, selon MacNeilage, la modification de la région corticale périsylvienne frontale, impliquée dans le contrôle des mouvements ingestifs (la mastication), a dû mener à capacité de production de la syllabe, c'est-à-dire au cadre. Et en conséquence, cette modification a sûrement dû rendre également les autres capacités relatives à l'ingestion disponibles pour la modulation de ce cycle, afin de former différentes consonnes et voyelles, c'est-à-dire le contenu. Sa théorie suggère alors que le contrôle de la production de la parole ait évolué par le phénomène darwinien de descendance avec modification, la modification en question étant celle du système du contrôle moteur cortical des primates. Un système médian et un système latéral, ce dernier incluant l'aire de Broca, associés respectivement au contrôle des vocalisations chez le primate et à la capacité d'apprentissage vocal. Le système médian serait donc le siège du cadre, où se logeraient contrôle du rythme et imitation de la production de parole. Le système latéral serait impliqué dans des mouvements plus fins et contrôlerait la production des contenus segmentaux, caractéristique spécifique à l'humain.

Dans son optique phylogénétique, MacNeilage explique que les trois composantes principales du système de production vocal des mammifères – respiration, phonation et articulation – ont subi des modifications avec l'arrivée de la bipédie chez les hominidés. En effet, les composantes respiratoire et phonatoire ont pris une orientation de type verticale. On peut noter ainsi que, chez les hominidés avancés, la partie postérieure du système articulaire prend une configuration verticale, au contraire de la partie antérieure : ce qui mène schématiquement à l'obtention d'un conduit vocal coudé à deux tubes. L'intérêt de cette évolution vers un conduit à deux tubes aurait été d'augmenter considérablement le potentiel acoustique de l'hominidé, lui permettant ainsi de produire une gamme de sons plus étendue (contre cette idée trop répandue depuis la théorie de Philip Lieberman, qui ne résiste même pas à l'acoustique de base du conduit vocal, cf. Heim et al., 2002 ; notons que depuis 1999, MacNeilage a renoncé à cette position). La principale différence entre la parole humaine et les systèmes de cris des autres mammifères implique donc cette composante articulaire. Chez tous les mammifères, les composantes respiratoires et phonatoires peuvent être décrites en termes de cyclicités biphasiques modulées. Chez les mammifères non-humains, le système

articulatoire est typiquement utilisé seulement dans une configuration ouverte pendant la production d'appel. Sauf chez les animaux qui aboient, comme le chien ou le babouin, et dans la parole humaine, où le conduit vocal alterne de façon plus ou moins régulière entre une configuration relativement ouverte et une configuration relativement fermée (ouverte pour les voyelles et fermée pour les consonnes), ce qui peut être utilisé comme une caractéristique de base de la parole. Rappelons que la syllabe, unité universelle de la parole, est définie en termes d'un nucleus avec un conduit vocal ouvert et d'une marge avec un conduit vocal fermé. Le contrôle de cette cyclicité d'ouvertures/fermetures chez les humains permettra de produire les unités de base que sont les consonnes et les voyelles. La modulation de ce cycle ouvert/fermé chez les humains permettra de produire des unités de base différenciées parmi les consonnes et les voyelles. De ce fait, la communication humaine se distingue des autres communications vocales de mammifères en terme de phases différenciées d'un mouvement cyclique, par le fait qu'un troisième niveau de cyclicité, de type articulatoire, co-existe avec les deux niveaux —respiratoire et glottique ou laryngé— déjà présents chez les mammifères.

### **I.1.1.2. Une explication ontogénétique**

Après une première vision phylogénétique pour éclairer le chemin évolutif de la parole, MacNeilage propose de se pencher sur son ontogenèse. Il suit donc le principe cité précédemment que l'ontogenèse récapitule la phylogenèse. Cela signifie que le développement du contrôle de la parole chez le bébé humain suivrait les mêmes étapes fondamentales que le développement de ce contrôle chez l'espèce humaine.

MacNeilage note qu'à l'âge de 5 mois, les bébés sont déjà capables de jouer à faire osciller leur mâchoire (*jaw wags*) sans phonation, gestes qui peuvent être rapprochés des claquements-succions de lèvres (*lipsmacks*) qui se rencontrent chez les primates non-humains. Ce rapprochement est permis par les fonctions similaires dont ils font preuve. Les *lipsmacks* sont observés dans de très nombreuses interactions sociales, notamment dans les communications d'affiliation, très tôt, dès 3 jours chez le petit rhesus. Plus généralement cette relation d'affiliation entre deux êtres implique un contact sensoriel olfactif, tactile, visuel. Ce comportement est donc plutôt similaire à ces mouvements de mandibule observés chez l'enfant, et pourrait

ainsi être considérés comme un précurseur phylogénétique de la parole, un type de geste communicatif intermédiaire entre ingestion et parole.

Mais plus spécifique encore dans le développement de l'enfant est le comportement appelé *babillage canonique* (pour ses premières définitions comportementales précises, dépassant le vague de termes antérieurs à la Spitz, comme *lallation*, cf. Oller dès 1978<sup>4</sup>, puis Stark en 1979<sup>5</sup>). C'est un type de babillage rythmique qui émerge soudainement vers l'âge de 6-7 mois, quelle que soit l'origine linguistique de l'enfant. Le babillage est caractérisé par une coordination oro-laryngée (larynx + conduit vocal) pendant la phonation. Il est produit sous la forme de suites répétitives de proto-syllabes de type CVCVCV... soit de cycles rythmiques d'ouverture-fermeture de la mandibule avec phonation. Typiquement le babillage correspond à des productions de type [bababa...] ou encore [dadada...]. Qu'il soit bien entendu, en dépit de ces notations en symboles phonétiques, qu'il ne s'agit aucunement d'une suite de segments *b+a*, ni même de syllabes *ba+ba* : à qui l'idée viendrait-elle de noter les flexion-extensions du bras du bébé qui gigote en consonnes-voyelles ou en syllabes-syllabes? Un train de ces « protosyllabes » est mieux conçu comme un train de détentes du conduit vocal : à ce stade, l'enfant se montre capable d'un début de maîtrise de ses deux ensembles de coordinations glottiques et supra-glottiques pour produire un rythme de ce type. Les énoncés canoniquement babillés alternent généralement un seul type de « consonne » (un même lieu de closure ou contact dans le conduit vocal) et une « même voyelle » (une zone d'ouverture du conduit vocal beaucoup moins déterminée que la zone de butée d'un articulateur contre le plafond de la bouche, qui permet classiquement de définir un lieu consonantique).

### **1.1.1.3. La dominance du cadre et la notion de « cadre pur »**

Dans l'étude des productions liées au babillage, on peut observer une certaine variabilité dans la coarticulation entre l'apparition du babillage autour de 7 mois et les premiers mots autour de 12 mois. Cette variabilité peut être attribuée à la seule oscillation de la mandibule, ce que MacNeilage nomme la « dominance du cadre ». Les proto-syllabes du babillage sont constituées de consonnes et voyelles qui sont

<sup>4</sup> OLLER D.K. (1978). Infant vocalization and the development of speech. *Allied Health and Behavioral Science*, 1, 523-549.

<sup>5</sup> STARK R.E. (1979). Prespeech segmental feature development. In P. Fletcher and M. Garman (Eds.). *Language Acquisition*, 15-32. New York: Cambridge University Press.

de même lieu d'articulation, comme par exemple une consonne coronale [d] accompagnée d'une voyelle antérieure [æ], ou alors une consonne labiale [b] avec une voyelle centrale de type schwa ou [a]. L'auteur suggère alors que la mandibule soit le seul articulateur actif dans ce type de réalisation, tandis que tous les autres articulateurs sont passifs, à l'image de la langue qui reste en position de repos, ce que MacNeilage appelle le « cadre pur ». Ceci est illustré par le fait que dans la plupart des cas, le babillage canonique va être constitué d'une syllabe donnée, suivie par la même syllabe, soit un babillage rédupliqué de type [baba]. Les syllabes successives qui diffèrent –le babillage varié comme [badi]– sont expliquées en termes de contrôle du cadre, reflété dans les changements d'élévation de mandibule entre syllabes. Il n'existe pas vraiment de chronologie précise dans l'apparition du babillage rédupliqué et varié, l'un pouvant survenir avant l'autre indifféremment, ou l'un pouvant survenir sans l'autre également, selon les enfants.

Les changements dans la dimension verticale liés à l'amplitude de la mandibule sont plus fréquents que les changements dans la dimension horizontale. Dans le cas du changement de lieu de la consonne ou de la dimension avant-arrière de la voyelle, la langue serait déjà en place. La langue permettrait alors la variabilité avant une séquence de babillage, expliquant alors que l'on trouve différents lieux d'articulation consonatiques dans ces mêmes séquences. A l'intérieur d'une même configuration d'articulation, plusieurs voyelles et consonnes pourront être réalisées lors des différentes phases de l'oscillation mandibulaire. Le fait que la langue ne soit pas contrôlée implique alors que l'articulation de la consonne et de la voyelle soit fortement coarticulée, résultant alors dans le choix d'un même lieu d'articulation.

Comme MacNeilage et Davis (2001) le précisent, les consonnes occlusives et nasales sont favorisées largement par rapport aux consonnes avec une closure incomplète comme les fricatives et liquides, formes jugées plus difficiles à réaliser par l'enfant. Plus particulièrement trois types particuliers de formes CV tendent à prédominer :

- consonnes labiales [p b m], co-occurrentes avec des voyelles centrales de type schwa ;
- consonnes réalisées avec la partie antérieure, la lame de la langue, ou coronales [t d n], co-occurrentes avec des voyelles antérieures de type [æ] ;
- consonnes réalisées avec le dos de la langue ou dorsales [k g], co-occurrentes avec des voyelles d'arrières, de type [o].

MacNeilage et Davis soulignent ici le phénomène d'« inertie linguale », c'est-à-dire « un manque de mouvement actif de la langue dans la transition de consonne à voyelle »<sup>6</sup>.

Ils expliquent ainsi la dominance du cadre du contrôle moteur du babillage chez l'enfant de la façon suivante : le cycle mandibulaire d'ouverture-fermeture seul comme dans le patron labial-central appelé « cadre pur » (*pure frame*) s'opposent au patron coronal-avant (« cadre avant » ou « *fronted frame* ») qui correspond à un mouvement de langue vers l'avant avant le début de l'oscillation mandibulaire, la langue étant pré-disposée (*pre-setting*) ; ou s'opposant encore au patron dorsal-arrière (« cadre arrière » ou « *backed frame* ») de positionnement de la langue. Le côté rédupliqué des syllabes du babillage s'explique alors par la répétition pure du cadre sans aucun changement actif dans la position de la langue.

#### ***1.1.1.4. Différents « cadres purs » idiosyncrasiques***

Vilain et al. (2000) ont cherché à décrire l'acquisition de la parole, des syllabes de babillage canonique aux syllabes coarticulées, en se posant des questions sur les capacités des modèles computationnels de production de la parole. Pour cela, ils ont utilisé des modèles articulatoires, plus à même de mettre à disposition les degrés de liberté des processus articulatoires dans la parole, afin de mettre en valeur les contraintes et la variabilité dans la production. Leur méthode consiste à analyser les patrons adultes de coarticulation lors de la production de séquences VCV, de type [aba]. Cette phase d'acquisition du modèle leur a permis également d'extraire les contours du conduit vocal global, et de donner une vision précise des actions individuelles de chaque degré de liberté du conduit vocal (mandibule, lèvres, langue...). Les auteurs proposent que, lors de la production d'une séquence [aba], la configuration de la voyelle [a] réclame une action de la mandibule, mais réclame également une action des muscles linguaux. C'est en fait la production de la consonne qui met largement bien en évidence l'activation du muscle de la langue. Le contrôle de la voyelle [a] était supposé à la base ne nécessiter qu'une ouverture de la mandibule, et les auteurs ont observé que la production est bien plus complexe, recrutant la langue également. C'est la perturbation engendrée par le geste consonantique surimposé qui a pu révéler l'existence de cette commande motrice.

---

<sup>6</sup> « ... A lack of active tongue movement in the transition from consonant to vowel. » (MacNeilage et Davis, 2001:697)



De plus, l'hypothèse d'un *cadre pur labial* a été testée par Vilain et al (1999). Des études précédentes avaient montré que dans les productions de babillage, on trouvait indifféremment des closions labiales et coronales parmi les enfants. MacNeilage avait ajouté que la production bilabiale constituait le cadre pur ou « *pure frame* » dans la mesure où cette production était le produit direct de l'élévation de la mandibule. Les auteurs ont alors réalisé des simulations de babillage par modélisation articulatoire, afin de tester l'hypothèse de l'émergence des cadres de la parole. Ils ont utilisé trois modèles articulatoires anthropomorphiques (*Gentiane*, *Maeda*, *Bergame*) et basés sur des locuteurs français adultes, et également un modèle articulatoire de croissance du conduit vocal, le modèle *Growth*, de 0 à 21 ans. Ce dernier modèle possède « la capacité de simuler la maturation des dimensions du conduit vocal avec les conséquences acoustiques correspondantes »<sup>7</sup> (Vilain et al., 1999). Les simulations effectuées par Vilain et al. consistaient à bloquer tous les degrés de liberté du modèle articulatoire sauf le degré de liberté de la mandibule, qui correspond à l'apparition du cadre du babillage canonique chez l'enfant. Les trois modèles *Gentiane*, *Maeda* et *Bergame* ont donné, respectivement, une configuration *labiale* [baba], *coronale* [dada], et *labio-coronale* [bdabda]. Ces trois cadres obtenus par la seule commande mandibulaire montrent que les « *pure frames* » dépendent du locuteur, de son anatomie plus précisément, et peuvent très bien être de nature coronale, pas seulement labiale comme MacNeilage l'avait suggéré. Ainsi, le lieu d'articulation que l'on trouve dans tel ou tel cadre pur serait simplement lié aux caractéristiques morphologiques idiosyncrasiques de tel ou tel bébé.

Vilain et al. (1999), expliquent en outre la variation intra-individuelle de la façon suivante :

Cela pourrait expliquer une variation intra-individuelle dans le babillage de l'enfant, simplement due à une contrainte biologique, c'est-à-dire des approximations accidentelles dans le positionnement de la mâchoire et des articulateurs portés, sans présupposer un contrôle précoce des degrés de liberté du conduit vocal<sup>8</sup>. (1999 :2499)

<sup>7</sup> "... the capacity of simulating the maturation of the dimensions of the vocal tract with the corresponding acoustic consequences." (Vilain et al., 2002:2498).

<sup>8</sup> "This could explain an intra-individual variation in infant babbling, merely due to biological noise, i.e. accidental approximations in the positioning of the jaw and the carried articulators, without presupposing a precocious control of the degrees of freedom of the vocal tract" (Vilain et al., 1999:2499)

### ***1.1.1.5. Le contrôle neural du « cadre »***

D'un point de vue du contrôle moteur, MacNeilage propose que l'apparition du babillage corresponde à l'apparition du cadre porteur de la parole, c'est-à-dire du contrôle de la porteuse mandibulaire. A ce stade, le babillage consiste en des cadres rythmiques, totalement dépendants des oscillations du seul articulateur contrôlé, la mandibule. Ce cadre ne sera rempli que plus tard par un contenu segmental, au moment où les autres articulateurs portés seront à leur tour contrôlés. MacNeilage (1998) propose alors que ce contrôle des oscillations mandibulaires dépende d'un système d'initiation motrice, l'aire motrice supplémentaire (SMA). Il défend l'idée que la construction articuloire fondamentale de la parole réside dans le cycle mandibulaire, soutenue par le fait que la région du lobe frontal inférieur contenant l'aire de Broca, soit le locus cortical principal du contrôle de processus ingestifs chez les mammifères. Chez le singe, la communication vocale est contrôlée dans la région principale du cortex, dans le cortex cingulaire antérieur, sur la surface médiane de l'hémisphère. Si on stimule électriquement cette zone chez le singe, des vocalisations sont obtenues. D'autre part, si cette aire se trouvait endommagée, cela l'empêcherait alors d'émettre des cris volontairement. MacNeilage prend alors en considération le rôle de l'aire motrice supplémentaire (SMA), une aire prolongeant le cortex cingulaire antérieur et étroitement liée à celui-ci, dans la corticogénèse. Si on stimule électriquement cette aire, on obtient alors du sujet des productions involontaires de séquences de syllabes CV simples de type [dadada...] ou [lalala...], des palilalies. MacNeilage a recueilli des données de la littérature portant sur la stimulation corticale et les lésions irritatives de cette zone, et il a particulièrement mis en valeur le recueil de données de Jonas<sup>9</sup> rapportant ce phénomène. Ce qui a permis à MacNeilage de suggérer fortement que SMA soit impliquée dans la génération de ce cadre chez les humains modernes.

C'est une candidature neurale que soutiennent Abry et al. (2002). Dans leur étude, les auteurs soulignent le fait que de nombreux patients, qui ont une aphasie globale accompagnée de lésions périsylviennes sévères, arrivent encore à produire des séquences récurrentes de type CV. Ces études permettent également aux auteurs de s'opposer aux propositions soutenant le rôle prépondérant du cortex

---

<sup>9</sup> Jonas, S. (1981) The supplementary motor region and speech emission, *Journal of Communication Disorders*, 14 :349-73.

latéral (soit de Broca) dans le contrôle de la production de la parole, voire du babillage canonique, en faveur d'un contrôle de l'aire motrice supplémentaire SMA, soutenant ainsi la proposition de MacNeillage. Abry et al. repartent du patient Leborgne, rencontre heureuse de Broca. Ce dernier a pu effectuer un examen post-mortem du cerveau de ce patient, atteint d'une aphémie<sup>10</sup> datant de 21 ans. Ce patient ne répondait que par la monosyllabe 'tan' répétée deux fois de suite 'tantan', aux questions qu'on lui posait ; ce qui lui avait valu son sobriquet devenu célèbre de Tan. L'examen de son cerveau —scanné bien plus tard par Castaigne et al.<sup>11</sup>—a fait apparaître que sa lésion s'étendait largement autour des aires 44 et 45 de Brodmann. Mais personne ne s'est demandé comment dans cette aphasie le sujet pouvait produire encore 'tantan', puisque ce n'était plus semblait-il de la parole... Et tout un chacun étant naturellement frappé par la lésion périsylvienne gauche dans la zone réputée du langage, personne n'a noté que SMA était intacte... Ces données, ajoutées à celles recueillies par Jonas, renforcent l'idée qu'il existe un contrôle dans l'hémisphère gauche *non-latéral* pour ces syllabes rédupliquées. SMA gauche reste ainsi le candidat neural le plus plausible pour le chargement, l'initiation, et le *self-pacing* de la séquence motrice de base de la parole.

Abry et al. citent le plus grand nombre de données disponibles qui relèvent de cette aphasie qu'ils ont baptisée « aphasie du cadre » (ou *frame aphasia*, terme repris par l'éditeur-fondateur de la revue *Aphasiology*, Chris Code dans ses publications, dont Code, 2005). Ces dernières incitent à faire un rapprochement avec les données du babillage, qui comme on l'a vu précédemment, présentent une dominance de séquences labiales et coronales. Ces syllabes répétées montrent surtout la prédominance de consonnes dentales et alvéolaires (soient coronales). MacNeillage et Davis (2001) ont bien montré par ailleurs qu'on ne retrouvait pas dans cette aphasie de nos Tantan, les co-occurrences CV caractéristiques du babillage canonique ([dododo] est possible). Il faut se rappeler qu'on a affaire à des sujets qui ont déjà acquis la coarticulation ou coproduction CV, avec indépendance de C et V, que nous allons voir émerger (ci-dessous) chez l'enfant au bout d'un an... Et ils l'ont

---

<sup>10</sup> Aphémie : « Arrêt de la parole d'une durée relativement longue (2 à 7 jours). Trouble transitoire survenant habituellement à la suite d'une intervention chirurgicale intéressant l'aire motrice supplémentaire ou d'une thrombose de l'artère cérébrale antérieure » (Définition tirée de *Terminologie de neuropsychologie et de neurologie du comportement*. Recherche et réd. Louise Bérubé, 1991, p. 58)

<sup>11</sup> Castaigne P., Lhermitte F., Signoret J.L., Abelanet R. (1980). Description et étude scannographique du cerveau de Leborgne. La découverte de Broca. *Revue Neurologique*, 136, pp. 563-583.

préservée dans leur « aphasie du cadre » qui coproduit un train *harmonisé* de contacts et d'ouvertures stables.

Ainsi, cette comparaison entre productions monosyllabiques de patients aphasiques et productions de babillage d'enfants permet de mieux appréhender quelle est la structure neurale sous-jacente au cadre de MacNeilage. De Bleser et Poeck<sup>12</sup> s'étaient déjà posés cette question :

Rather than comparing nonpropositional sounds to frequency counts in normal propositional English, it would have been better to investigate the inventory of sounds occurring in recurring utterances in comparison to the early acquisition of language specific sounds in child language (Bleser et Poeck, 1983)

De plus, en reprenant les études scanner du grand aphasiologiste allemand Klaus Poeck<sup>13</sup>, on peut montrer que ses patients pouvaient produire des énoncés récurrents de type CV, de la même façon que Leborgne et ses 'tantan', sans le support du système latéral. Cette observation se pose donc en faveur d'un contrôle des séquences CV rédupliquées non latéral. Ziegler et al. (1997)<sup>14</sup> ont publié des données qui soutiennent SMA dans l'hémisphère gauche comme meilleur candidat.

Nous venons ainsi de voir qu'il existe une première étape fondamentale dans l'émergence de la parole chez l'enfant, sous la forme du babillage canonique vers 6-7 mois. Ce babillage correspond à l'apparition du cadre, soit au contrôle de la porteuse proximale, la mandibule, produisant les oscillations d'ouverture/fermeture du conduit vocal dues au rythme mandibulaire. Une fois le contrôle du cadre installé chez l'enfant, va alors émerger le contrôle des articulateurs portés ou effecteurs distaux, ce que MacNeilage nomme l'émergence du « contenu » segmental.

## **1.2. Le contrôle du « contenu » ou contrôle segmental**

Le contrôle plus fin des articulateurs que sont la langue et les lèvres va permettre à l'enfant de varier plus largement ses productions. L'enfant va acquérir

---

<sup>12</sup> De Bleser R., Poeck K. (1983). Comments on paper « Neurolinguistic analysis of recurrent utterance in aphasia » by C. Code (*Cortex* 18, 141-152, 1982). *Cortex*, 19, 259-260.

<sup>13</sup> Poeck K., De Bleser R., Graf Von Keyserlingk D. (1984). Neurolinguistic status and localization of lesion in aphasic patients with exclusively consonant-vowel recurring utterances. *Brain*, 107, pp. 199-217.

<sup>14</sup> Ziegler W., Kilian, B., Deger, K. (1997). The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia*, 35, 9, pp. 1197-1208

dans un premier temps le contrôle des contacts au niveau local, avant d'acquérir par la suite le contrôle des postures, c'est-à-dire un contrôle global au niveau syllabique.

### I.2.1. Vers une lèvre de moins en moins compliant...

La différence entre contrôle proximal et contrôle distal –ou entre contrôle de la porteuse mandibulaire et articulateurs portés comme lèvres et langue– est bien mise en évidence par l'exemple donné par Munhall et Jones (1998), dans la figure I.1 ci-dessous.

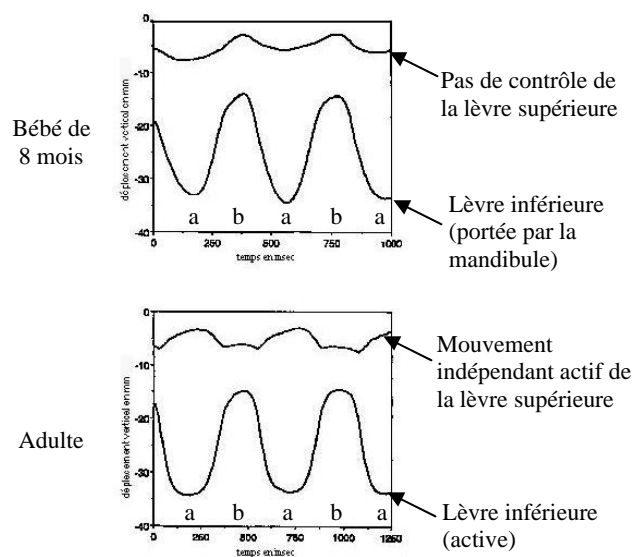


FIGURE I.1 : Déplacements verticaux des lèvres inférieure et supérieure, au cours de la production répétitive de [bababa] par un bébé de 8 mois et un adulte (Munhall et Jones, 1998 ; p. 525)

Munhall et Jones ont comparé les enregistrements des lèvres supérieure et inférieure d'un bébé de huit mois et d'un adulte, lors de la production de la séquence [bababa]. Ces données ont été obtenues via un système Optotrak, en plaçant des LEDs au milieu de la lèvre supérieure et à côté du bord vermillon de la lèvre inférieure. Sur le tracé du haut, les mouvements enregistrés pour la lèvre supérieure sont passifs, uniquement dus à la poussée de la lèvre inférieure, qui est elle-même portée par la mandibule. Le cycle alterne ouverture de la mâchoire qui correspond à la production de la voyelle, et fermeture de la mâchoire qui correspond à la production de la consonne, sans mouvement actif de la lèvre supérieure. Les auteurs soulignent le fait que ce patron de mouvement est cohérent avec l'idée de MacNeilage, selon laquelle le babillage implique uniquement le mouvement de la mandibule. Par comparaison, sur le tracé du bas, on voit chez l'adulte : (i) que la

lèvre supérieure résiste activement à la poussée de la lèvre inférieure, pour produire la closure de la consonne ; (ii) et que cette même lèvre supérieure se relève activement en coordination avec l'abaissement de la lèvre inférieure, pour produire l'ouverture de la voyelle.

Munhall et Jones mettent bien en évidence par ces tracés les différences de patrons de contrôle articulatoire entre un bébé au stade du babillage et une production adulte.

Cependant, dans ces données, nous ne disposons pas du tracé du mouvement mandibulaire. Ce mouvement est cependant bien présent au cours de ce cycle. C'est même le seul actif, puisque si l'on soustrayait le mouvement mandibulaire de celui de la lèvre inférieure, nous n'observerions plus aucun mouvement de cette lèvre, et nous venons de dire que la lèvre supérieure ne doit son mouvement qu'à la poussée de l'inférieure, elle-même portée par la mandibule.

On peut observer typiquement, dans le cours du développement, pour les mouvements de type [bababa], que la lèvre supérieure devient de moins en moins compliant à la poussée de la lèvre inférieure portée par la mandibule. Ce qui aboutira finalement à un contrôle des contacts autonome.

## **I.2.2. Un développement séquentiel pour le contrôle de la parole ?**

Les découvertes de Munhall et Jones soulèvent une autre question fondamentale, celle de l'ordre d'apparition des différents contrôles impliqués dans la parole. Ces auteurs ont pu démontrer que lors de l'émergence de la parole chez l'enfant, à l'étape du babillage canonique, le seul articulateur contrôlé est la mandibule, jouant comme la porteuse de la parole. Mais qu'en est-il des autres articulateurs ? Quels articulateurs vont développer leur contrôle autonome avant les autres ? Le développement de la production de parole s'étend sur une longue période, et il est donc intéressant de déterminer la chronologie des différentes étapes charnières dans son déroulement.

L'étude de Green et al. (2002) nous apporte certains éléments de réponse. Les chercheurs tentent de découvrir si le contrôle des différents articulateurs se déroule séquentiellement ou non, en gardant à l'esprit que « la formation des gestes

articulatoires est guidée par des contraintes biologiques efficaces »<sup>15</sup>. L'idée derrière leur expérience est la suivante :

For instance, if control over the lips, tongue, jaw, velum and larynx develops sequentially, young children would be obligated to rely on "best suited" or most developmentally advanced articulator(s), with the less developed articulators contributing to a large portion of behavioral instability." (Green et al., 2002, p.67)

Ils ont en fait comparé les patrons de mouvement des lèvres supérieure et inférieure et celui de la mandibule chez l'adulte, à ceux d'enfants de différents âges (1, 2 et 6 ans), afin de tester les différences entre articulateurs, et d'observer leurs évolutions propres. En accord avec la théorie « *Frame, then Content* » de MacNeilage, qui propose que les oscillations mandibulaires qui produisent les constrictions du conduit vocal soient le patron moteur pour la parole précoce, Green et al. cherchent à tester spécifiquement l'hypothèse d'une stabilité précoce des patrons de mouvements de la mandibule par rapport aux patrons de mouvements des lèvres, dans des productions de type [baba] [papa] ou [mama] –énoncés que l'on peut trouver spontanément dans les productions d'enfants de façon dominante. Les résultats observés montrent alors que chaque articulateur possède son propre schéma développemental, et que le patron de mouvement de la mâchoire arrive à maturation bien avant celui des lèvres supérieure et inférieure, chez les enfants de 1 et 2 ans. De façon intéressante, chez l'enfant de 2 ans, les coefficients pour la lèvre inférieure étaient significativement plus grands que ceux pour la lèvre supérieure. Les enfants de 6 ans produisaient, quant à eux, des mouvements articulatoires similaires à ceux des adultes. De plus, les patrons de mouvement de la lèvre supérieure étaient significativement plus stables chez ces enfants de 6 ans que chez les enfants de 2 ans. En ce qui concerne précisément l'évolution des lèvres, nous nous référons ici à l'étude de Green et al. (2000). Après avoir soustrait le mouvement de la lèvre inférieure de celui de la mandibule afin de dégager son mouvement propre, ils observent un mouvement indépendant de la lèvre inférieure autour de l'âge d'1 an chez l'enfant. Ils observent un mouvement propre de la lèvre inférieure qui descend lorsque la mâchoire remonte (une étrangeté que nous expliquons par le fait que le bébé n'ait pas encore de dents), tandis que la lèvre supérieure est passivement poussée par la lèvre inférieure (le contour de la lèvre supérieure est inversé sur leur figure). A l'âge de 2 ans cette dernière n'est toujours pas contrôlée par l'enfant dans cette étude. Il faudra attendre l'âge de 6 ans pour observer un

<sup>15</sup> « ...the formation of articulatory gesture is guided by potent biologic constraints » (Green et al., 2002, p.67)

patron plus stable et moins chaotique, indiquant un début de contrôle de celle-ci, avec toutefois une certaine variabilité qui indique que ce contrôle n'est pas encore totalement ajusté.

En d'autres termes, leurs observations nous permettent de confirmer que pendant les premières années de vie, la mandibule est en avance sur les lèvres, puisque les patrons de mouvement de la première sont quasi-similaires à ceux des adultes. En revanche, les patrons des lèvres, supérieure et inférieure, chez les enfants de 1 an ne ressemblent pas aux patrons observés chez les adultes. Ces résultats amènent Green et al. (2000) à soutenir que « le développement de la parole implique une intégration du mouvement des lèvres dans un patron de mouvement mandibulaire relativement bien établi »<sup>16</sup>. Les auteurs expliquent alors que la stabilité trouvée dans ces oscillations mandibulaires pourrait jouer le rôle d'un « équilibre prélinguistique » (*prelinguistic equilibrium*, p.76), les facteurs biomécaniques contraignant alors les mouvements oro-moteurs. Ces découvertes soutiennent leur hypothèse d'un développement séquentiel des différents contrôles articulatoires, avec la prédominance de la porteuse proximale avant les effecteurs distaux contrôlés plus tardivement chez l'enfant, comme il est de mise dans le contrôle moteur (« loi proximodistale », rappelée par J. Vauclair, Développement du jeune enfant. Motricité, perception, cognition, Belin, 2004, p. 65).

### **I.2.3. Quel développement du contrôle du velum pour la nasalité ?**

Outre le développement du contrôle des articulateurs tels que la mandibule, les lèvres et la langue, quel est le développement du contrôle du velum chez l'enfant ? Lalevée et Vilain (2003) ont montré dans une étude que le velum semble n'être quasiment pas contrôlé à sept mois chez l'enfant. Pour mieux appréhender le contrôle du velum l'étude de Rossato et al. (2003) chez l'adulte nous propose quelques principes généraux sur le contrôle du velum pour les consonnes et les voyelles. Rossato et al. (2003) ont étudié les distributions de la position du velum d'un sujet français, lors de la réalisation de séquences VCV, impliquant des consonnes orales et nasales, et des voyelles orales et nasales. Les mesures de ces distributions ont été obtenues à partir d'un articulographe électromagnétique, avec

---

<sup>16</sup> « ... speech development involves integrating lip movement into a relatively well-established mandibular movement pattern » (Green et al., 2002, p.75)



une bobine collée sur le vélum, ce qui permettait de récupérer les déplacements essentiellement verticaux du velum. La figure I.2 ci-dessous donne les histogrammes de hauteur de velum obtenus pour les quatre catégories.

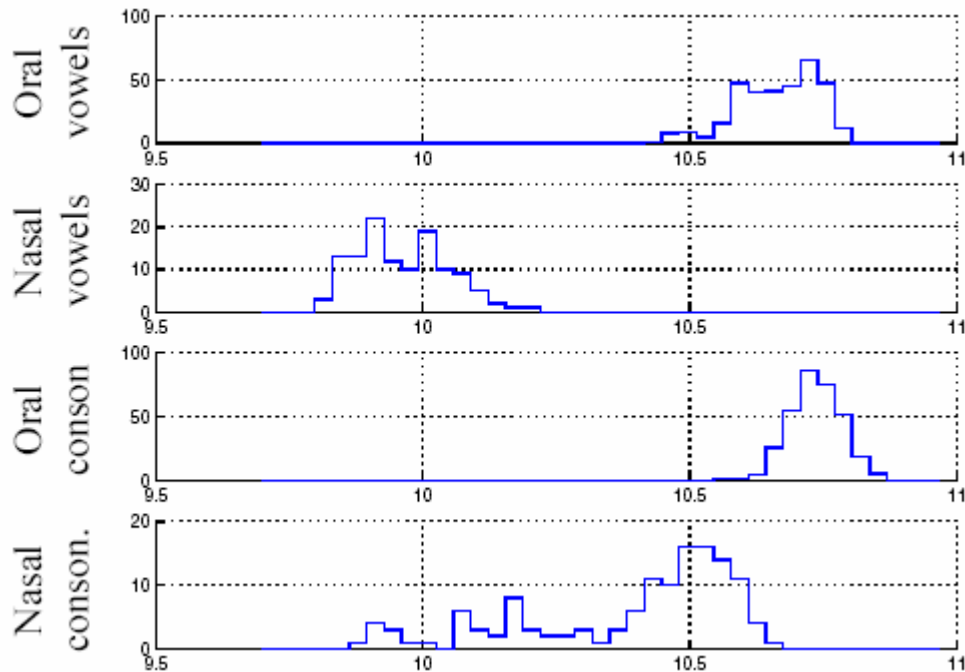


FIGURE I.2 : Distributions de la hauteur du velum pour les voyelles orales, nasales et les consonnes orales et nasales (Rossato et al., 2003, p. 3142)

Nous pouvons remarquer que les valeurs d'ouverture sont très différenciées pour les réalisations des voyelles. Une voyelle orale obtiendra des valeurs autour de 10.66 cm, tandis qu'une voyelle nasale verra ses valeurs se distribuer autour de 9.96 cm. Ainsi, les voyelles nasales réclament clairement une plus grande ouverture du conduit nasal. Les auteurs observent toutefois que la position du velum n'est pas toujours fermée pendant les voyelles orales, mais que celui-ci peut être au contraire légèrement abaissé, sans que pour autant la voyelle soit automatiquement perçue comme nasale (interviennent des questions d'impédance).

Lorsque l'on étudie la distribution des valeurs des consonnes, on remarque que les consonnes orales sont proches de celles des voyelles orales. Les consonnes nasales ont des valeurs variant largement : même avec un velum aussi haut que celui d'une voyelle orale, le rapport d'impédance entre les deux conduits fait que la colonne d'air sera mise en vibration dans les fosses nasales, même avec une ouverture vélopharyngée relativement petite, du moment que la fermeture du conduit oral est complète.

Ces données montrent qu'il existe une gamme de valeurs pour la hauteur du velum, comprise entre 10.4 cm et 10.65 cm, où voyelle orale et consonne nasale sont produites. Ainsi, les auteurs concluent que le velum ne doit pas nécessiter un contrôle très fin afin de produire une séquence redoublée comprenant consonnes nasales et voyelles orales, du type [mama]. C'est ce que Christian Abry a baptisé « jaw mama for free », pour expliquer l'universelle appellation de la maman : il suffit de bouger la mandibule pour produire [mamama] avec des contacts labiaux nasals (« cadre pur labial ») et des voyelles orales (pour des nasales : il faudrait baisser davantage, soit activement le velum ; de même [nanana] pour un enfant avec un cadre pur coronal).

Cette découverte revêt un caractère fondamental pour l'étude de la nasalité chez l'enfant, puisqu'elle expliquerait alors pourquoi les enfants sont capables de produire de telles séquences lors du babillage, période à laquelle ils ne sont pas sensés savoir coordonner leurs mouvements du velum. C'est précisément cette idée qui a été formulée par Lalevée (2003), permettant ainsi à l'auteur d'émettre des hypothèses quant au développement ontogénétique de cet articulateur.

Une petite ouverture du passage vélopharyngé n'entraînera donc pas de changements acoustiques radicaux pour une voyelle orale alors que dans le cas d'une consonne, la structure acoustique changera radicalement. Une séquence [mama] pourrait donc être produite avec une même position « par défaut », relâchée, soit non contrôlée du velum.

En résumé, on peut émettre l'hypothèse que dans les premières étapes du développement de la parole, la position par défaut du vélum serait celle d'une légère ouverture du passage vélopharyngé telle que pour une consonne nasale, alors que la position activement basse du velum pour les voyelles nasales impliquerait un contrôle complexe de celui-ci, qui se développerait plus tardivement dans l'ontogenèse. (Lalevée, 2003:25)

Lalevée (2003) a étudié l'évolution de la proportion des nasales chez une enfant, Célia. La proportion des consonnes nasales diminue régulièrement entre 6 et 11 mois, baissant de 25% à 10%. Cet enfant a soudain un pic de consonnes nasales à 12 mois, que l'auteur explique par l'apparition des premiers mots dans sa production, constitués principalement des mots « non » et « maman ».

Ces différentes études ont illustré l'évolution du contrôle des différents articulateurs engagés dans la naissance de la parole chez l'enfant. Clairement la mandibule est le premier articuleur contrôlé par l'enfant, bien avant que ne se mette en place l'indépendance des effecteurs distaux, tel que les lèvres inférieure et supérieure. Une fois le contrôle du cycle acquis dès 6-7 mois, et une fois le contrôle des contacts établi, nous arrivons à une étape charnière dans l'évolution de l'enfant, le contrôle postural des configurations vocaliques.

#### **I.2.4. Quel développement du contrôle des articulateurs dans la coproduction des consonnes et des voyelles ?**

Nous venons de voir, à travers ces études, le développement des contrôles des articulateurs pour les consonnes ou « closants » chez l'enfant, dans la phase de fermeture ou closance du cycle syllabique. Pour le développement des voyelles ou « vocants » chez l'enfant, il nous semble intéressant de commencer par une étude de Buhr (1980). Ces « vocants » font appel à des contrôles plus fins que ceux des contacts des premières consonnes, notamment pour le développement du contrôle de la langue et des lèvres dans la phase ouverte du cycle syllabique (sans parler des coordinations linguo- et labio-mandibulaires qui caractérisent une maîtrise des voyelles achevée bien plus tardivement).

Chez l'adulte, les voyelles sont produites grâce à la combinaison des contrôles de différents articulateurs, qui comprennent : (i) la posture de la langue dans la cavité orale qui par les constriction qu'elle peut produire dans la cavité buccale modifie la longueur effective du tractus vocal, mesurée du larynx à la sortie aux lèvres; (ii) la configuration des lèvres, qui va pouvoir allonger ou raccourcir la longueur du conduit vocal et modifier la taille de la sortie du conduit ; et (iii) la hauteur du larynx qui a aussi un effet sur cette longueur totale. On peut considérer que l'ensemble de ces contrôles pour la configuration des voyelles de la langue apprise aboutit à un contrôle de type postural.

Chez l'enfant, à la naissance, le conduit vocal se positionne différemment de celui de l'adulte. Ainsi le larynx se positionne afin que la partie supérieure de l'épiglotte soit opposée aux 2<sup>ème</sup> et 3<sup>ème</sup> cervicales. Ensuite, le larynx et le tiers postérieur de la langue vont descendre dans le cou, pour arriver, vers l'âge de 4-5

ans dans une configuration où la portion arrière de la langue devient la paroi antérieure du pharynx. Buhr pose ainsi deux facteurs pour le développement et l'intégration pour la production des voyelles : la forme changeante de la configuration du conduit vocal et le contrôle neuromusculaire de la langue, des lèvres et de la mâchoire. Pour étudier l'évolution de la production des voyelles chez l'enfant, l'auteur a suivi un enfant longitudinalement de 16 à 62 semaines. Les résultats montrent qu'à partir d'un abaissement uniforme des valeurs formantiques à travers le temps, on peut observer un allongement du conduit vocal de façon graduelle. A 16 semaines, on observe un chevauchement considérable parmi quasiment toutes les voyelles. A 24 semaines, un espace vocalique rudimentaire commence à émerger. A 41 semaines, l'espace vocalique est plus défini, bien que certains chevauchements persistent. Enfin, à 62 semaines, l'étendue des voyelles individuelles se rapproche sauf pour les voyelles [ɛ] et [i]. Selon l'auteur, il semblerait que l'enfant persiste dans le babillage jusqu'au moment où il commence à produire ses premiers mots. L'espace vocalique va se développer ainsi petit à petit jusqu'aux premiers mots, et l'auteur suggère même que son développement puisse être une condition nécessaire pour la production des mots. Buhr a pu observer que lors des premières semaines, l'enfant produit beaucoup les voyelles [i] [e] et [ɛ], ce qui biomécaniquement nous paraît normal, puisque le principal articulateur requis pour ces voyelles est la mâchoire, premier articulateur contrôlé par l'enfant. La voyelle [u] semble survenir plus tard puisqu'elle réclame une coordination des lèvres, de la langue et de la mâchoire. L'auteur propose également d'expliquer ce retard par le fait que l'orbicularis oris, muscle facial permettant la protrusion entre autre, ne soit pas encore totalement développé plus tôt. Ainsi il semblerait que l'enfant doive acquérir les facultés neuromusculaires nécessaires de la langue, de la mâchoire et des lèvres, afin de moduler la forme de la cavité orale. Les conséquences de ce développement se retrouvent dans la remarque que la partie avant du triangle vocalique semble précéder la partie arrière. Le triangle vocalique semble se développer pendant le babillage, et son émergence et sa stabilité semblent dépendre largement du développement neuromusculaire et anatomique du conduit vocal pendant les premières années de vie. Le fait de trouver des occurrences comme [a], [i] et [u] à 17, 18 et 24 semaines respectivement, indique la nécessité d'une restructuration du conduit vocal assez précocement dans l'enfance, avec des formes

qui vont se stabiliser vers 36 semaines pour les deux premières et un peu plus tard pour le [u].

A l'âge de 13 mois naît la coarticulation dans le babillage canonique, en même temps que naissent les premiers mots eux aussi coarticulés d'après les données acoustiques de Sussman et al. (1999, cf. aussi Sussman et al., 1996), que nous allons exposer ci-dessous.

Du point de vue articulatoire, chez l'adulte, et selon Vilain (2000), dans une séquence [aba] (cf. figure I.3), la langue en montant de [a] vers [b], ne monte pas autant qu'elle monterait si elle était passivement portée par la mandibule, laquelle porte la lèvre inférieure à la rencontre de la lèvre supérieure pour effectuer la closure de [b]. Autrement dit, la langue résiste par un contrôle musculaire (celui de l'hyoglosse) à la montée mandibulaire. C'est en fait pour nous la véritable définition du contrôle de la coarticulation ou coproduction (pour une revue, cf. Bonnot, 1990) : comment co-produire un [a], avec la langue suffisamment basse, dans un [b], soit préparer une ouverture de la bouche, cependant que la bouche est fermée par la production du [b]. La possibilité de préparer la position ouverte du [a] pendant la fermeture permettra d'avoir des formants caractéristiques de la voyelle qui seront atteints très vite après l'ouverture du conduit vocal, sans passer par un intermédiaire de type schwa.

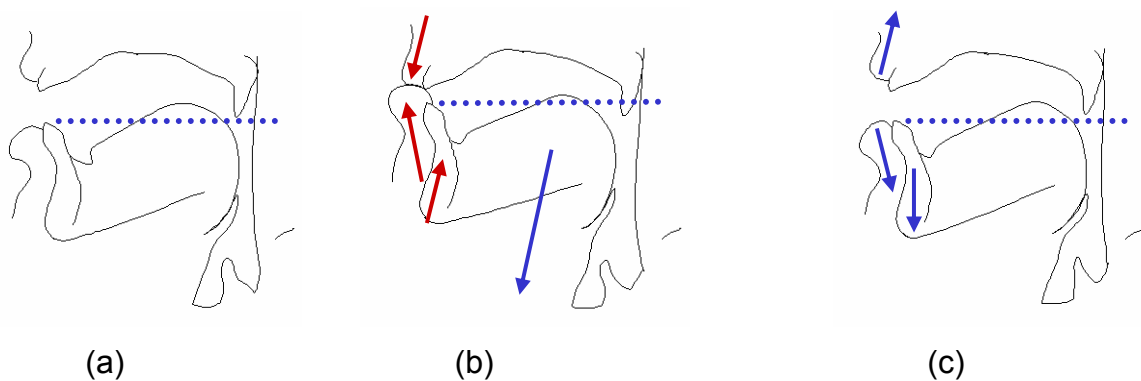


FIGURE I.3 : Radiocinématographie d'une séquence [aba] chez un adulte (tracés aux centres des réalisations). On remarquera en (a) qu'à partir d'une posture basse de la langue (dont le dos émerge au-dessus de l'horizontale pointillée, positionnée sur les incisives de la mandibule comme repère), la langue va monter, transportée par la mandibule, laquelle est recrutée par la coordination labio-mandibulaire (flèches en b pour la closure, en c pour la détente). Mais si le dos est plus haut, plus près du plafond de la bouche en (b), il est plus bas par rapport aux incisives qu'en (a) et (c). L'action du muscle hyoglosse a abaissé le corps de la langue (flèche vers le bas en b) pendant que la mandibule montait pour aider à la closure bilabiale. (D'après Vilain, 2000).

Sussman et al. (1999) apportent quelques éléments de compréhension de ce phénomène, dans leur étude sur le développement de la coarticulation d'un enfant suivi par Barbara Davis de 7 mois à 40 mois. Ils s'intéressent particulièrement au moment où les enfants commencent à établir un contrôle moteur différentiel pour les gestes concernant l'occlusion d'une consonne, tandis que simultanément (en coproduction) ils configurent le corps de la langue afin de former le geste vocalique. Pour étudier le développement de la coarticulation chez l'enfant à partir du signal acoustique (des trajectoires des formants), les auteurs ont utilisé les pentes de l'équation du locus. Il s'agit d'une régression linéaire obtenue à partir des fréquences de transition du deuxième formant (F2), mesurées en début et en milieu de voyelle, voyelle précédée d'une consonne occlusive, qui peut être de lieu labial, alvéolaire ou vélaire. Les pentes d'équation du locus varient comme une fonction directe du lieu d'articulation de l'occlusion, et constituent un index basé quantitativement pour évaluer le degré de coarticulation CV (Cf. Figure I.4). Plus les pentes sont fortes, plus grand est le degré de coarticulation anticipatoire. Inversement, plus la pente est faible et plus le chevauchement coarticulatoire est réduit.

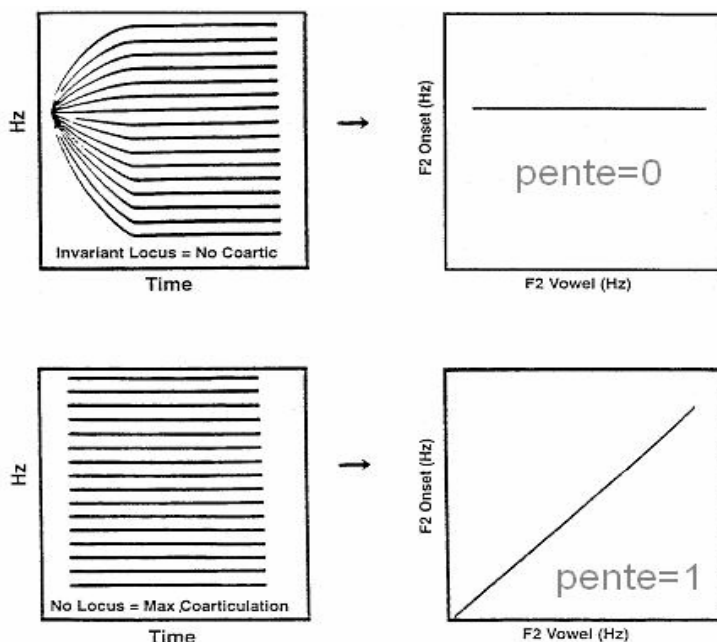


FIGURE I.4 : Extrêmes théoriques des pentes d'équation du locus. Les figures du haut illustrent la représentation de la transition F2 lorsqu'il n'y a aucune coarticulation entre la voyelle et la consonne (le locus, point de convergence des transitions qui n'est pas influencé par le contexte vocalique sur les F2), et la pente zéro qui devrait en résulter. Les figures du bas illustre la coarticulation maximale entre voyelle et consonne avec aucun locus consonantique fixe (celui-ci étant entièrement dépendant du contexte vocalique de F2) et une pente d'équation résultante de 1. (D'après Sussman et al., 1999).

L'étude des loci donne accès indirectement à l'évolution des patrons de coarticulation chez l'enfant. Dans le cas de l'occlusive labiale (sur la figure I.5 ; nous ne retiendrons que ce lieu qui permet de comparer clairement dans ce travail babillage et premiers mots), de 7 à 10 mois, les pentes du babillage canonique sont en dessous du niveau adulte, augmentant progressivement de 11 à 13 mois, et se rangeant autour la norme adulte de 13 à 16 mois. Ce qui suggère que l'enfant commence à contrôler les configurations de sa langue pour réaliser la partie vocalique de la syllabe, pendant le contrôle des lèvres.

Mais ce qui est encore plus intéressant, c'est de constater qu'à 13 mois de babillage les premiers mots sont là et avec une pente qui commence nettement plus haute que là où a commencé le babillage. Cette pente de l'équation du locus va se maintenir relativement haute jusqu'à la fin du suivi après 3 ans, le babillage canonique coarticulé ayant disparu à 16 mois.

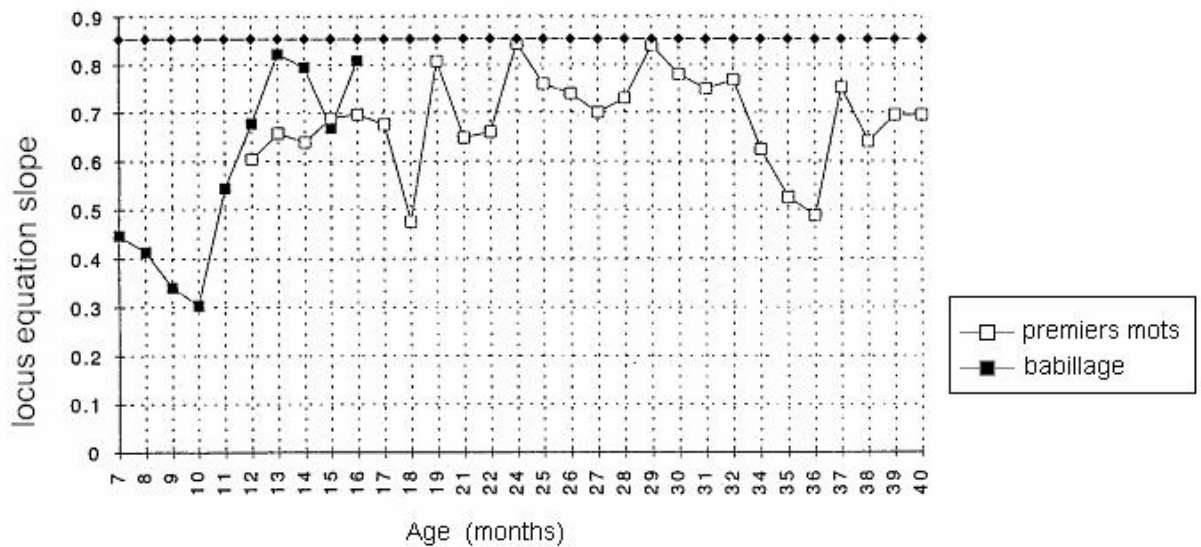


FIGURE I.5 : Pentas d'équation du locus pour des production contenant [bV] d'une enfant suivie de 7 à 40 mois. Les valeurs des énoncés de babillage (présents jusqu'à 16 mois) sont les carrés noirs, les mots (apparaissant à 12 mois), les carrés blancs, et la norme adulte pour la parole spontanée est donnée par la ligne des petits losanges noirs. (D'après Sussman et al., 1999).

### I.3. Conclusion

Au cours de cette première partie, notre volonté a été de mettre en évidence plusieurs mécanismes nécessaires à la mise en place de la parole chez l'enfant, des mécanismes qui sont plutôt peu connus des recherches linguistiques qui postulent

des principes et s'intéressent peu à cette question de savoir comment la phonologie de la syllabe s'ancre dans ses contrôles moteurs. C'est ce que nous avons tracé depuis le contrôle du rythme du babillage canonique jusqu'à la coarticulation dans le premier mot, en passant du contrôle du cadre rythmique à celui du contenu segmental.

Nous voulons tirer profit de cette coïncidence surprenante, de ce rendez-vous développemental entre le phénomène de la coarticulation et les premiers mots, sans pour autant que notre explication du moment aille au-delà d'une dominance du cadre qui fournirait comme bénéfice gratuit une harmonie vocalique et consonantique sur tout le gabarit du mot.

Nous allons, dans les parties suivantes, nous interroger sur ces phénomènes que nous annonçons en introduction, ceux qui doivent nous permettre de lier l'émergence du contrôle de la parole au contrôle du signe. Pour cela, il paraît nécessaire, tout comme nous venons d'énoncer les étapes fondamentales pour la mise en place de la parole, de traiter maintenant des propositions que nous pouvons emprunter, aménager, synthétiser, sur les mécanismes pertinents pour le cadre du signe, mécanismes qui vont permettre au bout du compte à l'enfant d'accéder à la sémantique de sa langue ambiante.



## SECONDE PARTIE: LE CADRE DU SIGNE

Lune ?



S'agissant du contrôle de la production des premiers mots chez l'enfant, nous aurons avantage à ne pas le distinguer dans un premier temps du contrôle d'un autre signe, le signe de la langue des signes. Dans la recherche de ce cadre du signe, il nous semble nécessaire d'intégrer les principaux mécanismes utiles au contrôle de la sémantique des premiers mots comme des premiers signes (Cf. Figure B.1, flux de droite).

Le système audiovisuel humain mérite d'être étudié dans un premier temps, afin de poser les principes de base essentiels pour créer une analyse de scène audiovisuelle robuste, qui prendra en considération les capacités et habiletés cognitives de la communication naïve ou intuitive. Il convient également d'adopter une position développementale afin d'expliquer la naissance de ces mécanismes fondamentaux qui se développent très tôt dans l'enfance, pour les intégrer par la suite dans les systèmes utilisant la communication interactive face-à-face. En tenant compte de la Théorie du Contrôle, nous pouvons présager de la naissance d'une biocybernétique de la communication face-à-face intuitive compatible avec les cerveaux des agents communicants (ou les comportements infographiques ou mieux robotiques créant pour nos cerveaux des illusions partielles mais prégnantes d'agentivité).

## **II.1. Le système attentionnel : orientation et alerte**

### **II.1.1. Vision et attention**

Le terme « attention visuelle » fait référence à plusieurs aspects liés à la vision, et aux procédés que cette dernière utilise pour définir les traits pertinents dans l'environnement. L'attention va sélectionner un panel d'informations qui vont être dirigées au cerveau et qui seront potentiellement disponibles pour une action ultérieure. Dans leur article, Harris et Jenkin (2000), recensent les différents types d'attention, tels que l'attention sélective, l'attention analytique, l'attention dirigée ou encore l'attention de vigilance. Une aire cérébrale particulière semble être activée exclusivement lorsque l'attention est requise, il s'agit du cortex cingulaire. Cette aire constituerait donc un lieu de contrôle exécutif, qui ne serait plus activée dès lors qu'une tâche serait automatisée et ne nécessiterait plus une attention accrue.

L'attention peut être considérée comme l'outil d'une conscience sélective ou, en termes encore plus utilitaires, elle est l'ensemble des processus de haut niveau qui guident la perception. Une façon de mieux appréhender le rôle de l'attention dans la vision (sans entrer dans la pathologie de l'héminégligence), et son pouvoir sélectif dans l'exploration d'une scène, est sans aucun doute le phénomène de « cécité au changement » (*change blindness*). Pour prendre un exemple simple, si on introduit une distraction mineure ou une pause dans l'attention, et qu'à ce moment là on modifie la scène visuelle (on enlève un élément du décor même très évident), on observe que le changement est difficilement repérable par le sujet testé, voire même impossible à percevoir (plusieurs sites vous font jouer à ce jeu de la différence en cliquant sur deux images successives).

Le rôle de la vision dans le mécanisme d'attention semble donc prépondérant, et a il été traité largement dans la littérature. Pourtant il ne faut pas pour autant négliger le rôle des autres capteurs sensoriels pour l'attention, comme nous allons le voir dans la partie qui suit.

### **II.1.2. Notre première orientation d'alerte**

La première orientation dont nous héritons se fait par le bruit, il s'agit d'un réflexe d'orientation bien développé. En effet, le système attentionnel semble naître de cette première modalité auditive. Lorsqu'une porte claque à l'intérieur d'une pièce où se trouve un bébé, ce dernier sursaute face à cet événement brutal... même in utero. Dès son premier jour de vie, dix minutes après sa naissance, le bébé oriente sa tête pour localiser un bruit qui lui parvient, en général celui de sa maman qui lui parle. C'est bien la preuve que les êtres humains bénéficient d'un système de traitement de signal efficace, et qui pourtant, au niveau du nombre de capteurs externes, n'est semble-t-il pas d'une grande complexité. L'audition humaine est en fait tributaire de deux pavillons fixes, non-orientables individuellement et qui sont portés couplés par le système céphalo-moteur.

### **II.1.3. Un handicap avantageux**

Il pourrait paraître en effet handicapant pour un être humain de ne posséder que deux oreilles, alors qu'il semblerait beaucoup plus avantageux de posséder une rangée de microphones (d'oreilles), disposée tout autour de la tête comme le montre

l'illustration suivante (Figure II.1), favorite en visio-conférence (tirée du site Medialab du MIT pour illustrer l'environnement de télécommunication Handy 21).

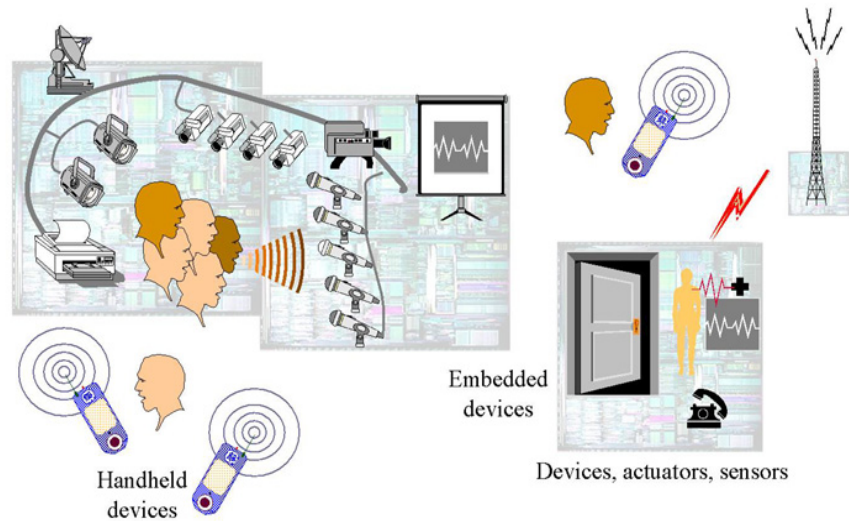


FIGURE II.1 : Une tête entourée de micros et de caméras (site Medialab du MIT : Projet Oxygen)

Dans ce cas-là nous serions certes « omniscients » mais encombrés d'une surcharge de signaux. Avec ce type de système auditif, le sujet humain, en présence d'une ou plusieurs sources, serait amené à traiter une multitude de signaux et ne serait pas capable de sélectionner-localiser le signal pertinent au milieu des autres. En fait, chaque source nous est connue par deux signaux, un par oreille. L'entrée sensorielle est certes réduite (deux oreilles), mais l'efficacité est ailleurs. Car le sujet va bouger activement sa tête pour mettre en phase, à quelques microsecondes près, les deux signaux qui lui parviennent. Loin d'être passive, cette perception est une perception active, comme l'est l'*active vision* en robotique, ce qui met d'emblée au premier plan le couplage perception-action dont bénéficie notre système auditif, et aussi visuel (ne serait-ce que par les actions motrices des muscles oculaires, de direction, vergence et accommodation). Avec nos seules deux oreilles, nous pouvons clairement parler d'un « handicap avantageux » : dans la mesure où les signaux ne nous arrivent pas simultanément aux deux oreilles, la tête s'oriente du côté où le signal a la plus grande amplitude pour localiser la source, amenant les deux signaux en phase dès les tympans. Notre système possède donc seulement deux oreilles mais se trouve être extrêmement simple en traitement du signal : il suffit de tourner la tête vers la source, pour obtenir la coïncidence de deux signaux de sensations différentes en amplitude et en phase pour nous placer en face d'une source unique,

en face d'un objet unique dans le monde. Il n'y a sans doute pas liage (*binding*) plus simple.

#### **II.1.4. Le principe de la coordination en perception active**

La vision de l'enfant va très vite rattraper et surpasser la modalité auditive. Comme nous bénéficions de deux oreilles, nous bénéficions également de deux yeux, et non pas d'un ensemble de caméras disposées autour de la tête. Les yeux vont également se positionner en face de la source par la coordination oculo-céphalo-motrice, qui est intégrée dans le système nerveux (Prablanc et Pelisson, 1990).

Le réflexe d'évitement illustre parfaitement ces mécanismes. Si l'on est en situation de recevoir un projectile, le déroulement des actions va se passer de la manière suivante. L'alerte est d'abord auditive si l'objet est audible (*looming* ou effet Doppler), déclenchant l'orientation oto-céphalo-motrice. L'œil, porté par le système oculo-céphalo-moteur, va bénéficier de cette coordination en récupérant d'abord les informations par la vision périphérique, ce qui lui permet de « sentir visuellement » qu'un objet approche (il existe des neurones bimodaux visuels et tactiles pour cela). Enfin un dernier mécanisme entre en jeu, il s'agit du réflexe de projeter le bras pour se protéger, geste accompagné généralement de l'émission d'une vocalisation. Ces systèmes se déclenchent quasi simultanément, la commande étant parallèle : sur l'oreille (la tête), l'œil (le système oculo-moteur), le bras (le système brachio-manuel) et la voix (le système oro-laryngé), en tenant compte des constantes de temps différentes de ces systèmes dans un déroulement de temporisation. Toutes ces actions servent en fait à intégrer les sensations grâce au principe de coordination, comme l'illustre la célèbre illusion d'Aristote. Ce principe est clairement illustré par la coordination bimanuelle : quand je porte un vase de fleurs plein d'eau des deux mains, je perçois un vase et non deux moitiés de vase ou dix sensations de mes dix doigts. De même, quand j'ai deux signaux identiques en phase dans les oreilles, je n'en perçois qu'un en face de ma face, même si je sais que ces signaux viennent des oreillettes de mon casque. Cette illusion irrépressible provient de mes capacités de perception active.

Nous venons de voir que notre système attentionnel est un outil puissant, et il est multimodal pour l'analyse de la scène audiovisuelle de la communication parlée.

Ce système joue évidemment un rôle très important dans le développement de l'enfant, puisqu'il va permettre à ce dernier de porter attention à des choses qui l'entourent dans son environnement, d'abord par l'œil, puis par l'index, dans son cours développemental, afin d'établir du sens sur les objets/agents/événements d'intérêt. Nous allons maintenant voir à quelles propriétés particulières l'attention de l'enfant est le plus sensible dans son environnement, et comment se gère l'indexation des ces focus d'intérêts.

## II.2. L'indexation des objets dans le monde

### II.2.1. La notion d' « objectitude »

Nous nous situons toujours dans une optique développementale, et c'est donc logiquement que nous allons nous intéresser à l'émergence de la notion d'objet chez l'enfant, afin de découvrir les mécanismes « infantiles » qui sont toujours les premiers « réflexes » activés chez l'adulte. De nombreuses études, qui ont été menées sur ce sujet, établissent à ce propos un parallèle entre le concept d'objet chez l'enfant, et l'attention orientée sur l'objet chez l'adulte. Ces deux littératures sont concernées par des problèmes parallèles, mais ne sont pas pour autant toujours en interaction. Elles s'intéressent aux bases de l'individuation de l'objet et à l'identité numérique. D'après l'article de Leslie et al. (1998), notre modèle de représentation de l'objet résiderait dans la notion-clé d'*indexing* (indexation). Il s'agit d'un mécanisme d'attention sélective, le fait de pointer un objet physique dans un lieu. Dès que le pointer visuel est acquis chez l'enfant, la permanence de l'objet pourra exister et on pourra lui rattacher des informations spécifiques.

La notion d' « objectitude » (*objecthood*), est un concept essentiel, présent dès les premières semaines de vie –déjà à 2 mois selon Cohen et Cashon (2001). Dès les premières années de la vie d'un enfant, le système attentionnel fait davantage que simplement focaliser sur les objets du monde physique. Il les sélectionne en nombre restreint : l'attention ne s'exerce que sur un tout petit ensemble d'objets simultanément. L'attention permettra plus tard de récupérer des propriétés sur ces objets en même temps. Notre attention première ne peut être attirée que par trois ou quatre index au maximum. L'indexation de l'objet est considérée comme le mécanisme qui limite l'énumération de l'enfant. Cognitivement,

le bébé a la capacité de suivre plusieurs objets simultanément dans une même scène, en assignant à chaque objet un « pointeur » appartenant à un ensemble limité d'index mentaux. Le fait d'apposer un index sur un objet ne véhicule pas les informations relatives aux propriétés de l'objet, ces propriétés devront être liées plus tard à ce dernier.

Cette notion d'objectitude constitue, selon Leslie et al. (1998), une structure tout à fait fondamentale de la pensée humaine, et nous nous en inspirons car elle peut nous fournir une base pour l'apprentissage sémantique du langage. L'attention se focalise sur les objets du monde physique, et ceux-ci vont en retour structurer l'attention visuelle. Les objets vont être transformés en représentations internes, qui vont permettre d'attribuer à chaque objet un index relié à un objet du monde physique. Cette fonction d'indexation en tant que telle sert à établir un *mapping* direct entre représentation interne et objet du monde réel, base possible du *fast mapping* lexical.

Rappelons qu'en aucun cas elle ne va véhiculer initialement des informations de propriétés sur l'objet en question. En effet, comme le soulignent Cohen et Cashon (2001), les jeunes enfants perçoivent en premier lieu la permanence d'une « chose », bien avant de pouvoir individuer plusieurs objets dans une scène par leurs différentes propriétés de forme ou de couleur. Cette dernière capacité (qui semble survenir entre 4 et 7 mois selon ces chercheurs) peut être retardée par l'incapacité de l'enfant à intégrer l'information de la permanence de l'objet avec l'information de ses propriétés. Les enfants progresseraient alors développementalement à travers une série de niveaux ou paliers de traitement de l'information. Mais dès les premiers mois le suivi et la ségrégation d'objets (dans les tâches de MOT ou *Multiple Object Tracking*) requiert simplement que les enfants perçoivent ou comprennent que les items sont des objets permanents qui peuvent être suivis dans leurs déplacements, sans qu'ils distinguent également ces objets sur la base de leurs différentes propriétés perceptives.

## II.2.2. L'indexation chez l'adulte et l'enfant

Cette indexation, des objets, des événements, des agents, qui semblent être les mieux reconnus des objets chez les enfants, constitue en outre une théorie permettant d'ancrer la capacité de suivi de une à trois cibles (« proies » ou

« prédateurs ») dans l'*analyse de scène*, comme le proposent entre autres Leslie et al. (1998). Selon nos auteurs, l'indexation de l'objet entraîne un mécanisme d'attention sélective, qui requiert un nombre de ressources limité dans l'analyse de scène. Ils suggèrent alors que le mécanisme d'indexation de l'objet soit limité à 3-4 objets. Chez l'enfant particulièrement, cette indexation pose les fondations du concept d'objet, avec la capacité d'attribuer des index ou pointeurs aux objets que l'enfant rencontre. Progressivement, l'enfant pourra alors affecter à l'index désiré, des informations supplémentaires, comme des informations sur les propriétés de celui-ci.

Selon Leslie et al. (1998, cf. aussi Tremoulet al., 2000), un « index-objet » ne fait référence qu'à un seul objet, et une fois assigné, peut le suivre dans différentes localisations spatiales. L'enfant ne pourra établir des index distincts que si chaque objet possède une localisation distincte dans la scène. Et les index-objets ayant la propriété d'être en ensemble limité, ils devront donc être « dé-assignés » avant d'être ré-assignés à un nouvel objet. Ce système d'indexation ne permet ainsi qu'à un nombre restreint d'objets d'être assignés (suivis) dans l'analyse de scène. Contrairement à Piaget, qui pensait que l'enfant n'était pas capable d'indexer un objet s'il ne le percevait pas en présence, Leslie et al. (1998) argumentent que les index permettent de garder l'assignation même si l'objet est en mouvement, et même s'il est occulté. Dans ce dernier cas, les enfants pointeront alors sur une localisation proche derrière ce qui obstrue l'objet. Le suivi de l'objet, dont nous parlerons plus tard, semble ici complètement nécessaire à l'index pour continuer à traquer l'objet derrière son occultation, aidé par des principes de trajectoire de physique intuitive.

Ce suivi est réalisé via le système *Where* qui détecte la permanence de l'objet et suit sa trajectoire, opposé au système *What*, qui est un système permettant à l'enfant d'individuer et identifier des objets par les informations de traits. Nous reparlerons en détail de ces systèmes dans la sous-partie suivante. Retenons simplement ici que l'information spatiotemporelle est fondamentale pour l'indexation des objets/agents/événements. En revenant à cet ensemble limité d'objets indexés dans une scène, Leslie et al. (1998) proposent, avec le soutien d'études sur les temps de fixation, que l'enfant puisse détecter clairement une différence entre des ensembles de 2 et 3 items, et détecte également la différence entre 3 et 4 objets.



Toujours selon les auteurs, il semblerait que la numérosité soit acquise par l'assignation des index aux objets, c'est-à-dire une assignation par localisation.

Hauser et al. (2002) ajoutent, après d'autres auteurs comme Stanislas Dehaene, que la discrimination des nombres est limitée dans le système disponible à l'enfant, suivant la loi de Weber : nous observons une plus grande discriminabilité parmi de petits nombres que parmi de grands nombres, et plus entre les nombres qui sont le plus éloignés (par exemple discriminer 7 vs. 8 est considéré comme plus difficile que 7 vs. 12). Cette sensibilité approximative au nombre (numérosité) est couplée à un mécanisme plus précis, mais qui est limité à des valeurs inférieures à 4. Ce système permet, lui, de distinguer précisément 1 de 2, 2 de 3 et 3 de 4, et il semble être recruté dans le contexte de suivi indexical de l'objet, subissant les contraintes de la mémoire de travail.

### **II.2.3. Indexation dans le cadre robotique**

Cette notion d'indexation de l'objet sur la base de la localisation n'est pas sans nous rappeler également les théories récentes de mécanisme d'attention visuelle basées sur l'objet, dans un cadre orienté robotique. Dans une conception du tout-déictique en perception-action-cognition, Pylyshyn (2000) s'est posé la question de l'inadéquation de certaines formes de représentation pour la modélisation d'agents tels que les humains ou les robots. Selon l'auteur, certaines théories ne cherchent pas à établir une forme de connexion préconceptuelle directe entre les objets du monde physique visuel, et les liens qu'ils établissent au sein du système visuel. Cette connexion souffre également de terminologies diverses, nous trouvons la notion de référence chez les philosophes, d'indexicaux et démonstratifs pour les sémanticiens, voire de pointeurs déictiques chez les théoriciens cognitivistes. Pylyshyn préfère voir cette connexion sous le terme d'index visuel, selon lui plus souvent usité, et il argumente que la théorie de l'indexation offre une synthèse qui fournit des implications importantes pour expliquer un grand nombre de découvertes psychophysiques.

Les représentations descriptives courantes, en propriétés des objets, échouent, selon Pylyshyn, à prendre en considération les relations indexicales. Ces relations sont dépendantes du contexte et sont critiques pour déterminer certains types d'action. En Sciences Cognitives, on se réfère de plus en plus à une cognition

située. C'est un effort pour diminuer le rôle, et surtout la taille, des représentations au profit d'une intelligence active avec le mot d'ordre courant dans une robotique « à la Brooks »: « before mapping the world, first move... with 3 indexes ». Cette théorie, on le voit, est étroitement liée au besoin de référence indexicale. Certains chercheurs pensent ainsi que l'environnement est utilisé simplement comme une extension de la mémoire de travail, car les gens ne stockent pas tout ce qu'ils voient en mémoire lorsqu'ils agissent dans une scène.

Prenons l'exemple proposé par Pylyshyn concernant les trois formes de pensée (représentation) possibles chez un robot. La question principale est de savoir quelles représentations de connaissances doit posséder ce robot afin de réaliser des actions dans le monde réel. La première forme de représentation est le formalisme logique de l'intelligence artificielle, une suite de commandes de navigation. Il se trouve être inefficace puisqu'il ne permet d'avoir une référence aux items individuels que par leurs propriétés (formes, couleurs, etc.). Ce qui est trop contraignant pour des actions rapides (de type réflexe d'évitement d'un objet qui vous vient dessus en loomant auditivement ou qui croît soudain sur votre rétine), car l'action doit être trop détaillée et on ne peut rentrer toute la carte infiniment détaillée d'un monde même limité avant de faire exécuter le moindre mouvement au robot dans une pièce. Une autre forme de représentation est le modèle internalisé du monde visuel (photographie de toute la pièce sous tous ses angles), qui a les mêmes inconvénients d'incomplétude que le formalisme logique. On a donc besoin d'une troisième forme de connaissances qui se définit par l'utilisation des indexicaux : le robot a en fait simplement besoin pour naviguer, d'un moyen de diriger son attention sur des objets individuels (des obstacles, un passage, etc.). Voici l'illustration (Figure II.2) de ces trois « formes de pensée » du robot :

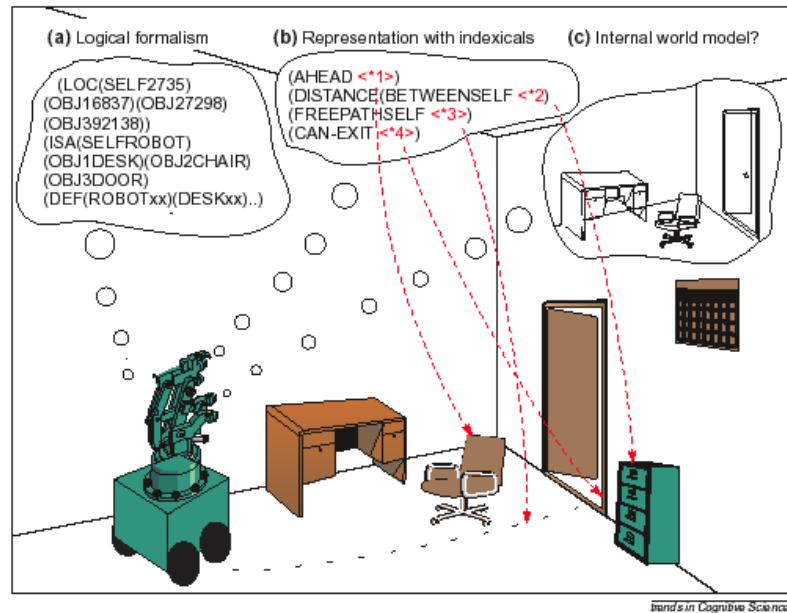


FIGURE II.2 : Trois façons différentes par lesquelles le robot peut se représenter son monde (D'après Pylyshyn, 2000).

L'exemple des robots permet de démontrer que moins de computation est nécessaire si les pointeurs aux objets dans la scène sont utilisés comme une partie de la représentation, car cela permet aux objets pertinents d'être sélectionnés directement. La référence démonstrative est très importante en robotique où la vision doit se connecter avec les actions. Pour le système visuel en analyse de scène la seule façon d'accrocher une nouvelle propriété est avant tout de la lier à une objectitude spatiale. Lorsque les yeux explorent une scène, très peu d'informations sont retenues d'une fixation à une autre. Les changements dans une scène sont rarement mis à jour à moins que l'attention soit focalisée sur l'objet qui change. La solution réside dans un pointeur, un pointeur liant une représentation d'un objet à un objet réel dans la scène, un pointeur qui pourra agir comme une référence démonstrative.

Pylyshyn (2000) propose donc qu'un pointeur appelé index visuel (*FINST* = *FIN*ger of *INST*antiation) désigne le fait de pointer un objet. Le système visuel possède certaines façons de sélectionner ou d'individuer un petit nombre d'éléments visuels marqués sans reposer uniquement sur le codage détaillé de chacun avec ses propriétés. Et c'est bien ce que propose la théorie de l'indexation, le système visuel étant dans sa conception même organisé pour garder avant tout le suivi de l'individualité de certains types d'objets.

En résumé, Pylyshyn soumet deux hypothèses importantes pour notre recherche: (i) les objets apparaissant soudainement dans le champ visuel sont assignés à des index visuels, et (ii) une fois l'objet indexé, on peut y accéder directement sans avoir à le chercher sur la base de ses propriétés. C'est dans ce sens que Pylyshyn voit ces index comme des démonstratifs (*demonstratives*) ou pointeurs.

Nous allons maintenant nous intéresser plus en détail aux grands systèmes pragmatico-sémantiques du cerveau, qui vont entrer en jeu dans le suivi des objets, des agents, des événements, dans le monde physique. De quels mécanismes avons-nous besoin pour détecter la présence de quelqu'un, pour différencier le soi de l'autre ? Quels mécanismes sont nécessaires pour suivre un objet dans l'espace et l'identifier ? Il est fondamental ici de s'intéresser à ces systèmes, dans la mesure où ils peuvent être considérés comme de véritables racines pour l'acquisition lexicogrammaticale du langage.

### **II.3. Les grands systèmes sémantiques cérébraux de suivi d'agents/objets/événements dans le monde**

Nous allons maintenant nous intéresser aux systèmes permettant de suivre les objets, agents et événements dans le monde physique.

Afin d'étudier l'émergence de la notion d'objectivité chez l'enfant, de nombreuses expériences sur des bébés ont été réalisées afin de mettre en évidence les mécanismes intuitifs qui sont engagés dans son apparition : qu'est-ce qu'un objet de manière intuitive pour un bébé ? Quels sont les systèmes dont il dispose pour les suivre dans le monde ? Ces expériences ont donc tenté à travers des « tours de magie », de comprendre ce qui entraîne la surprise dans le comportement de l'enfant, découvrant ainsi ses mécanismes naïfs dont sa « physique naïve » ou intuitive, et de même sa biologie et sa sociologie naïves.

Notre intérêt pour la deixis va se porter en premier sur les circuits cérébraux qui traitent spécifiquement de la localisation (voie *Where*) et de l'identification (voie

*What*) des objets, agents, événements. Ce sont deux systèmes, ou deux voies spatiales bien documentées dans la littérature, pour leur séparation neurale et leurs fonctions diverses. En 1969, Schneider (cité dans Goodale et Milner, 1992) postulait déjà une séparation anatomique entre le codage visuel de la localisation d'un stimulus, et l'identification de ce même stimulus. Dès les années 80, Ungerleider et Mishkin (1982) ont donné les arguments pour cette séparation en deux voies pour la vision, en s'appuyant sur les données comportementales, électrophysiologiques et neuro-anatomiques, obtenues sur des singes rhésus (*Macaca Mulatta*).

Les auteurs remarquent qu'une altération de la partie postérieure du cortex inférieur temporal interfère principalement avec la faculté de discrimination visuelle, tandis qu'une altération à la partie antérieure affecte la mémoire visuelle. En fait, chaque aire striée transmet l'information visuelle, relayée à travers le cortex préstrié, en direction du cortex temporal inférieur. Ce dernier constitue la dernière « station » le long d'une voie visuelle corticale allant du cortex strié au préstrié. Ce système est primordial pour analyser et coder des dimensions physiques des stimuli visuels nécessaires pour l'identification et la reconnaissance.

Le cortex pariétal postérieur, comme l'inférieur, dépend également lourdement des inputs envoyés du cortex strié. Des dommages situés dans cette aire entraînent une constellation d'altérations spatiales visuelles, dont une désorientation spatiale visuelle, des négligences contralatérales, par exemple.

L'hypothèse des auteurs était de montrer que l'appréciation des qualités d'un objet, et de sa localisation spatiale, dépendent du traitement des différents types d'information visuelle véhiculés dans les cortex temporal inférieur, et pariétal postérieur, respectivement. Les études d'ablation apportent un fort soutien à cette dichotomie. Il semble en effet que la trajectoire ventrale (le fasciculus longitudinal inférieur à partir du cortex occipital suit une voie ventrale dans le lobe temporal) soit spécialisée dans l'identification de l'objet ; tandis que la trajectoire dorsale (le fasciculus longitudinal supérieur à partir du cortex occipital suit une voie dorsale traversant la région pariétale postérieure dans son trajet vers le lobe frontal) serait spécialisée dans la perception de l'objet situé dans l'espace.

Une analyse de scène visuelle peut ainsi être réalisée par au moins deux voies, la voie ventrale à travers le cortex temporal inférieur (IT), qui traite l'information sur les traits qui identifient les objets, comme la forme et la couleur (soit l'information

What) et une voie dorsale à travers le cortex pariétal postérieur (PP), qui traite l'information sur la localisation et les relations spatiales entre objets (soit l'information Where).

Ces deux grands systèmes sémantiques constituent des modes puissants de connaissance sur le monde et nous allons les traiter en les couplant terminologiquement selon les couples linguistiques anciens de l'indoeuropéen (ces couples existent dans d'autres groupes linguistiques, cf. Diessel, 2003) couples fonctionnels encore bien audibles phonologiquement en anglais, pour l'interrogation Wh- vs. la démonstration Th-, soit Where/There et What/That pour commencer.

### **II.3.1. Le système Where**

La notion d'objet est fortement liée chez l'enfant au concept de numérosité. A cinq mois, un bébé sait déjà compter. Lors d'une expérience, on montre deux poupées à un bébé âgé de cinq mois, puis on les dissimule derrière un paravent. Lorsque le paravent est levé, il ne reste qu'une poupée. Le bébé est étonné, surpris, par cet événement inattendu (tout comme l'est par ailleurs un chimpanzé). En revanche, si on cache trois poupées et que, lorsque le paravent est enlevé, l'enfant découvre trois nounours à leur place, il n'est pas surpris. Pour lui, c'est la numérosité qui compte et non pas l'identification de l'objet, pourvu que celui-ci soit resté permanent spatialement. A cet âge-là, le bébé possède ce que l'on nomme le système « Where » (Cf. Figure II.3), c'est-à-dire le système cérébral dorsal qui permet d'individuer un objet et d'établir les notions d' « objet seul » et « de plus d'un objet ». Les bébés semblent ainsi posséder la « bosse des maths » très précocement, mais leur numérosité se restreint à réagir à des différences entre 1, 2 ou 3 objets.

Piaget (selon Houdé, 1998) avait émis l'idée que la notion de nombre chez l'enfant apparaissait après que ce dernier eut acquis des capacités telles que classer, inclure et sérier. Cependant les découvertes de Wynn (citée dans ce qui suit d'après Houdé, 1998), que nous venons de reprendre via Leslie et al. (1998), ont permis en 1992 de reposer l'émergence de la notion de nombre dans un autre contexte. Wynn (Houdé, 1998) s'est posé la question de savoir si un bébé de 4 ou 5 mois était capable de calculer précisément un résultat d'opérations arithmétiques simples. Ses observations montrent que celui-ci possède l'aptitude à additionner

$1+1=2$  mais également de soustraire  $2-1=1$ . La situation d'expérimentation consiste toujours à présenter un événement attendu ou possible, deux peluches de Mickey ; puis un événement inattendu ou impossible, un seul Mickey, une fois le paravent soulevé ; et à mesurer le temps de fixation visuelle de l'enfant. Wynn (Houdé, 1998), met en évidence le fait que les bébés perçoivent ce qu'elle nomme « l'erreur de calcul », c'est-à-dire  $1+1=1$ . Les bébés fixent plus longuement l'événement impossible. Le bébé serait donc apte à établir une distinction entre la notion « d'un seul » et la notion de « plusieurs ». Wynn (Houdé, 1998), a également montré qu'il peut discriminer deux quantités comme deux et trois, car il est surpris de l'événement  $1+1=3$ . Selon cette étude, le bébé posséderait donc un mécanisme cognitif qui permettrait de calculer le résultat d'opérations telles que celles énoncées précédemment, et Wynn (Houdé, 1998) va plus loin en émettant l'hypothèse que le bébé serait déjà doté de concepts numériques. Le débat sur ces capacités numériques précoces reste ouvert.

Dans une expérience menée sur des enfants âgés de 10-14 mois, Carey (2004) place dans une boîte 3 objets, un à la fois ou tous en même temps, et permet ensuite aux enfants de manipuler cette boîte afin qu'ils récupèrent un objet à la fois. Le patron des attentes confirme que les enfants s'attendent à trouver 3 objets dans la boîte, créant ainsi une représentation mentale pour chacun d'eux. Toutefois cette performance s'arrête à 4 objets. Dès lors que l'ensemble excède 3 objets, les enfants ne peuvent plus retenir un modèle d'items distincts dans leur mémoire à court terme.

Un enfant possède donc la notion de numérosité, mais restreinte à 2 ou 3 objets. Et c'est précisément, ainsi que nous l'avons vu, le nombre d'objets (proies, prédateurs) que notre attention dans une scène peut suivre au maximum simultanément. Ce chiffre 3 se retrouve dans les expériences d'alerte auditive ou d'alerte visuelle : au-delà de 3 ou 4 événements (coups de feu ou flashes) l'alerte n'est pas plus importante ou plus dramatique. Le système neural en jeu semble être occipito-pariétal. Ce serait la base de cette « bosse des maths » que l'enfant détient déjà à l'âge de 5 mois, ce qui permettrait ainsi de justifier ce fameux chiffre trois. Très précocément ces trois objets vont être indexés, tout d'abord par le suivi du regard ; puis ensuite par le geste du pointer de l'index, lorsque l'enfant utilisera vers l'âge de neuf mois le système bras-main.

Face à la résolution d'un problème d'individuation de l'objet, l'adulte va pouvoir se servir de plusieurs informations qui lui sont accessibles spontanément, à savoir la permanence de l'objet (l'information spatio-temporelle), les propriétés perceptives de l'objet (couleur, taille, texture, forme), et le type d'objet (catégorisation d'objets sous des concepts). La question que nous sommes alors amenée à nous poser est celle de l'individuation chez l'enfant. Dès son plus jeune âge, l'enfant s'intéresse aux objets. Mais possède-t-il dès tout petit l'accès à ces trois types d'informations, ou pouvons-nous remarquer la précocité de l'une d'elles, qui sera plus utilisée par l'enfant ?

De nombreuses études ont mis en avant le fait que l'enfant établit une assignation de l'objet à un index par l'information spatiotemporelle (Leslie et al., 1998). Celle d'Aguiar et Baillargeon (1999), montre que l'identité de l'objet est établie chez des enfants aussi jeunes que 2 mois ; mais pour ceux-ci la détection d'une discontinuité spatiotemporelle entraîne l'établissement d'une représentation de deux objets numériquement distincts.

Xu et Carey (1996) ont proposé une expérience permettant de mettre en relief ce phénomène, en présentant à de jeunes enfants un canard jaune émergeant d'un paravent, puis retournant derrière celui-ci, suivi par une balle rouge qui émerge du même paravent, puis y retourne aussi. Le paravent est ensuite relevé proposant un événement attendu, 2 objets, ou un événement inattendu, 1 objet. Les auteurs observent alors que des enfants âgés de 10 mois ne regardent pas plus longtemps l'issue inattendue d'un seul objet, suggérant qu'ils n'utilisent pas les différences de propriétés ou de type pour conclure à 2 objets distincts. A l'inverse, les enfants âgés de 12 mois réussissent dans cette tâche, avec un temps de regard plus long à l'issue inattendue, révélant ainsi une maturité dans les mécanismes de suivi de l'objet que ne possèdent pas les enfants de 10 mois. Cette expérience a été répliquée par la suite notamment par Wilcox et Baillargeon (1998) dans leurs expériences 1 et 2. Et de la même façon, d'après les travaux de Mareschal (2000), si on montre à un enfant de dix mois un canard en jouet et une voiture en jouet apparaissant et disparaissant, un jouet à la fois, derrière un écran les occultant, ces enfants vont s'attendre à un seul objet lorsque l'écran est levé. Et pareillement, si on leur montre le canard et la voiture qui apparaissent et disparaissent simultanément, les enfants de 10 mois attendent deux objets derrière l'écran relevé. Les bébés ne sont pas étonnés car ils



considèrent que ces objets sont spatialement constants (en fait ils ne sont jamais vus en deux lieux différents, par exemple côte à côte, en même temps). De même, ils ne vont pas être surpris si une grenouille se transforme sur place en prince charmant, car pour l'enfant il s'agit du même objet. Ce qui nous amène bien à confirmer l'idée que les enfants paraissent individuer les objets par l'information spatio-temporelle, du moins dans un premier temps.

Van de Walle et al. (2000), dans une tâche différente de celle des auteurs précédents, ont également démontré que les enfants utilisent l'information spatio-temporelle pour établir des objets distincts avant même de remarquer l'information de type ou de propriété. En effet, des enfants âgés de 10 mois ou 12 mois devaient relever des objets dans une boîte, sans pouvoir les voir et simplement en les touchant. Les expérimentateurs leur ont donné auparavant des informations de type différent concernant le nombre d'objet (1 ou 2 objets). Dans une condition propriété/type, une balle rouge est sortie de la boîte puis replacée à l'intérieur, suivie d'un canard que l'on sort de la boîte et que l'on replace à l'intérieur. Dans une seconde condition dite spatio-temporelle, on montre la balle et le canard simultanément avant que chaque objet ne soit enlevé puis remis dans la boîte. Après avoir retiré le premier objet de la boîte, si sa recherche persistait avec une certaine durée dans la boîte vide, c'était considéré comme l'évidence de l'établissement d'une relation entre 2 objets distincts. Les résultats de cette expérience de recherche manuelle ne montrent pas d'évidence de représentation de 2 objets distincts dans la condition propriété/type pour des enfants âgés de 10 mois : ils ne recherchent pas le second objet de façon persistante, à l'inverse des enfants de 12 mois testés. Les auteurs concluent donc que les enfants utilisent l'information spatio-temporelle pour établir des objets distincts bien avant de relever les informations de traits ou de propriété, dès l'âge de 12 mois.

La conclusion avancée par Van de Walle et al. (2000) concernant l'incapacité d'utiliser les informations de traits et de propriété pour établir une distinction entre 2 objets avant 12 mois a été remise en question par l'étude de Wilcox et Baillargeon (1998). Ces derniers argumentent que cette incapacité pourrait être due à un facteur de complexité dans la réalisation même de l'expérience, et qu'une simplification de la tâche d'individuation pourrait conduire à l'utilisation de cette information par l'enfant. Les auteurs testent donc l'individuation avec un seul facteur de trajectoire : une boîte

bouge d'un côté, disparaît derrière un écran et une balle émerge de l'autre côté de l'écran. Lorsque l'écran est relevé, l'enfant voit seulement la balle, constituant un événement attendu. Les résultats de cette expérience tendent à prouver que des enfants aussi jeunes que 9.5 mois regardent plus longtemps l'événement inattendu et qu'ils réussissent ainsi dans cette tâche d'individuation bien avant 12 mois.

Dans une autre expérience menée par Xu (2002) utilisant une procédure similaire à celle de Xu et Carey (1996), l'auteur montre que des enfants âgés seulement de 9 mois peuvent établir une assignation de 2 objets distincts si on leur fournit des étiquettes distinctives impératives, type « Look ! a duck ! », « Look ! a ball ! ». Les enfants âgés de 10 mois réussissent sans l'étiquetage à la condition que le contraste entre les 2 objets soit très fort, entre un visage de poupée et un objet inanimé par exemple, se basant sur la préférence naturelle de l'enfant pour ce qui est agentif.

Ces différentes études sur l'individuation de l'enfant présentent des résultats assez contrastés dans un domaine d'étude en plein essor, comme le soulignent Xu et al. (2004) dans leur article. Pour résumer, il semblerait que : (i) les enfants ne réussissent pas avant 12 mois dans une tâche d'individuation d'objets complexes ; (ii) les enfants réussissent vers 9.5-10 mois dans une tâche d'individuation simple ; et enfin, (iii) ils réussissent aussi tôt que 9 mois si on leur présente une tâche d'individuation simple avec étiquetage verbal, en fait un guidage incitatif.

Ces réflexions ont alors mené Xu et al. (2004) à conduire une nouvelle série de quatre expériences afin de tester tous les paramètres, qu'ils soient liés à la complexité, à la nature des objets, etc., ceci afin de parvenir à des résultats cohérents en ce qui concerne l'individuation. Ces expériences testent la couleur, la taille, la combinaison taille-couleur-modèle, et les différences de forme pour l'individuation des objets, en utilisant une procédure assez similaire à celle de Xu et Carey (1996), mesurant le temps de fixation à l'issue d'un événement inattendu lorsque le paravent est relevé.

L'expérience 1 vise à tester la couleur, avec une condition couleur différente (une balle rouge et une balle verte par exemple) et une condition même couleur (2 balles rouges ou 2 balles vertes par exemple). Dans la condition de couleur différente, les enfants de 12 mois échouent à utiliser les différences de couleur pour établir une relation distincte entre ces 2 objets. Dans la condition de même couleur,

ils échouent également à établir une relation d'un seul et même objet. Ils ne montrent aucune attente particulière d'un ou deux objets lorsque l'écran est relevé.

L'expérience 2 teste la taille de l'objet. Ils utilisent la même procédure que l'expérience 1 avec cette fois-ci une condition de taille différente (une petite balle rouge vs. une grande balle rouge par exemple) et une condition de même taille. Les résultats montrent que, bien que les enfants codent des différences de taille entre les objets, cette information n'est pas suffisante pour établir la présence de 2 objets distincts derrière le paravent.

Ces deux premières expériences mettent donc en relief l'échec des enfants de 12 mois à utiliser la couleur et la taille pour établir 2 objets numériquement distincts. Ce qui va à l'encontre des résultats obtenus par Xu et Carey (1996). Toutefois, les stimuli de Xu et Carey (1996) diffèrent de ceux de Xu et al. (2004) dans la mesure où ils présentent des informations de propriété et de type (canard vs. voiture), et diffèrent également sur plusieurs propriétés. De ce fait, l'expérience 3 de Xu et al. (2004) va explorer si la présentation d'objets variant sur de multiples propriétés (couleur, taille, patron de surface) mène ou non à une individuation. Les résultats montrent que les enfants échouent une fois encore à utiliser la combinaison de différentes propriétés pour établir cette individuation de 2 objets distincts, suggérant alors que le succès de l'expérience de Xu et Carey (1996) ne repose pas sur cette idée. Néanmoins, les objets de Xu et Carey (1996) diffèrent également en forme et type alors que les objets de Xu et al. (2004) sont du même type et partagent une seule forme.

L'expérience 4 teste alors la différence de forme seule comme soutien à l'individuation d'objets à 12 mois. Les objets utilisés sont de même couleur, taille et possèdent le même patron de surface mais diffèrent simplement dans leur forme globale (tasse vs. balle, bouteille vs. boîte par exemple). Les résultats montrent ici que les enfants réussissent à utiliser les différences de forme pour déterminer l'existence de 2 objets. Ils regardent plus longtemps l'issue inattendue de 2 objets derrière le paravent.

En conclusion, les auteurs montrent que similairement à Xu et Carey (1996) et Van de Walle et al. (2000), les enfants réussissent à utiliser l'information spatiotemporelle pour établir une représentation de 2 objets distincts, mais qu'à contrario, ils échouent à établir cette distinction à partir d'informations de couleur,

propriété, bien qu'ils perçoivent ces différences, la forme servant tout de même plus tôt à individuer.

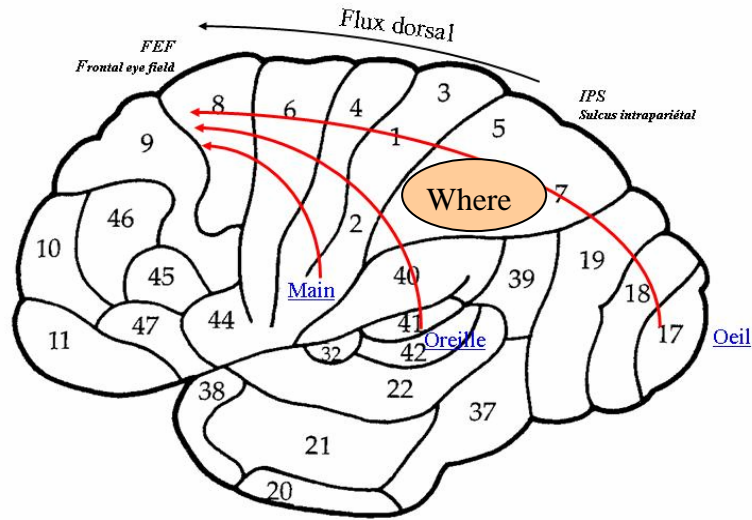


FIGURE II.3 : Les flux multimodaux du système sémantique dorsal Where.

L'indexation de l'objet permet ainsi d'établir une objectitude très tôt avec l'information spatio-temporelle. Il est clair qu'il faut établir une distinction entre l'individuation par la localisation et l'individuation par traits ou propriétés, soit l'identification. Les comportements intuitifs des enfants tendent à prouver qu'ils individuent dans un premier temps un objet spatiotemporellement, et que lorsqu'ils se trouvent dans des situations ambiguës, ils font intervenir l'identification par traits, mais beaucoup plus tard dans l'enfance. Pour que l'enfant puisse acquérir ce dernier mécanisme, il doit arriver à modifier son premier modèle. C'est ce que résumant Leslie et al. (1998) par le schéma suivant (Figure II.4).

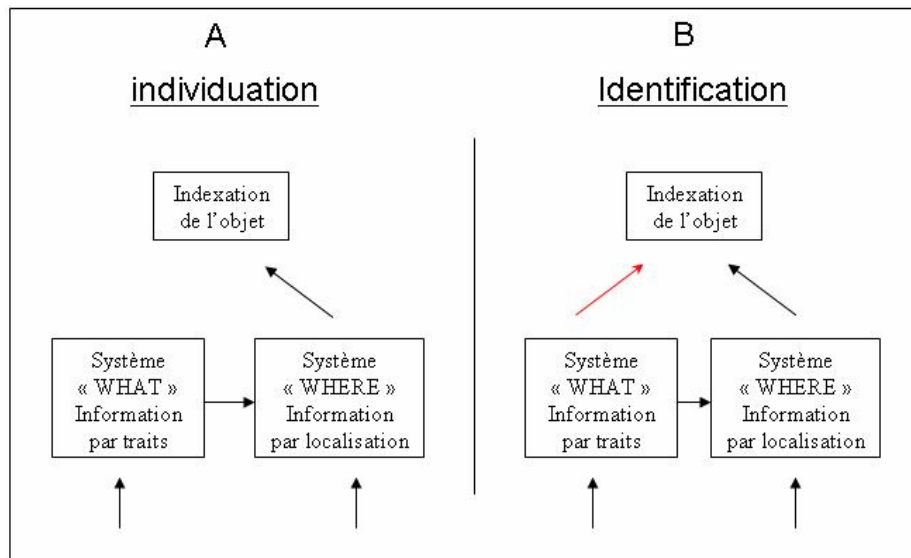


FIGURE II.4 : Modèle pour l'indexation de l'objet (d'après Leslie et al., 1998). Dans le panel A, les objets sont indexés uniquement par la localisation. Dans le panel B, l'indexation de l'objet est réalisée aussi bien par la localisation que par l'information de traits. L'indexation par voie double dans B se développe plus tardivement.

Une expérience proposée par Carey (2001) permet de bien mettre en évidence ces deux systèmes. Dans la première condition (cf Figure II.5), on vous présente le panel 1, puis après un certain temps le panel 2.

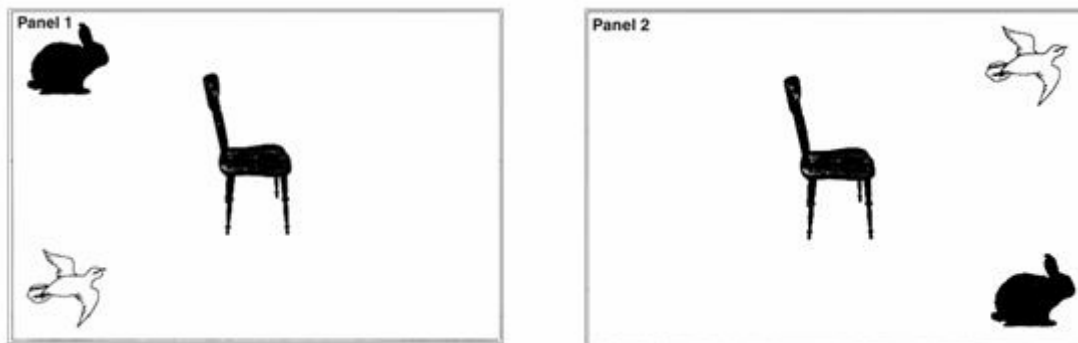


FIGURE II.5: Expérience de mouvement apparent illustrant la préemption du système Where sur le système What (D'après Carey et Xu, 2000).

À la question « qu'est-ce qui a changé ? », vous répondrez sans aucun doute : « Le lapin qui était en haut à gauche est maintenant en bas à droite et l'oiseau qui était en bas à gauche est maintenant en haut à droite ». Dans la seconde condition, on vous demande de regarder la chaise comme point de fixation de votre regard puis le panel 1 et le panel 2 vous sont montrés alternativement à une vitesse assez rapide pour que se produise comme au cinéma ou avec un stroboscope un phénomène de mouvement apparent. Dans ce cas, compte tenu de la distance régissant le

groupement gestaltiste entre objets, il suffit d'après l'expérience que nous avons refaite, pour voir le flux global alterner de droite à gauche, avec en haut le couple lapin noir=oiseau blanc, et en bas l'oiseau blanc=lapin noir est, d'une cadence de quatre images par seconde. Dans cette condition, les sujets voient deux objets se déplacer parallèlement et aucun ne voit le lapin et l'oiseau traverser l'écran en diagonale.

Comme une abondante littérature relate fréquemment l'existence de cette séparation *Where* vs. *What* surtout pour le domaine visuel, nous citerons brièvement ici une étude de Zatorre et al. (2002) dans le but de montrer que cette relation existe aussi dans d'autres domaines, et plus particulièrement pour l'audition. Le rôle de l'information spatiale dans la désambiguïsation de sources auditives qui se chevauchent est un rôle crucial, facilité par la séparation spatiale. Les auteurs suggèrent que l'activité qu'ils observent dans le STG (gyrus temporal supérieur) reflète ce processus. On sait d'autre part que la partie antérieure du STG s'occupe moins de localisation spatiale que de reconnaissance de structures sonores.

Toutes ces ontologies ontogénétiques qui viennent d'être évoquées se retrouvent dans l'attention visuelle chez l'adulte. Il fallait donc en priorité poser ces principes dans leur ontologie naïve, afin de ne pas tomber dans les courants d'idées qui veulent que les objets soient identifiés tout de suite par les enfants par les propriétés ou traits qui leur sont propres. Ceci pourrait être une erreur commune pour mal-poser les principes d'une communication face-à-face efficace, qui s'inspire des mécanismes humains.

Leslie et al. (1998) concluent que la notion centrale autour du concept d'objet consiste, abstraitement, en un pointage mental, soit un « this » (ceci) ou un « that » (cela). Ce qui souligne l'aspect crucial de la notion de déictique bien connue en linguistique, et que nous allons évoquer. Nous verrons que l'avance du système *Where* comparé au temps mis à mettre en place un système *What*, peut peut-être se compenser par l'apparition précoce d'un système *There* précurseur d'un système démonstratif *That*. Mais, on s'en doute déjà, nous n'aurons sur un sujet aussi complexe que des suggestions à apporter.

### II.3.2. Le système What

Plus tard, après 10 mois, l'enfant peut-il se passer de l'information spatio-temporelle ? Complètement, certes non, mais à la place d'individuer un objet par cette seule information, l'enfant va individuer cet objet et l'identifier par l'information de trait, en priorité par sa forme, semble-t-il. On se situe au cœur de l'autre grand système sémantique : le système cérébral ventral « What » (Cf. Figure II.6). Dans le développement de l'enfant, ce système se forme parallèlement au système SAM (*Sharing Attention Mechanism*) que nous verrons plus loin.

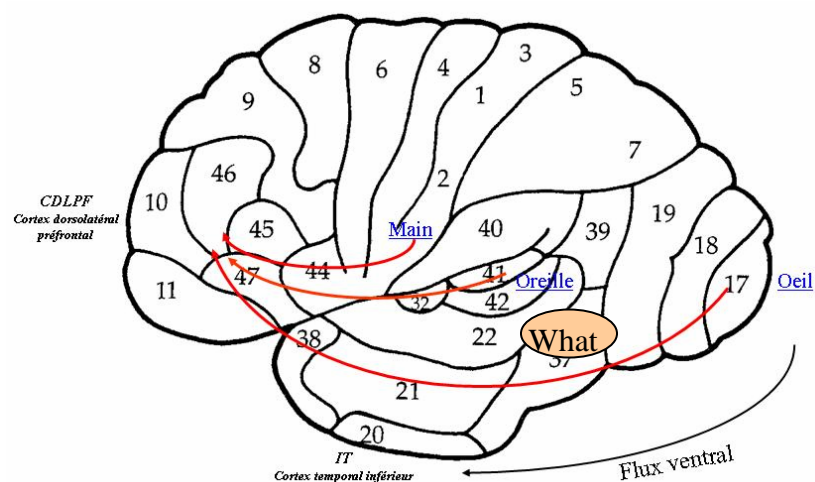


FIGURE II.6 : Les flux multimodaux du système sémantique ventral What.

Le circuit cérébral *What* se situe dans la partie ventrale, partant du lobe occipital, traversant le lobe temporal jusqu'au lobe frontal. Le système *What*, tout comme le système *Where*, prend racine dans chaque aire correspondant à la vision (région visuelle, aire 17), à l'audition (région de la cochlée, aire 41) et au toucher (région de la main, aire SII). Le circuit *What* reçoit les informations de ces différents capteurs sensoriels et permet d'identifier les formes (cf. le système *How* ci-dessous). Il permet également de récupérer les propriétés de couleur dans la reconnaissance des objets. Ce système concerne la faculté de reconnaître des objets significatifs. Le circuit *What* est activé lors d'une monstration et constitue un dispositif naturel pour acquérir des noms et des notions linguistiques. Il se développe chez l'enfant après 12 mois, après la mise en place du système *How*, dont nous allons parler ci-dessous.

Prenons l'exemple d'une chaussure et d'une tasse tournant (se poursuivant) autour d'un paravent sans jamais qu'on puisse voir les deux ensemble. Avant dix mois, l'enfant n'est pas étonné de voir un seul objet restant derrière le paravent quand on le relève. Après dix mois, l'enfant va s'en étonner. A douze mois, il est capable de se passer de la localisation, il va individuer et identifier l'objet par sa forme, Mais à douze mois il n'est pas encore apte à identifier un objet par sa couleur. On peut penser que les formes sont plus précocement significatives pour le système de préhension que les couleurs.

L'idée générale de cette séparation *Where* vs. *What* repose sur le fait que la voie dorsale sélectionne en premier la localisation d'un objet et la voie ventrale intervient ensuite afin de l'identifier. Toutefois, certains auteurs penchent pour une séparation moins nette entre voie dorsale et voie ventrale. C'est le cas de Hamker (2002) qui suggère, selon son propre modèle, que la voie ventrale contribue à la sélection de la localisation d'un objet par des connexions réalisées en feedback. Ce ne serait plus ainsi la propriété unique du système dorsal. Dans son étude il démontre que la voie ventrale code un objet d'intérêt aussi bien que sa localisation. Selon lui, la reconnaissance de l'objet et l'attention recruterait la même architecture neurale et l'effet de *feedback* qu'il prédit permet au système *What* de discriminer très précocement les traits de l'objet.

### **II.3.3. Le système How**

Le système d'identification par la forme semble bien être plus précoce que le système *What* par les propriétés, comme la couleur. Nous le nommons le système pragmatico-sémantique « How » (celui d'une voie pragmatique, pour reprendre le terme de Jeannerod, qu'il oppose à une voie sémantique comme *What*). Il se situe au cœur du gyrus supramarginal (aire de Brodmann 40) (cf. Figure II.7).



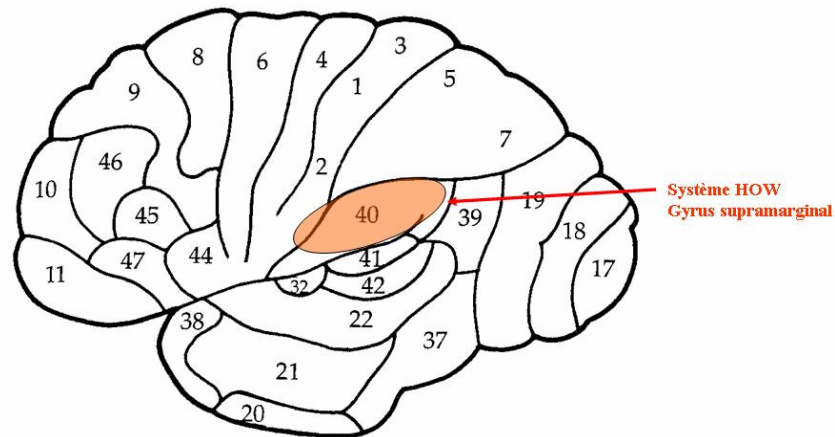


FIGURE II.7 : Le système pragmatico-sémantique *How*.

Quelle évidence pour une distinction entre voie *What* et voie *How*? Des études ont montré qu'après une lésion de la voie *What*, les patients bien qu'incapables de pouvoir reconnaître un objet courant (on parle alors d'agnosie visuelle), restaient tout à fait aptes à saisir cet objet. Il existe, une preuve de cette double dissociation entre ces voies. Un patient atteint d'une ataxie optique, (pour nous disons une lésion du système *How*) pourra diriger sa main dans l'espace vers le lieu de l'objet mais sera incapable de pouvoir ouvrir les doigts correctement afin de préparer la prise qui correspond à la forme de l'objet. En effet, lorsque l'on se saisit d'un objet, il s'opère une anticipation de la prise (*preshaping*) que l'on va avoir sur cet objet : c'est l'affordance de l'objet.

Goodale et Milner (1992) ont proposé une division des systèmes *What* et *How* pour le cortex cérébral postérieur du primate, comme une alternative supplémentaire à la division *What* et *Where* bien connue. Ces auteurs soulignent également l'existence de cette double dissociation *What* vs. *How* chez des patients, avec d'un côté des patients atteints d'agnosie aperceptive ou agnosie de forme visuelle (incapacité à reconnaître les objets) et d'un autre côté des patients atteints d'ataxie optique (inhabileté à agir sur les objets).

Selon les auteurs les patients atteints d'agnosie de forme visuelle ne sont donc plus capables d'identifier les objets suite à un dommage au cortex visuel postérieur. De plus, ces patients partagent des problèmes d'identification d'objets visuels similaires, comme par exemple distinguer un cercle d'un triangle, ou établir une distinction entre « M » et « W ». Ainsi, la faculté à copier visuellement des

formes présentées, qui requiert une représentation de la forme relativement intacte, semble altérée suite à ce type de lésion. En revanche, malgré une faculté de perception visuelle et d'identification des formes altérées, ces patients agnosiques visuels possèdent encore un contrôle visuomoteur intact. En d'autres termes, ils peuvent atteindre et saisir des objets de formes variées de façon appropriée. La question qui se pose alors est celle de savoir comment un patient agnosique pour la forme visuelle peut-il interagir sur un objet qu'il ne peut pas percevoir ? Vecera (2002) tente d'apporter des explications plausibles pour expliquer cette dissociation *What vs. How*.

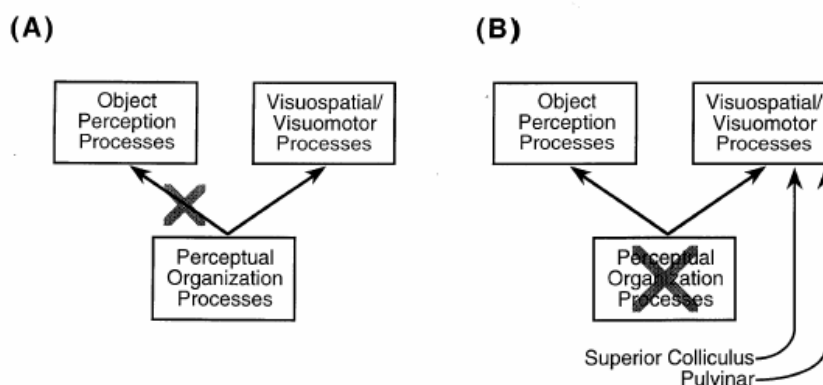


FIGURE II.8 : Agnosie visuelle et ataxie optique: (A) explication de la déconnexion ventrale ; (B) explication de l'input épargné (D'après Vecera, 2002).

La première explication de Vecera (2002) (Figure II.8 (A)) concerne la déconnexion ventrale. L'agnosie de forme visuelle serait causée par un dommage le long de la voie de traitement ventrale (voie *What*). Ce type de dommage entraîne alors une altération de la connexion entre vision de niveau précoce et perception de la forme, empêchant ainsi la représentation de la forme appropriée. Toutefois, cette explication ne lui paraît pas satisfaisante et une explication davantage biologique paraît nécessaire, en invoquant les modèles de réseau neuraux afin de mieux appréhender les processus cognitifs normaux. Dans cette seconde explication (Figure II.8 (B)), le dommage principal survient dans les aires visuelles de bas niveau qui fournissent l'input aux deux voies. La fonction visuomotrice reste intacte grâce à des inputs existant dans le système dorsal. Le colliculus supérieur se projette sur les aires visuelles du lobe pariétal via le pulvinar, et ces inputs « préservés » peuvent permettre une action visuellement guidée restant intacte. Toutefois, Vecera souligne que ce type d'explication devrait impliquer que le traitement ventral soit relativement conservé, et cette explication n'en dit pas plus sur la fonction même de ces fameux

inputs de « rechange » préservés. L'auteur propose donc d'avancer une troisième explication tentant de palier aux imperfections des deux explications précédentes. Vecera admet l'hypothèse d'un input commun aux flux dorsal et ventral. Toutefois, il suggère que les différences entre représentations perceptives et visuomotrices puissent expliquer la dissociation observée dans l'agnosie de forme visuelle. Il estime que la tâche d'identification de l'objet est une tâche statistiquement plus complexe que ne l'est la tâche de localisation spatiale. Selon Vecera, les inputs dégradés le long de la voie « What » empêchent de percevoir la forme. Ces mêmes inputs dégradés le long de la voie « How » interrompent mais n'empêchent pas la localisation.

### **II.3.4. L'intégration du système What et du système Where**

Il paraît opportun, une fois ces systèmes exposés, de se poser la question de leur liage ou intégration. Dans quelles conditions et à quel moment ces informations convergent-elles : à quel moment l'information de l'identité de l'objet rejoint-elle celle sur sa localisation ? Rao et al. (1997) proposent de souligner le rôle du cortex préfrontal dans ce but. Les neurones à l'intérieur du cortex préfrontal contribuent à la mémoire de travail et constituent alors de bons candidats pour intégrer des signaux provenant de diverses régions. Selon les auteurs, l'information spatiale du cortex pariétal postérieur et l'information sur l'objet du cortex inféro-temporal sont reçues par des régions séparées du cortex préfrontal, le cortex préfrontal dorsolatéral (aires 46 et 9) et le ventrolatéral (aire 12). Ces aires possèdent toutefois des interconnexions qui devraient amener à l'intégration des systèmes *What* et *Where*. Des études physiologiques ont prouvé que différentes régions du cortex préfrontal véhiculent des informations distinctes : soit l'information sur l'objet (dans le cortex préfrontal dorsolatéral) ou soit l'information spatiale (dans le cortex préfrontal ventrolatéral), mais aucun neurone n'a été rapporté comme véhiculant les deux à la fois. Afin de tester cette idée, les auteurs ont procédé à une expérience menée sur deux singes, en enregistrant l'activité des neurones du cortex préfrontal latéral. L'expérience consistait à ce que le singe maintienne son regard fixé sur un spot. Un objet-échantillon était brièvement présenté au centre du regard. Après un délai, 2 objets-tests étaient brièvement présentés à 2 ou 4 localisations extrafovéales possibles. Un des objets-tests était relié à l'objet-échantillon, l'autre pas. Après un nouveau délai, le singe devait réaliser une saccade vers la localisation mémorisée du

lien. Cette tâche requérait donc de la part des 2 singes qu'ils lient le système *What* au système *Where*. En effet, ils devaient dans un premier temps se souvenir de l'identité de l'objet lors du premier délai (information *What*) et utiliser ensuite cette information pour trouver le lien, puis se souvenir de la localisation lors du second délai (information *Where*). Il fut enregistré que les neurones *What* et *Where* véhiculaient l'information sur l'objet et sur la localisation lors des différents stades du test. Les neurones *What*, les neurones *Where* et les neurones *What-et-Where* étaient effectivement distribuées entre le cortex préfrontal dorsolatéral et ventrolatéral. Cette étude montre que lorsque l'information d'objet et l'information de localisation sont utilisées ensemble, l'information sur ces attributs converge dans le cortex préfrontal. Ces résultats soutiennent donc la notion qu'une fonction du cortex préfrontal est d'intégrer l'information disparate des deux voies. Ainsi les signaux *What* et *Where* ont la possibilité d'être intégrés par des interconnexions entre les cortex préfrontaux dorsolatéral et ventrolatéral, soit à travers des projections convergentes du cortex temporal et pariétal sur le cortex frontal via une combinaison de ces voies. De ce fait les neurones présents dans le cortex préfrontal peuvent traiter les signaux *What* et *Where* et contribuer à les lier pour diriger l'action.

### **II.3.5. Le système *When* prioritaire sur le *Where* et le *What***

Ces deux systèmes sémantiques fondamentaux, les systèmes *Where* et *What* se trouvent préemptés par un autre système sémantique précoce chez l'enfant, à savoir le système *When*. Ce système se situe au départ ou à la racine de chaque division de flux, comme nous pouvons l'observer sur l'image II.9. Ce circuit *When* est spécialisé dans le traitement du mouvement, la détection des événements, et il est au cœur du traitement du mouvement biologique. Ce circuit dédié au mouvement biologique détecte les intentions de mouvement, avant même que le mouvement ne soit réellement initié et avant que l'objet en mouvement ne soit reconnu et analysé.

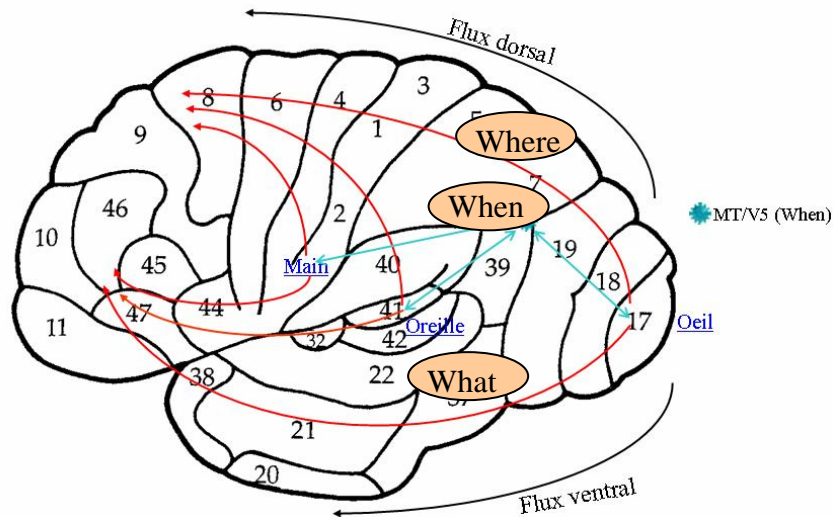


FIGURE II.9 : Le système *When* (étoile MT/V5) à la racine de chaque ségrégation des flux *What* et *Where*, pour les informations auditives, visuelles et tactiles.

Le circuit *When* détecte donc le mouvement biologique de parties du corps écologiquement et socialement significatives, tels que les mouvements (ou des intentions de mouvement sur des images figées) de la bouche, de la main, des yeux. Il permet de détecter si la cause du mouvement est d'origine biologique ou non (muet sur une barre de fer vs. déchargeant ses *spikes* sur une main). Pour détecter l'agentivité, la notion d'autopropulsion est importante. Mais les enfants préfèrent regarder comme un agent ce qui possède les éléments d'un visage, surtout des yeux, une bouche, ou une réactivité quelconque (une masse informe qui couine ou recule quand on la touche). En effet, ce circuit carte ne fonctionne que sur le mouvement d'un être qui peut être assimilé à un être vivant, comme par exemple un robot qui bouge de façon autonome et qui semble doté d'intentions.

Comment démontrer concrètement sous quelle condition ce système *When* préempte les deux autres systèmes ? Une preuve du caractère prioritaire du système *When* a été mise en évidence par Shimojo et al. (2001). Ces auteurs ont réalisé une expérience où deux boules identiques se déplaçaient en diagonale, se croisant à un moment donné. Simplement sur la vision de ces deux boules, les sujets les distinguaient comme se croisant. Les auteurs ont ensuite ajouté lors du croisement des deux boules, un son bref. Ils ont pu remarquer que cet événement altérait fortement la perception visuelle des sujets testés, puisque ceux-ci percevaient alors les deux boules comme se heurtant et divergeant l'une de l'autre (voir Figure II.10)

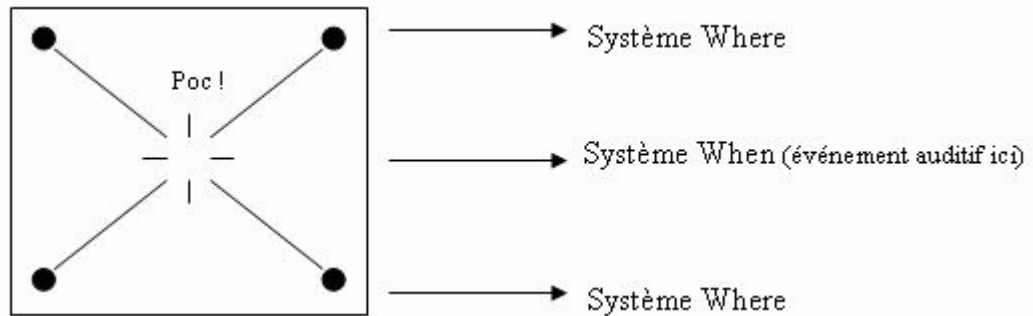


FIGURE II.10: Le système *When* préempte le système *Where* (D'après l'expérience de Shimojo et al., 2001). Suivi de deux boules, qui partent du haut, (i) se croisent et continuent leurs trajectoires vers le bas; ou (ii) changent de direction après un événement produit à leur rencontre (choc auditif « Poc ! », flash visuel, retour tactile). Par ailleurs, en jouant sur les changements (échanges) de forme (boule qui devient triangle) ou de couleur (boule rouge qui devient verte), on peut constater que le système *Where* préempte bien le système *What* (comme dans l'expérience supra de la Figure II.5) : on suit l'objet même si en fin de course on peut constater qu'il a changé de forme et/ou de couleur.

Shimojo et al. se posent la question du mécanisme neural sous-jacent et argumentent, d'après des études psychophysiques portant sur le mouvement visuel, que la perception du croisement des trajectoires de deux boules est dû au recrutement temporel des signaux de mouvement locaux suivant cette trajectoire. De plus, le fait de suivre attentivement les objets augmente la perception visuelle du croisement. L'introduction d'un événement perturbateur, ici un son bref –ou un flash visuel ou le retour d'une impulsion tactile, peu importe la modalité utilisée, seules les fenêtres d'intégration sont différentes– va rompre ce suivi. En d'autres termes, lorsque les boules se croisent, on est en présence de deux flux directionnels *Where*. Les sujets vont ainsi suivre les objets grâce à leur permanence par leur trajectoire de mouvement inertiel dans l'espace. Par contre, lorsque l'on introduit un événement, comme ici le « Poc ! » d'un choc, les deux flux *Where* vont être réorientés par cet événement *When* et les deux boules diverger après leur rencontre, au lieu de se croiser l'une occultée sous l'autre. De cette façon, cet événement *When* va changer le cours des directions des flux *Where*. Il faut ajouter (ainsi que nous l'avons testé) que si les mobiles changent de forme (de boule à losange) ou de couleur, les trajectoires ne changent pas plus de direction que dans l'expérience du lapin noir et de l'oiseau blanc rapportée plus haut : elles se croisent et on s'aperçoit que l'objet a changé seulement en bout de course. Seuls les événements (choc, flash) font diverger les directions.

### II.3.6. Le système *That*

Le système *That* –le circuit du pointage et plus généralement le circuit pour la fonction « monstrative »– serait-il localisé dans l'aire de Broca (zones 44, plus 45 et 47) ? Nous allons partir d'un couple attentionnel bien connu, le circuit oculaire fronto-pariétal FEF-LIP, pour tenter de le translater plus bas dans le cortex, vers un circuit fronto-pariétal vocal Broca-Gyrus Supramarginal. Entre l'établissement de ces deux circuits, un article peu remarqué d'Astafiev et al. (2003, comprenant les grands spécialistes du cerveau visuel, Van Essen, et de l'attention, Corbetta), nous servira de premier guide pour aller du pointage oculaire au pointage digital.

Lorsque l'on pointe de l'œil dans l'espace, on observe un flux dorsal entre le LIP (*Lateral Intraparietal Sulcus*), qui s'active dès qu'on porte son attention à un objet situé dans l'espace, et le FEF (*Frontal Eye Field*) qui permet de diriger attentionnellement les yeux vers cet objet. FEF fait partie de la zone de Brodmann 8, dans la partie préfrontale du cerveau, et il est de même structure cytoarchitectonique que le LIP auquel il est fortement connecté.

Quand il s'agit de pointer avec le doigt le couple LIP-FEF est translaté somatotopiquement plus bas (latéralement). AIP (*Anterior Intraparietal Sulcus*) et F5 (l'homologue de Broca chez le singe) forment le nouveau couple recruté. La liaison AIP-Broca permet de guider le bras correctement dans une direction et avec une prise préalablement choisie en fonction des informations sensorielles. L'étude d'Astafiev et al. (2003) montre que le pointer (de l'index seulement, dans la contrainte couchée du scanner, où l'on évite en IRM les artefacts de grands mouvements comme le bras), produit une dominance gauche, quelle que soit la main, avec une activation du SMG.

Si à présent on désire pointer dans le monde, non plus avec le doigt, mais sur sa propre voix, c'est-à-dire réaliser un focus sur une partie de son propre énoncé (il s'agit en fait d'un pointer sur son propre corps dans l'action de son conduit vocal vers l'autre, à la manière dont on pointerait pour montrer à l'autre un de ses doigts), que va-t-il se passer ?

Cette question est d'abord partie de la mise à disposition neurale d'une condition « baseline », celle apportée par l'étude de Murphy et al. (1997) portant sur les différentes aires cérébrales recrutées dans le contrôle moteur de la parole. Leur

tâche consistait à demander à 6 sujets normaux de répéter de façon continue la phrase classiquement utilisée pour son aspect phonétiquement réitérant dans les études prosodiques « Buy Bobby a poppy ». Les activations ont été observées en tomographie par émission de positons dans quatre conditions : phrase parlée, phrase murmurée silencieusement (sans phonation voisée), sans articulation supraglottique (juste « hummée ») et enfin pensée silencieusement. Ils trouvèrent des activations associées au contrôle de la respiration pour la parole, la vocalisation et l'audition et une activation bilatérale dans les cortex moteur et sensori-moteur, des activations dans le thalamus, le cervelet et l'aire motrice supplémentaire. Ce qui ressort d'intéressant dans cette étude c'est la découverte que dans aucune de ces conditions l'aire de Broca n'était recrutée. Ce résultat a été soutenu depuis par d'autres de l'équipe de Wise.

Que se passe-t-il maintenant si nous comparons ce type de condition répétée à des conditions introduisant un focus sur un élément de la phrase, que ce focus soit intonatif ou syntaxique ? Intonatif : « Buy BOBBY a poppy » ; syntaxique « It's to Bobby that you (have to) buy a poppy ». C'est précisément ce qu'ont testé Løevenbruck et al. (2005) dans leur étude IRMf sur le français.

Cette étude (Løevenbruck et al., 2005) a porté, ainsi que nous allons l'exemplifier, sur l'exploration contrastée des réseaux du cerveau activés dans la deixis prosodique et dans la deixis syntaxique. Il est bon de rappeler que l'utilisation d'une construction présentative pour l'extraction « C'est à Bobby qu'on achète un coquelicot » est particulière au français dans certains cas (comme « est-ce que... ? » pour l'interrogation ; comparer l'anglais « He is the one who did it » au lieu de « It's he/him who... » toujours hésitant). Cette construction est une ancienne construction démonstrative (dont il resterait des traces vivantes dans la syntaxe du français actuel, Mireille Piot, comm. pers.).

L'expérience a été réalisée sur seize hommes droitiers et avec un paradigme dit « bloc » alternant quatre conditions : (i) la phrase de base « Madelein' m'amena » ; (ii) la tache de deixis prosodique « MADELEIN' m'amena », avec un focus intonatif sur l'agent Madeleine ; (iii) la tache de deixis syntaxique « C'est Mad'lein' qui m'am'na » ; enfin (iv) les deux deixis prosodique et syntaxique cumulées « C'est MADELEIN' qui m'am'na » (ces deux dernières avec élisions du e d'amena pour gabariser tous les énoncés à 6 syllabes).



Les résultats montrent que le patron commun des activations significatives dans les trois conditions déictiques, comparées à la phrase de base, est de nette dominance gauche, contrairement à la prosodie dite affective de dominance droite. Ce patron d'activités inclut l'operculum gauche LIFG (*Left Inferior Frontal Gyrus* dont Broca), les aires prémotrices et l'insula gauche. On a observé en outre que la deixis prosodique active en plus le gyrus cingulaire antérieur gauche, le gyrus supramarginal gauche (Left SMG) et le gyrus temporal antérosupérieur gauche (Wernicke). La comparaison directe entre deixis prosodique et deixis syntaxique met clairement en évidence les activations significatives de l'aire de Wernicke et de LSMG pour la prosodie syntaxique. On a donc en noyau commun de ces tâches déictiques, Broca (dont la réputation syntaxique n'est plus à faire), qui est actif dans les deux cas. L'activation additionnelle de LSMG dans la deixis prosodique (que nous avons considérée comme un pointage vocal) suggère que LSMG serait recruté dans la deixis intentionnelle, c'est-à-dire la deixis « forte » ou expressive très corporellement orofaciale. La deixis grammaticalisée, plus « légère » ou abstraite en termes de mobilisation corporelle, serait traitée seulement par LIFG (Broca). Des études complémentaires sont envisagées avec une grammaticalisation gradiente et des tâches émotionnelles (Figure II.11 ci-dessous).

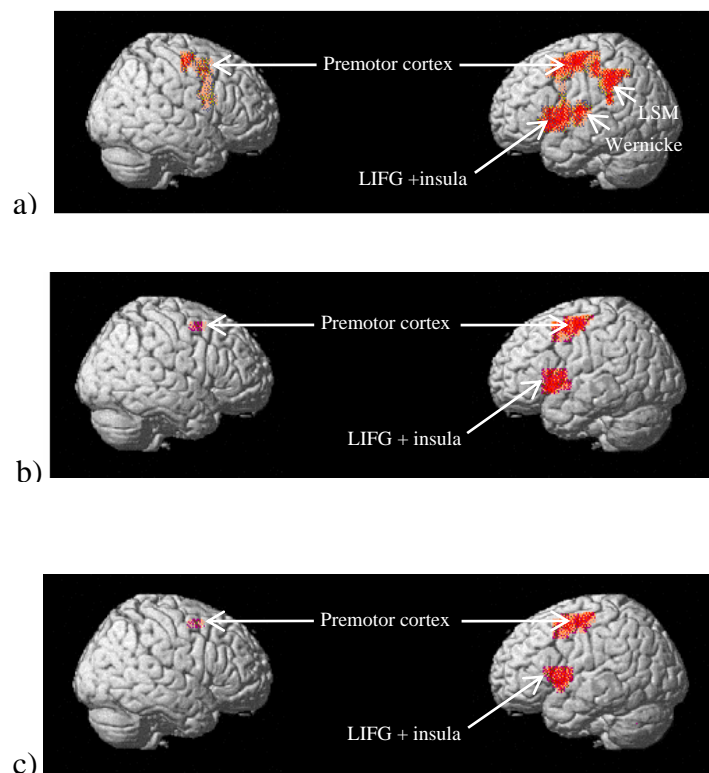


FIGURE II.11: a) Condition de deixis intonative « MADELEINE m'amena » ; b) condition de deixis syntaxique « C'est Mad'leine qui m'am'na » ; (c) condition de deixis syntaxique + intonative « C'est MAD'LEINE qui m'am'na » (Løevenbruck et al., 2005)

Cette étude nous permet de mettre en valeur le fait que lors de tâches impliquant le focus prosodique –soit le fait de montrer sur sa propre voix– l'aire de Broca est activée, ce qui n'était pas le cas dans une tâche simple de répétition de la phrase. On a noté que, somatotopiquement, nous observons bien une activité de l'aire somatosensorielle associative 40 (bouche, larynx) à gauche.

Enfin cette étude monte qu'à contrario de la deixis prosodique, dans la deixis par extraction syntaxique uniquement, c'est-à-dire par la grammaire, l'aire 40 n'est cette fois-ci plus activée. Ce qui indique que lorsque la monstration est grammaticalisée, on perd en quelque sorte le « feeling » ou en d'autres termes les sensations d'attente (*expectation*) pariétale des résultats de l'action frontale organisée par Broca, telle qu'on peut l'imaginer pour la bouche et le larynx linguistiques.

En outre, les résultats nous procurent un résultat étonnant : si nous ajoutons un focus prosodique à l'extraction syntaxique, alors l'aire de Broca est activée, mais le focus prosodique n'active plus, dans ce cas précis, l'aire 40. Autrement dit, dès que la grammaticalisation est installée, elle n'est plus pénétrable par ce « feeling » . Au niveau cérébral la grammaticalisation entraîne une dépariétalisation définitive, puisque l'on ne s'attend plus du tout aux résultats de ses actions orofaciales, une fois celles-ci grammaticalisées, un processus irréversible.

En conclusion, l'activation de l'aire de Broca, à gauche, montre que la tâche de focus déictique, même agie sous la modalité uniquement prosodique, requiert un contrôle de la syntaxe, et par conséquent utilise dans ce contrôle de la monstration, un système *That* (Figure II.12). Ce dernier est intégré par l'intonation pour toutes les langues (qui ne possèdent pas toujours des outils grammaticalisés : cf. allid « PETER hat es gemacht », et pas « Es ist Peter der... ») dans un circuit frontopariétal Broca-SMG que C. Abry (Abry et al., 2004) a baptisé le *That-Path* ou la « ça-voie », un nouveau système déictico-syntaxique qui traite la deixis orofaciale linguistique dans le cerveau.

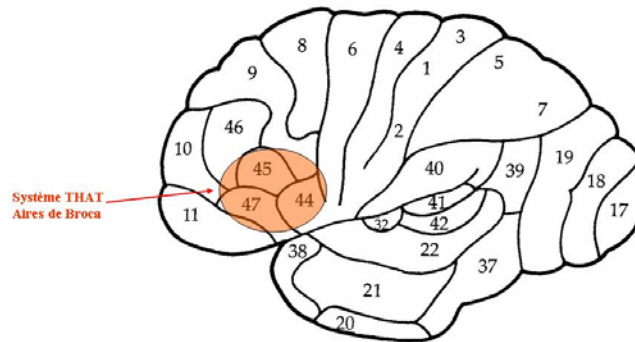


FIGURE II.12 : Le système déictico-syntaxique That.

### II.3.7. Le système Then

Le circuit cérébral préfrontal que nous avons baptisé *Then* (à l'autre bout du système multimodal *When*, de la jonction pariéto-occipito-temporale) se situe au cœur des aires 46 et 9, qui constituent le cortex dorso-latéral préfrontal (CDLPPF) (Figure II.13). Ce système est connu, entre autres, pour sa capacité à ordonner des événements. Il convient de différencier l'ordre de plusieurs actions dans une histoire, d'une succession d'agents dans une phrase. Prenons une illustration anglaise tirée des rimes enfantines, celle de la formule dite des doigts qu'on prend pour des agents (ici *little piggies*), en les montrant, tirant ou secouant l'un après l'autre, avant la *punchline* ou chute qui finit généralement en *tickling* ou chatouillis de l'enfant (Diaferia et Abry, 2005) :

This little piggy went to market;  
 This little piggy stayed at home;  
 This little piggy had roast beef;  
 This little piggy had none;  
 This little piggy cried, Wee, wee, wee,  
 All the way home.

La première phrase prise seule, « This little piggy went to market », active l'aire de Broca. Mais, lorsque l'on énonce la suite (qui répond à un « et alors ? » de suspense ou au « *then* ; et alors... » de liaison des épisodes courant dans la narration) une énonciation dans l'ordre des séquences, cela devient une histoire chronologico-causale, avec un séquençage narratif activant la partie pré-frontale du cerveau, notamment l'aire 46.

Le circuit *Then* agit comme un séquenceur d'actions situées dans l'appréciation (*appraisal*) du contexte et des conditions locales, qui pourront conduire éventuellement à une inhibition conditionnelle. Pour comparer les différentes fonctions de Broca en syntaxe vs. celles du système narratif préfrontal (cf. Sirigu et al., 1998, Mar, 2004), prenons l'exemple de patients ayant une lésion de l'aire de Broca. Ils vont se révéler incapable de construire une phrase, ni de comprendre des phrases ambiguës, mais n'auront cependant aucun problème pour comprendre un récit. Une lésion de Broca n'altère pas en soi le caractère sémantique d'une phrase, mais pose problème dans des cas d'ambiguïté, comme dans les exemples suivants :

- (1) la fille en vert conduit le camion rouge
- (2) le camion rouge est conduit par la fille en vert

En s'aidant de la sémantique de ces énoncés, les patients atteints d'une lésion de Broca ne montrent pas de difficultés à comprendre, puisque toute ambiguïté peut être levée dans la mesure où il est impossible qu'un camion puisse conduire une fille ! A présent qu'en est-il pour les exemples suivants :

- (3) le camion vert pousse le camion rouge
- (4) le camion rouge est poussé par le camion vert

Les patients vont échouer puisque, étant donné que l'agent et le patient désignent une entité similaire, ils ne seront plus en mesure de se reposer sur l'ordre des mots, et ils vont ainsi penser à tort, dans l'exemple (4), que c'est le camion rouge qui pousse le camion vert.

Ce qui nous permet maintenant de démontrer que le système *Then* possède une existence propre en dehors de Broca, est le fait même que ces patients n'aient aucune altération dans la compréhension du séquençement des histoires. Si nous prenons maintenant l'exemple de patients cérébro-lésés au niveau du CDLPF, nous observons que les exemples (1), (2), (3) et (4) seront parfaitement compris et désambiguïsés, dans la mesure où ils comprennent le passif, et que la syntaxe est intacte. Mais à l'inverse de notre cas précédent, ces patients ne seront plus capables de produire ou de suivre de façon ordonnée une histoire. Cette perte d'appréciation du contexte entraîne une perte de la « morale », celle de l'histoire au moins (la chute ou *punchline*). Des lésions dans cette zone préfrontale engendrent une perte de la

compréhension sémantique d'une histoire séquencée, mais pas de la syntaxe : c'est le patron inverse des patients cérébro-lésés de Broca.

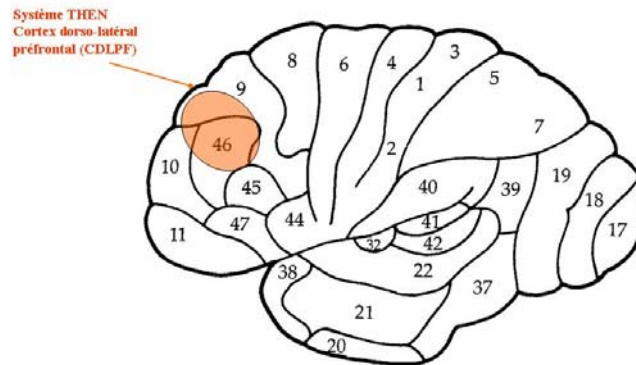


FIGURE II.13 : Le système *Then* localisé dans le cortex dorsolatéral préfrontal.

### II.3.8. Un cerveau pour 7 systèmes sémantiques

Nous venons de repérer fonctionnellement et neuro-anatomiquement les différents grands systèmes pragmatico-sémantiques cérébraux qui interviennent dans notre cadre du signe afin de suivre les agents, objets, événements, dans le monde, et nous en donnons la représentation résumée dans la Figure II.14 ci-dessous :

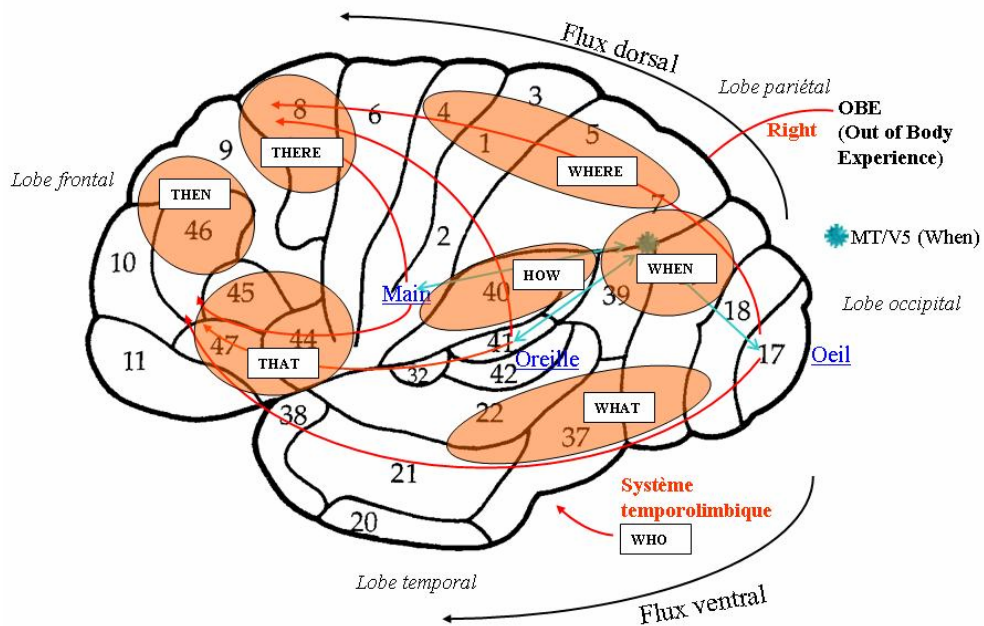


FIGURE II.14 : Les 7 grands systèmes cérébraux pragmatico-sémantiques

On notera immédiatement que ces systèmes sont certainement plus nombreux que le chiffre magique 7. Il nous faudrait plus de temps pour développer les systèmes de suivi de soi et de l'autre. Face à la disparité des résultats en ce domaine, nous avons choisi de supprimer toute la partie que nous avons consacrée à ces systèmes importants pour le développement de l'enfant. Nous avons gardé deux expériences parlantes de *cortical mapping* sur des patients épileptiques réalisées par l'équipe d'Olaf Blanke à Genève. La première (Blanke et al., 2002) reproduit le déclenchement de différentes OBEs (*Out-of-Body-Experiences* ou expériences hors du corps), une première déjà réalisée à Montréal par le grand Penfield dès les années 40, avec une stimulation du gyrus angulaire (BA 39) droit. Ce qui reproduit à Genève une dissociation (dédoublement) du corps senti (proprioceptif) et du corps vu (visuel), chez une patiente qui n'avait jamais éprouvé cet état. Il est fort probable que le liage ou l'intégration de ces deux schémas corporels du Self, se produise dans ce carrefour pariéto-occipito-temporal connu pour sa multimodalité. Une autre expérience publiée cette fin d'année 2006 par la même équipe (Arzy et al., 2006) a obtenu la perception d'un alien plus ou moins envahissant (intruder plus ou moins menaçant) interférant dans les actions du sujet, sous la forme de la présence d'une ombre dans le dos de la patiente. Les travaux sur l'épilepsie du lobe temporal avaient déjà noté, notamment autour de Jean Bancaud, cette sensation d'étrangeté (l' *Unheimlich* selon le mot de Freud) ou plus précisément de présence étrangère sentie (*sensed presence*) due à la stimulation du système temporo-limbique. Ce que nous avons proposé comme racine du système *Who*, en nous appuyant sur les cas d'hypervigilance dans la détection d'un prédateur. Le groupe de Genève, qui a obtenu cette sensation que nous baptisons d' « alien shadower », précise qu'elle résulte cette fois-ci d'une stimulation dans le gyrus angulaire gauche. Il y a certainement un long chemin pour aller de ces cas au suivi de soi et au suivi de l'autre (cf. entre autres, Chaminade et Decety, 2002, etc.), notamment si l'on constate que la production par *cortical mapping* de cet autre-soi-non-reconnu-comme-tel (cf. p. ex. le même problème du non-soi pas défini à partir de ce qui est le Self, traité par Georgieff et Jeannerod, 1998), pose la question complémentaire de l'autre-bien-reconnu-comme-tel, un système essentiel pour le développement de toute empathie agentive dans la cognition sociale (Decety, 2003).

De même que pour les systèmes du *Who* (et de celui du *Self* des schémas du corps) nous avons gardé une étiquette pour le système *There*, que nous proposons appuyé sur le FEF. Ainsi que nous le rappellerons plus loin, des arguments linguistiques et philosophiques ont été avancés par plusieurs philosophes (résumé dans Hurford, 2003) comme quoi il n'y aurait jamais de locatif dans le démonstratif. Pour nous la meilleure preuve empirique linguistique du contraire est sans aucun doute la possibilité pour les langues de faire évoluer un locatif en démonstratif (« ici » devenant « cet », Diessel, 1999), comme *ken* en buang de Nouvelle Guinée. Outre les arguments de Rao et al (1997) déjà évoqués ci-dessus pour une intégration des voies *What* et *Where* dans le préfrontal, nous mentionnons qu'une expérience est en cours de dépouillement à l'ICP (dirigée par Hélène Løevenbruck et Coriandre Vilain) qui devrait apporter des lumières sur la dissociation possible du « là » et du « ça », ce dernier dominant le locatif quand les deux sont en co-occurrence.

Les retombées linguistiques de ces systèmes, dont beaucoup sont déjà fort bien développés dès la première année de vie, sont assez évidentes pour l'indexicalisation et le pointer (pour une exploitation linguistique de la distinction *Where/What*, cf. Hurford, 2003, et avant Givón, 1998), et pour la permanence et l'objectitude, dont l'agentivité. En ce qui concerne la grammaticalisation, celle de la forme linguistique vedette pour le pointer, *that*, est évidemment plus tardive que la première année (sans parler des formes dérivées de la démonstrative, comme la relative et la complétive, qui ne s'installent que vers 4 ans, Lieven et al., 1997). Et il en va de même de la grammaticalisation par dépariétalisation que nous venons d'évoquer ci-dessus. Lieven et al. (1997) ont montré que, dans la lignée de Lambrecht (1988) pour les adultes francophones (cf. aussi Katz, 2000), les constructions présentatives qui nous ont servi pour l'extraction (deixis syntaxique) étaient utilisées plus précocément par les enfants que les constructions réputées sources (sans extraction), de manière formulaire (grammaticalement toutes « packagées »). Pour fixer les idées nous avons en français « Maman ! » —> « C'est Maman ! » —> « C'est Maman qu'est méchante » —> « C'est Maman qui dit qu't'es méchante » (ces dernières constructions bien maîtrisées après 4 ans).

En ce qui concerne l'objectitude il semble que la maîtrise de la forme (pour les affordances de préhension, cf. le système *How*), bien avant la couleur pour l'identification dans le suivi, corresponde à la grammaticalisation possible de cette

forme dans les langues, avec l'exemple classique des prépositions *around*, *along*, etc. (« autour de minuit », « tout le long de l'eau ») ; et celui des classificateurs, comme en japonais : *nihon* pour « deux » stylos (objets allongés) contre *nimai* « deux » assiettes (objets plats... Et deux CD ? *mai* pour le disque, *hon* pour la piste). Alors que ce stade n'est jamais atteint pour les couleurs (le vin blanc n'est certes pas de couleur blanche, mais d'une autre catégorie que les rouges et les rosés ; mais nous n'avons pas affaire ici à un classificateur véritablement grammaticalisé). Nous poserons en conséquence la proposition que **les propriétés qui ne sont pas traitées dans les mécanismes de suivi de l'objet avant l'âge d'un an ne peuvent plus être grammaticalisées par la suite**. Cette remarque nous permet de souligner l'importance du développement de ce cadre du signe pour le développement ultérieur du langage.

Ces grands systèmes étant repérés, nous proposerons d'envisager maintenant les différents mécanismes requis pour le développement de la sémantique chez l'enfant, avec un focus tout particulier sur les quatre modules d'attention tels que les définit Baron-Cohen (1995).

## **II.4. Quatre modules dans la voie développementale vers une théorie de l'esprit pour la communication, l'attente et l'attention**

Pour le suivi du mouvement des objets/agents/événements intégré dans une théorie de l'esprit, attribuant des contenus mentaux à l'autre, nous reprenons ici le modèle présenté par Baron-Cohen (1995), implémenté par la suite en robotique par Scassellati (2001), afin de décrire les quatre mécanismes ou modules qui se développent parallèlement à l'indexation des objets du monde. Ces mécanismes sont à la base même de la capacité humaine, capacité universelle, à lire l'esprit de l'autre. Selon Baron-Cohen (1995), ils reflètent quatre propriétés fondamentales de notre relation au monde, qui sont la volonté, la perception, l'attention partagée et les états épistémiques.

### **II.4.1. Le détecteur d'intentionnalité (ID)**

Ce premier mécanisme perceptif représente le mouvement autopropulsé de stimuli en termes d'états mentaux volontaires primitifs de buts et désirs. Ce module



est appelé détecteur d'intentionnalité (ID, *Intentionality Detector*). Ce module est nécessaire pour donner le sens de ces mouvements universels que sont l'approche et l'évitement. ID est activé à chaque fois que le système perceptif détecte quelque chose qui peut être un agent. Il produit des représentations dyadiques qui décrivent ces mouvements de base d'approche et d'évitement. Ce module opère sur les stimuli qui ont un mouvement autopropulsé et fournit un critère pour distinguer les stimuli qui sont potentiellement animés (agents) de ceux qui ne le sont pas (objets).

Son rôle de module de base est d'importance pour la lecture de l'esprit de l'autre, car il reçoit une large gamme de signaux provenant de toutes les modalités, audition, vision, toucher. Il devient totalement prépondérant chez des personnes aveugles, où il se déclenche continuellement pour se représenter les expériences tactiles vécues par le sujet. Chez l'adulte, et même chez l'enfant, des formes en mouvement seront spontanément traduites en terme d'agents conduits par des états mentaux, comme ont pu le démontrer Heider et Simmel dès 1944<sup>17</sup> (cité par Baron-Cohen, 1995).

Baron-Cohen souligne que le module ID est similaire au mécanisme ToBy (*Theory of Bodies*) proposé par Leslie (1994), dont la fonction est de permettre notre compréhension de la causalité physique. ID est également similaire à ce que Leslie nomme ToMM1 (*Theory of Mind Mechanism system 1*), qui interprète les actions réalisées par les agents comme étant dirigées vers un but. ToMM1 est ainsi exclusivement centré sur l'agent, ce qui s'oppose d'une certaine façon à la vision de ID par Baron-Cohen, qui va également détecter le caractère dirigé vers le but de mouvements non-agentifs.

## II.4.2. Le détecteur de direction du regard (EDD)

Le second module appelé détecteur de direction de l'œil (EDD, *Eye-direction Detector*, que possède déjà le serpent) traite les stimuli uniquement visuels (contrairement à ID) pour déterminer la direction du regard. Il a été très bien étudié du point de vue neural par Perrett (1999) et collaborateurs. Il a été intégré par Baron-Cohen (1995) comme étant un module appartenant à un très haut niveau modulaire, celui de la théorie de l'esprit. Ce module est composé de neurones qui sont

---

<sup>17</sup> Heider F., Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

spécifiques de la détection de la direction de l'œil ou à défaut de la face. Dans une situation où les yeux vont être cachés, par des lunettes de soleil par exemple, le système EDD va attribuer l'orientation du regard à la direction de la face. Plus généralement, les neurones déchargent à chaque fois qu'il existe un risque qu'un individu soit en train de vous dévisager. Il peut être schématisé de la façon suivante :

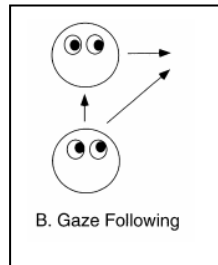


FIGURE II.14 : Le système EDD (D'après Emery, 2000).

Ce module a trois fonctions de base. Premièrement, il détecte la présence de stimuli qui ressemblent à des yeux. Les enfants humains aussi jeunes que deux mois préfèrent regarder des visages humains et passent significativement plus de temps à regarder les yeux que les autres parties du visage. Voilà pourquoi nous notions précédemment que les agents étaient les objets les mieux reconnus par les enfants.

Deuxièmement, EDD vérifie si les yeux sont en train de vous regarder ou s'ils regardent autre chose. Baron-Cohen (1995) émet l'idée que le fait qu'une personne extérieure attire le contact de l'œil produit un déclenchement psychologique tout à fait naturel, qui donne du plaisir à l'enfant. Ce qui n'est pas forcément le cas chez les autres primates, chez les autres mammifères de notre entourage, où un contact de l'œil prolongé peut parfois entraîner des alertes négatives (comme chez les babouins, qui se sentent menacés, comm. pers. de J. Vauclair). On peut noter que des enfants âgés de 6 mois vont regarder deux à trois fois plus longtemps un visage qui les regarde, plutôt qu'un visage regardant dans une autre direction. Chez l'enfant, le contact de l'œil présente des conséquences plutôt plaisantes, liées à des émotions positives, allant parfois jusqu'à déclencher le sourire. L'enfant possède par ailleurs un régulateur du degré de contact de l'œil, mis par exemple à contribution dans le jeu du « coucou » ou « peekaboo », où il suit le contact de l'œil à travers des phases d'occultation et de dévoilement des yeux de la maman.

Enfin, troisièmement, EDD interprète la direction du regard comme un état perceptif, c'est-à-dire que ce module code des états représentationnels dyadiques

agentifs de la forme « cet agent me regarde » ou « cet agent ne me regarde pas » pour reprendre les exemples de Baron-Cohen (1995). Cela implique évidemment la compétence de l'enfant à s'imaginer que des yeux peuvent voir, compétence acquise très tôt par le simple fait de fermer et ouvrir les yeux par l'enfant. L'enfant distingue très tôt les notions de voir/ne pas voir, mais également de voir A et voir B. EDD permet ainsi à l'enfant de lire les comportements en terme d'un petit ensemble d'états mentaux, comme le but, le désir et le fait de voir.

Ce système, qui est présent dans notre cerveau, sert donc bien plus qu'à suivre de façon précise la direction dans laquelle se dirige le regard d'autrui.

Dans la communication face-à-face, celui des deux qui parle à l'autre personne, ne va pas constamment suivre le regard de son interlocuteur. Il va simplement vérifier que l'attention de l'autre est suffisamment soutenue, en jetant de temps à autre un regard sur ses yeux. Par contre, l'interlocuteur attentif ne dévie pas son regard du visage de l'autre au niveau des yeux. Le bénéfice gratuit de cette différence de comportement est pour l'interlocuteur de pouvoir traiter les mouvements rapides présents dans la parole. Suivant du regard les yeux de l'autre (et non la direction du regard de l'autre) il est en vision dite focale, centrale ou fovéale, ce qui lui donne une image nette des yeux de son interlocuteur. Mais, en plus de cette vision centrale, le sujet va pouvoir récupérer les mouvements des lèvres, très rapides dans la parole (de l'ordre de six mouvements d'ouverture de la bouche par seconde), par le système de la vision périphérique, qui donne une image floue mais qui est spécialisé dans la détection de la vitesse. Cette « double vision » permet ainsi de récupérer l'information visuelle donnée par les mouvements de lèvres mais également de donner une preuve à son interlocuteur que l'on suit bien sa conversation.

Toutefois, les représentations dyadiques fournies par ID d'une part et EDD d'autre part sont plutôt limitées. Ce type de relation spécifie uniquement la relation d'intention qui peut exister entre deux entités ou objets (agent et objet ou agent et self), mais ne représente pas une relation où deux agents portent attention à un même objet, caractéristique fondamentale d'une réalité partagée. Plus simplement, si nous ne possédions que ID et EDD, nous serions alors incapables de partager des expériences avec autrui, et ainsi toute communication se verrait réduite à un état autistique. Il nous faut donc développer une nouvelle compétence qui nous permettra

d'entrer dans une représentation triadique, qui spécifiera une relation entre soi, un agent et un objet (ou un troisième agent) d'intérêt commun, ce qui offrira un partage d'attention entre les deux premiers agents. Cette relation triadique apparaît avec le troisième mécanisme proposé par Baron-Cohen (1995).

### **II.4.3. Le mécanisme d'attention partagée (SAM)**

Les deux premiers modules que nous venons d'explorer, ID et EDD, forment ainsi à eux deux la base d'un début de détection d'agentivité. Ainsi, l'existence de ce troisième module se base sur les représentations apportées par ID et par EDD. Ce troisième module se nomme le mécanisme d'attention partagée ou SAM (*Shared Attention Mechanism*). Ce dernier prend les représentations dyadiques de ID et EDD et produit des représentations triadiques de la forme : « John regarde [que] (je regarde la fille) ». Cet état attentionnel partagé est le produit d'un enchâssement d'une représentation dyadique à l'intérieur d'une autre. Le module SAM rend la sortie d'ID disponible au module EDD, ce qui permet l'interprétation de la direction de l'œil comme un état-but. En permettant à l'agent d'interpréter le regard des autres comme des intentions, le module SAM fournit un mécanisme pour créer des représentations emboîtées de la forme : « John voit [que] (je veux le jouet) ». Ce module nous apporte ainsi la spécification que simultanément un agent externe et un individu peuvent prêter attention à un même objet, agent ou événement. Cet état attentionnel est permis par le fait que le détecteur d'intentionnalité se rend accessible au détecteur de direction de l'œil comme pointant sur un événement ou sur un objet-cible. Selon Baron-Cohen (1995), ce mécanisme devient disponible chez l'enfant entre neuf et dix-huit mois.

L'approche modulaire proposée par l'auteur est soutenue ici par une évidence liée à la cognition sociale comparative. En effet, Baron-Cohen a pu noter que plusieurs espèces d'invertébrés utilisent la direction du regard comme un déclencheur social, mais que seuls peu d'espèces non-humaines semblent inférer l'intentionnalité. Cette thèse soutient alors le point de vue que EDD et SAM soient deux modules bien dissociés. De plus, la question de la spécificité de SAM reste toujours d'actualité : une évidence grandissante tend à montrer que certains primates non-humains portent attention au regard de leur congénère et utilise ce déclencheur

social pour prédire leurs intérêts, leur attention, voire même peut-être leurs intentions.

Baron-Cohen et ses collègues<sup>18</sup> ont voulu tester la véracité d'états mentaux inférés par les yeux chez les enfants, et ont trouvé que des enfants âgés de neuf à dix-huit mois répondent à des actions ambiguës en regardant instantanément les yeux de l'adulte, afin de désambiguïser le but dans les yeux de la personne. C'est exactement ce que les enfants font lorsqu'ils sont confrontés à la figure II.15 présentée ci-dessous : ils se servent de la direction du regard indiquée par le visage pour répondre correctement à la sucrerie que le visage indique.

Il convient ici de faire une parenthèse explicative pour différencier des termes souvent utilisés indifféremment dans la littérature pour référer à un même concept, mais qui nécessitent pourtant d'être quelque peu nuancés. Suivant l'idée exposée par Emery (2000), il convient ainsi de différencier la notion d' « attention partagée » de la notion d' « attention conjointe », même si cette différence demeure toutefois subtile. Toujours selon Emery, l'attention conjointe est un mécanisme qui va requérir que deux individus, X et Y, portent attention à un même objet, l'objet Z. L'individu X va détecter que le regard de Y n'est pas dirigé vers lui, et va suivre la ligne de vue de Y afin de repérer le focus de son attention. Ainsi X et Y regardent le même objet, comme nous pouvons le voir sur la figure C tirée de Emery (2000) (Cf. Figure II.15). L'attention dite partagée est un mécanisme un peu plus complexe que l'attention conjointe. En effet, l'attention partagée semble être une combinaison entre attention mutuelle et attention conjointe, où le focus d'attention de X et Y porte à la fois sur l'objet d'intérêt Z mais également sur chacun d'eux que l'on traduit par « Je sais que tu regardes Z, et tu sais que je regarde Z », comme le montre la figure D tirée elle aussi de Emery (2000) (Cf. Figure II.15).

---

<sup>18</sup> Phillips W., Baron-Cohen S., Rutter M. (1992). The role of eye-contact in the detection of goals: Evidence from normal toddlers, and children with autism and mental handicap, *Development and Psychopathology*, 4, 375-383.

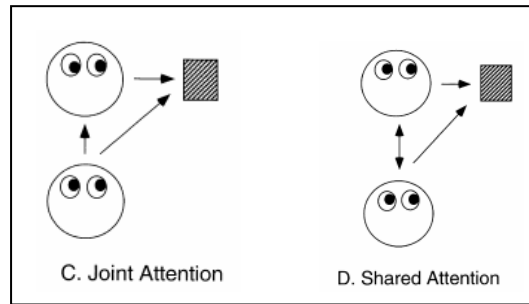


FIGURE II.15 : Attention conjointe vs. attention partagée (D'après Emery, 2000).

Selon Emery, le suivi du regard ainsi que l'attention conjointe sont représentés à l'intérieur du module EDD, et non pas dans SAM, qui ne contiendrait alors que l'attention partagée. Dans un commentaire proposé à Baron-Cohen, Perrett et Emery ont essayé de différencier plus finement les modules, en proposant un module de détection de direction de l'attention nommé DAD (*Direction of Attention Detector*) qui traiterait les déclencheurs d'attention potentiels provenant de la tête, des yeux ou même du corps. Ils proposent un second module, le mécanisme d'attention mutuel MAM (*Mutual Attention Mechanism*) utilisé pour détecter le regard mutuel. Le recours à ces deux nouveaux modules permettrait alors d'expliquer plus finement la différence entre attention conjointe et attention partagée : l'attention conjointe requiert l'activation de EDD et de DAD, tandis que l'attention partagée (ou SAM) requiert l'activation de EDD, DAD et MAM.

L'attention partagée constitue réellement un processus fondamental pour l'interaction et la cognition sociale, puisqu'elle se caractérise par le fait qu'elle peut ré-orienter l'attention vers une cible, du moment que celle-ci demeure l'objet d'attention d'une autre personne. Ce module semble alors incontournable pour faciliter la communication et pour inférer des états mentaux à partir du comportement d'autrui.

L'attention partagée se manifeste par la faculté à suivre le regard de l'autre, une faculté qui émerge vers 9 mois chez l'enfant. Toutefois le concept même d'attention partagée va bien au-delà du simple suivi du regard. Outre son rôle de guide d'attention au sein de l'environnement, ce module va également déterminer si la personne partage une expérience avec quelqu'un d'autre, et va déterminer quel est l'objet d'intérêt qui a déclenché cette attention commune.

Que sait-on de l'attention conjointe chez les primates non-humains ? Sont-ils dotés des mêmes capacités présentes chez l'enfant à ce stade du développement ? Pour répondre à cette question, nous nous appuyerons sur l'étude menée par Kumashiro et al. (2003) sur des singes japonais (*Macaca fuscata*).

Chez l'enfant, ce mécanisme est très étudié et consiste à mettre en évidence l'émergence d'un échange triadique, qui constituerait une première tentative de la part de l'enfant à intégrer simultanément l'objet d'intérêt et un engagement humain sur le focus attentionnel. Baron-Cohen (1995) et Tomasello (1995) ont, entre autres, bien montré que les enfants s'engagent dans une variété de comportements indiquant qu'ils perçoivent bien un engagement de l'agent dans les actions guidées vers un but.

Il semble aussi que chez les petits singes, il existe une compétence de reconnaissance de l'être humain comme agent intentionnel. Selon Kumashiro et al. (2002), le petit singe serait capable d'acquiescer le pointer communicatif et les gestes de fixation du regard qui sont nécessaires pour réaliser le phénomène d'attention conjointe. Kumashiro et al. (2003) ont voulu tester l'existence d'un tel comportement d'attention conjointe chez ces petits singes, en testant leur aptitude à imiter naturellement ou non les actions humaines. Leur étude s'est déroulée en trois parties : une action-test, un mouvement puis à nouveau une phase de test, sur 4 singes. L'action-test observe le comportement du singe en réponse à plusieurs présentations d'actions humaines : une action orale (protrusion de la langue) et quatre actions manuelles (par exemple toucher un bouton). La phase « mouvement » examine si le singe, qui montre une utilisation unidirectionnelle du pointer et qui ne montre pas le suivi du regard, est capable ou non de reproduire le mouvement humain et de contrôler avec précision sa reproduction. Enfin une dernière phase test examine l'imitation naturelle du singe après que celui-ci ait acquis la capacité d'attention conjointe. Les résultats pour cette étude montrent que les deux singes qui possédaient déjà l'attention conjointe étaient capables d'imiter les actions de protrusion et les actions de main dans la première partie « action-test » du dispositif expérimental. Le singe qui possédait une attention conjointe incomplète (il utilisait le pointer impératif et la fixation du regard, mais n'utilisait pas cette même fixation dans un contexte communicatif) pouvait reproduire la protrusion de la langue, mais pas toutes les actions manuelles (comme l'action de trier du coton). Ce singe était

intéressant dans la mesure où il offrait le potentiel de faire abstraction de la similarité entre humain et singe quand l'attention était dirigée vers le modèle d'action. Et, après acquisition du suivi de regard triadique (avec pointer), il devint capable d'imiter l'action humaine naturellement, similairement aux deux singes précédents.

Les deux types de singes (avec ou sans attention conjointe comme pré-requis) reproduisent les protrusions de la langue. Il semble qu'ils possèdent une tendance naturelle à diriger leur attention sur les mouvements de lèvres des autres individus. Cette attention pourrait alors faciliter la construction de représentations internes de mouvements de bouche, qui pourraient être ensuite utilisées pour reproduire le même mouvement. Le mouvement de bouche semble en tous cas être plus facilement reproductible que les mouvements d'autres parties du corps. Cette idée trouve également un soutien dans les études sur les enfants humains, réalisées notamment par Meltzoff et Moore (1983, 1989), qui montrent cette même tendance d'imitation plus précoce de mouvements de bouche que de mains. Toutefois, la plupart des singes qui ne possèdent pas l'attention conjointe avec les humains se montrent aversifs au regard, et n'observeraient ainsi pas assez les lèvres de l'individu humain pour reconstruire une représentation, entraînant de ce fait un manque de comportement imitatif.

La découverte importante de cette étude de Kumashiro et al. (2003), présentée ici, réside dans le fait que l'action des singes suit l'action des hommes dans les phases de test, et que le mouvement des singes suit le mouvement des hommes dans la phase « mouvement ». Ceci tend donc à prouver qu'il existe chez ces singes le suivi de l'action et le suivi du mouvement. Les auteurs spéculent alors sur le fait que l'imitation soit basée sur un mécanisme de suivi, et l'imitation naturelle (suivi de l'action) serait induite par l'attention conjointe. Ainsi, l'acquisition par les singes du geste communicatif œil-regard permettrait à ceux-ci d'acquérir la capacité d'imitation naturelle.

Selon les auteurs, il pourrait exister deux types d'approches comportementales pour l'imitation des actions et mouvements des autres. Une première approche qui relèverait de l'imitation naturelle, sur la base de l'attention conjointe, à travers l'utilisation du regard et du pointer communicatif. L'observateur porte attention à celui qui exécute l'action comme un agent attentionnel manipulant un objet cible. Une seconde approche qui concerne la reproduction du mouvement,



sur la base d'une attention conjointe incomplète, en portant attention sur un mouvement local. Ce type d'observateur peut porter attention au mouvement en étant indifférent à la présence d'un agent attentionnel.

Les singes japonais sont donc capables de reproduire les actions et les mouvements des êtres humains en l'absence d'apprentissage d'instructions par le langage. Ils peuvent enrichir le répertoire de leurs actions par l'imitation naturelle, à travers l'attention conjointe exprimée par l'utilisation bidirectionnelle du pointer communicatif et des gestes de suivi du regard. L'imitation naturelle semble donc constituer une forme fondamentale d'apprentissage social nécessaire pour partager, utiliser et comprendre une action commune.

#### II.4.4. Le mécanisme de la théorie de l'esprit (ToMM)

Suite au module SAM, l'enfant acquiert plus tard dans l'enfance le quatrième et dernier module : le module de la théorie de l'esprit (ToMM, *Theory of Mind Mechanism*). Ce terme de théorie de l'esprit est emprunté ici par Baron-Cohen à Alan Leslie (1994), désignant le système d'inférence d'états mentaux extraits à partir des comportements. Selon Baron-Cohen (1995), ToMM requiert : (i) la représentation de l'ensemble des états mentaux épistémiques (penser, croire, imaginer, exprimer...); (ii) un moyen de lier les concepts d'états mentaux (de volonté, de perception et épistémiques) dans un tout cohérent visant à comprendre comment états mentaux et actions peuvent être reliés. Il peut être schématisé de la façon suivante :

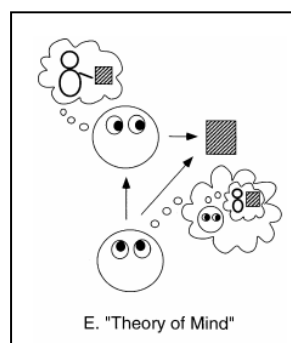


FIGURE II.16 : Le mécanisme de la théorie de l'esprit (D'après Emery, 2000).

Le module ToMM permet la construction de représentations telles que « John croit [que] (il pleut) ». Il permet également la suspension de relations de vérité de propositions (opacité référentielle), ce qui offre une possibilité de signification pour

représenter des états de connaissance qui ne sont pas nécessairement vrais, comme « John pense [que] (Elvis est [encore] en vie) ».

Une explication des deux niveaux (ordres) de ce mécanisme est clairement donnée dans Tourette et al. (2000) :

L'enfant commence par comprendre, en partageant avec lui des actions et des co-références, que les actions d'autrui sont guidées par des états mentaux. Simultanément il prend conscience de ses propres états mentaux et découvre que les états mentaux d'autrui peuvent être différents de ses propres états mentaux. Lorsqu'il devient capable d'attribuer à autrui des états mentaux différents des siens propres, cela signifie qu'il dispose de capacités méta-représentationnelles (ou représentations de second ordre, c'est-à-dire distinctes des représentations du réel, dites de premier ordre). (2000:62)

Selon Baron-Cohen, ID et les fonctions de base de EDD deviennent disponibles aux enfants dans les 9 premiers mois de leur vie. Le module SAM se développe plus tardivement, entre 9 et 18 mois. Enfin ToMM se développe de 18 à 48 mois. Afin de visualiser cette progression développementale, voici un schéma, qui a été repris dans la robotique de Scassellati (2000) :

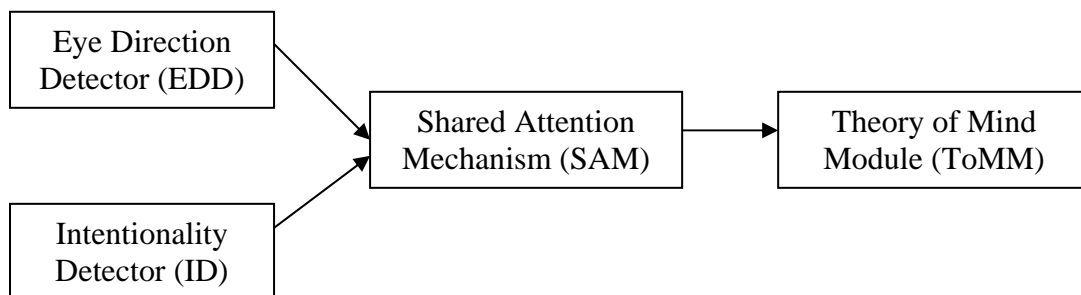


FIGURE II.17 : Vue d'ensemble du modèle de Baron-Cohen du développement de l'attention conjointe et de la théorie de l'esprit (D'après Scassellati, 2000).

Pour reprendre le cas de l'autisme chez l'enfant, Baron-Cohen (1995) considère que les déficiences observées pour cette pathologie ne touchent ni le détecteur d'intentionnalité, ni le mécanisme de détection de direction de l'œil, mais en revanche dans la plupart des cas c'est le mécanisme de la théorie de l'esprit qui est altéré (parfois chez quelques patients ce sont à la fois le mécanisme de la théorie de l'esprit et le mécanisme de l'attention partagée qui sont altérés). Tourette et al. (2000) soulignent que les recherches auprès de jeunes enfants atteints d'autisme ont mis en relief le fait qu'ils ont un déficit en attention conjointe préverbale et en théorie de l'esprit. Ces deux altérations semblent liées dans la mesure où la première paraît nécessaire pour le développement de la mentalisation, fonction qui constitue un pré-requis à la théorie de l'esprit. Selon les chercheurs, ce sont les capacités méta-représentationnelles qui paraissent faire défaut aux enfants autistes.

Si l'on compare ces différents modules à ceux existant chez les animaux, le système EDD semble être présent le plus fréquemment chez certains animaux comme les invertébrés. Quant au détecteur d'intentionnalité, il est présent chez de nombreux primates, à la différence du système d'attention partagée, qui semble n'être disponible que chez les grands singes. En ce qui concerne le module de la théorie de l'esprit, il paraît pour l'instant non prouvé d'attribuer cette capacité aux autres primates, car ils ne paraissent pas inférer aisément des croyances ou états mentaux chez les autres individus.

#### **II.4.5. Pourquoi les agents sont-ils les mieux reconnus ?**

Les enfants sont très tôt dans leur petite enfance très attirés par le caractère agentif des choses qui les entourent. Ils semblent posséder précocement une conscience de ce qui est biologique, ou de ce qui n'est pas biologique. Pour l'enfant toute chose s'apparente à un agent lorsque celle-ci possède un visage, ou lorsque la chose réagit quand on l'excite. Dans le cas d'un mouvement d'auto-propulsion, comme une boule poussant une autre boule, l'enfant considère qu'il s'agit d'un agent. De la même façon, si une plante a une bouche (carnivore), ou si un objet magnétique quelconque possède un œil (deux exemples tirés de la vague, très appréciée par Guillaume, un enfant de C. Abry, vers ses 8 ans, des Pokémons ou plutôt des Digimons, ces derniers se métamorphosant ou « digivolvant »), l'enfant les considère comme agents. Les enfants semblent ainsi posséder très tôt une sémantique des classes, où des catégories comme les agents, les outils, etc. sont très clairement distinguées. Et dans cette sémantique, les enfants semblent connaître les agents, bien avant les objets.

De plus, les notions d'objet, de numérosité, d'agent, sont des notions essentielles à saisir afin de comprendre les mécanismes sur lesquels reposent nos capacités cognitives. Le débat se situe toujours sur l'innéité ou l'acquisition des-dites-capacités. Les études de Johnson (2000) s'attachent à ce propos à découvrir l'existence du concept d'agent dans la phase pré-linguistique infantile. Les enfants étant considérés comme encore non-verbaux jusqu'à l'âge approximatif de deux ans, il sera intéressant de distinguer le rôle que peut avoir le langage pour la représentation de cette notion, à travers le pré-linguistique chez l'enfant.

La notion d'agent est en elle-même très pertinente dans la mesure où elle permet de récupérer des informations sur les états mentaux, qui en général sont plutôt inférés que perçus. La spécificité de ces états mentaux consiste en plus à entretenir une relation avec le monde, par rapport à d'autres qui ne sont pas observables non plus et qui n'entretiennent pas ce type de relation. Le concept d'agent permet d'avoir des prédictions et des explications sur le comportement des autres individus. Par ailleurs certains chercheurs vont jusqu'à supposer que l'absence de cette notion importante peut être considérée comme une cause principale de l'autisme (voir plus haut).

Une notion cruciale qui régit les compétences humaines, sociales et linguistiques repose sur cette capacité que nous avons à construire les autres comme des agents, et à nous construire nous-mêmes en tant qu'agent, qui possédons des états mentaux, tels que la perception, l'attention, les désirs, les croyances. La connaissance de cette capacité nous entraîne naturellement à nous demander à partir de quel moment elle est rendue disponible, et par quels moyens elle l'est. Afin de trouver des réponses à ces questions cruciales, de nombreuses études ont été menées sur les enfants en bas âge, en utilisant ce que l'on nomme le paradigme des « fausses croyances », qui consiste à entraîner des comportements qui ne sont pas prédits. Selon Dennett (2000), le raisonnement entraîné par ces fausses croyances tient à prouver qu'en s'intéressant à certains comportements, tel que le pointer, on pourrait récupérer des indices sur l'attribution d'états mentaux. Dennett (2000) se tourne plus particulièrement vers l'existence ou non de linéaments d'une « théorie de l'esprit » très tôt dans l'enfance. C'est lui-même qui a entraîné le courant de la théorie de l'esprit, justement sur la base du développement de ces fausses croyances, qui sont un succès chez les enfants de deux à quatre ans. Cependant, le fait que les enfants avant quatre ans fassent preuve d'une absence de raisonnement de fausses croyances ne signifie pas que l'enfant est dénué d'aptitude à attribuer des états mentaux aux autres individus ou à d'autres agents. Il existe des traces avant l'âge de deux ans chez les enfants à travers l'émergence du vocabulaire mentaliste, tels que les verbes « vouloir » et « voir ».

Sans nous intéresser plus en détail à ce paradigme tardif des fausses croyances, que pouvons-nous dire des attributions d'états mentaux de l'enfant lors des interactions précoces qu'il entretient avec les personnes qui s'occupent de lui ?

Afin de retrouver l'émergence de cette capacité, Johnson (2000) oriente sa recherche dans trois domaines : les interactions entre l'enfant et autrui ; les observations passives entre l'enfant et autrui ; et enfin les interactions et observations entre les enfants et les agents non-humains.

#### **II.4.5.1. Interactions entre l'enfant et autrui**

Les premières traces d'attributions mentalistiques s'effectuent à travers les premiers gestes de communication et d'attention, qui se déroulent dans la période d'âge de neuf à douze mois. A ce stade, les enfants commencent petit à petit à comprendre puis produire des gestes, comme par exemple le pointer, le fait de montrer, de demander. Ces comportements sont typiques du désir de l'enfant de diriger l'attention des autres sur un objet, un agent ou une autre partie du monde. De même à cette même période, les enfants commencent à suivre l'attention des adultes, en alternant leur propre regard entre les personnes adultes et les objets présents dans la scène.

Baldwin (cité par Johnson, 2000) a montré qu'à l'âge de dix-huit mois, les enfants interprètent un mot nouveau logiquement en fonction du référent qui est suggéré par le regard fixe de celui qui s'adresse à l'enfant, lorsqu'il prononce le mot, et non pas en se référant à leur propre regard. En revanche, entre quatorze et seize mois, l'enfant n'arrive pas à apparier correctement un nouveau mot avec un référent. Le problème repose sur le conflit qui existe entre l'objet qui relève de leur regard propre et celui montré par leur interlocuteur. Mais l'enfant utilise tout de même l'attention de celui qui s'adresse à lui pour ne pas identifier ou apparier de façon incorrecte les objets auxquels il prête attention. Carpenter et ses collègues (Johnson, 2000) ont par ailleurs mis en avant le fait qu'entre 14 et 18 mois, les enfants peuvent volontairement imiter les actions réalisées par un adulte, si l'action de ce dernier a été préméditée et n'est pas survenue de façon accidentelle. A ce propos, Meltzoff (d'après Johnson, 2000) confirme que les enfants montrent une capacité à comprendre les buts, quand cette technique d'imitation est utilisée, dès 18 mois.

#### **II.4.5.2. Observations passives entre enfant et autrui**

Nous pouvons également faire appel à l'étude de Woodward (1998), qui par le biais de l'accoutumance visuelle, cherche à tester si les bébés effectuent un codage

des actions réalisées par les individus humains en termes de buts ou de mouvements spatio-temporels ou trajet (*path*) L'expérience rapportée repose sur trois conditions d'exposition :

- a) celle de l'événement-contrôle, pour l'accoutumance ;
- b) celle de l'événement-test, dans lequel le trajet spatio-temporel de la main est changé, mais où l'objet-cible reste le même ;
- c) celle de l'événement-test, dans lequel l'objet-cible de la main est changé, mais où le trajet spatio-temporel reste le même.

La moitié d'un groupe d'enfants va être habituée à l'événement d'une main qui s'approche d'un ou de deux objets disposés sur une petite estrade. Une situation d'événement-test va ensuite consister à changer soit le trajet spatio-temporel de la main ou soit l'objet-cible. L'hypothèse de Woodward (1998) est que les enfants coderaient les actions de la main comme des buts, reflétant essentiellement une relation agent-monde. Dans ce sens, les appréciations provoquées lorsque l'objet-cible (ou le but) est changé devraient être plus inattendus, donc plus étonnants pour les enfants, que lorsque c'est le trajet de la main qui change. Ce qui se trouve être le cas : les enfants de 5 mois et de 9 mois sont étonnés par le changement de cible de la main et non par le changement de trajet de la main.

Woodward (1998) a réalisé la même expérience avec des enfants qui ont été placés dans les mêmes conditions, à la différence près que l' « agent » a été remplacé par une baguette. Les résultats montrent des patrons assez différents : aucune classe d'âge d'enfants n'est étonnée par le changement de l'objet-cible de la baguette. Ceci a pour conséquence de démontrer qu'ils n'établissent pas un codage de la relation entre la baguette et l'objet, qui n'est alors pas considérée comme un fait saillant de l'événement.

Cette expérience a pu ainsi démontrer que dès l'âge de cinq mois, les bébés sont capables de reconnaître, d'un côté, que le comportement vaut parce qu'il est dirigé vers le monde et, d'un autre côté, que l'individualité des concepts-cibles est saillante. Par conséquent, au regard de ces conclusions, Woodward (1998) pense que les bébés ont une relation intentionnelle entre l'objet et le monde.

Phillips et ses collègues (cités par Johnson, 2000, repris de Spelke et al., 1995) ont de leur côté posé l'hypothèse qu'à l'âge de 12 mois, un enfant comprend

que les désirs prédisent des actions. Ainsi leur temps de regard est prolongé lorsqu'il voit un adulte sourire en direction d'un objet et —au lieu de se saisir de l'objet en question— se saisir d'un objet différent.

Ces résultats montrent que les enfants reconnaissent des agents mentalistes. Mais cette capacité précoce peut être traitée de façon alternative en envisageant la présence de réponses conditionnées chez l'enfant. Dès ses premières années de vie, l'enfant peut observer les individus et comprendre des actions sur le monde. Dans ce sens, cela signifierait que les enfants auraient acquis un nombre d'associations appropriées, sans spécialement avoir attribué des états mentaux. Sur cette idée, Corkum et Moore (cités par Johnson, 2000) ont démontré qu'on peut développer, chez l'enfant entre 8 et 9 mois, un conditionnement à suivre du regard ; alors que sans ce conditionnement, il échoue à suivre le regard de façon spontanée.

#### **II.4.5.3. Les interactions et observations entre les enfants et les agents non-humains**

Cependant un certain nombre de travaux convergent vers la présupposition du rôle crucial de l'attribution d'agentivité pour l'attribution d'états mentaux. Ce dernier champ d'étude concerne les interprétations que font les enfants sur des agents non-humains. Plusieurs postulats de « psycho-physique naïve » se mêlent pour l'identification de la présence d'agents mentalistes. Par exemple l'importance de spécificités morphologiques telles que le visage et les yeux, l'asymétrie le long d'un axe antéro-postérieur, l'autopropulsion ou la capacité à s'engager dans des interactions avec d'autres agents. Ordinairement, les enfants sont très sensibles aux spécificités morphologiques.

Johnson (2000) a prouvé que la présence du visage ou un comportement interactif, même sans visage, ont tous deux pour conséquence la reconnaissance ontologique d'un agent. Mis en présence d'objets nouveaux, et si ces objets ont un visage, les enfants sont enclins à en suivre le « regard ». De même s'ils réagissent de manière autonome (« contingente ») aux enfants. Johnson montre que c'est à l'âge de 12 mois que les bébés semblent capables d'attribuer des états mentaux (en perception, en attention vers autrui) ; et ils sont alors également capables d'attribuer ces mêmes états mentaux à de nouvelles entités, qui ne sont pas des agents.

Les études de Johnson, Slaughter et Carey (cité par Johnson, 2000) ont utilisé un objet nouveau pour l'enfant, pour explorer l'attribution d'états mentaux chez des enfants de 12 mois, dans le paradigme du suivi de l'objet. L'objet en question se présente sous différentes formes proposées comme agents mentalistes, avec des variations qui consistent en la présence ou absence : de spécificités faciales ou d'un comportement interactif avec l'enfant. Les auteurs ont trouvé que le bébé suit le « regard » de l'objet en déplaçant sa propre attention dans la direction de l'objet, plutôt que dans la direction opposée, direction stimulée lors des conditions de familiarisation. Ce regard de l'objet est suivi dans les conditions suivantes : (i) si l'objet a un visage ; (ii) si, lorsque l'enfant bouge ou babille, l'objet émet des bips et des flashes (situation d'interaction) ; (iii) si les deux conditions précédentes sont réunies. Les observations montrent en outre que l'enfant suit le regard de l'objet de façon beaucoup moins certaine lorsque ce dernier ne possède pas de visage, comme lorsque l'objet a une forme banale, en l'occurrence une boule de poils ou un jouet en peluche sans visage.

Les enfants semblent donc utiliser des répliques sélectives pour prendre la décision d'attribuer à un objet le fait qu'il dispose d'un « esprit », qu'il perçoit et prête attention, et ces répliques concernent en particulier la présence d'un visage et l'envie d'interagir réciproquement. En résumé, d'après Johnson (2000), que l'on considère les études de Meltzoff ou celles de Woodward, on aboutit au fait que les bébés traitent les contrôles des « agents » humains et non-humains différemment.

Quelques problèmes d'interprétation ont été soulevés par ailleurs à ce sujet, en apportant l'idée qu'une interprétation mentaliste des objets chez l'enfant n'engage pas forcément l'existence de processus d'identification de l'objet qui sont spécialisés pour cela (les répliques spécifiques). Certains chercheurs ont posé l'existence de libérateurs (déclencheurs ou *cues*) de signaux pour le suivi précoce du regard, sans forcément que l'enfant attribue au regard une relation d'un « esprit » au monde.

Povinelli et Eddy (Johnson, 2000) ont discuté l'idée qu'il peut exister une faculté qui se forme par évolution pour suivre un regard sans attribution de perception. Cette hypothèse repose sur des expériences menées sur le chimpanzé et montrant que ce dernier a la capacité de suivre le regard d'un humain, mais qu'il n'est pas capable de faire passer ses propres désirs à l'humain.



Les mécanismes qui ont été exposés ici se posent comme des bases pour la reconnaissance de l'objet chez l'enfant. Mais un certain nombre d'incertitudes demeurent, en particulier l'interaction des agents et les significations qu'ils véhiculent. Quelle que soit la théorie qui sera privilégiée, et comme l'ont déjà souligné Tomasello et Dunham (repris par Johnson, 2000), les attributions mentalistes seront argumentées plus fortement, si on arrive à démontrer que les différents comportements contextuels trouvent leur origine à l'intérieur d'une même fenêtre développementale.

Toutes ces notions sont encore une fois d'une importance primordiale pour notre système audiovisuel, car elles posent les principes de l'attention pour la communication, dans le sens où s'il faut intégrer des animats qui s'adressent à leur utilisateur, il faudra qu'ils possèdent de préférence un visage, sur lequel on pourra percevoir des intentions et des états mentaux. Ce qu'exploitent les roboticiens de l'affectivité, notamment pour l'autisme, comme ceux que nous allons rencontrer maintenant.

#### **II.4.6. Une étude des modèles de développement social via une robotique humanoïde**

Un des pionniers dans ce domaine, notamment pour le pointer, Scassellati (2000), propose d'utiliser sérieusement des robots anthropomorphes pour évaluer les modèles de développement social humain, ceci dans l'héritage de l'approche d'un Rodney Brooks au MIT. C'est dans cette même optique qu'avait pris naissance, dans cette équipe pionnière en la matière, le robot humanoïde COG (jeu de mot sur cognition et rouage). COG a été construit avec des systèmes sensoriels et des facultés motrices inspirés de l'humain, et il a été doté d'une architecture d'apprentissage originale, dite à « *subsumption* » (dès Brooks, 1986), soit avec des niveaux de connaissances qui pouvaient largement « s'ignorer » les uns des autres, tout en aboutissant à des comportements cohérents. Afin d'approcher le mouvement humain, COG possédait un total de 21 degrés de liberté, répartis sur un torse, une face et deux bras. Cette approche a été continuée par Cynthia Breazal avec les célèbres robots expressifs de tête (ou « affectifs ») *Kismet* puis *Leonard*. D'autres projets ont bien entendu cours : en Grande-Bretagne, où se signale particulièrement la robotique sociale de Dautenhahn avec les enfants autistes ; et au Japon, où dans

la même veine développementale, et avec un accent particulier sur le pointer, *Infanoid* de Kozima (2002) évolue depuis des années (cf. dernièrement un autre projet de communication face à face de Shiomi et al., 2006) ; jusques et y compris aux laboratoires Sony de Paris, avec le groupe de Steels et Kaplan (tous les sites de ces réalisations sont facilement accessibles sur le web). Cette implémentation robotique offre des opportunités uniques pour évaluer les modèles du développement, particulièrement en ce qui concerne le développement du comportement d'attention conjointe, qui regroupe généralement les systèmes de monstration et d'interrogation primitifs que l'on trouve chez les enfants. Cette attention conjointe est en effet une caractéristique critique pour l'apprentissage social au cours du développement.

Afin d'évaluer la performance du système de façon incrémentale, et de lier le comportement observé du robot avec le comportement observé chez les humains, l'ossature du robot doit être décomposable en un ensemble de comportements représentatifs des systèmes EDD, ID et SAM, que nous avons détaillés précédemment, par les deux modalités de contact sociales que sont l'œil et le pointer du doigt. Les quatre comportements qui doivent être testables et observables sont : (i) le maintien du contact de l'œil; (ii) le suivi du regard; (iii) le pointage impératif et (iv) le pointage déclaratif. La figure II.18 ci-dessous offre une représentation de ces quatre comportements. Ils ont été choisis car ils représentent les comportements représentatifs qui peuvent être compatibles avec la technologie cognitive computationnelle et systémique du robot. Ils apportent des améliorations non-négligeables pour alimenter le répertoire comportemental du robot humanoïde.

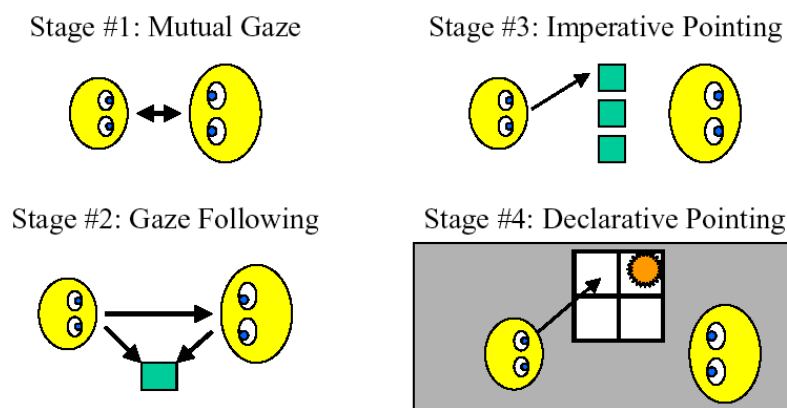


FIGURE II.18 : Une décomposition basée sur la tâche en quatre parties des bases de l'attention conjointe (Butterworth 1991, d'après Scassellati, 2000).

Le second mécanisme d'attention conjointe après le maintien du contact de l'œil, est le pointage. Le développement de ce dernier pour diriger l'attention, est basé sur un contrôle sensori-moteur plus complexe que le suivi du regard. Il requiert forcément chez le robot l'utilisation du bras, et la reconnaissance des déclencheurs (*cues*) gestuels.

Un des avantages relevés par Scassellati pour cette approche réside dans le fait que les observateurs humains anthropomorphisent facilement leurs interactions sociales avec un robot, de la même façon qu'avec un humain. De plus, les bases qui sont implémentées pour tester ces modèles peuvent être utilisées pour d'autres tâches robotiques. Le fait que le robot doit avoir la contrainte d'être de forme humaine, agit de façon importante sur son comportement : par exemple, pour observer un objet avec attention, le robot doit pouvoir orienter sa tête et ses yeux vers une cible. Ce faisant ces contraintes permettent aux utilisateurs d'interpréter facilement le comportement du robot et cela facilite ainsi les interactions entre humain et robot humanoïde.

De plus, le fait de pouvoir implémenter la reconnaissance et la production des comportements d'attention conjointe, va permettre des interactions personne/système qui n'étaient pas possibles auparavant. Le robot doit être capable d'avoir l'apprentissage d'un observateur qui utilise des signaux sociaux normaux, ceci de la même façon que le ferait un enfant humain en situation d'apprentissage. De plus, le robot doit alors être capable d'exprimer ses états internes lors d'interactions sociales sans s'appuyer sur un vocabulaire artificiel.

#### **II.4.6.1. L'œil comme premier outil d'indexation dans le monde**

L'œil est le premier outil disponible à l'enfant pour explorer l'environnement externe qui s'offre à lui, mais également pour l'indexer. Le contact de l'œil constitue une composante fondamentale de l'interaction avec le monde social qui l'entoure, et forme un bon indicateur des états émotionnels ou mentaux de l'autre. C'est pourquoi il est nécessaire d'en dégager les principaux mécanismes afin de mieux appréhender son rôle dans le développement de l'enfant et de mieux modéliser ses caractéristiques chez un robot anthropomorphe.

Nous avons mis en évidence l'importance de la mise en place de nombreux mécanismes divers pour créer ce cadre du signe, ou cadre sémantique chez l'enfant. Des notions telles que l'attention et l'intention demeurent des notions cruciales dans la compréhension des relations qu'entretiennent les agents avec les objets et événements de leur environnement. Nous avons expliqué l'importance des 7 grands systèmes pragmatico-sémantiques pour le suivi des agents, objets, événements. Nous avons expliqué le lien fort qui existait entre ces mécanismes et les différents modules d'attention qui se développent chez l'enfant. Mais l'idée sous-jacente ici est que toutes ces étapes charnières constituent la pierre angulaire de l'indexation du monde chez l'enfant, qui va lui permettre de mettre du sens sur les objets, les relations entretenues dans son monde.

Dès ses premiers mois de vie, l'enfant va explorer le monde qui l'entoure en exploitant son système sensoriel, et ce que ses différents capteurs pourront lui fournir comme indices l'aidera progressivement à se représenter cet univers. L'enfant va donc partir en exploration, en indexant ce monde avec les yeux en premier lieu. Les yeux permettent à l'enfant d'obtenir des connaissances très directes sur les objets de son monde, sur les agents de son monde, et lui permettent en outre de « lire » les états mentaux et émotionnels d'autrui, ne serait-ce que par les expressions du visage et du regard. Les yeux sont le point de focus principal dans le visage d'un autre pour un enfant et ils restent l'indicateur le plus précis de l'orientation de l'attention. Evidemment si ceux-ci sont occultés, l'enfant aura recours à d'autres indicateurs, que sont la direction de la tête, voire même l'orientation du corps tout entier, afin de dégager la direction d'attention ou d'intérêt de son interlocuteur.

Ce n'est pas seulement pour l'enfant en culture que le regard joue un rôle communicatif très fort : il est dans la nature vital pour plusieurs espèces. Emery (2000) cite, dans sa revue de question, le cas de plusieurs espèces de papillons qui ont développé un patron de forme d'œil, des leurres, sur leurs ailes afin de repousser leurs prédateurs potentiels, des oiseaux, leur faisant détecter qu'il pourrait s'agir d'yeux appartenant à un rapace, pour eux le prédateur. Quant à la détection d'un danger suite à la fixation par le regard, les iguanes, par exemple, détalent plus rapidement si vous les regardez plutôt que si vous les ignorez. De la même façon, nos poulets (*Gallus gallus*) répondent au regard maintenu d'un homme en adoptant

une immobilité tonique, un phénomène réputé, dans le règne animal, donner des chances de survie à celui qui fait le mort. Nous pouvons citer, plus spectaculaire encore, une espèce de serpent à nez de hérisson (le *hog nosed*) qui simule la mort en réaction à des yeux dirigés vers lui.

Des enfants aussi jeunes que 2 ou 3 mois sont très attentifs aux visages, reconnaissent les visages familiers, et sont plus sensibles à certaines expressions faciales que d'autres. Mais le trait social le plus important pour l'apprentissage précoce du partage d'attention est sans aucun doute l'œil, et des études ont montré que des enfants âgés de 3 mois fixent de façon disproportionnée les yeux d'un visage fixe.

#### **II.4.6.1.a) Implémentation de l'attention conjointe**

Les mécanismes d'attention conjointe requièrent une grande coordination entre plusieurs processus —perceptif, sensori-moteur, attentionnel, cognitif— qui incluent eux-mêmes les bases motrices de l'œil, ceux de la détection du visage, des yeux, de la direction de l'œil, de la reconnaissance d'un geste, enfin tous les systèmes d'attention dont la propriété est de pouvoir offrir le comportement social adéquat à la situation, et qui peuvent se traduire par un contrôle moteur du bras, une réponse émotive, etc.

#### **II.4.6.1.b) La maintenance du contact de l'oeil**

Détecter et répondre au contact de l'œil constitue la première étape de la structure développementale à implémenter, et requiert principalement des facultés perceptives. Le robot devra donc être capable de détecter les visages, de déterminer la localisation de l'œil à l'intérieur du visage, et de détecter ensuite si ce même œil regarde ou non le robot. Pour reconnaître, par exemple, si une nourrice regarde ou non le robot, on devra tenir compte de la position de l'œil à l'intérieur de la tête, et de la position de la tête à l'égard du corps.

#### **II.4.6.1.c) Le suivi du regard**

Le suivi du regard est un mécanisme crucial pour l'interaction sociale, et joue un rôle prépondérant dans la communication référentielle mère-enfant. Pour reprendre les termes de Butterworth et Grover (1990), l'attention visuelle conjointe constitue un « regard déictique » (*deictic gaze*) qui, selon eux, est un précurseur nécessaire à la mise en place des gestes déictiques, tel que le pointage manuel, qui

visé à amener l'attention d'une personne, en général les parents ou la nourrice, sur un objet d'intérêt :

Deictic gaze is thought to pave the way in development for deictic gestures, such as manual pointing, that draw attention to a particular object by locating it for another person. (1990:605)

Le mécanisme d'attention conjointe partagée poserait ainsi une brique fondamentale dans le partage de l'expérience entre mère et enfant, qui permettra l'acquisition du langage ultérieurement.

Une fois que le système est ainsi capable de détecter le contact de l'œil, il faut ensuite posséder trois bases supplémentaires pour réaliser le suivi du regard (d'après Butterworth et Grover, 1990), pour pouvoir : (i) extraire l'angle du regard ; (ii) extrapoler l'angle du regard par rapport à un objet distal ; et (iii) faire jouer les routines motrices pour effectuer une alternance entre l'objet distal et la nourrice.

A 6 mois (cf. Figure II.19), l'enfant possède la faculté de regarder correctement le côté de la pièce où sa mère porte son regard. En d'autres termes, le mouvement de la tête de la mère indique simplement la direction générale (droite/gauche). Et il ne pourra localiser que la première cible qui sera rencontrée dans son champ visuel. Il est incapable d'attribuer le signal de la mère à l'espace hors du champ visuel immédiat. Ainsi, à 6 mois, l'attention conjointe est restreinte aux cibles disponibles à l'intérieur du champ de vision de l'enfant.

A 9 mois entre en jeu le mécanisme d'attention conjointe visuelle nommé le mécanisme « écologique » (cf. Figure II.19). L'environnement naturel et sa structure différenciée permettent à l'enfant de réaliser la fonction communicative du signal adulte. Naturellement, ce qui attire l'attention de la mère et l'amène à se tourner dans cette direction, amène également l'enfant à porter son attention sur le même focus. Selon les auteurs, le mécanisme écologique permet ainsi « une rencontre des esprits » convergente vers l'objet. Les enfants utilisent ici une stratégie particulière en cherchant sur la ligne du regard un objet saillant. Ce qui est pertinent ici, c'est que même lorsque l'objet réel de l'attention est plus loin dans l'angle de vue, l'enfant reste néanmoins bloqué sur le premier objet qu'il rencontre. Ce mécanisme écologique paraît donc fortement dépendre des propriétés intrinsèques de capture d'attention des objets dans l'environnement, et des changements de direction dans le regard de la mère, dans une situation d'attention partagée. Bien que l'enfant ne cherche pas d'objets cachés avant 8 ou 9 mois, le comportement de la mère signale

la possibilité permanente d'un objet, potentiellement accessible dans son champ de vision.

A 12 mois, l'enfance commence à localiser les cibles correctement, soit la première, soit la seconde le long de la trajectoire visuelle, lorsque la cible est stationnaire. Il s'agit du mécanisme « géométrique » d'attention visuelle conjointe (cf. Figure II.19). Ce n'est pas avant 12 mois que l'enfant prêterait attention à l'objet le plus distal de la mère, il acquiert alors à ce stade-là ce que l'on nomme la *vergence*. Butterworth et Grover (1990) utilisent le terme « géométrique » puisque ce mécanisme semble impliquer l'extrapolation d'une ligne invisible entre les yeux de la mère et le référent du regard, qui sera mis en relation à partir de la position de l'enfant. Le mécanisme géométrique disponible à l'enfant semble encore, à cet âge-ci, se restreindre à son espace de perception (l'enfant ne se tourne vers la cible qu'à l'intérieur d'un angle visuel de  $40^\circ$ ), et les bébés de 12 mois échouent encore à porter attention à des cibles se situant derrière eux, hors de leur champ perceptif.

Ce n'est qu'à partir de 18 mois que l'enfant commence à entrer dans l'étape représentationnelle, dans laquelle il suit les angles du regard hors de son propre champ de vision, c'est-à-dire convergent sur un objet qui le forcerait à tourner la tête. A ce stade, l'enfant est aussi précis dans sa détection de cible, que celle-ci se situe en première ou en seconde position dans sa trajectoire visuelle. Il faut remarquer que lorsqu'une cible est présente dans son champ visuel, l'enfant ne va pas aller chercher une cible qui pourrait se trouver derrière lui. Néanmoins, lorsque son champ visuel perceptif est vide, l'enfant va être capable d'accéder à la portion invisible de l'espace.

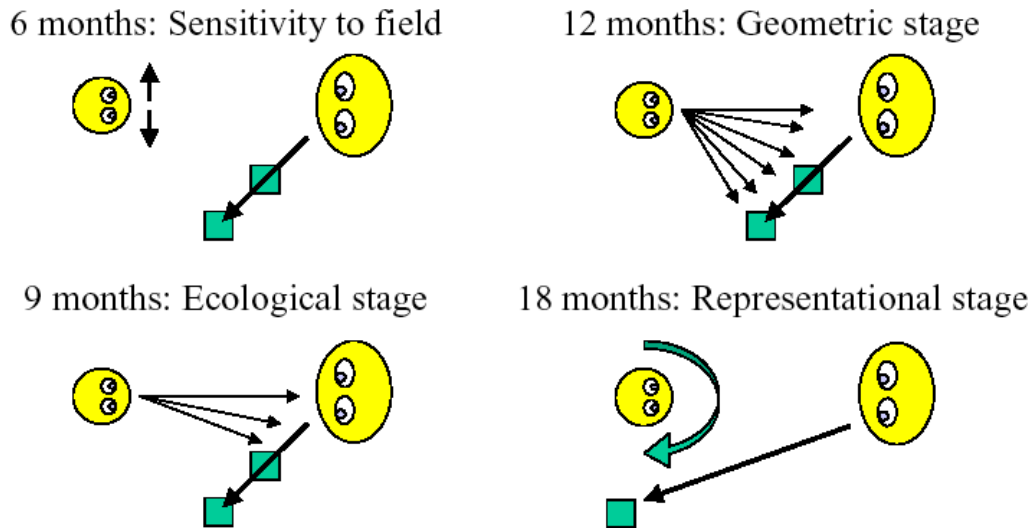


FIGURE II.19 : Progression développementale du suivi du regard (adapté à partir de Butterworth 1991, d'après Scassellati, 2000).

Ces différentes étapes développementales humaines sont cruciales et il est nécessaire d'implémenter cette progression dans un système robotique. Si le robot est capable de suivre l'angle du regard, il va pouvoir regarder les objets qui possèdent une couleur, une intensité ou un mouvement saillant, et il pourra alors imiter la stratégie écologique.

Deak et al. (2001, Fasel et al., 2002) ont proposé de comparer cette explication proposée par Butterworth et al. à celle proposée par Baron-Cohen. Les auteurs argumentent que la théorie de Butterworth souffre d'un problème majeur dans la séquence développementale même des mécanismes de changement. Selon Deak et al., tous les groupes d'âges d'enfants établissent l'attention conjointe comme une fonction de déclenchement social (fixation du regard), mais également comme une fonction écologique. Cela signifie que même lorsque les enfants ont acquis la faculté à extrapoler un vecteur de regard, les déclencheurs écologiques restent toutefois critiques et prédominants.

De la même façon, Deak et al. proposent une ré-interprétation des mécanismes proposés par Butterworth et al. Ils remettent en doute le fait que les changements qui s'opèrent entre six et douze mois chez l'enfant justifient l'apparition d'un nouveau mécanisme, le mécanisme représentationnel. Deak et al. penchent plutôt pour un mécanisme qui serait déjà présent et qui permettrait d'inférer des indices spatiaux, et qui par la suite, serait suffisamment arrivé à maturation pour fonctionner dans le contrôle du suivi du regard de l'enfant. Selon eux, il existerait



alors une évidence pour un apprentissage de suivi de déclencheur du regard soutenu par le rôle prépondérant des interactions sociales, critiques pour la croissance de l'attention partagée.

Deak et al. se sont également intéressés aux modules d'attention proposés par Baron-Cohen, dont nous avons parlé précédemment, une approche modulaire, comme celle exposée par Butterworth et al. Cependant Deak et al. semblent vouloir poser une alternative à ces deux approches modulaires en suggérant la pertinence d'approche de systèmes dynamiques émergents. Selon eux, la mise en place du système attentionnel pourrait être expliquée par l'existence de systèmes d'apprentissage dynamiques, qui seraient auto-organisés, et qui ne se mettraient pas en place de façon aussi modulaire, comme proposé par Baron-Cohen ou Butterworth et al.

Quelle que soit l'approche utilisée pour étudier l'émergence de l'attention chez l'enfant, l'utilisation du regard constitue ainsi un outil incontournable pour la mise en place de la perception sociale chez l'humain. Ce dernier va également l'utiliser dans des contextes mentalistiques, ce qui est spécifique à son espèce, ainsi qu'à quelques espèces de primates non-humains. Il nous semble tout à fait pertinent de référer ici (Figure II.20) au tableau présenté par Emery (2000), qui détaille la présence ou absence des compétences de traitement de regard chez différentes espèces d'animaux, chez l'enfant au cours de son développement. Ce tableau présente aussi la grande qualité d'indiquer les mécanismes défailants chez des enfants atteints de diverses pathologies.

Subject Group	Eye Presence	Simple Gaze	Gaze Following	Joint Attention	Mental Attribution
Fish	✓	?	?	?	?
Reptiles	✓	✓	?	?	?
Birds	✓	✓	?	✓ (?)	?
Rodents	✓	?	?	?	?
Dogs (domestic)	?	✓	✓	?	?
Prosimians	?	✓	X	?	?
Monkeys	✓	✓	✓	✓	X
Great Apes	✓	✓	✓	✓	X (?)
Human					
—3 months	✓	X	X	X	X
—9 months	✓	✓	X	X	X
—12 months	✓	✓	X	✓	X
—18 months	✓	✓	✓	✓	X
—24 months	✓	✓	✓	✓	X
—48 months	✓	✓	✓	✓	✓
—Autism	✓ (?)	✓	?	X	X
—Down syndrome	✓	✓	✓	✓	✓
—Schizophrenia	✓	✓	?	?	?
—Amygdala damage	✓ (?)	X	X (?)	?	X (?)

FIGURE II.20 : Tableau résumant les différents types de traitement chez différents êtres vivants (✓ indique une évidence positive, X indique aucune évidence, ? indique que cela n'a pas été testé ou que l'évidence est controversée) (D'après Emery, 2000).

## II.4.6.2. Le doigt comme second outil d'indexation dans le monde

### II.4.6.2.a) Codage du pointage impératif

Ce comportement est réalisé par l'implémentation de la tâche générique d'atteinte d'une cible visuelle. Chez l'enfant, la première étape se développe autour de 5 mois et se caractérise par une atteinte qui, partant d'une position des yeux de l'enfant, se déplace le long d'un angle du regard, directement à travers l'objet-cible. Si l'enfant échoue dans cette tentative, le bras se retire à la position de départ et l'essai est renouvelé.

Pour implémenter ce mécanisme dans le système robotique, l'apprentissage doit procéder de la manière suivante : (i) localisation d'une cible visuelle, (ii) saccade oculaire vers cette cible, en utilisant la cartographie bimodale (audiovisuelle) de la saccade, déjà apprise dans l'analogie computationnel du colliculus supérieur, en relation avec les neurones « omnipause » du tronc cérébral (neurones « *when?* »,

silencieux juste avant et pendant la saccade) et les neurones « burst » (neurones « *where ?* ») qui codent la direction et l'amplitude de la saccade ; (iii) conversion de la position de l'œil en une atteinte balistique en utilisant la cartographie balistique; (iv) comme le bras bouge, il faut utiliser la détection de mouvement pour focaliser la fin du bras; (v) utilisation de la cartographie de la saccade pour convertir le signal d'erreur des images coordonnées dans des positions du regard; et (vi) retirer le bras (opérations à répéter pour l'apprentissage).

#### **II.4.6.2.b) Codage du pointage déclaratif**

Au lieu de considérer le vecteur formé par l'angle du regard pour atteindre l'objet distal, on considère plutôt, pour le pointage déclaratif, le vecteur formé par la position du bras par rapport au corps. On utilise les mêmes mécanismes, mais on a besoin en plus d'un système pour reconnaître les gestes rudimentaires. Scassellati propose que le fait de produire des gestes de pointage déclaratif repose sur l'imitation des pointages déclaratifs qui se déroulent dans un contexte social approprié. Sur le robot, un mécanisme de suivi a été ajouté à la sortie du détecteur de visage, et les sorties ont été ensuite classées. L'auteur note, au passage, que certains robots possèdent la capacité d'imiter des signes de hochements de tête pour dire oui ou non à la « nourrice ». Lorsque la « nourrice » répond « oui ! », le robot reprend simplement par un hochement de tête qui signifie « oui ! ».

Nous venons de poser les modules d'attention nécessaires au développement de l'enfant dans notre cadre du signe ou cadre sémantique, et nous venons de comprendre plus précisément sur quoi reposait le mécanisme d'attribution d'intentions. L'implémentation dans le cadre robotique nous a permis, de plus, de mieux appréhender le développement des bases sociales chez l'enfant humain. Il convient de garder à l'esprit que l'enfant est doté d'un premier outil, l'œil, pour attirer l'attention de l'adulte, et qu'à ce premier mécanisme vient s'ajouter un second outil robuste, le pointer de l'index. Dans son étude, Masur (1983) a montré pour 4 enfants anglais testés, que la faculté à envoyer des signaux coordonnés où le regard et le geste sont dirigés de façon divergente dans une interaction triadique, n'émerge seulement qu'à 12 mois ou plus tard. A présent, attardons-nous sur la finalité même de ces mécanismes d'attention, la compréhension de l'intention.

## **II.5. La compréhension de l'intention à partir des mécanismes d'attention et d'imitation**

Il semble nécessaire ici de parler plus amplement du mécanisme de compréhension des intentions, car il est de première importance si l'on désire mieux appréhender dans quel sens ces modules menant à la théorie de l'esprit ont une incidence si particulière dans le développement de l'enfant. Les modules cités précédemment jouent un rôle crucial pour la communication mère-enfant, et dans l'attente que l'un et l'autre peuvent avoir sur leurs comportements respectifs, et aussi sur la façon dont ils se servent des mécanismes d'attention. A ces points fondamentaux, vient s'ajouter la compréhension de l'intention chez l'enfant. Comment l'enfant va-t-il, en se reposant sur ces pierres angulaires, apprendre à comprendre la notion d'intention, et à s'en servir ensuite de façon adéquate dans son développement, tout particulièrement via l'imitation ?

### **II.5.1. La notion d'intention via l'observation des actions**

Une étude de Blakemore et Decety (2001) porte sur le mécanisme sous-jacent de l'attribution d'intention aux actions, un mécanisme qui peut reposer sur la simulation de l'action observée et son *mapping* direct comme représentation de nos propres intentions. Les humains possèdent une faculté inhérente à comprendre les intentions d'autres personnes. Cette capacité est une composante de la théorie de l'esprit, que nous venons d'exposer précédemment. Les auteurs s'attachent ici à la faculté de comprendre les intentions d'autres personnes en observant leurs actions. Ce niveau de la théorie de l'esprit semble être un pré-requis, selon Blakemore et Decety, pour la compréhension, de niveau plus haut, des contenus mentaux des autres.

Les auteurs démontrent que le mouvement biologique constitue une catégorie spéciale définie à partir de la perception visuelle du mouvement, qui constitue une source cruciale dans l'input sensoriel des actions. Il est essentiel d'être capable d'identifier le mouvement de formes biologiques par rapport à d'autres types de mouvement dans l'environnement naturel, afin de prédire les actions d'autres individus. C'est une capacité que l'on retrouve chez les animaux. Leur survie dépend

en effet de leur faculté à identifier les mouvements des prédateurs et compagnons. Comme les animaux sociaux, les humains se comportent largement sur la base de leurs interprétations et prédictions sur les actions des autres.

## **II.5.2. L'imitation motrice au service de la compréhension de l'intention**

Suite à la découverte par l'équipe de Giacomo Rizzolatti à Parme, des neurones miroirs dans le cortex prémoteur ventral chez le singe —des neurones « miroirs » qui déchargent aussi bien quand le singe accomplit des mouvements de la main dirigés vers un but spécifique, que lorsqu'il observe un autre individu faisant les mêmes mouvements— il existe actuellement une évidence forte que chez les humains, plusieurs régions du cerveau incluant le cortex prémoteur, le cortex postérieur pariétal et le cervelet, sont activées pendant la génération d'action et pendant l'observation et la simulation d'autres actions.

### **II.5.2.1 Un système miroir chez le singe macaque**

De façon intéressante, l'attention spatiale semble être gérée par une zone corticale précise, au niveau de l'operculum frontal, au carrefour entre la partie ventrale de l'aire 6 et les aires 44-45. Selon Rizzolatti et al. (1996), cette zone corticale est homologue à l'aire F5 du singe, qui est engagée dans la représentation du mouvement, et qui contient une catégorie spécifique de neurones. Ceux-ci déchargent, chez le singe, lorsque l'animal exécute un mouvement de la main dirigé vers un but. Mais ils déchargent également lorsque ce même singe voit son congénère réaliser le même mouvement. Certains de ces neurones vont décharger pour une pince pouce-index de précision par exemple pour de petits objets, tandis que d'autres neurones vont décharger pour une saisie impliquant la main entière, pour des objets plus grands. Cette zone corticale contient deux types spécifiques de neurones. Le premier type concerne les neurones dits « canoniques », pour lesquels une simple vision de l'objet suffit pour leur activation. Ils semblent coder l'« affordance » d'un objet ou sa préhensibilité, soit l'aspect *pragmatique* des connaissances lui permettant de savoir comment se saisir de l'objet cible, plutôt qu'un contenu *sémantique* permettant de le reconnaître dans sa catégorie. En cela ils sont bien différents du deuxième type de neurones, celui que nous venons d'évoquer, les « neurones miroirs ». Ces derniers ne vont pas décharger à la simple

vision d'un objet passible d'une action, mais déchargent au contraire à la vue d'une action ou d'une composante de l'action (phase, but, etc.). Ce type de neurones décharge quand on saisit un objet, quand on voit quelqu'un d'autre saisir cet objet, mais ne décharge pas à la vue de l'objet seul. Ils semblent donc fortement liés à l'exécution et observation d'une action.

Cette découverte permet aux auteurs de supposer une relation étroite entre représentations motrices de soi et de l'autre. Williams et al. (à paraître) font même l'hypothèse suivante sur le fonctionnement de ce type de neurones :

MNs [mirror neurons] appear to have the capacity to embody a "supramodal representation" of action, functioning as a bridge between higher visual processing areas and motor cortex (between seeing and doing). (p. 11)

Arbib (2005) s'est également penché sur la question des mécanismes cérébraux pour le contrôle visuel des mouvements de la main, et spécifiquement ceux qui lient le cortex pariétal au cortex frontal prémoteur (cf. aussi Roy et al., 2000). Il retrace ainsi l'activité corticale mise en jeu pour le phénomène de la saisie : l'information motrice est transférée de la zone F5 du singe macaque (voir Figure II.21) au cortex moteur primaire (F1), auquel F5 est directement connectée, aussi bien qu'aux centres sous-corticaux variés impliqués dans la réalisation du mouvement (système fronto-striatal). Les neurones situés dans la partie rostrale de l'aire inférieure 6 (soit l'aire F5) déchargent lors des mouvements actifs de bouche et/ou de main. Rizzolatti et al. ont montré que F5 possédait un vocabulaire de schémas moteurs prédéfinis. Arbib argumente que la situation est plus complexe et que l'exécution de la saisie engage de nombreuses boucles et de nombreuses autres régions cérébrales que AIP (le sulcus antérieur intra-pariétal) et F5. Arbib avance que les cellules AIP codent des « affordances » pour saisir à partir du flux visuel et qu'elles envoient ensuite cette information à F5.

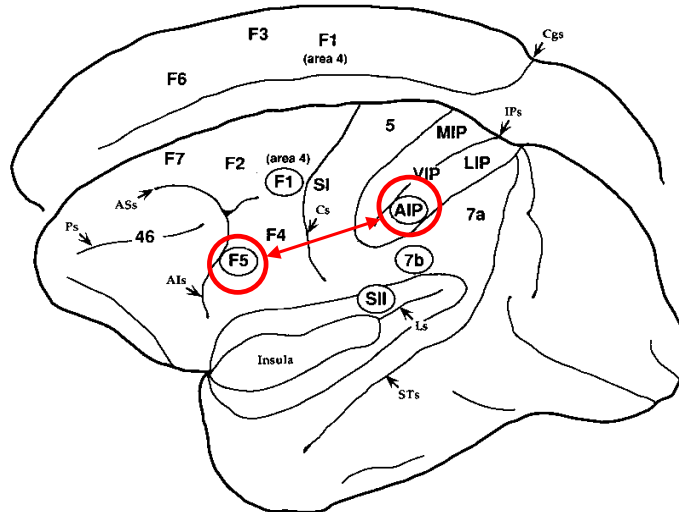


FIGURE II.21 : Vue de l'hémisphère gauche du cerveau du macaque, et notamment du couple fronto-pariétal F5-AIP pour la préhension ; avec un aperçu en miroir de la partie supérieure mésiale (D'après Arbib, 2005).

L'auteur rapporte le rôle crucial d'IT (cortex inféro-temporal) et de PFC (cortex préfrontal) dans ce mécanisme de saisie. Le flux dorsal (du cortex visuel primaire au cortex pariétal) apporte l'information nécessaire à AIP pour la reconnaissance des différentes parties de l'objet qui peuvent être saisies de façons distinctes. Ces informations sont envoyées à F5 où une sélection est opérée pour la saisie réelle. Le flux dorsal ne « sait » pas ce qu'est l'objet (système *What*), il envisage simplement un ensemble d'affordances possibles. Le flux ventral (du cortex visuel primaire au cortex inféro-temporal IT) prend alors le relais pour caractériser l'objet et envoie l'information au cortex préfrontal qui va « prévenir » F5 du choix optimal de l'affordance, la plus appropriée à la tâche réalisée par la main. Ce modèle représente alors de façon adéquate comment F5 accepte les signaux provenant des aires F6 (pré-SMA), 46 (cortex préfrontal dorsolatéral) et F2 (cortex prémoteur dorsal) en réponse aux contraintes liées à la tâche, à la mémoire de travail et aux stimuli, qui tous instruisent l'action respectivement.

En sus de ce modèle cortical, Arbib (2005) souligne l'importance de la découverte d'un système miroir pour la saisie. Ce système miroir, ou « système de correspondance entre l'observation et l'exécution », décharge quand le singe saisit un objet ; mais aussi également lorsqu'il observe un expérimentateur réalisant un geste identique. La majorité des neurones miroirs sont sélectifs pour un type d'action particulier (saisir, manipuler, arracher, etc.) et leur existence implique dans

quasiment tous les cas un lien entre le mouvement effectif observé et le mouvement effectif exécuté. Ces neurones reçoivent un input provenant de la région PF du cortex pariétal (homologue de BA 40), codant les observations des mouvements de bras et de main.

### **II.5.2.2 Un système miroir chez l'humain**

Daprati et al. (1997, Daprati et Sirigu, 2006) pensent que chez les humains, un mécanisme similaire peut opérer pour la reconnaissance de l'action, incluant même la reconnaissance des gestes de parole.

Arbib (2005) relate l'existence d'un tel système chez l'homme en citant des expériences PET. Ces différentes études ont mis en lumière le fait que l'observation de la saisie chez l'homme a activé de façon significative, et toujours dans l'hémisphère gauche, le sulcus temporal supérieur (STS), le lobule pariétal inférieur et le gyrus frontal inférieur (aire 45). Les aires 44 et 45 constituant l'aire de Broca, cela a donc ouvert la possibilité de l'existence d'un système miroir pour la saisie dans Broca. Chez le singe, F5, considérée comme étant l'homologue de l'aire de Broca chez l'homme, est de la même façon impliquée dans le mécanisme de saisie. Le fait qu'il existe une correspondance aussi claire que celle-ci entre primate humain et non-humain semble donc indiquer tout naturellement qu'il existe un système fondamental pour la reconnaissance de l'action. Le système miroir constitue un point d'ancrage important puisqu'il nous informe de nos propres mouvements et des mouvements des autres en générant un code pour l'action qui ne se résume pas seulement aux mouvements mais aux mouvements **et** aux buts de ces mouvements.

Il existe selon Blakemore et Decety (2001) un lien naturel entre *observation* et *génération* de l'action, qui est fourni par l'imitation motrice. En effet de jeunes bébés peuvent imiter des gestes faciaux, ce qui témoigne d'un système précoce pour le couplage de la perception et de la production des actions. Une série d'expériences a amené à la découverte, que la compréhension pré-verbale de l'enfant au sujet des personnes, est comprise à l'intérieur d'une charpente qui inclut buts et intentions.

Chaminade et al. (2002) ont mené une étude en ce qui concerne précisément l'imitation. Ils la considèrent comme un mécanisme naturel engageant le couplage perception/action, lequel joue un rôle considérable dans le développement humain, tout particulièrement afin d'extraire l'intention du comportement d'autrui. Les



auteurs ont utilisé, pour leur étude, une imagerie PET, afin d'observer les bases neurales de l'imitation des actions orientées vers l'objet, chez des adultes normaux. L'expérience repose sur deux conditions : (i) le stimulus-événement est montré aux sujets pendant l'observation du modèle ; et (ii) la manipulation-réponse est réalisée par le sujet. Les résultats ont montré des recouvrements des activations quand les sujets imitent les deux actions. Mais une activité spécifique a été détectée dans le cortex préfrontal médial pendant l'imitation des gestes significatifs, alors que l'imitation du but a été associée à une activité augmentée dans le cortex prémoteur gauche.

Williams et al. (à paraître) postulent que l'imitation et ses mécanismes associés pourraient être les précurseurs de la théorie de l'esprit. En effet, l'imitation et l'attribution d'états mentaux réclament toutes deux de transposer la perspective de l'autre à soi-même. En d'autres termes, l'autre a des croyances que je m'approprie via mon propre système cognitif afin de les comprendre et de les prédire. De plus, des expériences réalisées en imagerie cérébrale ont montré que lors d'une tâche d'imitation simple, on constate une activation dans l'aire 44, et dans le cortex pariétal, suggérant alors que le système des neurones miroirs est impliqué dans l'imitation chez l'homme. Les auteurs notent alors l'importance ontogénétique de ce système pour différencier les pensées de soi et de l'autre :

MNs [Mirror Neurons] provide a key foundation for the building of imitative and mindreading competencies. (...) We should thus expect that MNs play important roles in the whole ontogenetic cascade from early imitation to elaborated ToM [Theory of Mind]. (Williams et al., à paraître:16)

Arbib (2005) insiste sur le fait que cette hypothèse du système miroir, ou cette faculté unique à copier des actions, constitue tout simplement le pré-requis, la première étape vers l'imitation. En effet, l'imitation implique l'analyse de mouvements complexes en premier lieu. Cette extension de l'hypothèse du système miroir vers l'imitation semble être une nouveauté-clé dans le cerveau des chimpanzés et des hommes, par conséquent de leur ancêtre commun pour le langage. La différence entre primate humain et non-humain réside dans le fait que ce dernier ait une imitation dite « simple », en d'autres termes il peut imiter de nouvelles séquences courtes lorsque celles-ci sont présentées de façon répétée. Tandis que l'homme bénéficie d'une imitation « complexe » dans le sens où il peut acquérir des séquences nouvelles et plus longues en une seule exposition. La forme que l'imitation prend chez le singe constitue un processus très long et difficile par rapport

à la rapidité avec laquelle les humains peuvent acquérir de nouvelles séquences via l'imitation. Arbib souligne le fait que cette extension du système miroir d'actions simples à des actions composées est une pierre d'angle pour expliquer que le cerveau des hominidés soit devenu « prêt-pour-le-langage ».

Iacoboni (in press) propose dans ce sens une architecture neurale minimale pour décrire l'imitation, que nous décrivons plus loin (en reprenant sa première publication, celle de Dubeau et al., 2002). L'auteur suggère à ce propos un lien fort entre imitation et langage, rappelant qu'il est reconnu dans la littérature que les enfants peuvent imiter beaucoup plus tôt qu'ils ne parlent. Si l'aire de Broca joue un rôle essentiel dans l'imitation, comme il le pense, alors on peut sans doute conclure que cette aire n'est pas strictement réservée au traitement linguistique. Ce qui laisse la porte ouverte à l'idée que le circuit pour la reconnaissance de l'action chez le singe ait évolué par la suite pour soutenir l'imitation (nous dirons en *scaffolding*) et plus tard le langage.

### **II.5.3. Le développement du SIM (*Shared Intention Mechanism*) chez l'enfant ou l'intentionnalité selon Tomasello et al. (2004)**

Ce qui différencie principalement la cognition humaine de la cognition animale réside dans l'évidence que les humains ont la faculté à prendre part à des activités collaboratives, en partageant des buts, et des intentions, soit, selon Tomasello et al. (2004), en témoignant d'une intentionnalité partagée. Cette faculté repose sur la compréhension des intentions, qui nous permet alors de discerner clairement ce qu'autrui perçoit, désire, sait, croit, ou encore de détecter ses objets d'intérêt ou d'attention.

Dans sa voie développementale, l'enfant va être progressivement équipé de compétences lui permettant de collaborer et d'interagir avec les autres personnes. Les intentions vont commencer à émerger vers la fin de sa première année de vie, tandis que les croyances, intimement liées au développement de la théorie de l'esprit, ne vont pas émerger avant quatre ans chez l'enfant.

Une étude réalisée par Melis et al. (2006) tente d'expliquer les origines phylogénétiques de telles compétences chez l'enfant. Certaines observations suggèrent qu'à l'état sauvage, les chimpanzés posséderaient de telles compétences

collaboratives. Toutefois, il paraît difficile de déterminer précisément quelles compétences cognitives sont réellement en jeu lors d'activités coopératives à travers l'observation écologique, et les chercheurs sont également confrontés à l'argument que ces observations interactives peuvent constituer des sous-produits d'attraction ou aversion symétriques entre individus, ce qui n'impliquerait alors pas nécessairement une quantification précise des coûts et bénéfices que peuvent engendrer le choix d'une collaboration avec un congénère. Les auteurs ont donc cherché à tester si les chimpanzés reconnaissent quand la collaboration est nécessaire, et s'ils choisissent le collaborateur le plus efficace entre deux de leurs congénères. Les résultats de cette étude démontrent que les chimpanzés peuvent en effet comprendre à quel moment le besoin de recrutement d'un collaborateur est exigé, et ils peuvent en outre identifier et ainsi choisir le meilleur des collaborateurs après un petit nombre d'interactions entre eux. Ces découvertes amènent les auteurs à défier l'hypothèse que les comportements coopératifs chez les chimpanzés ne représentent pas une collaboration active dans laquelle les individus choisissent intentionnellement avec qui et quand ils veulent travailler. Cette vision du problème amène les auteurs à mettre en évidence l'éventuelle implication dans les formes humaines de collaboration de « building blocks » déjà présents chez le chimpanzé et d'autres espèces primates.

### **II.5.3.1. L'action intentionnelle humaine**

Tomasello et al. postulent alors que les êtres humains uniquement bénéficient d'une adaptation biologique leur donnant l'opportunité de participer à ces activités collaboratives socialement coordonnées et ainsi à s'impliquer dans des intentions conjointes. Dans la figure II.22 ci-dessous, les auteurs proposent, dans une approche mêlant systèmes de contrôle et intentionnalité partagée, une représentation schématique de l'action intentionnelle humaine.

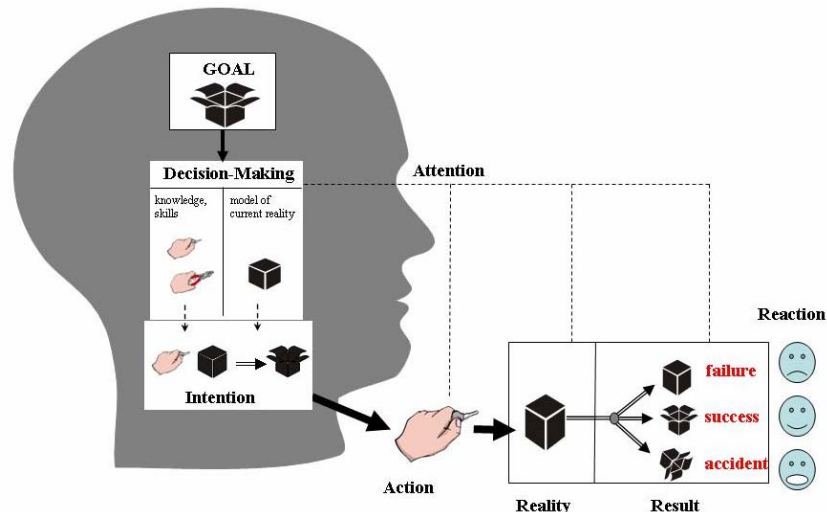


FIGURE II.22 : L'action intentionnelle humaine. Le but est une boîte ouverte. L'acteur choisit un moyen (plan), représenté par les mains agissantes, qui forme une intention. L'action qui en résulte cause un résultat, qui amène à une réaction émotionnelle de l'acteur (d'après Tomasello et al., 2004)

Les auteurs définissent clairement la notion d'intention comme une planification de l'action qui se réalise dans la poursuite du but. Dans la figure, l'intention inclut le but –la boîte ouverte– et l'action planifiée choisie afin que le but soit réalisé. Une fois ce dernier réalisé, on obtient alors le résultat de l'action, qui va avoir une conséquence sur l'état du monde, et qui va être accompagné d'une réaction émotionnelle en fonction du résultat obtenu. L'organisme va ainsi porter attention à un aspect particulier de la situation, pertinent pour le but, entraînant ainsi une perception intentionnelle ou attention sélective.

### II.5.3.2. Comprendre cette intention

Nous venons de lire la définition même d'une action intentionnelle réalisée par un sujet, mais qu'en est-il de la compréhension de cette intention par une tierce personne ? Mais afin de poser les fondements de cette compréhension, effectuons un retour ontogénétique, et intéressons-nous aux niveaux de compréhension émergents chez l'enfant, afin d'en expliquer la phylogenèse, et avant de parler plus en détail de l'intentionnalité partagée. Il existe, chez l'enfant, trois niveaux de compréhension des actions des autres : (i) comprendre l'action animée ; (ii) comprendre la poursuite de buts, et (iii) comprendre les choix de planification.

Dès la naissance, l'enfant est sensible au mouvement biologique auto-produit, l'exemple le plus frappant étant celui du bébé qui se tourne pour regarder dans la

même direction que d'autres personnes. Vers six mois, l'enfant sera même capable de prédire ce que d'autres personnes vont faire dans des situations qui lui sont familières. A ce stade de compréhension de l'action animée, l'enfant n'a besoin de comprendre que les comportements produits spontanément par les autres, sans avoir recours à une compréhension plus fine de la structure interne des actions intentionnelles. L'état des capacités de l'enfant à cet âge pourrait être schématisé dans la figure II.22 comme étant une « tête vide ».

Les enfants souffrant d'autisme comprennent également les autres personnes comme des êtres animés qui vont produire des comportements spontanés. Il semble que ces enfants témoignent d'une compréhension que les autres ont des buts et qu'ils voient des choses. Certes les enfants autistes présentent des déficits clairs dans les tests de suivi de regard spontanés, mais quand on leur demande, ils peuvent rapporter ce que l'autre personne regarde. Même si certains enfants peuvent comprendre les autres comme possédant des buts, ils ne peuvent toutefois pas encore comprendre le processus de prise de décision par lequel un acteur va choisir de façon rationnelle parmi des choix comportementaux afin de produire une action intentionnelle.

On peut remarquer que les primates non-humains comprennent eux-aussi leurs congénères comme des agents animés, qui produisent des comportements spontanément. Deux lignes de pensées s'opposent ici dans le domaine de la cognition animale en ce qui concerne la compréhension de buts et d'intentions. D'une part Povinelli et Vonk (2003) estiment que la compréhension des buts et de la perception est un exemple évident de compréhension des états mentaux. Dans leur étude, ils se posent la question de la faculté à concevoir un monde mental chez le chimpanzé. Les auteurs pensent que les humains et les chimpanzés ont hérité de structures mentales communes afin de former des conceptions abstraites, mais qu'au cours de l'évolution, notre espèce a tissé un nouveau système de théorie de l'esprit. Le scénario qu'ils proposent ici, nommé « hypothèse de ré-interprétation », suppose que la capacité d'abstraction comportementale soit déjà présente chez l'ancêtre commun des primates humains et non-humains, mais que les humains aient ajouté à cela un autre système dans le but de coder un plus grand nombre de comportements, et de manière plus mentaliste. Ainsi, les humains attribuent une existence à des états mentaux comme penser, savoir, vouloir, etc. (des états

mentaux de premier ordre), mais ont également la capacité à attribuer cette même capacité d'attribution d'états mentaux à eux-mêmes et à d'autres (états mentaux de second ordre). Il semble que les études menées sur le sujet tendent à démontrer que l'esprit des chimpanzés contient des représentations mentales. Mais la vraie question est si ces représentations mentales sont des représentations d'abstractions comportementales et d'états mentaux, ou alors des abstractions purement comportementales ? Les auteurs s'accordent à dire qu'il est difficile de répondre à cette question en invoquant la nature même des expériences menées sur les chimpanzés. En effet, les techniques utilisées vont dans la plupart des cas présupposer que le chimpanzé a accès à un invariant comportemental, et ainsi il demeurera très difficile d'établir si un codage mental est également utilisé ou non.

D'autre part, Tomasello, Call et Hare (2003) considèrent bien que la capacité à comprendre l'action intentionnelle doit être attribuée en termes de buts et perception. En effet, Tomasello et al. éprouvent plus de difficultés à mettre en évidence que les chimpanzés, ou tout autre grand singe, comprennent les intentions ou actions dirigées vers le but des autres. A partir d'études menées sur les chimpanzés, les auteurs s'accordent à dire que les chimpanzés ont un accès au contenu de ce que les autres voient, et dans une moindre mesure sur le fait que ceci gouverne leurs comportements. Tomasello et al. ont alors questionné la compréhension même de l'action intentionnelle chez le chimpanzé, en réalisant une expérience où un humain présente à ce dernier de la nourriture dans sa main, en se comportant soit de façon réticente, ou comme étant incapable de lui donner la nourriture en question. Dans les deux cas, les chimpanzés testés n'obtenaient pas de nourriture. De fait ils étaient plus impatients, allant jusqu'à frapper leur cage violemment, et ils partaient plus rapidement de la zone où se trouvait l'humain, lorsque les humains étaient réticents à leur donner la nourriture, que lorsque ceux-ci se montraient dans l'incapacité de le faire. Cette expérience met en évidence que les déclencheurs, que les singes utilisent pour identifier le comportement intentionnel qui leur est perceptible chez l'humain, impliquent des signes physiques —comme un effort, un essai, une frustration ou une satisfaction. Et cela démontre que les chimpanzés discriminent les actions intentionnelles des actions accidentelles. Tomasello et al. formulent alors une nouvelle hypothèse qui consiste à dire que : (i) les chimpanzés savent contrôler le regard des autres ; (ii) ils ont la connaissance

qu'un individu a vu quelque chose dans le passé immédiat, et saisissent ainsi, en fonction du comportement, une intention dans l'action ; et (iii) ces chimpanzés peuvent utiliser l'information sur ce que leurs congénères voient afin de prédire les comportements qu'ils vont adopter. Pour les auteurs, même s'il reste clair que (i) les chimpanzés ne possèdent pas de théorie de l'esprit similaire aux humains, (ii) et qu'ils ne possèdent pas non plus les mêmes capacités que l'humain en perception visuelle (l'attention et la perspective par exemple), (iii) ou qu'ils ne comprennent pas les intentions qui précèdent les comportements et les intentions communicatives, il n'en reste pas moins que ces primates possèdent un schéma socio-cognitif leur permettant de discerner une partie de la structure intentionnelle du comportement du congénère et en conséquence d'être influencés par la perception du comportement de l'autre.

Les chimpanzés comprennent bien que leurs congénères voient des choses. Ils suivent la direction du regard du congénère vers des cibles externes et vérifient en retour avec le regard de l'autre qu'il n'y a rien d'intéressant dans la zone en question. Et le fait de savoir ce que les autres peuvent voir va affecter les réactions du chimpanzé. Certains grands singes comprennent donc au moins certains aspects de la perception liés à l'action intentionnelle, même si cette compréhension n'implique pas les dimensions mentales les plus intentionnelles de cette action.

Si nous revenons maintenant aux enfants, ceux-ci autour de 10 mois vont segmenter le flux de comportements en des unités que les adultes devraient percevoir comme des actes distincts dirigés vers un but. Dans sa compréhension de la poursuite du but, l'enfant va rechercher l'information qui concerne ce but, en tirant profit du regard de l'autre, ou encore de son état émotionnel. Les enfants comprennent alors que les autres personnes ont des buts et qu'ils persistent dans leur comportement jusqu'à ce qu'elles voient que leur but est atteint, provoquant une réaction émotionnelle en rapport avec le résultat du but. Les enfants de douze mois vont même pouvoir suivre la direction du regard de l'adulte dans des situations plus complexes, soient dans des localisations hors champ visuel, indiquant la compréhension de la part de l'enfant, que l'adulte voit quelque chose que l'enfant ne peut pas voir. Dans la figure II.22, les compétences de l'enfant lui permettraient d'avoir recours, à l'intérieur de la « tête » de Tomasello, au but, et à son contrôle perceptif.

Lorsque l'enfant atteint le stade de compréhension du choix de planification, vers 14 mois, il commence à saisir que dans une poursuite de but, l'autre peut considérer des plans d'actions variés, et en sélectionner un pour l'action intentionnelle. L'enfant comprend que l'autre perçoit et évalue la réalité de façon rationnelle avant d'opter pour un plan d'action donné, pour accomplir son but. Dans la figure II.22, l'enfant serait doté de toutes ces compétences présentes dans la « tête » du schéma.

### II.5.3.3. Intentionnalité partagée

La notion d'intentionnalité partagée réfère à des actions collaboratives, dans lesquelles les participants ont un but partagé. Les buts et intentions de chaque participant doivent évidemment prendre en considération les buts et intentions de l'autre. A l'intérieur de cette relation, chacun est sensible à l'autre, tout en partageant un but réalisé mutuellement, en coordonnant les plans d'actions et intentions de chacun. Les aspects de cette intentionnalité partagée ont été transposés très explicitement dans la figure II.23 par Tomasello et al.

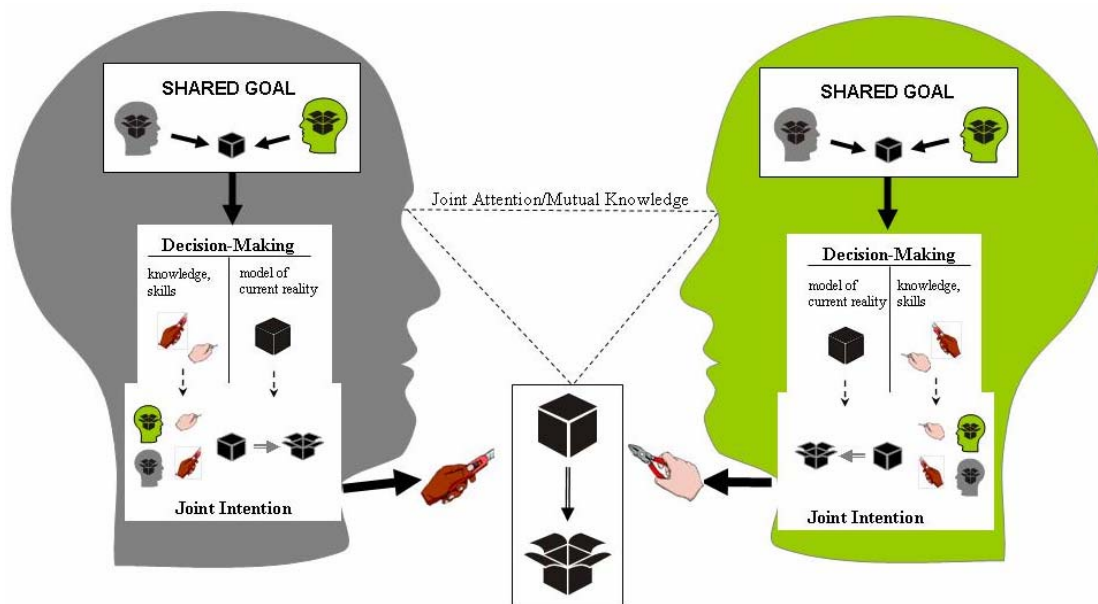


FIGURE II.23 : Conception de chacun des participants dans une activité collaborative dans laquelle un but partagé et une intention conjointe, avec rôles complémentaires, sont formés. (D'après Tomasello et al., 2004)

Ce diagramme permet aux auteurs de mieux représenter la compréhension de l'interaction de chacun des participants. Le but ici –toujours ouvrir la bouteille– va être partagé par les deux participants, sa représentation cognitive contenant le soi et l'autre. Le but va alors concerner l'action mutuelle des deux participants, qui vont



témoigner d'une motivation partagée dans une activité collaborative. Outre cet engagement partagé, la figure II.23 témoigne également d'une intention conjointe, dans la mesure où chaque participant se représente cognitivement les deux rôles de la collaboration dans un seul et unique format représentatif, qui va permettre de concevoir un rôle inverse et une aide mutuelle.

De la même façon que nous avons présenté les trois étapes nécessaires à la mise en place de l'action intentionnelle humaine dans la partie précédente, nous allons voir les trois modules entrant en compte dans l'émergence de l'intentionnalité partagée, tels qu'ils sont cités par Tomasello et al. : (i) l'engagement dyadique ; (ii) l'engagement triadique ; et (iii) l'engagement collaboratif.

L'engagement dyadique, qui préempte les deux autres, vise à partager les buts et la perception. Au-delà d'une sensibilité évidente pour les contingences sociales, les enfants humains et les adultes interagissent les uns avec les autres de façon dyadique dans des proto-conversations. Ces proto-conversations consistent en un engagement direct entre un enfant et un adulte, où l'enfant fixe du regard l'adulte dans un échange de regard mutuel. Cette activité est caractérisée comme dyadique dans le sens où l'enfant n'opère pas de contrôle du regard de l'adulte vers son objet d'intérêt, mais établit simplement une relation directe entre lui et son interlocuteur. Tomasello et al. schématisent cette idée sur la figure II.23, en argumentant que la « tête » serait vide à ce stade du développement.

En ce qui concerne les primates non-humains, et toujours dans une optique visant à différencier la cognition humaine de la cognition animale, ceux-ci semblent manquer de motivation et de compétences pour les formes les plus basiques d'états psychologiques partagés avec les autres. Ils ne s'engagent pas dans des proto-conversations entre adulte et enfant, comme les humains. Les bébés singes ont très peu de comportements réclamant un contact face à face. On peut également ajouter les études en pathologie du comportement, particulièrement l'autisme : toutes soulignent que les enfants souffrant de cette affection ne s'engagent pas dans les proto-conversations.

Plus tard dans le développement, autour de 9-12 mois, intervient l'engagement triadique, où on observe un partage des buts et des intentions. Parallèlement à la compréhension des autres personnes comme ayant des intentions dirigées vers des buts, les enfants commencent également à s'engager dans des

relations triadiques impliquant l'enfant, l'adulte et une entité externe, un objet d'intérêt par exemple —soit le « triangle référentiel » nous rappelle Tomasello (2003b)— un objet vers lequel l'enfant va diriger l'attention de l'adulte. Dans ce type de relation, on peut observer une coordination des regards des deux protagonistes sur une même cible. Pour reprendre la terminologie utilisée et citée précédemment par Baron-Cohen, nous nous situons dans le mécanisme d'attention partagée (SAM), que Tomasello et al. nomment ici « perception conjointe ». Sur la figure II.23, nous n'aurions donc que buts partagés et contrôle perceptif. Tomasello (2003a,b) propose que la communication symbolique soit le processus par lequel un individu essaie de manipuler ou partager l'attention avec un autre individu. Cette tentative implique généralement la référence (inviter l'autre à partager l'attention sur une entité extérieure) et la prédication (diriger l'attention de l'autre en direction d'un trait appartenant à une entité extérieure sur lequel l'attention n'a pas été encore portée). Comme le souligne Tomasello, c'est à cet âge que les enfants commencent à régler l'attention de l'autre, et les comportements qui en résultent, sur des entités de leur environnement. Toujours selon lui, ces différents comportements attentionnels conjoints émergent au cours de la même période ontogénétique, reflétant un changement cognitif qui s'opère chez l'enfant : la compréhension des autres en tant qu'agents intentionnels.

Cet engagement triadique constitue de plus un outil fiable pour diagnostiquer les cas d'autisme, car les enfants autistes présentent de gros déficits dans l'engagement triadique partagé et l'attention conjointe. Ces enfants s'engagent très peu dans une situation conjointe coordonnée, et invitent très peu les autres à entrer dans une attention conjointe en utilisant par exemple notre geste de pointer déictique pour montrer un objet d'intérêt.

De la même façon que l'enfant autiste, le singe ne s'engage que très rarement dans des relations triadiques avec d'autres congénères et entités externes. Dans une interaction, il semble que le regard du singe vers l'adulte constitue plus un outil de vérification qu'un regard de partage d'intérêt. Les singes montrent peu de désir de partager. De plus, il a été observé dans de nombreuses études que sous des conditions naturelles, les singes ne pointent pas, ne montrent pas des choses à leurs congénères. Outre la production de gestes, des études ont souligné le fait que les

grands singes ne comprennent pas les intentions communicatives comme évidentes dans des actes de pointers indiquant un emplacement de nourriture.

Les singes et certains enfants autistes paraissent donc comprendre en premier lieu les actions dirigées vers le but, même si celles-ci ne sont pas intégralement intentionnelles. Mais, par la suite, ils ne suivent pas la voie développementale classique dans l'engagement social de la même façon que les enfants en développement typique. Ni les singes, ni les enfants autistes ne s'impliquent dans des engagements dyadiques partagés (protoconversation), et dans des engagements triadiques partagés (avec intention et attention conjointe). Enfin, ni les uns, ni les autres ne présentent une quelconque manifestation de motivation à partager des états psychologiques avec l'autre. Il est pourtant vraiment nécessaire d'aller voir quelles sont les capacités des primates non-humains, car elles nous donnent des indices sur l'adaptation biologique dont les primates humains ont fait preuve, tout en gardant à l'esprit les racines robustes d'engagement social pré-existant chez les primates non-humains. L'étude sur des pathologies comme l'autisme nous informe également sur l'émergence développementale des mécanismes nécessaires à l'intentionnalité partagée.

Le dernier stade de l'intentionnalité partagée est l'engagement collaboratif, concernant les intentions conjointes et l'attention, autour de 12-15 mois chez l'enfant. Il s'agit ici d'un changement important dans les relations triadiques établies entre adulte et enfant. Avant 12-15 mois, les enfants sont plus impliqués dans un engagement conjoint passif, tandis qu'à partir de cet âge, ils entrent dans un engagement conjoint coordonné. Cette différence est cruciale, car dans ce dernier engagement, l'enfant ne se contente pas simplement d'interagir avec un adulte sur une entité extérieure, mais il va diriger en quelque sorte le comportement de l'adulte et son attention, de façon partagée avec l'adulte. A ce stade, il semble que les enfants ne se contentent pas de partager des buts, mais vont en plus coordonner les rôles. Si nous revenons à la figure II.22, l'enfant possède à ce point du développement toutes les compétences inscrites dans la « tête » de Tomasello.

Enfin —chose qui est d'importance primordiale pour notre étude— c'est également à cette période que les enfants commencent à établir activement l'attention conjointe avec d'autres à travers des gestes comme le pointage. En effet, le pointage déclaratif possède un statut privilégié dans le partage d'attention, dirigé

par le geste produit par l'enfant. Des expériences ont par ailleurs montré que lorsqu'un adulte réagit au pointage d'un enfant de 12 mois en regardant simplement l'objet indiqué, ou en regardant l'enfant avec une émotion positive, ou encore en ne faisant rien, les enfants ne sont pas satisfaits, indiquant que le résultat ne leur convient pas. A l'inverse, si l'adulte répondait à ce pointage en regardant l'objet et en le commentant positivement, les enfants étaient satisfaits, indiquant que le partage d'attention et d'intérêt était le but rempli. Comme le proposent Tomasello et al., il semble que les enfants âgés d'un an aient comme but l'attention conjointe en elle-même, tout en aidant aussi les autres à atteindre leur but en dirigeant leur attention de façon pertinente.

#### **II.5.4. Quelles fonctions cérébrales pour la compréhension de l'action et de l'intention ?**

Restant toujours préoccupée par la notion d'intention, il nous paraît nécessaire d'avancer notre compréhension de ce phénomène d'attribution d'intention à l'action de l'autre au niveau des fonctions cérébrales. Quelles sont donc les fonctions cérébrales mises en jeu dans la lecture d'intentions ? Par exemple, comment interprète-t-on typiquement l'intention de quelqu'un qui nous fixe, au delà du commun « Qu'est-ce qu'il me veut ? Celui-là ! » ?

Certains neurones situés dans le sulcus temporal supérieur ou STS ont été découverts comme étant sensibles à des visions très spécifiques du corps, ou de parties du corps. La fonction de ces neurones, selon Emery (2000), suivant en cela les premiers travaux de l'équipe de Perrett, pourrait être de déterminer tout d'abord la direction de l'attention de l'autre. D'un point de vue neuropsychologique, il semble bien que le STS, avec l'amygdale et le cortex orbitofrontal forment un système plutôt complet pour coder la direction de l'attention visuelle de l'autre, soit typiquement l'objet du focus d'un autre agent.

Les études de Buccino et al. (2001) et Dubeau et al. (2002) nous permettent d'ancrer plus avant l'idée d'une somatotopisation en *body-parts*, pour la lecture des intentions grâce au système des neurones miroirs. Se pose ici la question des interactions en fonction du corps, et précisément en fonction de ces parties du corps très significatives pour l'interaction entre agents, que sont l'œil, la bouche et la main.

Dans l'étude de Dubeau et al. (2002), les auteurs ont cherché à l'intérieur de la région du sulcus temporal supérieur (STS), s'il existait une topographie pour les mouvements liés à la bouche, aux yeux, et aux doigts. STS semble en effet être le siège d'une action spécifique de traitement des actions des parties du corps. Ils ont demandé à leurs sujets d'observer des mouvements saccadiques des yeux, des mouvements de la bouche qui articulait des syllabes silencieusement, et des mouvements de doigts. Leurs résultats ont mis en évidence l'existence d'un tel type de topographie dans le STS, reflétant le dialogue fonctionnel entre les régions temporale supérieure, prémotrice et pariétale. La figure II.24 nous montre que, dans cette région du cerveau, et de façon bilatérale, la représentation des doigts est la plus dorsale, tandis que la représentation de la bouche est plus antérieure que l'aire des yeux.

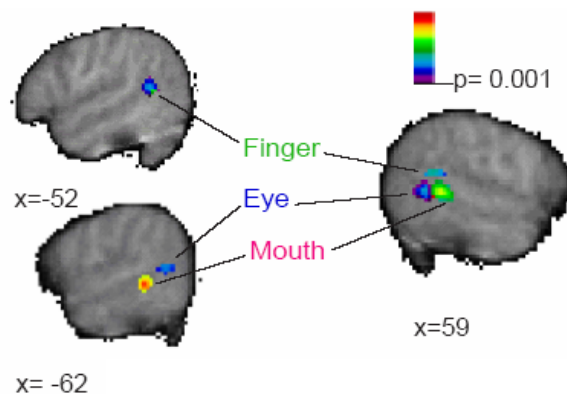


FIGURE II.24 : Représentations somatotopiques des mouvements de bouche, œil, doigts dans le Sulcus Temporal Supérieur. (D'après Dubeau et al., 2002)

Une étude de Iacoboni et al. (in press) suggère que l'activité accrue dans STS soit due à des copies efférentes de commandes motrices qui sont originaires des aires miroirs fronto-pariétales et qui sont envoyées à STS pour contrôler les buts. Le dialogue fonctionnel mis en évidence constituerait selon les auteurs une intégration de la description visuelle d'une action avec la planification motrice requise pour la répliquer. Dès lors, les conséquences sensorielles qui vont être prédites à partir des plans moteurs pour imiter l'action observée, vont être envoyées en retour au STS afin de les contrôler.

Dubeau et al. (dont Iacoboni) ont déjà proposé un circuit neural minimal pour la compréhension et l'imitation de l'action, qui représente des compétences fondamentales pour l'interaction sociale. STS va fournir une première description visuelle des intentions de l'action. Le cortex pariétal postérieur (PPC) va ensuite ajouter l'information somato-sensorielle à l'information visuelle. Et l'aire de Broca

(BA44) va coder le but de l'action. Une représentation de ce circuit est donnée dans la figure II.25 suivante :

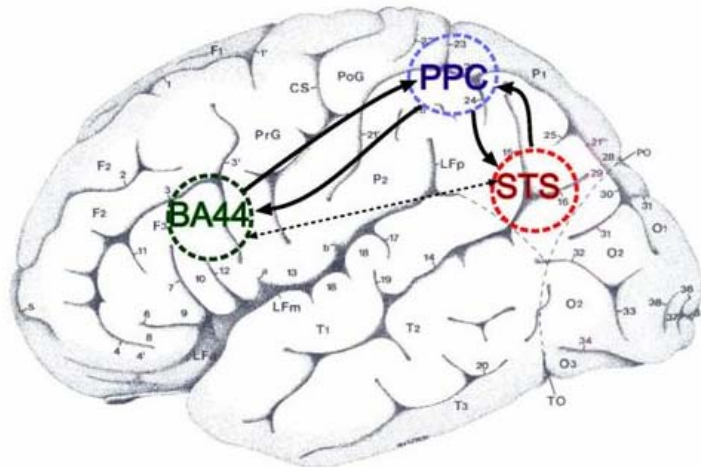


FIGURE II.33 : Le circuit minimal pour la compréhension et l'imitation de l'action. (D'après Dubeau et al., 2002)

Dans une tâche d'imitation, Iacoboni et al. (in press) détaillent cette architecture neurale similaire : le cortex temporel supérieur fournit une description visuelle de l'action observée —celle qui doit être imitée— aux neurones miroirs du cortex pariétal postérieur. Les neurones miroirs fournissent une information somatosensorielle supplémentaire sur l'action qui doit être imitée et envoient cette information aux neurones miroirs du frontal inférieur. Les neurones miroirs du frontal inférieur codent le but de l'action. Les copies efférentes de commandes motrices fournissent les conséquences sensorielles prédites des actions imitatives planifiées et sont envoyées à STS en retour. STS effectue un lien entre la description visuelle de l'action et les conséquences sensorielles prédites. Si le lien est bien établi, l'action imitative peut alors débuter.

Pour reprendre la proposition de Dubeau et al., et la reformuler selon nos propres termes, quand quelqu'un nous fixe du regard (« Qu'est-ce qu'il me veut, celui-là ? »), un circuit temporo-pariéto-frontal est activé, pour comprendre l'intention de l'action, avec la possibilité de décliner cette intention articulateur par articulateur (comme : « Qu'est-ce qu'il veut me faire avec ses mains ? Avec sa bouche... »). Dans ce but, les trois questions, liées aux trois fonctions cérébrales en jeu, vont être somatotopisées en *body-parts*. La première question est, dans le sulcus temporel supérieur (STS) : « Qu'est-ce que l'autre a l'intention de me faire par ce geste avec cette partie de son corps ? » (« You dig it ? »). La seconde question, au cœur du cortex pariétal postérieur (PPC), est du type : « Quelle sensation ça me ferait si on

me faisait subir cette action avec ce geste précis» (« You feel it ? »). Suivie de la troisième question, pour l'aire de Broca (BA44) : « Comment va-t-il s'y prendre précisément ? » ou « Au cas où j'aurais à le faire, comment ce geste est-il organisé ? » (« How to do it ? »).

Nous venons ainsi de voir que la représentation des différentes parties significatives du corps est somatotopisée. Les actions réalisées par la main, la bouche, l'œil sont cartographiées dans des zones cérébrales spécifiques. En se plaçant toujours dans la compréhension des actions/intentions de l'autre, une question survient alors quant à cette découverte. Une action reliée à un objet, transitive (par exemple je vais taper dans un ballon avec le pied), active-t-elle les mêmes aires cérébrales qu'une action qui n'est pas reliée à un objet, intransitive (par exemple, je vais taper avec le pied sans la présence du ballon) ?

Dans la lignée des travaux effectués quelques années plus tôt par l'équipe de Perrett, Buccino et al. (2001) apportent des éléments de réponse à cette question. Les auteurs ont testé si l'observation des actions réalisées avec différents effecteurs (bouche, main et pied) vont activer des zones spécifiques du cortex prémoteur en accord avec l'organisation motrice somatotopique de cette région. Ils ont également regardé l'influence de l'objet sur l'analyse d'une action observée.

Lors de l'observation de mouvements de bouche non reliés à l'objet (activité de mastication), les aires 6 et 44 sont activées bilatéralement et l'aire 45 est activée dans l'hémisphère droit. Lorsque l'action est reliée à un objet (mordre une pomme présente), on observe le même patron d'activation mais plus faible, auquel s'ajoute une activation dans le lobe pariétal plus proéminente à gauche, impliquant les aires 39 et 40.

Les mouvements de main/bras non reliés à l'objet (imiter une atteinte dans le but de saisir un objet absent) entraînent une activation bilatérale de l'aire 6, située de façon plus dorsale que l'activation liée à la bouche. Tandis que lors de l'observation de mouvements de main/bras reliés à un objet (mouvement d'atteinte pour saisir un objet présent), les auteurs notent une activation bilatérale du cortex prémoteur, une activation dans la zone 44, plus deux activations dans le pariétal, le sulcus intrapariétal (de façon plus rostrale et caudale que l'activation liée à la bouche), et le gyrus angulaire 39.

En ce qui concerne les mouvements de pied (imiter l'action de taper dans un ballon visible), le secteur dorsal de l'aire 6 est activé et la partie postérieure du lobe pariétal l'est également : l'aire 7, qui présente en partie un chevauchement des activations de la bouche et de la main.

Ces découvertes permettent aux auteurs de mettre en évidence une organisation somatotopisée des activations dans le cortex prémoteur et dans le lobe pariétal. Lorsqu'un individu regarde un autre individu réaliser une action, différentes parties du cortex prémoteur vont être recrutées selon l'effecteur utilisé, indiquant une topographie très nette dans les activations de son cortex. Ces résultats sont également un soutien fort pour l'idée que le système des neurones miroirs ne se limite pas simplement aux actions de la main, mais comprend un inventaire plus large d'actions corporelles, incluant le pied ou la bouche. Selon les auteurs ce système « constitue le substrat neural d'un mécanisme de liage cartographiant les actions observées sur les représentations motrices de l'observateur »<sup>19</sup>. L'observation de l'action recrute ici les mêmes structures neurales que celles impliquées dans l'exécution de l'action qui a été observée.

En ce qui concerne une différence d'activation entre une action liée à un objet (transitive) ou non (intransitive), leurs résultats apportent des éléments de réponse très intéressants. Lorsqu'un objet constitue la cible d'une action, les auteurs remarquent une forte activation du lobe pariétal, organisée elle aussi de façon somatotopique en fonction de l'effecteur utilisé. Ainsi le lobe pariétal semble jouer un rôle prépondérant dans la description de l'objet pour une action. Les actions qui sont observées sont cartographiées dans le lobe frontal en représentations motrices correspondantes. Les objets qui sont observés sont quant à eux cartographiés dans le lobe pariétal en représentations pragmatiques (affordances), qui vont être reliées à l'effecteur utilisé.

Nous venons de voir une activation spécifique forte du lobe pariétal pour la description d'objet lors d'une action. D'autres chercheurs ont montré une certaine spécificité neurale pour les actions liées à des objets cibles. Des neurones situés dans le sulcus temporal supérieur (STS) répondent à des stimuli biologiques en

---

<sup>19</sup> “[It therefore] constitutes the neural substrate for a matching mechanism mapping the observed actions on the observer’s motor representations” (2001:403)



mouvement (main, visage, corps), mais ne semblent pas répondre à des corps en mouvement ou à des parties du corps lorsque ceux-ci ne sont pas engagés dans des actions orientées vers le but. Certains neurones dans STS vont s'activer lorsqu'une main atteint et se saisit d'un objet mais ne déchargent pas si l'objet n'est pas saisi. Les neurones STS codent donc la vue d'une interaction entre un objet et un agent intentionnel.

Les études et la revue de question d'Allison et al. (2000) soutiennent le rôle prépondérant de STS dans les mouvements liés aux mains, aux yeux, à la bouche, et au corps plus généralement, en considérant cette région comme une composante essentielle qui entre en jeu dans la perception sociale. Par exemple, chez les singes, le mouvement de bouche constitue une partie essentielle des gestes oro-faciaux. En effet, des cellules dans STS répondent à des mouvements de bouche spécifiques comme l'ouverture de la bouche (indice de menace ou de peur), ou le « sourire » (indice de soumission ou d'affect positif). Cet exemple se suffit à lui-même pour démontrer l'importance de ce déclencheur social dans la relation entre congénères, et dans leur système de communication. Chez l'homme, nous pouvons observer que la région STS est activée par la parole et par la lecture sur les lèvres. Allison et al. se demandent alors si les gestes manuels liés à la parole activeraient de la même façon STS.

L'étude de Neville et al. (1998) montre dans un premier temps que, quelle que soit la modalité de langage, tous les sujets opèrent un traitement sur leur langue native, ceci étant mis en évidence par l'activation de structures cérébrales dans l'hémisphère gauche, traditionnellement impliquées dans le traitement du langage. Leur étude portait, entre autres, sur les enfants sourds congénitaux, dont la langue native est la langue des signes américaine (ASL), et leurs résultats en neuroimagerie fonctionnelle mettent en évidence que STS est activé bilatéralement pendant le traitement de la phrase ASL, indiquant que STS est impliqué dans l'analyse de la langue des signes américaine. L'activation dans STS liée à l'hémisphère droit semble résulter de l'analyse du mouvement biologique, tandis que l'activation liée à l'hémisphère gauche semble refléter le traitement linguistique, indépendamment du mode de transmission. Ainsi, STS semble être activé, chez les primates humain et non-humain, pour les gestes de la main communicatifs, ce qui ne semble pas être le cas pour les mouvements de main non significatifs.

En effet, on sait que les gestes semblent être une composante indissociable de la parole. Un exemple typique et très illustratif est celui proposé par Iverson et al. (1998). Les chercheurs ont étudié la communication spontanée chez 12 enfants aveugles congénitaux de 9 à 18 mois, comparée à des 12 enfants signant, ceci dans une série de tâches de raisonnement, qui visaient à éliciter des gestes. La première observation réside dans le fait que tous les enfants produisent des gestes en accompagnement de la parole, même les enfants aveugles, suggérant ainsi qu'aucune expérience ou aucun apprentissage ne sont requis pour produire spontanément ces gestes. Une expérience additionnelle de ces deux auteurs montre que des enfants aveugles congénitaux produisent plus de gestes que ne le font des enfants voyants signant, lesquels savent que leur interlocuteur est aveugle, et ainsi incapable de profiter de l'information véhiculée par les gestes. Les auteurs avancent alors que les gestes sont une composante essentielle du processus de communication parlée lui-même et soulignent le caractère robuste de cette composante dans la parole courante.

Allison et al. (2000), comme nous avons commencé de l'évoquer, ont suggéré que la région STS soit le siège privilégié de la perception sociale, là où s'opère l'analyse des mouvements biologiques, corporels, laquelle va ensuite impliquer les déclencheurs délivrant l'information qui sera pertinente socialement. Les auteurs citent parmi ces mécanismes déclenchés la théorie de l'esprit, la mentalisation, l'attention sociale, la cognition sociale. STS reçoit des inputs à partir des systèmes *What* et *Where* cités précédemment, ce qui implique que cette région intègre l'information sur la forme et le mouvement significatifs pour l'interaction. Allison et al. signalent le rôle d'autres régions cérébrales dans la perception sociale, l'amygdale et le cortex orbito-frontal (OFC, *Orbito-Frontal Cortex*), qui se projettent elles-mêmes dans le STS. Ces différentes régions amènent les chercheurs à proposer le système tripartite suivant (Cf. Figure II.26) pour refléter la cognition sociale chez le singe et le « mindreading » de l'homme :

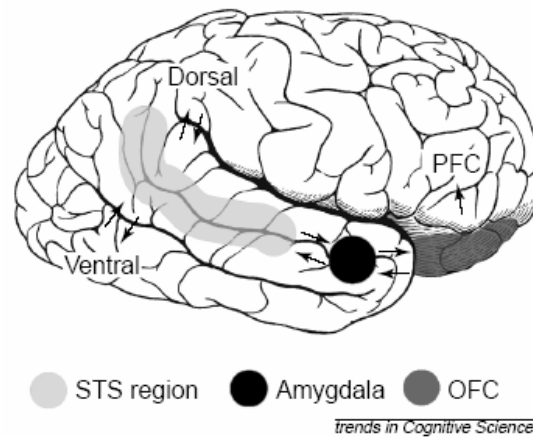


FIGURE II.26 : Les structures cérébrales fondamentales pour la perception et la cognition sociales chez l'homme ; le cortex orbito-frontal (en gris foncé : on y lit difficilement OFC) est en dessous du cortex préfrontal (PFC). (D'après Allison et al., 2000)

Dans ce réseau, STS n'envoie que des projections dites *feedforward* à l'amygdale, à laquelle il est connecté, et reçoit également des projections en feedback de cette dernière. Ce feedback peut induire une « amplification émotionnelle » de l'activité du STS. L'amygdale, sous-corticale, est quant à elle connectée au cortex orbitofrontal, lui-même connecté au cortex préfrontal (PFC). Ce dernier, connecté au cortex moteur et aux ganglions de la base, complète le réseau de la perception pour l'action. Emery (2000), que nous avons cité précédemment, avait déjà noté l'existence de ce réseau pour contrôler la direction de l'attention de l'autre, mais il semble donc que ce système rentre plus généralement dans le domaine de la perception sociale, englobant le contrôle du regard de l'autre.

L'expérience de Langton et al. (2000, Langton et Bruce, 2000) présentée dans la figure (II.26) ci-dessous, est particulièrement révélatrice du caractère véritablement irréprouvable (*mandatory*) du regard pour la deixis. Aussitôt qu'un visage est présent, même si l'on précise aux sujets que cette présence doit être négligée, un gain significatif de 20ms est obtenu si la cible apparaît dans la direction où le visage regarde.

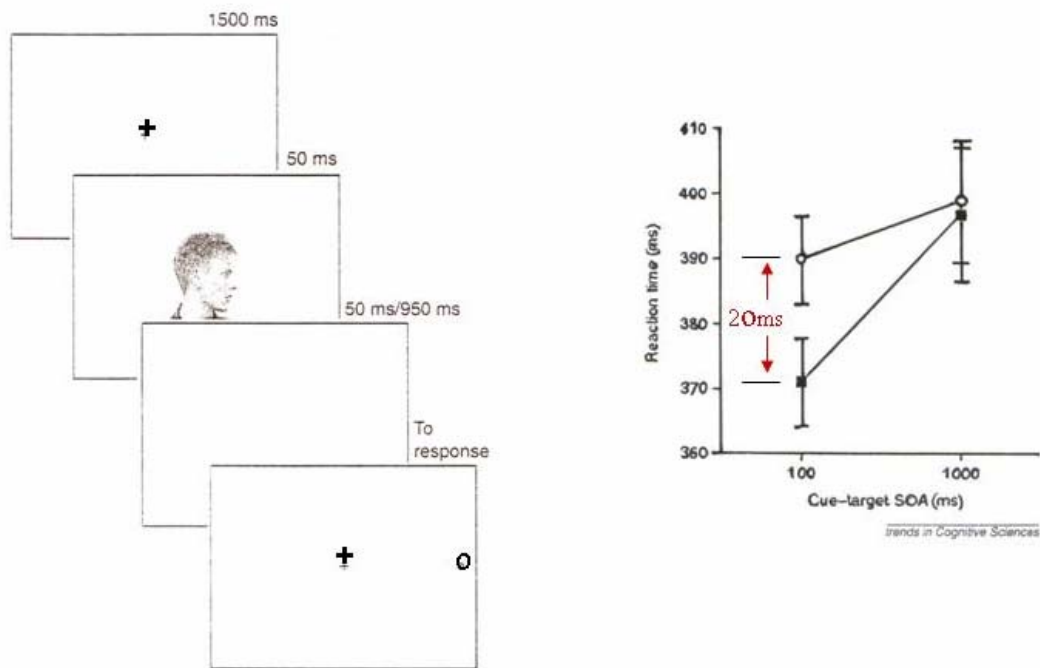


FIGURE II.26 : La présence d'un visage orienté vers l'apparition d'une cible (petit cercle) diminue irrésistiblement le temps de réaction de 20ms en moyenne, par rapport à une cible dans une autre direction, même si l'on instruit le sujet qu'il doit négliger la direction du visage. (D'après Langton et al., 2000)

### II.5.5. Comment passe-t-on de SAM à SIM ?

Nous avons présenté précédemment les différents mécanismes d'attention, avec une insistance particulière sur le mécanisme d'attention partagée SAM, et également les mécanismes d'intention, et particulièrement la notion d'intention partagée SIM (*Shared Intention Mechanism*) ou intentionnalité. Ces mécanismes revêtent une importance particulière dans le flux pragmatique-sémantique qu'est le cadre du signe. Mais comment s'établit le lien entre le SAM et le SIM ? En d'autres termes, comment l'enfant passe-t-il du mécanisme SAM au mécanisme SIM dans son développement ? Quel est le dispositif minimal qui va permettre l'interaction multimodale complexe mère-enfant ?

Pour répondre à ces questions, nous proposons l'idée que le mécanisme d'intention partagée minimal soit le pointer de l'index. Plus précisément, pour bénéficier d'une telle intentionnalité partagée, nous réclamons que la situation de communication face à face requiert les yeux de la maman et le pointer de l'index du

bébé. La notion de pointage déictique sera détaillée plus amplement dans la partie suivante.

Si nous reprenons la figure II.23 proposée par Tomasello et que nous l'appliquons à une situation de communication face-à-face entre mère et enfant, nous pourrions dire que l'enfant, dès dix mois, possède la compétence de partage d'attention (SAM, *Shared Attention Mechanism*), c'est-à-dire qu'il a la faculté de diriger le regard de sa mère vers un objet d'intérêt, mais qu'en outre il tire parti de cette aptitude pour partager avec l'autre ses intentions (SIM, *Shared Intention Mechanism*). L'enfant doté d'une capacité à (re-)diriger l'attention de sa maman, va sur cette base, partager son intention d'action avec elle, afin d'arriver à son but. Et le medium par lequel s'opère ce système SAM-SIM est le *pointing*. C'est le geste de pointage déictique intentionnel qui va remplir cette fonction cruciale pour le développement sémantico-pragmatique de l'enfant.

Une étude menée par Legerstee et Barillas (2003) apporte un soutien à notre idée que le suivi du regard, puis le pointage déclaratif soient indéniablement liés à la notion d'intention. Les auteurs soulignent le fait qu'entre 9 et 12 mois, les enfants subissent une transition-clé dans leur rapport au monde social via l'engagement dans des interactions triadiques. Ces échanges constituent alors les premières tentatives de l'enfant pour intégrer simultanément l'intérêt qu'il porte à un objet et l'engagement du partenaire de communication pour ce focus d'attention. Cette période transitoire est sans nul doute fondamentale pour le développement de facultés précoces indexant l'état intentionnel chez l'enfant. A ce moment précis, les enfants commencent à percevoir les gens comme agents intentionnels, agents qui peuvent avoir une perspective différente de la leur. Cette compréhension se manifeste à travers deux comportements très importants chez les enfants, le suivi du regard dès 6 mois et le pointage déclaratif autour de 12 mois.

Dans la littérature liée à ce domaine d'étude, il existe un certain consensus pour affirmer que ces deux comportements soient facilités par la maturation du concept de « personne ». Toutefois, certains détracteurs argumentent qu'ils n'impliquent pas pour autant la compréhension que les gens soient des agents intentionnels. Certains auteurs proposent que ce ne soit pas avant la fin de la seconde année de vie que le pointer de l'enfant reflète une compréhension conceptuelle du comportement intentionnel chez d'autres. C'est ce qu'ont voulu alors

tester Legerstee et Barillas (2003) au cours de deux expériences menées sur des enfants âgés de 12 mois à l'aide d'une procédure de conditionnement à tourner la tête (*head turn*). Les enfants sont conditionnés à suivre le regard ou tourner la tête soit vers une personne ou soit vers une poupée (classée comme « vivante »). Après ils sont soumis à un paradigme pour faciliter la production de gestes communicatifs sociaux (regard, vocalisation, pointer). Ces gestes sont observés : comment les enfants pointent vers des jouets actifs sonores en présence d'une personne ou d'une poupée ? Les résultats de la première expérience montrent que la plupart des enfants apprennent que le fait d'aligner leur tête avec celle du stimulus leur prédit un signe intéressant, que le stimulus soit une personne ou une poupée. Néanmoins, lorsque les enfants sont incités à générer de façon spontanée des gestes communicatifs, significativement plus d'enfants produisent des séquences particulières de communication qui impliquent le pointer, le regard, et des vocalisations qui s'adressent à la personne, par comparaison avec la poupée. Cela soutient l'idée que les enfants âgés de 12 mois construisent les personnes différemment des objets. Toutefois, dans cette expérience, les auteurs soulignent le fait que la personne regarde l'enfant et pas l'objet-cible, donc on ne peut réellement savoir si le fait que les enfants pointent différemment est une évidence du focus attentionnel de l'expérimentateur.

Ce qui a mené nos auteurs à conduire une seconde expérience dans laquelle ils testent si les enfants de 12 mois pointent différemment selon que l'adulte regarde l'objet actif sonore ou non. Cette expérience se décompose en deux conditions : une condition « in-focus », où l'enfant regarde le même objet que l'expérimentateur (2 chiens-objets, le focus se fait sur le même objet, le chien-objet regardé s'anime et aboie) ; et une autre condition « out-of-focus », où l'enfant regarde un chien-objet qui s'anime, tandis que l'expérimentateur regarde l'autre chien-objet. L'hypothèse des auteurs consiste à dire que si l'enfant comprend que l'expérimentateur porte attention à un autre objet que lui, alors il devrait regarder l'expérimentateur, pointer son objet d'intérêt, regarder à nouveau l'expérimentateur et vocaliser pour attirer l'attention vers lui et l'objet-cible. Les résultats de l'expérience font apparaître que les enfants de 12 mois différencient les deux types de condition et utilisent des réponses variées pour essayer de rediriger l'attention de l'expérimentateur vers l'objet-cible dans la condition « out-of-focus ». Ils voient ainsi les humains comme des agents

intentionnels dans la mesure où ils essaient de rediriger leur attention afin qu'ils puissent porter leur attention sur l'objet pointé par l'enfant, objet qui constitue l'événement d'intérêt pour celui-ci.

### ***II.5.5.1 Comprendre les intentions à 3 mois***

Sommerville, Woodward, Needham (2005) ont cherché à savoir si l'action altère la perception des propres actions d'enfants âgés de 3 mois. La compréhension des buts chez l'enfant est un domaine très intéressant dès lors que l'on sait que celle-ci guide l'apprentissage des premiers mots ou encore qu'elle gouverne l'apprentissage social et les essais de résolution de problèmes. Cette capacité à détecter des buts émerge dans l'enfance, et la première année de vie est cruciale pour le codage de certains événements interprétés comme dirigés vers un but. Ceci démontre une certaine faculté à se représenter l'action humaine et à se représenter certains mouvements d'objets.

Dans leur étude, les auteurs font la prédiction que l'expérience de l'action ait un impact sur l'interprétation même de l'action. Afin d'évaluer cet impact, ils ont focalisé sur la faculté de 30 enfants âgés de 3 mois à détecter la structure du but à partir d'un événement de saisie. Lors de l'expérience, la moitié des enfants était exposée à une tâche d'action avant une procédure d'habituation visuelle, et l'autre moitié au dispositif inverse. La tâche d'action consistait pour l'enfant à jouer avec deux jouets, une balle et un ours en peluche de façon libre, puis ensuite de mettre aux enfants des gants velcro auxquels peuvent s'accrocher les jouets, et de les laisser interagir avec ceux-ci librement également. La procédure d'habituation consistait pour l'enfant à regarder des événements de saisie qui lui étaient présentés. Les expérimentateurs mesuraient la durée du regard. L'enfant voyait dans un premier temps un expérimentateur atteindre et se saisir d'un des deux objets. Puis une fois que le taux d'attention de l'enfant chute, l'expérimentateur reposait l'objet et la position des jouets était inversée. La phase-test suivante présentait deux nouveaux événements-test. Un nouvel événement but : l'expérimentateur saisissait un jouet différent de celui qu'il avait manipulé pendant la phase d'habituation; ou un nouvel événement de trajectoire : l'acteur atteignait une localisation différente pour le même jouet pris lors de l'habituation.

Les résultats de cette expérience montrent que les enfants placés en condition d'action en premier regardaient significativement plus longtemps le premier test d'habituation que les enfants qui ont l'habituation en premier. Ce qui signifie, en d'autres termes, que le fait d'avoir manipulé les objets dans un premier temps va augmenter le potentiel d'attention de l'enfant à des événements d'atteinte similaires mais réalisés par une autre personne. Cet effet peut refléter la capacité d'un enfant à reconnaître des correspondances entre mouvements ou actions exécutées et observées.

Pour les deux groupes d'enfants, il apparaît également qu'ils regardent plus longtemps, de façon significative, un nouvel événement but plutôt qu'un nouvel événement trajectoire. Il semblerait ainsi que les enfants puissent détecter la structure dirigée vers le but d'un événement dans les actions d'une autre personne.

Le temps pendant lequel l'enfant dirige son regard et coordonne un contact manuel vers les objets paraît plus grand lorsque que les jouets sont manipulés à l'aide des gants.

Ainsi, à partir de l'expérience active, les enfants apprennent et transfèrent rapidement cette connaissance aux événements visuels observés. Ils sont donc capables de détecter la structure de l'action-but, dirigée vers un objet, et d'appliquer ensuite cette connaissance à leur propre perception des actions des autres. Cette découverte est, pour les auteurs, cohérente avec l'éventualité que les actions de soi et les actions des autres soient représentées de façon amodale dès très tôt en enfance, permettant ainsi à l'enfant de reconnaître la correspondance entre ses propres actions et celles des autres. En soi, le fait de former des représentations abstraites dirigées vers le but, constitue un mécanisme d'apprentissage puissant pour l'enfant, dans la mesure où les informations récupérées à propos de l'action pourront être transférées d'un agent à l'autre rapidement, quelle que soit la modalité.

#### **II.5.5.2. Compréhension de mouvements violant la biologie à 8 mois**

Que sait-on de la compréhension de l'enfant des actions humaines, et particulièrement les enfants sont-ils sensibles à des actions qui violent la biologie humaine normalement attendue ? Des études ont montré que des enfants aussi jeunes que 5 mois paraissent démontrer une sensibilité aux contraintes



biomécaniques du corps humain. Mais quelles sont les caractéristiques spécifiques auxquelles les enfants sont sensibles ? C'est à cette question que l'étude de Reid, Belsky, Johnson (2005) tente de répondre, en évaluant la sensibilité aux caractéristiques du mouvement biologique d'enfants âgés de 8 mois.

Leur première expérience comportementale présentait aux enfants un mouvement de membre (un torse humain avec un bras atteignant et saisissant un objet), que les adultes catégorisaient comme biologiquement impossible. Les enfants réagissaient à cette présentation en regardant plus longtemps le mouvement du corps impossible par rapport à un mouvement possible, démontrant ainsi leur surprise, et le caractère nouveau de ce mouvement. Ce résultat dénote l'existence de compétences motrices très fines leur permettant de discriminer ces stimuli, et cela dénote également l'existence chez l'enfant de 8 mois d'une association entre capacité à réaliser des actions motrices fines et capacité perceptive visuelle liée à l'observation du mouvement humain. Ainsi, il semblerait que la perception de l'action biomécanique fournisse une base pour le développement des ces compétences motrices fines.

Leur seconde expérience électrophysiologique, toujours sur des enfants âgés de 8 mois, confirme également l'association relevée précédemment, indiquant que l'expérience motrice est un bon prédicateur des facultés perceptives visuelles de l'enfant.

### **II.5.5.3. Compréhension de la structure intentionnelle et compréhension du pointer à 10 mois**

Nous venons donc de voir que les actions sont représentées sur la base du but, elles sont donc hiérarchiquement organisées et reflètent la capacité de l'enfant à saisir la pertinence causale entre une action et son but. Il reconnaît ainsi l'acte final d'une action comme le but ultime de la séquence engagée. Sommerville et Woodward (2005) notent l'existence de deux pré-requis pour la représentation de la structure du but lors de la première année de vie de l'enfant. Le premier pré-requis concerne le fait que les enfants représentent des actions simples comme étant dirigées vers le but, plutôt que de relever une trajectoire purement physique dans l'espace. C'est ce qu'ont démontré entre autres Reid et al. (2005) précédemment. Plus tard, vers 12 mois, les enfants vont interpréter des actions simples appartenant

à une séquence, comme étant dirigées vers un but, mais à un niveau plus haut. Les enfants vont ainsi lier les actions aux buts d'ordre plus hauts, sur la base de leur rôle causal dans l'atteinte du but. Cela démontre alors une capacité en développement qui donne des outils à l'enfant pour raisonner sur des situations causales particulières, contribuant à la représentation plus fine d'actions hiérarchiques. Les résultats des travaux menés par Sommerville et Woodward (2005) vont également dans ce sens. Cette faculté à comprendre l'action comme une figure hiérarchiquement organisée pour la représentation d'événement est très importante chez l'enfant comme chez l'adulte. Elle nous permet d'interpréter de nouvelles actions, ou des actions ambiguës, afin de prédire ou d'anticiper un but. Cette compétence est à la base de notre compréhension de l'action comme motivée et guidée par des états internes invisibles, les buts et intentions. Les auteurs montrent également que la période entre 10 et 12 mois chez l'enfant est primordiale dans la mesure où elle marque une transition dans la faculté de l'enfant à interpréter une action intentionnelle, compétence dont nous avons précédemment souligné le caractère crucial dans le développement de l'enfant. Il semble être en mesure d'apprendre la structure intentionnelle des actions des autres en les observant simplement, et en venant ensuite à appliquer cette connaissance dans leurs propres productions d'actions. Il semble donc très probable que le lien entre perception/compréhension et production de l'action soit réciproque ici.

Woodward et Guajardo (2002) nous apportent une piste supplémentaire à la lumière de leur étude sur la compréhension par l'enfant du geste de pointer, comme étant une action dirigée vers un but.

Les travaux précédents nous ont permis de mettre en valeur le fait que l'interprétation des actions des autres constitue une tâche critique pour les enfants. Critique dans le sens où cette compétence, et particulièrement la compréhension de l'intention de l'autre, permet à l'enfant de s'impliquer dans des actes communicatifs avec l'autre, et lui permet d'extraire alors des informations intéressantes sur les personnes, objets et événements qui l'entourent.

Cette compréhension implique de la part de l'enfant qu'il puisse tirer profit des gestes de pointage initiés par son partenaire d'interaction, et des composantes complexes que ce type de gestes engendre. Le geste de pointage est un élément pilier dans l'interaction et il fonctionne pour (i) accentuer l'objet d'intérêt, en

impliquant le suivi de la référence indiquée (et non pas le doigt qui pointe) ; (ii) diriger l'attention vers l'objet, dénotant une relation privilégiée entre la personne qui pointe et son objet-référent. La difficulté de compréhension de ce geste de pointage —et ce qui en fait toute sa puissance— réside dans sa nature dirigée vers l'objet, fournissant une base incontournable pour comprendre la nature communicative même du geste. Le pointeur ré-orienté l'attention de l'observateur vers son objet d'intérêt, dans un système d'attention partagé.

Que sait-on alors précisément de la compréhension de ce geste par l'enfant, et précisément que sait-on de la compréhension du caractère dirigé vers l'objet de ce geste par l'enfant ?

L'adulte extrait une quantité d'informations sur la relation particulière entre une personne qui pointe et son référent. L'enfant doit, quant à lui, saisir que cette réalisation instancie un tel type de relation, là réside toute la difficulté. Dans la littérature, il existe une évidence pour que, entre 9 et 12 mois, l'enfant commence à produire des pointers clairement dirigés vers l'objet, cela implique alors qu'il ait la connaissance préalable que le pointage se comporte comme une action intentionnelle, dirigée vers l'objet.

Woodward (1998), déjà cité, nous rappelle dans un premier temps que les enfants ne construisent pas toutes les actions manuelles comme étant dirigées vers le but. Ainsi, s'ils voient un acteur toucher un objet avec le dos de la main, qui reste inerte, ils ne l'interprètent pas comme dirigé vers le but mais comme étant simplement la représentation d'un événement de saisie. Cette découverte est importante car elle démontre le caractère très particulier et très pertinent des actions dirigées vers le but, dans le but communicatif.

Afin d'étudier la compréhension du pointage chez des enfants de 9-12 mois, Woodward et Guarjardo (2002) habitaient les enfants à un événement dans lequel un acteur pointait un des deux jouets (un ours et une balle). La moitié des enfants était exposée à la condition visage-main : l'expérimentateur cherchait le contact de l'œil avec l'enfant, disait « Hi » puis « look ! » en se tournant pour regarder, pointer et toucher de l'index un des deux jouets. Pour l'autre moitié, les enfants étaient exposés à la condition main seule, où l'acteur restait caché derrière le rideau, ne laissant apparaître que sa main pour pointer l'objet. Lors de l'habituation, l'acteur pointait toujours le même jouet, et après habituation, la position des jouets était

inversée à l'abri des regards. Les auteurs testaient ensuite deux paramètres : (1) les nouvelles trajectoires, où la trajectoire de la main de l'acteur avait changé, et (2) les nouveaux référents, où l'objet pointé était différent de celui lors de l'habituation (voir Figure II.27)

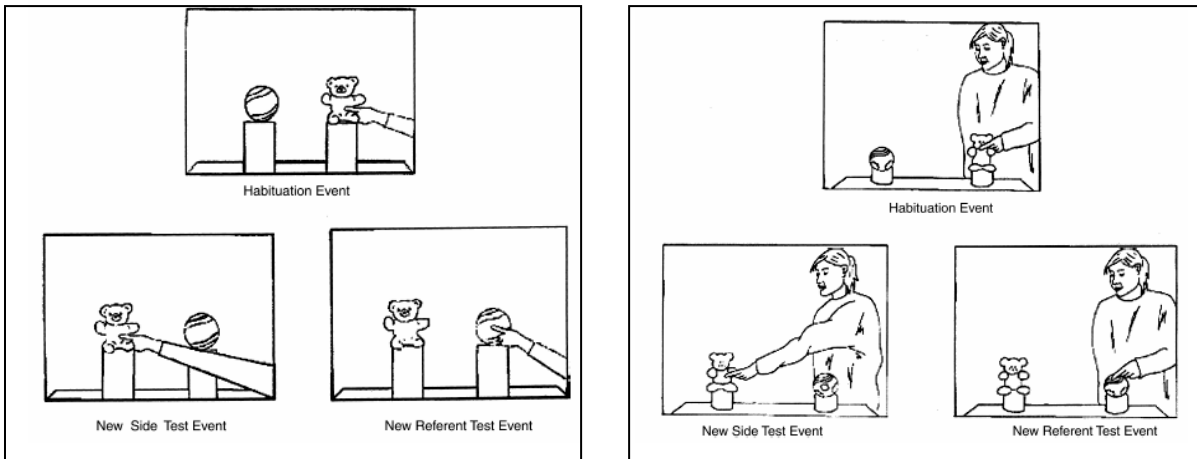


FIGURE II.27 : figure de gauche, exemple d'habituation et d'événements-tests pour la condition main seule ; figure de droite, exemple d'habituation et d'événements-tests pour la condition main et visage (d'après Woodward et Guajardo, 2002).

Les résultats mettent en évidence un temps de regard plus long sur les nouveaux tests référents plutôt que sur les nouveaux tests de trajectoire chez les enfants âgés de 12 mois, soulignant un intérêt dans le changement de la relation acteur/référent et un désintérêt dans le changement des propriétés de surface du mouvement de l'acteur. Ce qui est également intéressant, c'est que cette tendance n'est pas retrouvée chez les enfants âgés de 9 mois qui tendent à les regarder de façon égale.

Les auteurs remarquent que l'attention des enfants, qu'ils soient âgés de 9 ou 12 mois, est fortement dirigée par la main de pointage en direction de l'objet référent, mais notent qu'il est fort probable que ce focus attentionnel soit en lui-même responsable de la sensibilité des 12 mois au changement dans la relation acteur/référent. Si cette explication avait été valide, les 9 mois auraient alors également regardé les nouveaux tests référents plus longuement. Ce résultat significatif chez les enfants de 12 mois pourrait néanmoins refléter leur propre représentation de l'action de pointage, plutôt que simplement une réponse attentionnelle à cette même action. Ils auraient alors déjà codé l'action de pointage en termes de la relation entre acteur et objet.

De plus, l'étude de la production de pointer dans les deux groupes d'âge d'enfants suggère que leur émergence pourrait se chevaucher dans le temps avec le développement de la compréhension des pointers dirigés vers l'objet. Néanmoins, cette relation spécifique entre production de pointers dirigés vers l'objet et compréhension des pointers des autres comme dirigés vers l'objet ne semble pas si évidente à relier. Dans une seconde expérience étudiant cette relation sur des enfants de 9 mois, Woodward et Guajardo relèvent tout de même que les enfants qui pointent regardent plus longtemps la relation entre acteur et référent lorsqu'elle changeait, que les propriétés de surface des mouvements de l'acteur. Les non-pointeurs ne semblaient par contre ne pas différencier ces deux types de changement.

De ces expériences, les auteurs en dégagent le fait que les enfants âgés de 12 mois, par rapport à ceux de 9 mois, portent attention de manière spécifique à la relation entre acteur et objet du pointer. Les enfants de 9 mois diffèrent de ces derniers dans les traits ou représentations qu'ils affectent à la mémoire de l'action de pointage. Cette évidence semble donc suggérer une évolution dans les capacités de l'enfant entre 9 et 12 mois, période lors de laquelle il commence à comprendre la nature dirigée vers l'objet de ce geste de pointage. De plus, il semble bien, qu'au cours du développement, l'émergence de la production du pointage soit liée de quelque façon à la compréhension de ce geste, sous réserve de la mise en place d'autres ressources cognitives (celle de la mémoire à court-terme, par exemple). Toutefois, il n'est apparemment pas encore clair si c'est la compréhension de la nature dirigée vers l'objet du pointage qui entraîne la production de celui-ci, ou si la propre expérience de production de pointage dirigée vers l'objet amène à la compréhension de ce geste particulier. Butterworth et Grover (1990) citent l'étude de Schaffer, réalisée en 1984, qui soutiendrait la vision que la compréhension du pointage manuel survient à la fin de la première année chez l'enfant, en avance sur la production du geste. En effet, à 12 mois les bébés seraient capables de détecter le référent pointé par l'autre, alors que la production de pointer en tant que tel ne serait observée qu'à partir de 14 mois.

La compréhension du pointage a été liée de façon récurrente dans la littérature au développement cognitif de l'enfant. Le principe célèbre suivant « Quand le doigt montre la lune, l'idiot regarde le doigt » illustre parfaitement le lien existant

entre ces deux points de développement. Vers 6-9 mois, l'enfant ne comprendrait pas le geste de pointage (et le produirait peu en conséquence), fixant autant la main que la cible (comme le ferait l'idiote !). En revanche, à 12 mois, il le comprend très bien. Selon Butterworth et Grover, la compréhension du pointage manuel surviendrait donc à 12 mois, au moment même où le nouveau mécanisme géométrique qu'ils proposent serait disponible pour l'enfant.

Nous avons précédemment souligné l'importance de la mise en place de l'indexation dans le monde par l'œil, premier outil disponible pour l'enfant. Il est maintenant temps de nous intéresser en détail à la mise en place de l'indexation du monde par l'index, point crucial de notre thèse, afin d'en expliquer le développement propre, et pourquoi les gestes sont si importants dans la naissance de la parole.

## **II.6. La mise en place de la production du geste de pointage chez l'enfant dans le développement du cadre du signe**

“Pointing is the royal road to language for babies”  
(Butterworth, 2003)

Comme le notent entre autres Leekam et al. (1998), ou Deak et al. (2001), outre la direction de la tête et la direction du regard, d'autres déclencheurs, tels que le toucher, le pointer ou encore des vocalisations, rendent les objets et leur direction dans l'espace plus saillants aux enfants. Le pointer de l'index augmente la probabilité d'une réponse en magnifiant l'effet perceptif et en améliorant l'exactitude de la localisation spatiale à l'intérieur du champ visuel. Des enfants de moins de 18 mois ont plus de chances de répondre à un geste de pointer (déclencheur de haute validité selon Deak et al., 2001) d'une autre personne qu'à la direction du regard seul (déclencheur à validité modérée selon Deak et al., 2001), et ils ont plus de chances de répondre à la direction de la tête qu'à celle des yeux seuls. Les actes d'attention conjointe tels que les pointers fournissent ainsi des opportunités précoces pour la référence conjointe et la nomination de l'objet.

## II.6.1. Les origines ontogénétiques de la production de pointage

Le pointage n'est « rien sauf un mouvement de saisie abrégé » ?  
(W. Wundt, 1912)

Le thème des origines ontogénétiques de la production de pointage a été discuté à de nombreuses reprises dans la littérature, et les auteurs ne s'accordent pas toujours sur celles-ci. Une théorie répandue supposait que le geste de pointer de l'index dérive de mouvements d'atteinte ou de saisie, supposant que le mouvement lié à l'atteinte et la saisie serait réalisé de façon plus « économique » dans le pointer, le pointer les remplaçant comme geste référentiel. Une vision célèbre, dans cette même veine, est bien entendu la théorie vygotkienne du pointer (Vygotsky, 1988), dans laquelle l'auteur explique que le pointer se développe en dehors de l'interprétation de la mère, à partir d'essais ratés de l'enfant pour se saisir d'un objet. Autant d'essais avortés à prendre un objet conduiraient donc l'enfant à pointer vers celui-ci pour essayer de l'obtenir. Ces succès occasionnels feraient alors réaliser à l'enfant l'utilité communicative des atteintes échouées. Toutes ces hypothèses sur l'ontogenèse du pointer présentent comme point commun le fait que le développement cognitif autour du premier anniversaire de l'enfant induit l'enfant à apprendre ce nouveau comportement qu'est le pointer.

Nous réfutons ici cette vision vygotkienne, ainsi que celle de Wundt proposée au début du 20<sup>ème</sup> siècle, qui ne paraissent pas appropriées à la lumière d'études récentes qui démontrent que le geste de pointage n'est en soi aucunement un dérivé de gestes de saisie ou d'atteinte. Nous prendrons comme soutien l'étude de Masataka (2003) qui présente, dans son étude longitudinale sur des enfants japonais, l'évidence que le pointage de l'index émerge de l'extension de l'index mais pas de l'atteinte.

En attendant, une autre étude nous permet de mieux saisir la différence qu'implique chez l'adulte la saisie par rapport au pointage au niveau des circuits cérébraux mis en jeu par ces comportements. L'étude de Simon et al. (2002) nous paraît pertinente à ce sujet dans la mesure où leur but était de caractériser l'organisation fonctionnelle du lobe pariétal humain, alors que les sujets testés réalisaient plusieurs tâches différentes, dont le pointage et la saisie (toujours réalisés

avec la main droite). La tâche de saisie et de la tâche de pointage impliquent un chevauchement d'activités considérable dont une activation unilatérale du sulcus central gauche dans la région de la main, une activation post-centrale gauche et précentrale gauche, une activation du gyrus frontal supérieur mésial bilatéral au niveau de l'aire motrice supplémentaire et une activation cingulaire antérieure bilatérale. Toutefois, il subsiste d'importantes différences entre ces deux actions dans le lobe pariétal. En effet, le pointage cause en plus une activation bilatérale du segment horizontal du sulcus intrapariétal, mais peu ou pas d'activité supramarginale antérieure. Les chercheurs observent en plus une activation dans la partie postérieure du lobe pariétal supérieur bilatéralement, s'étendant jusqu'au précunéus gauche. Une activation symétrique du putamen antérieur et une activation thalamique gauche viennent également en sus. L'aire spécifique à la saisie chez l'homme paraît être l'homologue de l'aire intrapariétale antérieure (AIP) chez le singe, une région impliquée dans la forme visuellement guidée de la main. Par ailleurs, des lésions à cette aire entraînent un déficit sélectif de la coordination des mouvements de doigts requis pour la saisie.

Dans la recherche sur l'origine du pointer de l'index chez l'enfant s'opposent deux alternatives. La première alternative propose que les enfants communiquent initialement parce qu'ils ont des désirs et ont besoin des autres pour satisfaire leurs demandes. Cette proposition serait soutenue si nous présentions l'évidence que le pointage émerge des actes de saisie et d'atteinte, qui apportent les objets désirés à l'enfant. La seconde alternative consiste à ajouter à ces demandes, une nouvelle motivation de l'enfant pour communiquer par un désir intrinsèque de partage d'expériences intersubjectives. En d'autres termes, les enfants éprouvent le désir de partager avec les autres ce à quoi ils portent attention.

Les résultats de Masataka (2003) semblent indiquer que le début du pointer de l'index puisse être vu comme un produit développemental de l'extension de l'index, qui en serait alors son précurseur. Selon l'auteur, il existe une forte corrélation entre la production de vocalisations syllabiques et la production d'extensions de l'index chez des enfants aussi jeunes que 3 mois. Sur la figure II.28 réalisée à partir de Masataka (2003), nous observons que les enfants produisent un nombre croissant d'extensions de l'index au fur et à mesure du développement. Leur fréquence atteint son pic quand les enfants ont entre 11 et 12 mois. Selon Masataka,



cette période coïncide exactement avec celle du début du pointer de l'index. De plus, une fois que le pointer a émergé, sa fréquence augmente de façon nette tandis que la fréquence des extensions d'index chute brusquement. Par comparaison, le nombre d'occurrences d'atteinte n'a pas changé significativement en fonction de l'âge de l'enfant.

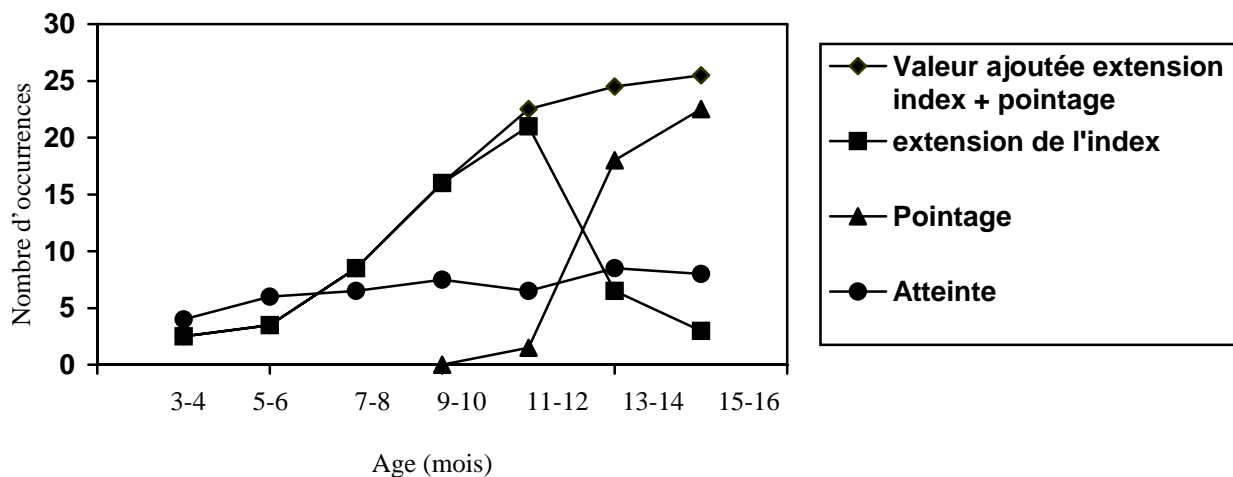


FIGURE II.28 : Nombre d'occurrences relevées pour les extensions de l'index (carrés), les pointers de l'index (triangles) et les mouvements d'atteinte (ronds) pour 8 enfants japonais suivis longitudinalement de 3 à 16 mois. Nous avons calculé les occurrences additionnées (diamants) des extensions de l'index et des pointers de l'index, pour montrer que les gestes du pointage prennent le relais en continuité des gestes d'extension de l'index, lorsque ceux-ci sont peu à peu remplacés par le pointer (D'après Masataka, 2003).

La seconde alternative est alors soutenue par ces résultats puisque le pointer surgirait d'autres actes manuels qui seraient reliés à l'exploration et l'auto-régulation de l'attention. Cette seconde proposition prédit alors que l'extension de l'index soit reliée à cette exploration et auto-régulation de l'attention et non pas au désir même de l'enfant d'apporter un objet à lui. Masataka argumente que lorsque les enfants développent le désir de partager l'exploration et l'attention de façon intersubjective, l'extension de l'index se développe en pointer de l'index. L'extension de l'index devient le pointer de l'index lorsqu'il est augmenté par une extension du bras intentionnelle vers un objet ou une localisation qui capture l'attention de l'enfant, probablement dans le but de partager l'attention avec quelqu'un d'autre. Selon Masataka (2003), il existerait donc une continuité développementale entre ces deux phénomènes. Ce qui nous intéresse particulièrement dans cette étude, outre le fait que nous soutenons également l'idée que le pointer de l'index n'émerge en aucun cas de mouvements de saisie et d'atteinte, ce sont les résultats de la valeur ajoutée

des extensions et des pointers de l'index que nous avons calculés sur ces données de Matasaka (Figure II.28), afin de montrer l'explosion de ce mécanisme gestuel dès 10 mois, peu après la mise en place du babillage canonique chez l'enfant. Contrairement aux mouvements d'atteinte qui restent assez stable à travers le temps et qui ne peuvent être en aucun cas corrélés à la naissance de la parole chez l'enfant.

Le pointage a été jugé par d'autres auteurs comme étant réalisé pour soi et non dans un but socio-communicatif. D'autres encore voient le pointage comme un simple mouvement exprimant à l'origine un souhait de saisir. Selon ce courant d'idées, le pointage n'aurait alors aucune fonction communicative. Enfin certains pensent également que le pointage manuel peut se développer hors de l'inspection visuo-tactile conjointe des objets.

Dans les années 80 émerge un nouveau courant d'idées sur le pointage, soutenu par des auteurs tels que Fogel et Hannan (1985), Fogel et Thelen (1987). Ceux-ci mettent en avant le côté universel du pointer, et son caractère typique à notre espèce, et soulignent surtout le dessein communicatif de celui-ci. L'acte de pointer devient alors une fonction communicative spécialisée dans un système d'attention partagée, qui pourrait être inné. Il n'apparaît plus comme un acte manqué de saisie, mais constitue un geste complexe à part entière qui implique une coordination précise entre extension du bras et de l'index vers un objet référent d'intérêt, en contrôlant si la personne en présence a bien localisé et dirigé son attention vers celui-ci.

L'idée que le pointage n'émerge pas d'un acte manqué de saisie est à présent fortement soutenue par de nombreuses études dont celle de Franco et Butterworth (1988, cité dans Butterworth et Grover, 1990) qui soulignent les différences entre l'action de pointer et l'action d'atteindre. L'atteinte n'est pas adressée à des stimuli distaux, généralement hors-de-portée, à l'inverse du pointer. Comme ils le remarquent si bien, l'atteinte est au service de la saisie tandis que le pointer est au service de la re-direction de l'attention, deux actions qui finalement sont ontogénétiquement très éloignées :

Reaching in the service of grasping, and pointing in the service of re-directing attention, are probably not ontogenetically closely related action. (Butterworth et Grover, 1988:620-621).

## **II.6.2. Le pointage déictique comme une action communicative de base chez l'enfant**

Le pointage de l'index nous apparaît donc comme un moyen de référence, plutôt spécifique à l'espèce humaine, étroitement connecté à l'acte communicatif, qu'il soit gestuel ou langagier. Butterworth (1998) postule même que celui-ci soit la « voie royale » menant de la communication pré-verbale au langage parlé. Il constitue l'émergence d'un contrôle moteur majeur pour l'apprentissage du langage. Tout comme le babillage, le pointer de l'index survient indifféremment à la culture d'origine (rien ne semble moins sûr selon Wilkins, 2003 ; mais ces données négatives semblent bien anecdotiques), mais surtout il ne semble pas être un phénomène d'imitation. Comme le souligne Masur (1983), il se développe initialement presque exclusivement comme un geste spontané, apparaissant sans comportement d'élicitation de la part de la mère. En moyenne, 93% des pointers dans les premiers mois de l'enfant sont spontanés, ce qui indique une fonction attentionnelle d'auto-direction des premiers pointages.

### **II.6.2.1 Pointage impératif, pointage déclaratif...**

La première unité métrique contrôlée correspondant à un signe pourrait être, comme nous l'avons annoncé, le « pied » (*foot*), qui reposerait sur le timing du geste de pointer porté par le bras. Très tôt l'enfant montre et interroge sans se servir du signe. On peut interroger avec les yeux, montrer également. On peut montrer avec le bras et interroger avec la voix (intonation). Pour acquérir le mot, l'enfant doit pouvoir contrôler une unité de la taille du signe, qu'il va amorcer et ancrer à l'aide des mécanismes de la monstration. Deux phases de cette monstration semblent très importantes pour l'établissement du signe : le pointage impératif et déclaratif.

À 9 mois, c'est-à-dire pendant la phase du système « Where/There », l'enfant établit un geste de pointer dit impératif dans la mesure où il est utilisé pour saisir un objet inaccessible. C'est ce pointage impératif que Vauclair (2002) attribue aux primates comme étant pour eux le seul possible à réaliser. Plus tard, à 12 mois, émerge le pointage déclaratif qui se définit par un bras étendu et un doigt, l'index spécifiquement, qui a pour fonction de faire porter précisément l'attention sur un objet distal. À la différence du pointage impératif, une requête n'est pas demandée pour

l'objet. De plus, ce type de pointer se spécifie par le fait que les enfants ne pointent pas en l'absence d'un observateur.

Cependant, si on revient sur l'idée qu'un bonobo ne sait pas déclarer, et qu'il pointerait uniquement en impératif, nous pouvons faire la remarque que 60 % des enfants font de même. En disant « yune » ils enjoignent un « Regarde la lune »; et ils ne produisent pas davantage du déclaratif en disant « (gade) a(v)ion! », même si la plupart des psychologues vont comptabiliser cette absence notoire de déclaration pure, type « ceci est un avion », dans les 40% de pointers déclaratifs du petit être humain.

Dans la littérature, le pointage communicatif est souvent considéré comme se décomposant en un pointage qui, d'une part, utilise plutôt l'adulte comme un outil pour obtenir un objet –c'est le pointage impératif ou proto-impératif– et d'autre part, un second type de pointage –le pointage déclaratif ou proto-déclaratif– qui considère l'objet comme un outil pour obtenir l'attention de l'adulte, avec une certaine compréhension de la notion de causalité. Le pointage impératif se voit interprété comme une explication comportementale assez pauvre, tandis que le pointage déclaratif est interprété comme une explication mentaliste beaucoup plus riche. C'est un des arguments qui peut être avancé lorsque l'on atteste du pointage chez les singes ou les autistes : ils peuvent pointer impérativement en l'absence de compréhension de l'agentivité mentale des partenaires de communication (Baron-Cohen et al., 1985 pour les autistes, Call et Tomasello, 1996 pour les singes). Faire un pointage impératif ne reviendrait qu'à stimuler les partenaires pour qu'ils entament une action.

Le pointage déclaratif refléterait quant à lui une sensibilité certaine à cette notion d'agentivité mentale partagée par les autres. Il n'aurait pas pour vocation de faire entrer quelqu'un dans une action précise, mais il viserait plutôt à changer l'état attentionnel de l'autre. Cette capacité nécessite donc une motivation première qui est celle de partager l'attention, appuyée par le suivi du regard, la monstration.

Camaioni (1993) fait partie des chercheurs qui proposent un décalage développemental net entre pointage impératif et déclaratif. Selon cet auteur, il serait nécessaire d'établir une transition socio-cognitive pour passer du premier au second, impliquant la compréhension que les personnes ne sont plus seulement des « agents d'action » mais également des « agents de contemplation ». Le pointage déclaratif

révèle alors la compréhension de l'intentionnalité de l'autre, qui n'est pas requise pour le pointage impératif. Selon Camaioni (1993), les enfants pointeraient seulement pour influencer le comportement d'autrui.

### **II.6.2.2. ... ou plutôt pointage coopératif ?**

Ces réflexions ont mené Liskowski (2005), dans un article absolument passionnant pour notre étude, à reconsidérer ce pointage chez l'enfant. Selon lui, le pointer n'est pas une action dirigée vers un but individuel dans l'environnement physique, comme manipuler un objet. Il s'agit plutôt d'une activité coopérative entre individus. Le pointage est un outil rapide et sans effort pour apporter par exemple une information à quelqu'un, aider à trouver un objet d'intérêt ou encore apporter une aide si le référent est mal identifié. Cette activité implique alors que l'interlocuteur puisse comprendre différentes relations, dont le partage d'attention entre locuteur et interlocuteur et des connaissances mutuelles entre partenaires de communication (comme défini par Tomasello, 1999).

Chez le primate non-humain, Leavens et Hopkins (1998) ont montré qu'ils produisaient des gestes de pointage en captivité (comportement que l'on retrouve rarement chez les singes à l'état sauvage). Masataka (2003) avance que le pointer de l'index soit apparemment unique aux humains, invoquant également le fait que les grands singes ne pointent pas à l'état naturel. Toutefois certains chercheurs soutiennent que leurs gestes manquent de facultés socio-cognitives nécessaires pour l'assimiler à du pointage déclaratif (Povinelli et al., 2003), et que les singes rencontrent des problèmes pour comprendre réellement l'intention communicative du pointage (Itakura, 1999). Selon Tomasello (sous presse), ils ne semblent pas s'engager clairement dans ce qui pourrait ressembler à de la communication humaine.

Ces considérations primatologistes ont amené Liskowski (2005) à se demander si, au cours de l'ontogenèse humaine, le pointer ressemble au pointage des grands singes, ne reflétant pas initialement de compréhension mentale, ou si, au contraire, ces pointers portent déjà des propriétés cognitives et motivées à portée communicative, qui seraient spécifiquement humaines.

Dans une étude antérieure (Liskowski et al., 2004), les raisons pour lesquelles les enfants de 12 mois pointent dans un contexte déclaratif classique ont

été étudiées. Les chercheurs s'attendent en fait à ce que les enfants aient un comportement distinct selon que l'adulte partage l'attention avec lui ou non. Leur étude porte sur 75 enfants allemands d'âge moyen de 12 mois et 6 jours, divisés en 4 groupes testés sur une des conditions d'expérimentations suivantes. (1) Condition d'attention conjointe (l'expérimentateur regarde vers l'arrière de façon répétée allant entre un événement et le visage de l'enfant, parlant avec excitation du stimulus qu'ils voient tous deux, et en utilisant des phrases comme « Oh wow ! What's that ? Are you showing Grover to me ? Yes, he is blue »). (2) Condition visage : l'expérimentateur regarde le visage de l'enfant sans jamais regarder l'événement et parle avec excitation à l'enfant. (3) Condition événement : l'expérimentateur ne regarde que l'événement, ne parle pas et ne montre aucune excitation. Et enfin (4) : la condition Ignorer, où l'expérimentateur regarde ses mains, sans jamais regarder ni l'enfant, ni l'événement, et cela sans excitation. L'analyse des données a fait ressortir le fait que les enfants de 12 mois ont plus souvent pointé parmi les tests où l'adulte partageait activement son attention et son intérêt (condition d'attention conjointe), comparé aux trois autres conditions. Toutefois, dans les conditions n'impliquant pas l'attention conjointe, les chercheurs ont observé que les enfants ont répété leur pointer plus souvent, indiquant vraisemblablement que l'enfant n'obtenait pas la réponse adulte qu'il désirait (l'attention conjointe de ce fait). Les auteurs relèvent en outre le fait que les pointers durent plus longtemps dans la condition d'attention conjointe, mais également dans la condition visage, indiquant que face à des émotions positives véhiculées par l'expérimentateur, l'enfant souhaite prolonger ou maintenir l'interaction. Dans la condition où l'expérimentateur ne regardait que l'événement, Liskowski et al. (2004) remarquent que les enfants regardent avec plus d'insistance le visage de ce dernier. Ceci peut être interprété par le fait qu'ils étaient déroutés que l'expérimentateur regarde l'objet sans montrer de réaction en rapport avec celui-ci.

Ces résultats tendent donc à prouver que les enfants pointent principalement pour partager l'attention sur un objet avec un partenaire. Les enfants pointent intentionnellement et communicativement, et adaptent leur comportement communicatif en fonction de la réponse comportementale du partenaire de communication. Ce partage l'attention implique toutefois que l'enfant soit capable de (re-)diriger l'attention d'une autre personne, mais également de recevoir un

commentaire sur l'objet d'attention mutuel. Le fait de rediriger l'attention implique en plus le fait que l'enfant soit sensible au fait qu'une autre personne soit focalisée ou non sur l'objet d'attention, et que l'enfant comprenne ainsi que l'état attentionnel d'une autre personne puisse être dirigé ou re-dirigé par le biais d'un geste communicatif, le pointer de l'index. Si l'attention de l'adulte est déjà portée sur l'objet d'intérêt, les enfants ont alors recours au pointage pour solliciter un commentaire sur l'événement auquel lui et le partenaire portent conjointement attention. En conclusion, à la base de toute forme de pointer, il existe un motif pour diriger l'attention d'une personne, une composante qui revêt un caractère particulièrement crucial pour notre propre étude des pointers d'enfants.

Dans son étude menée un an plus tard, Liskowski (2005) a voulu étudier ces deux derniers pré-requis au pointage déclaratif : diriger l'attention et recevoir un commentaire dans un contexte déclaratif, afin d'obtenir du pointage chez l'enfant, en expérimentant la violation des attentes de l'enfant dans 4 conditions : (1) condition attention conjointe : l'expérimentateur porte attention au référent de l'enfant, et montre des émotions positives, (2) condition malentendu : comme la précédente sauf qu'une barrière obstrue la ligne de vision de l'expérimentateur vers le référent de l'enfant, conduisant l'expérimentateur à référer de façon erronée à un morceau de papier insignifiant attaché à la barrière, (3) condition non-intéressé : l'expérimentateur réagit comme dans la condition d'attention conjointe, sauf qu'il commente le référent de façon neutre, montrant un certain désintérêt et (4) condition non-partage : utilisation d'une barrière et commentaire neutre sur le référent.

Il apparaît que les enfants sont plus satisfaits dans la condition d'attention conjointe, en effectuant plus de pointers que dans les autres conditions. Lors de la condition malentendu, les enfants ne sont pas satisfaits, ils persistent dans le message en pointant de façon répétée au référent avec significativement plus d'alternances du regard et de vocalisations. Dans la condition non-intéressé, ils ne répètent par contre pas le pointage, ne sont pas satisfaits de la réponse neutre et pointent moins que dans la condition d'attention conjointe. Ces résultats indiquent clairement une fois encore que les enfants pointent dans le but de diriger l'attention d'une autre personne vers un événement ou objet d'intérêt. L'enfant ne pointe donc pas simplement pour obtenir une émotion positive commentée sur un événement mais ils pointent comme une offre d'interaction mutuelle et de partage d'intérêt avec

un partenaire intéressé. Il semble donc très improbable que les enfants ne recherchent simplement qu'à ce que l'adulte s'oriente seulement comportementalement vers le référent.

A la suite de ces deux expériences et du point de vue de la motivation de l'enfant, il apparaît qu'à 12 mois, le pointer soit utilisé comme un acte communicatif social de façon inhérente, et cette structure coopérative mise en place dans ce type de relation triadique indique que ces actes communicatifs peuvent être interprétés comme un acte uniquement humain, comme le propose Tomasello (in press).

Une expérience toute récente de Liszkowski et al. (2006) explore les motifs de l'enfant dans le fait de diriger l'attention des autres, quand il utilise le pointer. Ils testent alors si les enfants peuvent être motivés pour aider quelqu'un lorsqu'ils pointent, pour lui apporter de l'information, tout comme le ferait un adulte qui pointerait différentes localisations pour apporter une information sur une route à prendre à autrui.

Liszkowski et al. (2006) soulignent que les adultes pointent impérativement tout en ayant acquis l'idée d'agentivité, intentionnelle et mentale, d'autrui, et que par conséquent, contrairement à ce qu'affirme Camaioni (1993), ils ne pointent pas pour influencer directement le comportement d'autrui, mais plutôt pour influencer les buts et l'attention d'autrui. Les enfants quant à eux ne pointent pas pour influencer le comportement de l'autre mais pour diriger leur attention et partager de ce fait une expérience d'un point de vue psychologique. Toutefois, les chercheurs soulignent qu'à ces deux motifs de pointage vient s'ajouter une troisième motivation, qui sort une fois encore de la dichotomie classique pointage impératif vs. déclaratif. En effet, le but n'étant pas d'obtenir un objet pour soi, ni d'obtenir l'attention pour partager un intérêt sur l'objet, mais plutôt ici de fournir une information à propos de l'objet à un partenaire de communication. Les auteurs ont donc testé cette dernière prédication dans leur étude sur des enfants âgés de 12 et 18 mois.

Lors de la première expérience, l'expérimentateur réclamait l'attention de l'enfant et commençait le test en prenant un objet (6 tests avec 6 objets différents). Les trois premiers objets étaient utilisés pour les tests passifs où l'enfant regardait l'expérimentateur utiliser les objets pour lui-même. Les trois derniers objets étaient utilisés de façon active, l'expérimentateur et l'enfant jouaient avec. Dans chaque test, après avoir utilisé ou joué avec les objets, l'expérimentateur, comme par



inadvertance, laissait tomber accidentellement l'objet sur le sol. Après ceci, il prenait un air surpris, regardant autour de lui en cherchant l'objet. Si l'enfant ne pointait pas, l'expérimentateur demandait « Where is it? Where is [la cible] now? ». Cette première expérience fit apparaître que les enfants dirigent l'attention d'une autre personne vers la localisation d'un objet, qu'elle se met à chercher lorsqu'il est tombé accidentellement. Ce qui suggère donc que l'enfant est capable de pointer communicativement dans un but autre que simplement demander ou offrir en partage.

Dans le but d'éviter que les pointers soient dus simplement au fait que, les enfants entendant l'étiquette de l'objet, cela induise une routine sociale, les auteurs effectuèrent une seconde expérience, dans laquelle ils contrôlèrent le plus possible les actions de requête ou de partage d'attention. Ils parvinrent à répliquer les résultats de la première expérience, démontrant que les enfants de 12 et 18 mois pointaient, même dans une situation plus centrée sur l'adulte, sans jouets intéressants, et même avant d'être verbalement questionnés sur la localisation de l'objet.

Leurs résultats montrent ainsi que les enfants ne pointent pas simplement pour diriger l'attention d'une personne qui est en train de chercher un objet, mais pointent aussi pour informer cette dernière. Ceci révèle clairement que les enfants possèdent la faculté de détecter quelle information est pertinente pour un adulte, et par là-même la faculté à considérer l'autre comme un agent intentionnel. Et cela révèle en outre qu'ils présentent une motivation pour fournir l'information à l'autre de façon coopérative.

Ces trois études récentes, absolument centrales pour notre travail, montrent ainsi que lorsque le pointage humain émerge, il consiste pleinement en un acte communicatif coopératif, impliquant, selon Liskowski (2005), la transmission intentionnelle d'information en dirigeant l'attention d'une autre personne vers un objet ou un événement. Liskowski offre une proposition qui à nos yeux est beaucoup plus pertinente qu'une simple distinction pointage impératif/déclaratif, à savoir l'idée qu'il faut s'intéresser aux motifs même du pointage plutôt qu'à ses différents types, motif qui sont humainement coopératifs par nature. A 12 mois, les pointers sont motivés par un intérêt mutuellement partagé dans le cadre d'un événement avec un partenaire coopératif : c'est donc un pointage plus informatif, comme montré par la

dernière expérience citée. Ainsi, le pointer humain, dans son ontogenèse la plus profonde est déjà fondamentalement différent de par sa fonction et son utilisation, des gestes observés chez les singes. Nous abandonnons en conséquence l'idée qui oppose classiquement pointage impératif vs. déclaratif, au profit de cette démonstration d'un pointage dit « coopératif » par Liszkowski.

## II.7 Le geste de pointer : « en route vers la syntaxe ! »

« Quand le sage montre la lune, le sot, lui, regarde le doigt... »

### II.7.1. Naissance de la syntaxe dans le bras...

Notre prédiction étant que le geste de pointer de l'index soit un pré-requis fondamental pour la mise en place de la syntaxe, dans le processus ontogénétique des premiers mots chez l'enfant, il nous paraît intéressant de reprendre un schéma proposé par Levelt (1998), qui expose les 4 étapes ontogénétiques dans la représentation lexicale de l'enfant, tout en tenant compte des potentialités intéressantes pour nous, des primates non-humains entraînés, comme Kanzi (Figure II.29).

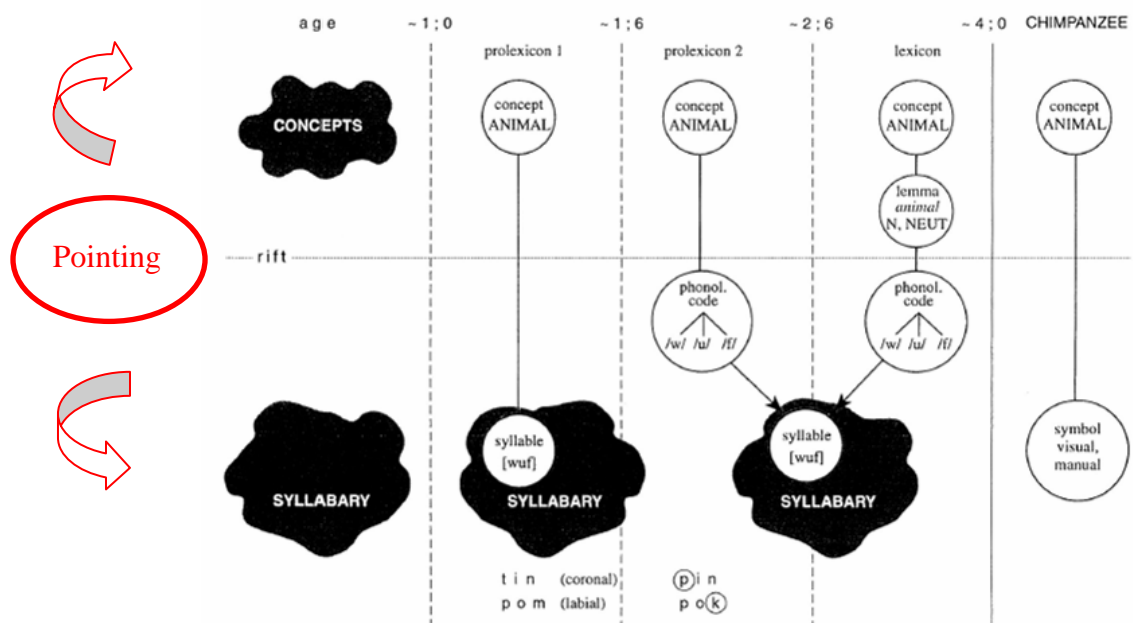


FIGURE II.29 : 4 étapes ontogénétiques dans la représentation lexicale des humains, et la représentation lexicale chez des chimpanzés entraînés (d'après Levelt, 1998). Nous avons ajouté à ce

schéma l'élément-clé, qui selon nous permet de lier ces deux systèmes, le système conceptuel et le système syllabique, par les flèches partant du système de « pointing » ou pointer de l'index.

Selon Levelt, lors de la première année de vie de l'enfant, les deux systèmes se situant de chaque côté de ce qu'il nomme un « rift » ou fossé (d'autres ont parlé de gap), traversent une première phase de maturation, indépendamment l'un de l'autre. Le système situé au-dessus du rift sur la figure concerne les concepts, l'enfant acquérant les concepts de base de lieu et temps, d'intention et de causalité dès sa première année de vie. De plus, il acquiert la notion de domaines sémantiques bien distincts concernant des concepts comme les gens, animaux, objets ou autres catégories. Le système situé en-dessous du rift est le système syllabique ou syllabaire, qui voit son développement s'accélérer dès 7 mois en moyenne par l'apparition du célèbre babillage canonique, dont nous avons précédemment parlé. Ce système s'étend très rapidement dans les derniers mois de la première année de vie, et se voit de plus réglé ou ajusté (on parle par ailleurs de *phonological attunement*) sur le patron syllabique de sa langue maternelle. Autour de son premier anniversaire, l'enfant utilise ensuite des proto-mots, c'est-à-dire qu'il réalise une syllabe ou une combinaison de 2 syllabes qu'il connecte systématiquement à un référent conceptuel. Sur la figure II.29, Levelt donne l'illustration de la syllabe [wuf] réalisée pour référer à un animal domestique présent dans l'environnement de l'enfant. Ce proto-lexique va considérablement se développer dans la deuxième année de vie. Cette explosion proto-lexicale va avoir des conséquences non négligeables sur l'enfant : il utilise de plus en plus de syllabes, parfois même des syllabes similaires, qui doivent être distinguées par leur fonction de référence. C'est à cette étape que va se mettre en place un nouveau processus, le processus de phonologisation du lexique. Prenant l'exemple du hollandais, Levelt explique que l'enfant va dans un premier temps posséder des proto-mots tendant à avoir un lieu d'articulation de la consonne uniforme, avant de le faire varier librement au début du mot, indifféremment au reste du mot. Il ne portera attention à la fin du mot qu'un peu plus tard, où il fera également ensuite varier le lieu d'articulation. A cette étape du processus, l'enfant commence alors à sélectionner le nucleus de la syllabe pour une libre assignation du lieu. Par ces mécanismes, l'enfant crée ainsi un système d'indexation phonologique génératif. La phonologisation se voit aboutie un an après l'explosion du protolexique, vers l'âge de 2 ans et 6 mois.

Nous sommes pour notre part tout à fait d'accord avec ces phases développementales et l'évolution maturationnelle de ces deux systèmes parallèlement, puis conjointement. Toutefois, nous restons sur notre faim quant à l'explication même de ce « rendez-vous » entre concept lexical et concept syllabique... Une rencontre qui reste « magique » dans le sens où nous n'en apprenons guère plus que sur le mécanisme permettant de faire sortir un lapin blanc d'un chapeau noir par le magicien... C'est pourquoi nous proposons d'ajouter à ce modèle de Levelt l'élément-clé sans lequel ce rendez-vous ne serait pas possible entre les deux systèmes, à savoir le pointer de l'index (ce *Pointing* fléché et que nous avons ajouté à la figure II.29 de Levelt). En effet nous postulons, et notre étude sur le pointer et le babillage des enfants de notre corpus nous permet de l'affirmer encore plus fortement, très empiriquement, que la métrique phonologique des premiers mots (pour un point de vue phonologique abstrait cf. Demuth, 1996 ; et son numéro spécial de *Language and Speech*, 49(2), 2006, sur le développement du *mot prosodique*) est une métrique qui prend naissance dans les rapports de contraintes du contrôle neuro-moteur par la coordination entre la mandibule et le bras, en d'autres termes que le pied est dans le bras. Ainsi le processus de syntactisation démonstrative co-émerge selon nous dans la phonologie du pied via la phonologie du geste de pointer, aboutissant à mettre en place la phonologie du mot. Nous proposons alors de résumer tout notre parcours théorique à travers ce dispositif développemental neural que nous avons tentativement lancé dès notre introduction.

### **II.7.1.1. Syntaxe par le geste ou la parole : même combat dans l'hémisphère gauche**

Nous avons donc mis en place un système attentionnel, à l'intérieur duquel nous travaillons sur la phonologie du mot à partir de la phonologie du pointer. Le pointer n'est ni une imitation, ni une émulation, mais reste en revanche un outil fondamental à la naissance de la syntaxe que l'enfant doit acquérir avant son premier anniversaire.

L'idée que le pointer est une racine neurale requise pour le développement de la syntaxe, et indépendamment de la préhension, s'est vue, par hasard, très bien illustrée par une vidéo familiale, cadeau d'un collègue de labo (Christophe Savariaux), de son petit garçon hors corpus. Rémi, âgé d'un an environ, pointe rapidement de la main droite avec une vocalisation « ha ! » depuis sa chaise haute

pour demander un bonbon Treetts à sa maman. Le fait qu'il pointe avec la main droite induit que c'est le cerveau gauche qui travaille, celui-là même qui est recruté généralement pour l'exercice du langage, dont la phonologie du mot. Et lorsque sa maman lui pose un bonbon sur la table de sa chaise, Rémi s'en saisit... cette fois-ci avec la main gauche ! C. Abry, invité pour une présentation dans l'équipe de Rizzolatti à Parme, a montré cette vidéo : juste pour signifier qu'en cas de concurrence pointer/saisie « Language is NOT within our Grasp » (cf. l'article programme *Language within our grasp* de Rizzolatti et Arbib, 1998). Un NOT qui réclamerait d'autres expériences que cette vidéo anecdotique familiale car, selon le bon mot d'un des confrères de Jacques Vauclair, le pluriel d'anecdote c'est 'no data'.

Un soutien pour le contrôle du pointer à gauche chez l'adulte vient de l'étude de Astafiev et al. (2003), déjà mentionnée comme remarquable lorsque nous avons défendu la présence d'une voie That. Sur la cartographie des signaux préparatifs pour l'attention, le regard et le pointage vers une localisation visuelle périphérique, leurs résultats font apparaître que le cortex pariétal postérieur et le cortex frontal contiennent plusieurs régions qui codent ces signaux préparatoires indépendamment de l'effecteur utilisé (œil ou bras), notamment le LIP ou sulcus intrapariétal (Figure II.30). Toutefois, les régions qui ont été activées plus spécifiquement lors de la planification de mouvements de main en pointage sont dominantes à gauche, et ceci quelle que soit la main utilisée, droite ou gauche. Et l'on retrouve un membre du couple utilisé dans le système That, le gyrus supramarginal.

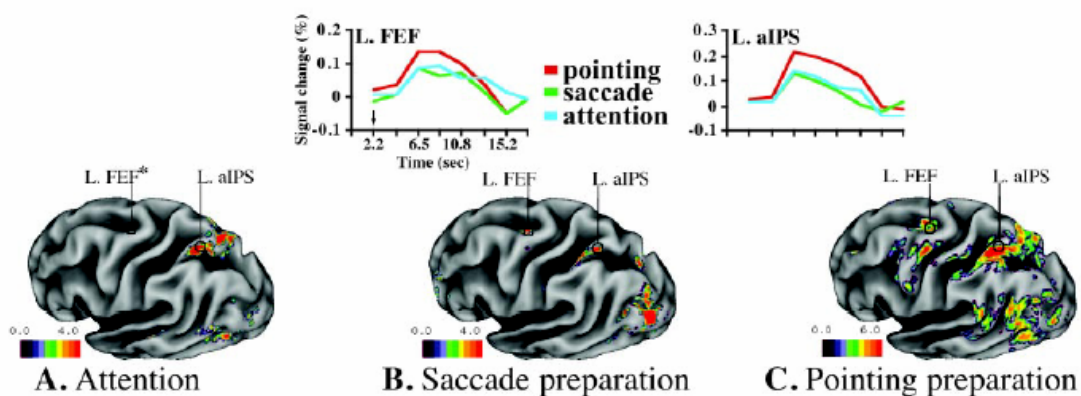


FIGURE II.30 : Activations cérébrales cartographiées sur l'hémisphère gauche du cerveau pour les tâches : attentionnelle, de préparation de pointer de l'œil (saccade) et de pointer du doigt. Le déroulement des réponses BOLD (soit la récupération métabolique de la consommation de glucose par les neurones) est donné en-dessus pour les régions du FEF (Left Frontal Eye Field ; l'étoile en A. signale l'absence d'activation au-dessus du seuil choisi) et du sulcus intrapariétal (Left anterior Intra-Parietal Sulcus, actif dans les 3 cas, et tout particulièrement pour le pointer) (D'après Astafiev et al., 2003).

En effet, la préparation d'une réponse de pointage implique un réseau fonctionnel bien distinct. Similairement au mouvement de préparation recrutant l'œil, le pointage a recruté des régions intrapariétales (aIPS, pIPS), mais il s'en démarque par l'activation dans le champ de l'œil frontal (FEF) et une étendue plus large du cortex pariétal postérieur et du cortex frontal, uniquement pour cette tâche donnée. Une région dans le lobule pariétal supérieur, le précuneus, était uniquement active pendant le pointage. Il n'est pas surprenant que le pointage recrute les aires FEF et IPS, ces aires étant traditionnellement recrutées pour les champs attentionnel ou oculomoteur. Mais la découverte la plus saisissante de ces auteurs, pour nous, réside bien dans le fait que, quelle que soit la main utilisée pour effectuer le pointage, la réponse de ces régions pariétale et frontale est latéralisée, dominante, à gauche, indépendamment de l'index utilisé pour répondre, que ce soit lors de la planification (préparation) du mouvement ou de son exécution. Ainsi, que le sujet pointe avec sa main ipsilatérale ou contralatérale, le circuit cérébral recruté reste essentiellement le même et surtout il reste actif du côté gauche du cerveau, le côté dominant du langage, ce qui vient directement appuyer notre idée que le pointage de l'index est étroitement relié au langage et à son développement syntaxique.

### **II.7.1.2. Syntactisation, grammaticalisation**

Revenons au modèle proposé par Levelt et particulièrement à la phase développementale qui débute dès 2 ans et 6 mois. L'enfant commence à s'intéresser aux mots de fonction. Avant cela, ces mots de fonction n'étaient pas utilisés sous une forme phonologique segmentalement repérable (sans parler de la question des « fillers », cf. Peters et Menn, 1993) ou alors simplement pour référer aux actions ou objets d'un certain type. Mais à ce stade, les mots de fonction deviennent tout à fait distincts des mots de contenu. Ils acquièrent les fonctions syntaxique et grammaticale qui les caractérisent. A cet âge, l'enfant entre dans l'étape de construction de syntagmes et phrases, et subit donc une explosion développementale du processus de syntactisation de la communication parlée. Les trois mois suivants cette explosion voient l'enfant augmenter de façon significative son débit de parole, d'auto-interruptions, d'auto-répérage et réparation des erreurs, de répétitions et d'hésitations, avant que son système morphosyntaxique ne se

stabilise. L'issue développementale de ce processus de syntactisation aboutit à la grammaticalisation du lexique. L'enfant sait assigner à un mot sa véritable fonction, il manipule les structures argumentatives et les rôles thématiques, et possède alors l'information de lemme dans le lexique. Vers 4 ou 5 ans, le système est pratiquement mis en place et un item lexical correspond pour un enfant, tout comme pour l'adulte, à: i) un concept lexical, ii) un lemme et iii) un code phonologique.

Ces processus de lexicalisation et de grammaticalisation attisant également notre curiosité vers le mot, nous proposons de nous attarder sur ce versant plus longuement, puisqu'ils découlent de la naissance de la syntaxe par pointer ou monstration, qui se situe au cœur de notre étude.

## **II.8. La monstration et l'interrogation aux origines de la parole**

Si l'on veut essayer de comprendre et d'étudier les mécanismes qui entrent en jeu dans l'émergence du langage chez l'enfant, à travers ses comportements en paroles et en gestes, il faut mettre l'accent sur deux mécanismes qui nous semblent cruciaux, en ce qu'ils enrachent profondément le départ du langage : les systèmes de monstration et d'interrogation. Ces deux systèmes montrent tous les jours leur robustesse dans une interaction humaine face à face, et la nécessité de les implémenter dans des outils de communication personne/système est une évidence.

Il nous a paru pertinent, dans un premier temps, de les étudier d'un point de vue primatologique, en soulevant la nécessité de posséder de tels mécanismes pour une organisation sociale dès les primates non humains. En montrant que certains de ces mécanismes existent chez les jeunes enfants, cela tend à soutenir l'hypothèse que ce sont des mécanismes très précurseurs. Nous parlerons ensuite des premiers systèmes de communication chez les bébés et les enfants, et nous montrerons comment tous ces mécanismes apparaissent à travers le mécanisme d'attention conjointe, ce dernier semblant se poser comme base de la monstration et de l'interrogation très tôt dans le développement. Nous discuterons également pour savoir comment ces deux systèmes s'ancrent dans le langage et quelle est

l'importance qu'ils ont par la suite dans le développement linguistique de l'enfant et l'histoire des langues.

### **II.8.1. Trois cris d'alerte pour les singes vervets**

Plusieurs études ont pu montrer que l'utilisation déictique du corps chez les primates non humains constitue en soi un tout premier système de communication : les systèmes d'alerte, qui ont pu être considérés comme une première performance linguistique, ce dont nous discuterons. Nous illustrerons leur présence chez différentes espèces étudiées dans leur milieu de vie, notamment les singes vervets et les suricates (sorte de mangoustes, non-primates, mais de l'ordre des carnivores, ce qui n'a pas grand-chose à voir avec leur régime plutôt omnivore). Nous parlerons également des bonobos et de leur champion Kanzi, pour ses performances langagières en laboratoire. Des études réalisées sur les cris d'alarme chez les singes vervets (*Cercopithecus aethiops*) ont prouvé qu'ils pouvaient véhiculer des messages spécifiques dans leurs cris, capacité qui a longtemps été présentée comme appartenant uniquement aux primates humains. Des recherches pionnières, notamment celles de Seyfarth et Cheney (Seyfarth et Cheney, 1990) faites sur ces singes vervets au parc national Amboseli (Kenya), ont mis en évidence que les vervets produisent des cris d'alarme différents quand des prédateurs différents sont proches. Les chercheurs ont alors procédé à des enregistrements de ces cris d'alarme et les ont ensuite diffusés à nouveau afin d'étudier le comportement des singes vervets. Les singes ont répondu en prenant une posture corporelle dirigée vers le haut quand il s'agissait de cris d'alarme pour les aigles, une posture corporelle dirigée vers le bas quand il s'agissait de cris d'alarme pour les serpents, courant se mettre à l'abri aux cris d'alarme pour les léopards. En dehors de cette dernière réponse de fuite, on peut donc observer que les singes vervets répondent par des mouvements posturo-corporels déictiques à une vocalisation, selon le type de prédateur auquel ils sont confrontés.



## II.8.2. Quel avantage évolutif à un système de communication avec monstration et interrogation chez l'animal ?

Un avantage évolutif certain, dans une conception darwinienne, est un bénéfice gratuit d'un système. Dans ce cas il réside dans le fait que l'on peut apprendre d'un congénère sa peur par son cri ou sur son visage. Comment un singe ou une mangouste apprendrait-il à avoir peur d'un serpent ? On peut penser que la peur n'est pas quelque chose d'inné chez eux, comme chez le petit d'homme (Mineka et al., 1984, Marks et Nesse, 1994). De fait, cet apprentissage peut se produire de deux façons. La première est l'apprentissage dit « traumatique », dans lequel l'animal, qui n'a pas peur du serpent, va apprendre à en avoir peur lorsqu'il se fera attaquer ou piquer. S'il en réchappe, c'est l'expérience traumatique qui va primer ici dans l'acquisition de la peur. Une autre voie est l'apprentissage via des mécanismes élaborés comme EDD et SAM, auxquels on peut ajouter les vocalisations d'alerte pour les vervets et autres espèces qui en possèdent. La mère ou la sentinelle va en fait montrer par son comportement qu'elle a peur du prédateur, et le bébé va simplement comprendre et imiter son attitude par empathie (mécanisme dit de résonance neuronale miroir). Il existe une expressivité de la peur, qui peut être une expression faciale chez les animaux comme les primates, qui ont une forte mobilité de la face (peu les vaches ?). Les jeunes singes arrivent ainsi à lire les intentions de leur mère sur leur visage.

Les jeunes vervets vont apprendre à associer les différents cris d'alarme avec les différents prédateurs. Au début, ils tendent à utiliser des cris d'alarme dans des circonstances inappropriées, des cris qui servent par exemple pour les aigles aussitôt qu'il y a un mobile en l'air, même s'il s'agit d'oiseaux inoffensifs pour eux, comme les vautours. Une voie pour accorder petit à petit leur comportement corporel et vocal au type de danger réclame de développer une inhibition, du système de peur panique, puis des peurs inadaptées, et pour cela le jeune interroge la réaction des congénères, les adultes plus expérimentés. Les vervets adultes n'ont pas peur des oiseaux sauf de l'aigle. Le bébé vervet va donc procéder selon plusieurs étapes d'acquisition : par le contrôle ou monitoring des réactions des sentinelles et du groupe, par la lecture de l'attitude des proches, sa mère, sa « baby-sitter » (nanny), etc. Il va donc « interroger » l'attitude corporelle, voire le visage des autres. Peut-on

vraiment parler d'une racine de l'interrogation à ce stade-là ? Encore peu d'études se sont focalisées sur ce point particulier pour pouvoir l'affirmer.

En ce qui concerne les précurseurs de la monstration, lorsqu'un singe a son regard dirigé sur un objet ou un être, on se retrouve dans un système attentionnel. Dans un partage déjà développemental de l'attention, un bébé singe face à un être inconnu va se tourner vers sa mère pour voir comment elle réagit, au cas où cet être serait pour eux dangereux. Par ailleurs, toujours en ce qui concerne le partage de l'attention, on peut remarquer que dans le cri du singe veilleur, il y a un « Watch ! », l'équivalent d'un simple signal d'alerte que traduit mieux le français « Attention ! » ; davantage qu'un « Look ! » dirigeant l'attention vers un objet. C'est un cri général comme « Timber ! » sur un chantier de bûcherons, qui donne bien l'objet courant du danger, mais pas son origine : il est comparable au sifflet des marmottes qui dit seulement « Prédateur ! » potentiel en vue. Une gradualité peut cependant s'établir entre les deux actions. L'orientation du guetteur attire l'attention des autres sur un danger et les amène à regarder dans la direction du danger. On ne peut pas vraiment considérer cela comme du déclaratif, mais plutôt comme de l'injonctif ou impératif, dans la mesure où il peut y avoir un fort degré d'urgence à regarder dans la bonne direction (comme pour fuir la chute de l'arbre dans « Timber ! »). Pour résumer, on peut dire que l'équivalent d'un « Watch ! » serait utilisé, dans un premier temps, pour dire « Attention (danger) ! » ; puis dans un second temps d'acquisition l'équivalent de « Look ! » serait utilisé pour dire « Regarde ! » dans cette direction et pas n'importe où. Il s'avère que cette vérification de la direction du regard du veilleur n'est pas toujours au rendez-vous avant la réaction de fuite appropriée, loin de là (Manser et Fletcher, 2004).

### **II.8.3. Peut-on parler d'un véritable langage chez les singes vervets ?**

Si on considère à nouveau les trois cris différents pour montrer trois différents types de prédateurs, on pourrait alors, de notre point de vue, suggérer que les singes vervets possèdent trois démonstratifs : un pour le serpent, un pour l'aigle et un pour le léopard. Cheyney et Seyfarth (1990) soulignent par ailleurs que les vocalisations des vervets ont une signification et que celles-ci forment un « langage » rudimentaire. De même Cook (1999) cite dans son article des recherches de Hauser

(1992) qui suggère que les vervets présentent certaines caractéristiques universelles du langage. Il note également une chute rapide d'amplitude et de fréquence en fin d'émission, ou « déclinaison de la fréquence fondamentale », chute qui signalerait la fin d'une déclaration particulière, qui est universelle dans les langues connues. Cette découverte soutiendrait alors la conception nouvelle que les vocalisations de primates non humains seraient bien plus similaires aux communications humaines qu'on ne l'avait dit jusque-là.

Les réponses sont, de plus, aussi distinctes et fréquentes lorsque des chercheurs jouent ces mêmes cris d'alarme –enregistrés au préalable– à travers un haut-parleur. Ceci a son importance dans la mesure où cela témoigne du fait que les réponses des singes vervets ne nécessitent pas obligatoirement la vue réelle d'un prédateur. Celui qui crie dirige l'attention des autres vers quelque chose qu'ils ne perçoivent pas mais qui est présent, en d'autres termes ils vocalisent pour localiser un prédateur, une caractéristique plutôt référentielle, qui pourrait alors constituer un précurseur du langage humain. Toutefois, comme le remarque Tomasello (1999, 2003b), certains peuvent objecter que les primates humains soient plus proches des grands singes que des petits singes, et que ces cris ne pourraient alors pas être un précurseur direct du langage. De plus, d'autres contre-arguments consistent à dire que ces vocalisations n'aient pas vocation à être référentielles, servant principalement à réguler les interactions sociales dyadiques, et que ces vocalisations ne sont pas apprises.

Sans aller plus loin dans le débat, nous supposons pour notre part qu'il pourrait s'agir d'un langage qui connaît des unités, en considérant les trois cris comme impliquant de 1 à 3 démonstratifs (cf. infra).

#### **II.8.4. L'apport des suricates**

En partant du fait qu'il existe une pièce manquante pour un phénomène de codage de l'information référentielle et de l'information affective dans les vocalisations animales, des travaux ont été réalisés sur des suricates (*suricata suricatta*), des mangoustes d'Afrique du Sud, qui montrent que leurs cris d'alarmes codent à la fois l'information sur le type de prédateur et la perception de l'urgence de la part de celui qui signale le danger. Citant les travaux de Marler (Marler et al., 1992), Manser et al. (Manser, Seyfarth et Cheney, 2002) disent que les signaux

d'animaux informent les auditeurs à la fois sur les événements spécifiques du monde, et le niveau d'excitation ou motivation du « signaleur ». Partant de cette idée, ils ont alors procédé à des enregistrements sur des suricates, et montré que le contenu de l'information dans ces cris d'alarme qui sont spécifiques aux prédateurs n'est pas altéré dans le taux d'appel ou contexte. Manser et al. se sont alors posé la question de savoir si les signaux de leurs suricates peuvent coder l'information référentielle et affective uniquement dans leur structure acoustique.

#### ***II.8.4.1. Les cris d'alerte***

Les suricates sont des mangoustes qui vivent en groupe d'individus et qui ont comme principale occupation de fouiller les alentours de cinq à huit heures par jour à une distance avoisinant les vingt ou cinquante mètres de leur terrier ou abri le plus proche. Lorsqu'ils effectuent ce travail de fouille, ils prennent la précaution de scruter fréquemment les alentours afin de ne pas se laisser surprendre par un prédateur. Ce sont les membres du groupe qui alternent la garde.

Lorsque ceux-ci repèrent un prédateur, ils émettent des cris différents. Certains cris paraissent être des cris d'alerte générale pour plusieurs prédateurs. Par contre ils utilisent d'autres types d'appel pour certaines classes spécifiques de prédateurs (voir Figure II.31 ci-dessous).

Ils donnent un premier type de cri d'alarme face aux prédateurs mammifères, essentiellement les chacals (*Canis mesomelas*), qui attaquent au sol. Ils utilisent un deuxième cri, acoustiquement distinct, en ce qui concerne les prédateurs qui tombent des airs, c'est-à-dire les aigles guerriers (*Polemaetus bellicosus*), les aigles fauves (*Aquila rapax*), et les autours pâles chantant (*Melierax canorus*), pour les attaquer. Et un troisième cri d'alarme face aux serpents comme le cobra cape (*Naja nivea*), la vipère heurtante (*Bitis arietans*) et les serpents taupes (*Pseudaspsis cana*). Ce qui est remarquable pour la sémantique de ce dernier type de cri, c'est qu'il est de la même façon utilisé pour les fèces, les échantillons d'urine et de poils venant de prédateurs ou de suricates étrangers. Quand ils se regroupent autour du signaleur, tous se coalisent pour attaquer le serpent : ils se trouvent collectivement en régime de « cris d'alarme de recrutement ».

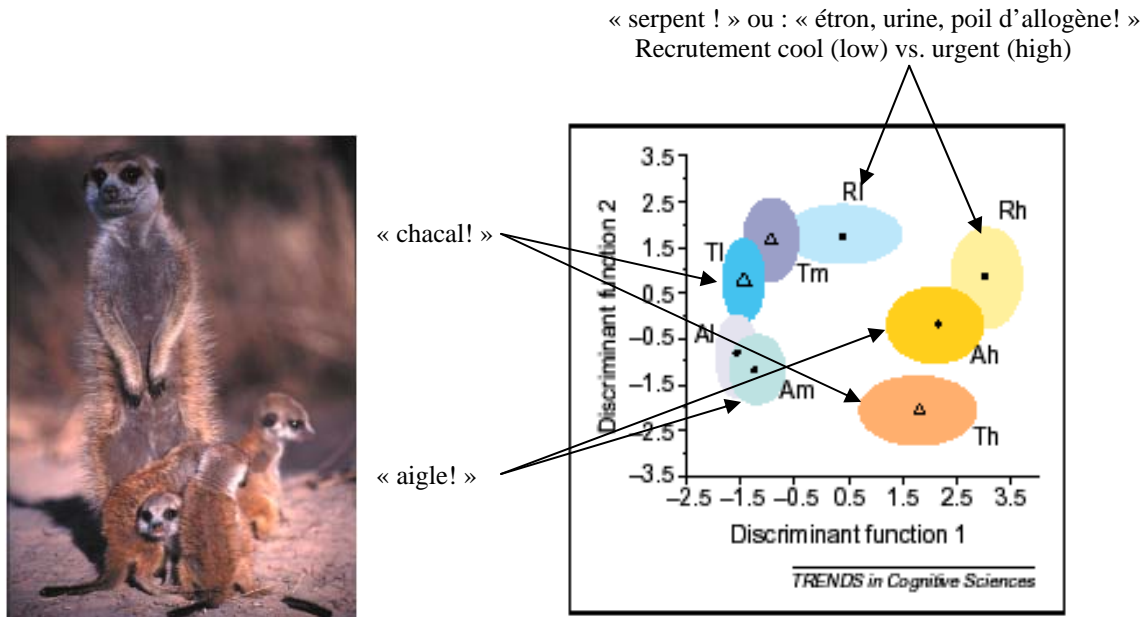


FIGURE II.31: Organisation des cris d'alarmes des suricates donnés dans différents contextes de prédateurs selon leurs dimensions établies par une analyse par fonctions discriminantes (DFA) des propriétés acoustiques des cris. T désigne les cris d'alarme destinés aux prédateurs terrestres, A désigne les prédateurs venant des airs et R désigne les cris de recrutement, avec respectivement l, m et h pour les cris de faible, moyenne et grande urgence. (D'après Manser et al., 2002)

On pourrait résoudre ces trois cris en trois démonstratifs de type « ça-là » de la façon suivante : (i) « Watch that ahead » (pour les chacals) ; (ii) « Watch that up in the sky » ; « Look thereabouts » (incitation à inspecter) ou bien « Look that shit down there » (dans ce dernier cas, un « ça » véritablement ancêtre du « ça » freudien !).

#### II.8.4.2. Réponses aux cris d'alerte

Outre ces trois types différents de cris, une autre caractéristique a été notée pour classer ces stimuli. Les cris d'alarme « chacal » et « aigle » ont été classés en fonction de la situation d'urgence : faible, moyenne et forte, correspondant respectivement à lointain, proche et très proche. Pour les prédateurs mammifères, les stimuli ont été classés comme lointain à plus de 200 mètres, proche de 20 à 200 mètres et très proches à moins de 20 mètres. Pour les prédateurs aviaires, les stimuli ont été classés comme lointain de 200 à 500 mètres, proche de 100 à 200 mètres et très proche à moins de 100 mètres.

Lorsqu'ils entendent un cri d'alerte pour les prédateurs mammifères, les suricates s'avancent rapidement dans la direction d'où provient le signal, se réunissent autour à 5 ou 10 mètres et battent en retraite vers l'abri le plus proche. Pour les cris d'alarme de prédateurs aviaires, on n'observe pas le même

comportement, ils restent accroupis, pétrifiés, fouillant le ciel du regard, puis courant vite au premier terrier le plus proche. Enfin pour les cris d'alarme de recrutement, ils dressent leur queue et approchent doucement du signaleur et reniflent les alentours. En plus de ces réponses comportementales, pour chaque classe de prédateur, les suricates réagissent en fonction de l'urgence. Ils s'arrêtent et s'ils ne détectent pas de prédateurs, reprennent vite leur tâche de recherche d'alimentation (*foraging*). À l'inverse, pour les cris d'alarme aviaires de faible et moyenne urgence, leur réaction est de courir immédiatement à un terrier, et l'activité de quête de nourriture met plus de temps à reprendre.

Cette étude a mis en évidence le fait que les suricates fournissent une preuve qu'ils donnent simultanément à leurs congénères l'information sur le référent spécifique externe et l'état motivationnel de l'alerteur dans leur performance acoustique. Manser et al. montrent ainsi que l'indexicalité est déjà présente dans les cris d'alarme des suricates. Ils font une discrimination qualitative lorsqu'ils reçoivent un cri d'alarme et, dans chaque catégorie de prédateurs, ils répondent quantitativement de façon différente aux cris dénotant les niveaux d'urgence, et cela suggère que l'on peut considérer que ces signaux sont codés simultanément. Ces trois cris peuvent constituer trois démonstratifs qui semblent de première importance pour leur survie. Il existe chez les suricates trois vocalisations pour la localisation.

De même que pour les singes vervets, il est possible de discuter d'un « langage ». Dans le cas des suricates, la « syntaxe » est plus complexe puisqu'ils combinent type de cri et niveau d'urgence pour communiquer. On peut parler d'un système plutôt complexe d'unités significatives, qui témoignent en faveur d'une racine d'un langage.

### **II.8.5. Les bonobos linguistes de 2 ans ½**

Vauclair (2002) cite une étude, menée par l'équipe américaine d'Atlanta, dirigée par Duane Rumbaugh et Sue Savage, sur des chimpanzés utilisant un langage visuel constitué d'une centaine de formes arbitraires, appelées « lexigrammes ». Chaque lexigramme représente une action, un objet ou un qualificatif. Ce langage est appelé « yerkish » (du célèbre centre de primatologie de Yerkes). Les chercheurs ont examiné les capacités de catégorisation et d'abstraction de trois chimpanzés. Ils ont fait apprendre aux chimpanzés à associer un lexique à

différents objets, et ils ont ensuite appris le lexique représentant la classe « nourriture » et la classe « outils ». Il en résulte que la manipulation des symboles par les chimpanzés équivaudrait à la propriété de dénomination caractérisant le langage.

D'autres travaux sur le célèbre bonobo Kanzi ont montré qu'à l'âge de six ans, il peut comprendre approximativement 400 mots d'anglais, en même temps que les symboles graphiques. Il est également capable de comprendre une succession de mots ordonnés dans des enchaînements complexes, comme :

- « Faire que le chien morde le serpent »
- « Faire que le serpent morde le chien »

Cette capacité de compréhension est importante : Kanzi est capable de combiner deux ou trois lexigrammes. Mais uniquement dans des situations où il recherche une satisfaction immédiate. Il faut noter également que lorsqu'il utilise des lexigrammes, il y associe fréquemment des gestes spontanés. Les auteurs concluent que le bonobo a une compréhension comparable à celle d'un enfant de 2 ans  $\frac{1}{2}$ .

Vauclair (2002) argumente que les chimpanzés, de même que les animaux de façon générale, qui font acte de communication, activeraient seulement des injonctions (autrement dits *mands*, dérivé de *command*, *demand*), de même que les enfants de très bas âge (y compris le premier pointage, dit « impératif » à 9 mois). Ce qui différencierait ensuite les deux, résiderait dans le fait que les enfants acquièrent par la suite la fonction déclarative, qui sert à véhiculer une information sur le monde et permet de l'échanger avec autrui. C'est ainsi cette fonction déclarative qui distinguerait la communication humaine de la communication non humaine. Vauclair conclue en suggérant que le langage des chimpanzés pourrait être en fait une étape primitive du développement du langage, et il reprend l'idée de Bickerton (1990) qui verrait en cela un « protolangage » qui se rapprocherait de celui des jeunes enfants et des pidgins.

## II.8.6. Conclusions

Les trois exemples que nous venons d'étudier, les singes vervets, les suricates et les bonobos, nous ont ainsi permis d'étudier les mécanismes de monstration et interrogation chez des primates non humains. Il en ressort qu'il existe

un système de communication élaboré, qui repose fondamentalement sur le système de monstration en particulier. Et l'on peut pousser l'idée qu'il s'agirait à ce stade d'un « protolangage ». En ce qui concerne le système d'interrogation, il paraît ne pas émerger comme signal vocal chez ces primates (voir pourtant l'interaction adulte-enfant). Mais des études plus poussées pourraient peut-être confirmer ou infirmer cette hypothèse.

L'étude menée sur les primates non humains nous a également permis de poser les mécanismes précoces utilisés pour communiquer en développement de la cognition sociale. Les cris d'alerte élaborés qu'utilisent les singes vervets et les suricates sont fréquemment sous-estimés par les linguistes qui ne s'émerveilleront jamais qu'un bonobo doué, en captivité, puisse atteindre le niveau de cognition linguistique d'un enfant de 2 ans et ½. Disons tout nettement que ce qui nous intéresse dans les primitives de la communication entre adultes, c'est ce qui nous reste du bébé. Alors jeter le bonobo avec l'eau du bain, c'est ne plus s'étonner que l'enfant de moins de 2 ans ½ puisse s'ancrer aussi loin qu'il le fait dans la langue de son milieu.

## **II.9. Quels avantages d'un système de monstration/interrogation dans la naissance du langage ?**

Si un primate non humain peut présenter des capacités de communication comparables à celles d'enfants de 2 ans ½, nous sommes en droit de nous poser la question de savoir comment cette fonction de vocaliser-pour-localiser évolue chez le bébé humain. Comment se fait l'intégration chez le petit homme des deux cadres que nous avons proposés en flux avec chacune des parties de cette fonction intégrative: le cadre de la parole, départ de la phonologie, et un cadre signe qui gabariserait la sémantique ? Ces deux cadres donnant naissance à la fin de la première année de l'enfant au mot, par l'intégration de la métrique de la syllabe dans la métrique du pied (Figure II.32).



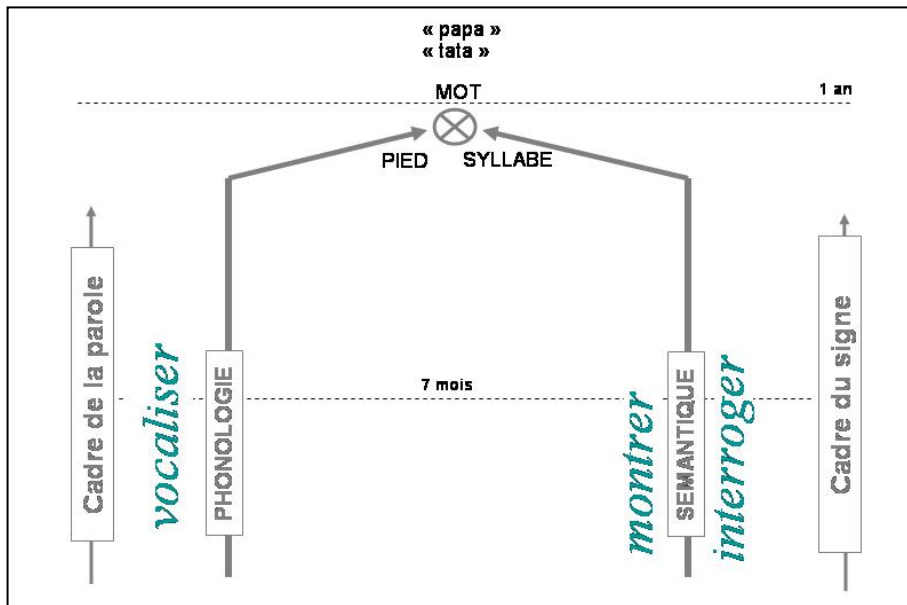


FIGURE II.32 : le cadre de la parole constitue la phonologie (vocaliser) et le cadre du signe constitue la sémantique (montrer / interroger) dans le cours du développement.

### II.9.1. La lexicalisation de la monstration

Les sections précédentes nous ont permis de mettre en relief la nécessité et l'importance d'avoir un système de monstration, que cela soit chez les primates non humains pour déclencher une alerte, ou chez les primates humains pour acquérir un système d'attention robuste qui permet ensuite d'ancrer le langage plus tard dans le développement. Le système de monstration peut constituer une primitive de la communication face à face, dans la mesure où il est nécessaire pour communiquer. Cette deïxis de monstration permet à l'enfant d'acquérir le lexique à travers les systèmes EDD et SAM par exemple, puis, par la suite, les gestes brachio-manuels de pointer.

De plus, il est important de noter que, de la même façon qu'un enfant peut assigner un index à trois objets maximum dans une scène, lorsque la deïxis se grammaticalise, nous constatons que la langue utilise peu de monstratifs : deux ou trois. Par exemple *this* et *that* en anglais, *este/ese/aquel* en espagnol, et *celui-là* (et peu *celui-ci*) en français. Le système de monstration est très enraciné et repose sur ce petit nombre d'unités, qui sont ancrées dans nos comportements attentionnels naïfs ou intuitifs présents depuis le tout petit enfant.

Nous avons déjà pu voir précédemment qu'il existe un « chiffre magique » 3 pour le suivi (l'indexation) d'objets dans une scène, et ceci très tôt dans les premiers mois. C'est un chiffre magique que l'on retrouve dans les systèmes linguistiques au niveau des pronoms. Ceci est synthétisé par Maillard (1987), dans sa grande thèse sur « ça », qui utilise le terme de « deixis à 3 places ». Par exemple cela peut nous donner les trois lieux suivants : *ici/là/là-bas*. Cette triade se retrouve dans d'autres langues comme en japonais, où l'on a *kore/sore/are*. Maillard s'intéresse également à la triade latine *hoc/istud/illud*, qui a subi quelques changements à travers le temps : *hoc* s'est affaibli déictiquement et a entraîné le glissement de *istud* en première position, lui-même remplacé en seconde position par *ipsum*, tandis que *illud* gardait sa troisième position, renforcé par *\*accu-illu*. Cela a donc donné une nouvelle triade *istu/ipsu/\*accu-illu*, qui se trouve être à l'origine des triades *este/ese/aquel* comme en espagnol.

On peut noter ensuite que la troisième place du tryptique neutre est souvent laissée inoccupée. Des études ont montré que *aquel* de l'espagnol, qui représente le pronom neutre, est jusqu'à dix fois moins employé que *este* et *ese* que cela soit à l'écrit ou à l'oral.

## II.9.2. Aux racines d'un système d'interrogation

Nous venons de voir que les langues utilisent généralement peu de monstatifs, environ trois. Il n'en est pas de même pour les interrogatifs, qui comptent un nombre beaucoup plus important d'éléments. Pourquoi cette disparité entre les deux systèmes ? Afin de mieux appréhender cette disparité entre nombre de catégories selon les domaines, intéressons-nous à l'acquisition de la peur chez l'enfant. Si l'on considère que l'homme peut retenir environ 500 espèces biologiques (animés/inanimés : plantes, oiseaux...), et qu'il peut retenir également environ 500 visages ou personnes (autre chiffre magique), comment alors faire l'apprentissage de ce qui est dangereux ou non ? Les différentes expériences que nous vivons, qu'elles soient traumatiques ou transmises par la mère, nous aident à établir différentes catégories de phobies. L'apprentissage de la peur fait partie de l'apprentissage des phobies, qui entraîne à son tour un apprentissage pour catégoriser les éléments du monde en goûts et dégoûts, à partir de certaines bases. Le goût pour le sucré des bébés n'est pas toujours cohabitant avec le goût pour les condiments ou celui pour

les boissons fermentées ou les viandes faisandées avant un certain temps. Afin que ces apprentissages aboutissent, nous avons besoin du mécanisme d'interrogation : il faut pouvoir interroger les goûts de la mère ou du père ou des pairs, pour répondre à leurs incitations à surmonter nos premiers dégoûts. Ce qui vaut pour l'apprentissage alimentaire, comme pour l'apprentissage sexuel. Pour cela nous possédons une capacité interrogative bien avant la liste d'interrogatifs des langues du monde, pour nous interroger sur l'expérience, le « ça fait quoi de se sentir comme ça ? » ou « d'éprouver ça ? » (le « What is it like to be a bat? » de Nagel, 1974). Ces interrogatifs grammaticalisés sont couramment au nombre d'environ 7 ou 8. Ainsi l'anglais a une racine très ancienne (indo-européenne \*kw-) son wh-, correspondant au qu- du français, héritier du latin (prononcé kw-) : quand (when), où (where), quoi (what), qui (who), lequel/laquelle (which), pourquoi (why) (sans parler de whose, whence, etc.).

Les racines de cette seconde notion-clé, l'interrogation, qui constitue un mécanisme élaboré dans toutes les langues, sont à rechercher dans les trois systèmes de suivi des événements/objets/agents When/Then, Where/There, What/That, qui correspondent à des circuits cérébraux bien spécifiques, comme nous l'avons vu. Ces systèmes sont interrogatifs par construction : ils sont prêts à différents âges pour servir l'apprentissage rapide (*fast mapping*) des valeurs répulsives/attractives des catégories en constitution dans l'expérience du monde. Les suricates ont un cri identique appelant à une mobilisation sociale, un cri dit de recrutement, quand il s'agit d'expulser un serpent, ou pour des traces corporelles d'étrangers au groupe : un étron, des traces d'urine, des poils, qu'ils soient d'un chacal ou de suricates d'un autre groupe. Faut-il y voir le signe précurseur d'une deixis *in absentia*, sinon *ad phantasma*, réputées absentes du règne sémiotique animal cloué au *hic et nunc* ? Ces suricates ont par ailleurs une cognition sociale développée, avec des sentinelles, des Kindergarten avec nannies, etc., et des gestes d'invite au grooming. Cette dimension répulsion/attraction, très générale dans les diverses théories des émotions, peut devenir maintenant aussi précise que possible, en se somatotopisant cérébralement. Ainsi le grooming active les neurones miroirs du Gyrus Frontal Inférieur (IFG, F5) chez le macaque, neurones qui répondent aux offres faites avec la succion de la langue entre les lèvres et/ou à un geste d'épouillage d'un sujet, une main grattant le dos de l'autre main (Ferrari et al., 2003,

Fogassi et Ferrari 2004a, 2004b). Rappelons ici que F5 contient des neurones de la bouche, de la main et de la coordination main-bouche, neurones parmi lesquels on découvre des neurones miroirs. Il existe d'autre part des zones activées dans le cortex antérieur de l'insula à la vue de « disgusting faces » ou visages à vomir (Phillips, 1997) : l'insula contient une carte des neurones du goût. La peur active sous-corticalement l'amygdale : pour mémoire traumatique avec l'hippocampe. Comme on le voit la surprise de l'interrogation ne peut pas être ramenée à un mécanisme Wh-?, trop général, opposé de manière aussi générale à l'attente ou expectation.

Nous pouvons maintenant nous poser légitimement la question de savoir s'il n'existe pas un avantage évolutif à posséder un système de négation ? La négation constitue sans doute une façon d'installer un système interrogatif complet. Répondre négativement à une question permet de faire un apprentissage, apprentissage qui n'est pas présent chez les animaux par exemple. Jackendoff (2002,2003), suivant en cela Bateson, a également posé qu'il n'existe pas de négation chez les animaux, et que ce soit un caractère propre aux humains. C'est vrai pour l'absence d'une vocalisation négative aussi spécifique qu'un cri d'alerte à l'aigle. Mais là aussi l'observation des animaux n'en est qu'à ses débuts et il existe de nombreux comportements de refus (p.ex. en réponse à l'invite, ne serait-ce qu'à jouer, d'un congénère), qui font plus généralement partie de la dimension d'évitement, y compris en co-occurrence avec des comportements vocaux, même s'ils ne déclenchent pas directement une stratégie de fuite.

Comme nous avons pu le voir pour les primates non humains, et les suricates, on pourrait entrevoir une racine de l'interrogation, lorsque les jeunes déclenchent des alertes inadéquates et testent le danger sur les réactions des congénères plus expérimentés. Cependant, il est peu probable que l'on puisse vraiment parler de négation, sauf si, lorsque le jeune déclenchait une alerte, sa maman ou les adultes venaient vers lui afin de lui envoyer un signal négatif, une punition, soit un renforcement pour l'apprentissage du coût d'une fausse alerte. La « pédagogie » utilisée est plutôt celle de la crédibilité du sujet : tant que son alerte n'est pas crédible le jeune vervet ou suricate ne sera pas « écouté ». Le système d'interrogation n'est donc pas aussi complet, que le mécanisme de monstration, dont les précurseurs semblent bien exister chez ces animaux.

## II.10. Du pointer à la morphosyntaxe via la prosodie

Le système d'interrogation est ancré dans un système intonatif de même que la monstration, et faire de l'interrogation ou de la monstration c'est bien faire de la grammaire, même sans formants grammaticaux autres qu'intonatifs.

### II.10.1. « Les origines de la grammaire » : *bootstrapping* chez l'enfant

Hirsh-Pasek et Golinkoff (1999) ont cherché à établir les hypothèses et les stratégies qui concernent cette partie de l'acquisition du langage qu'est la syntaxe. Comment les enfants arrivent-ils à induire les règles syntaxiques stables à partir de l'input linguistique qu'ils reçoivent, alors que ce même input covarie fortement avec le monde qu'ils voient autour d'eux ?

Les répliques prosodiques paraissent être l'aspect le plus saillant de l'input à un stade précoce dans l'acquisition du langage. A 2 ans, l'enfant apprend que les emphases sur la forme syntaxique, correspondant à des objets sémantiques sont cruciales pour la structuration du langage.

Généralement les enfants acquièrent les mots de classe ouverte qui étiquettent les objets et actions dans leur entourage, avant d'acquérir les mots de la classe fermée qui ont des significations abstraites.

Alors que certaines théories préconisent que les enfants sont déjà équipés de procédures d'apprentissage de domaine général, d'autres théories soutiennent à l'inverse que les complexités du langage ne peuvent pas être expliquées sans des contraintes innées de « faculté de langage ». Par exemple, les enfants aveugles semblent parfaitement capables de distinguer des verbes comme *look* et *see*.

Landau et Gleitman (cités par Hirsh-Pasek et Golinkoff, 1999), ont proposé comme mécanisme un « *bootstrapping* syntaxique ». Si, par exemple, un nom survient immédiatement avant et après un verbe, alors le verbe peut être causatif. Afin que ce *bootstrapping* syntaxique soit possible, il doit y avoir des corrélations entre syntaxe et sémantique, que l'enfant doit être capable d'exploiter à un âge relativement précoce.

Afin d'étudier « les origines de la grammaire chez les enfants », Hirsh-Pasek et Golinkoff vont utiliser le paradigme intermodal de regard préférentiel, adapté de Spelke. Ce paradigme consiste à ce que l'enfant soit assis sur un de ses parents, qui a lui-même les yeux bandés afin de ne pas interférer dans le choix de l'enfant, et placé au centre de deux écrans. Un haut-parleur à mi-chemin entre les deux écrans joue un stimulus linguistique qui est pertinent ou « relié » avec seulement une des deux présentations vidéo montrées sur chacun des écrans. L'enfant choisit alors de consacrer plus d'attention (la variable dépendante étant le temps de fixation visuelle) à l'événement vidéo qui est relié à ce qu'il reçoit comme stimulus audio, et moins à ce qui n'est pas relié au message linguistique. Les enfants ont ainsi accès à la syntaxe, à la sémantique, à la prosodie et au contexte. Ils peuvent alors tirer avantage de la coalition de tous ces indices normalement utilisés pour la compréhension du langage. La question dans ce type d'expérience intermodale est de savoir si les enfants peuvent trouver une régularité de lien entre les stimuli présentés dans deux modalités différentes.

Il faut partir d'une évidence dont tirent profit ces expériences : la parole dirigée vers l'enfant exagère les marques prosodiques aux frontières de constituants (fonction démarcative réalisée par une fonction contrastive). Les enfants sont directement sensibles à ces frontières prosodiques quand ils écoutent de la parole.

De plus, pour l'apprentissage grammatical, la détection des unités doit être couplée avec la faculté de relever les relations ou patrons corrélacionnels entre unités. On peut alors se demander si les locuteurs à mots simples (les enfants qui ont une longueur moyenne d'énoncé ou *Mean Length Utterance*, MLU, de 1 mot) reconnaissent la signification de l'ordre des mots dans certaines phrases qu'ils entendent. L'attention portée à l'ordre des mots paraît être critique pour l'apprentissage des langues. Pour cela, Hirsh-Pasek et Golinkoff ont cherché à savoir si les jeunes enfants peuvent détecter l'ordre des mots de phrases possessives réversibles (*mommy[']s baby*) et non réversibles (*mommy's face*). Une ambiguïté est à relever dans les résultats de cette recherche : il arrive que des phrases telles que « *mommy[']s baby* » ne soient pas réversibles du point de vue de l'enfant, qui préfère adopter par préférence la relation canonique (le bébé de maman). Outre ce problème, même les plus jeunes qui sont les moins avancés linguistiquement refusent de répondre quand une relation anormale du type « *face's mommy* » est

créée en inversant l'ordre des mots. Les auteurs interprètent cela comme étant une indication que les enfants notent que « l'ordre des mots inverse crée des relations possessives improbables ».

Hirsh-Pasek et Golinkoff ont testé si les enfants comprennent ou non l'ordre des mots, entre 16 et 18 mois, en projetant sur les écrans, un agent et un patient engagés dans des relations de rôles opposés. Les enfants furent capables d'utiliser l'ordre des mots des phrases stimuli pour guider leur regard en fixant l'écran qui dépeignait l'événement de façon plus signifiante que l'autre écran. Si les enfants font attention à l'ordre des unités à l'intérieur de l'énoncé, les arguments pour un « *bootstrapping* syntaxique » deviennent plus plausibles.

Hirsh-Pasek et Golinkoff proposent un modèle à trois étapes de la compréhension du langage. La phase I se situe entre 0 et 9 mois. Une fois que les enfants ont formé des représentations d'images schématiques d'événements et peuvent isoler des morceaux (*chunks*) acoustiques dans le flux de la parole, ils sont alors prêts à former des « paquets [*packages*] » ou formules acoustiques toutes prêtes. Ils doivent cependant postuler au préalable que le flux de parole est d'une certaine façon connecté aux événements qui co-surviennent.

Lors de cette phase, les enfants entendent les aspects acoustiques du langage qui accompagnent les événements et les utilisent pour diriger leur attention vers ces événements et autant que possible pour les « packager » ou pour les segmenter. Ces unités qui sont cadrées acoustiquement contribuent pour les enfants à internaliser les informations sur les événements, donnant un fondement pour des interprétations futures.

La phase II se situe entre 9 et 24 mois. À cette étape, les enfants vont basculer d'un cadre dominant de répliques prosodiques pour se pencher sur les répliques sémantiques. L'enfant commence à associer des packages acoustiques particuliers à des objets, événements ou actions particuliers de son environnement. En fait, ce qui se dégage de la phase I constitue l'input de la phase II. L'enfant commence une segmentation de ces paquets acoustiques déjà extraits, ce qui lui permet de réaliser une analyse linguistique plus fine. Lors de cette phase II, l'enfant commence à faire un mapping des mots avec leurs référents, un *mapping* souvent aussi rapide (*fast mapping*), que celui qui lui fait lire la peur sur le visage de sa mère.

La tâche de l'enfant consiste à mapper les unités acoustiques en unités linguistiques, le son avec la signification.

Lorsque la phase I prend fin et que la phase II débute, l'enfant commence à comprendre ses premiers mots et ceci bien quatre mois avant que n'apparaissent ses premières productions.

Alors que dans la phase I, l'enfant devait procéder avec une coalition de tous les indices (sémantique, prosodique,...), dans la phase II, les connaissances syntaxiques de l'enfant sont assez robustes pour que la compréhension survienne même lorsqu'on présente à l'enfant des informations conflictuelles venant de différentes sources d'input.

Enfin la phase III survient entre 24 et 36 mois. Les enfants sont alors capables d'effectuer une analyse syntaxique plus complexe. À ce stade, ils paraissent avoir une appréciation des relations grammaticales données par les constructions passives, les relations hiérarchiques, etc. Les enfants se trouvent ici au seuil de l'analyse de discours : ils peuvent désormais se reposer plus indépendamment sur l'information syntaxique pour construire leurs modèles mentaux et apprécier des interprétations plus poussées de l'input de langage.

Ces trois grandes phases ont été schématisées par les auteurs par les figures suivantes, qui mettent particulièrement bien en relief les différents acteurs principaux de ces stades développementaux, que sont la prosodie, la sémantique puis la syntaxe (Figure II.33).



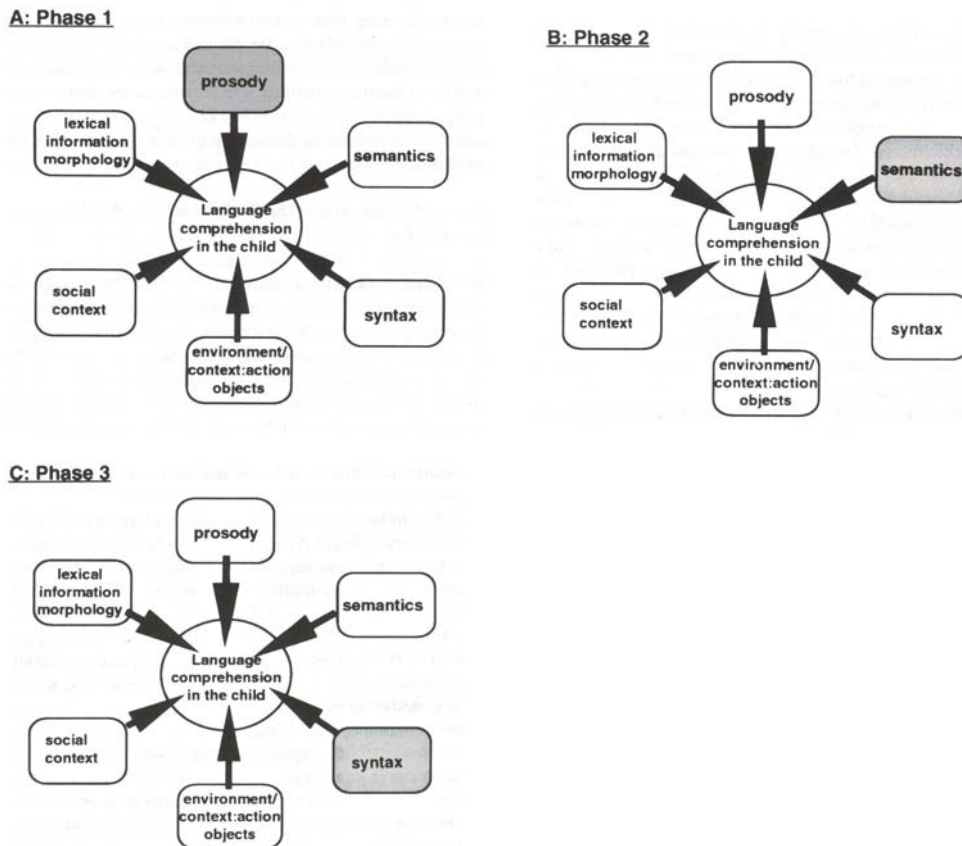


FIGURE II.33 : Les trois phases d'apprentissage du langage. En grisé, les composantes les plus activées en fonction de la phase d'acquisition. (D'après Hirsh-Pasek et Golinkoff, 1999)

En soutien à cette idée de « *bootstrapping* syntaxique », nous citerons également l'expérience de Gillette et al. (1999), Snedeker et Gleitman (2004), qui utilisent des vidéos de mères et jeunes enfants en interaction. Deux conditions d'indexation étaient testées : une condition *Word-to-World pairing* où lorsque la mère prononce le nom cible, un bip est substitué au mot (Exemple « Regarde le chat » où « le chat » est remplacé par un bip) et une condition *Structure-to-World pairing* où lorsque la mère prononce le verbe cible, un bip est substitué au verbe (Exemple « Regarde le chat » où « regarde » est remplacé par un bip). La tâche des adultes dans cette expérience était de deviner quel nom et quel verbe la mère produisait à ce moment-là. Le résultat est que, tandis qu'ils réussissent très bien pour récupérer les noms, les adultes échouent lamentablement pour prédire correctement le verbe. L'observation du contexte situationnel n'est donc clairement pas suffisante pour l'apprentissage des verbes. L'hypothèse du « *bootstrapping* syntaxique » n'est une théorie de l'apprentissage des verbes, que dans le sens où les enfants peuvent

potentiellement déduire une partie de la signification du verbe en notant la structure de la phrase, particulièrement le nom et la position des arguments du verbe. Il existe donc deux bootstrappers du langage : un bootstrapper qui est l'interrogatif (quoi vs qui, comme agent animé) et un *bootstrapper* qui est le démonstratif (deixis, anaphore : ex. ça vs. celui-là, celle-là, comme agents avec genre). Mais ces 2 bootstrappers n'existent pas sur le verbe ! Après avoir soulevé ce questionnement, nous proposons de nous intéresser plus amplement à cette défektivité spéciale du verbe.

### II.10.2. Le système omniprédicatif selon Launey

Mais avant, nous allons découvrir un système unilexical où le nom comme le verbe peuvent être utilisés sans formants grammaticaux obligatoires comme « prédicat nu ». Launey (1990, 1994) a étudié la langue des Aztèques, le nahuatl classique, Cette langue a pour nous la particularité intéressante de présenter un type grammatical qu'il qualifie d'omniprédicatif. On peut illustrer un tel fonctionnement aussi chez l'enfant en proposant que lorsqu'un enfant apprend un mot par son œil et son doigt (par exemple « yune » pour « lune »), il ne s'agit pas en réalité d'un mot, mais d'un prédicat du type « ça, c'est la lune ». Pour le nahuatl, Launey parle d'énoncés unilexicaux sans argument syntagmatique et qui peuvent aussi bien se réaliser avec n'importe quel item lexical d'une catégorie qui peut être marquée comme nom, verbe ou circonstant. Il cite l'énoncé *tisitl* (« médecin », signifiant « cet homme, il est médecin ») qui constitue un énoncé prédicatif saturé, c'est-à-dire qu'il représente ce qui est dit à propos d'un terme sémiotiquement bien identifié dans l'acte de communication. Nous insistons sur le fait que l'homme qui est désigné comme argument, n'est pas un élément du contexte, tel que le voit une certaine conception de la deixis. Pas plus que dans *tchoka* (« pleure » signifiant « cet homme, il pleure »). Dans ces deux cas, l'homme en question est aussi clairement capturé par l'acte sémiotique de pointer qu'une pomme est prise, la pomme n'étant pas le contexte de la prise, mais son objet, dans un geste qui est classiquement qualifié de *goal directed (oriented) movement or gesture*.

Si on revient à l'enfant, lorsqu'il montre l'argument —par un acte sémiotique qui peut être aussi tangible qu'une prise (en tirant sa mère par son vêtement ou en lui tournant la tête) ou qu'un contact (l'index peut, lorsque cela est physiquement

possible, aller jusqu'à toucher du doigt le référent, p.ex. « Tu TE le payes », quand on touche la poitrine de l'interlocuteur— l'émission « lune » ou « yune » de cet enfant en est bien le prédicat. Par contre, si on prononce « cette lune » (ou *in tisitl* « ce médecin »), alors il devient argument, et il faut forcément dire quelque chose de lui comme « ce médecin, il pleure » *tchoka in tisitl*. Une fois que le démonstratif est manipulable par l'enfant, s'il le place devant (en nahuatl, en français..., mais pas dans toutes les langues) un nom appris au préalable, alors il devient un argument et il faut impérativement lui donner un prédicat. C'est le cas, bien avant cette habileté sur un formant grammatical, de l'habileté intonative de « *pati papa?* » ou « *chaussure papa ?* » vs. « *papa ?* » ou « *chaussure ?* » : ces deux derniers énoncés sont des prédicats (énoncés omniprédicatifs) vs. les premiers où « — *papa ?* » et « *chaussure —* » sont en fonction d'arguments.

En résumé, on commence par avoir un prédicat avec un geste de monstration et un mot pour dire « ceci est la lune, un papa, etc. ». Enfin, le geste sera accompagné d'une grammaticalisation avec la maîtrise des démonstratifs issus des questions « *what's that ?* » (« qu'est-ce-ça ? »). Aussitôt que « ce » est utilisé sur un nom, alors il faut dire quelque chose de ce nom, devenu topique ou thème : le rhème, l'information « à propos » du thème (*aboutness*), devient obligatoire. Cette monstration grammaticalisée est donc bien une opération grammaticale d'argumentation.

L'action prédicative « yune » se résume ainsi en : (i) un argument-référent, topic-thème, qui est ici l'objet attentionnel <lune> « stické », soit suivi avec un des trois index mentaux disponibles, un index qui en constitue l'objectitude ; (ii) un outil sémiotique spécifique qui, dans un acte d'orientation, peut montrer en plus de l'oeil, c'est l'index-doigt ; (iii) enfin ce qu'on en dit avec la bouche (face-conduit-vocal et larynx), « c'est la lune », le prédicat-rhème.

### **II.10.3. La défektivité spéciale du verbe**

Comme nous l'avons déjà indiqué plus haut, le verbe, contrairement au nom, possède des défektivités fondamentales dans les opérations sémiotiques de base — monstration et interrogation— mises en évidence par Denis Creissels (Creissels, 1995, p. 46-47) :

Une observation capitale pour justifier de considérer les constituants nominaux d'une phrase comme valeurs prises par les variables d'une expression prédicative est l'existence d'interrogatifs correspondant à chacune des positions structurelles que peut occuper un constituant nominal. Il est d'ailleurs intéressant (même si cela ne correspond pas à des phrases correctes de la langue) de représenter les expressions prédicatives en symbolisant chaque argument nominal (ou quasi nominal) par l'interrogatif correspondant. Par exemple :

Expression prédicative :	qui	donne	quoi	à	qui
Valeurs affectées aux variables :	Michel		un livre		Jean

Il s'agit là de quelque chose de très important pour une caractérisation aussi générale que possible du contraste entre nom et verbe : dans les débats sur la possibilité de reconnaître dans toute langue un contraste entre nom et verbe, on s'enlise généralement dans des discussions sans fin sur la question de savoir s'il existe ou non des universaux sémantiques au niveau des morphèmes flexionnels qui (au moins dans la plupart des langues) permettent de délimiter les espèces grammaticales nom et verbe ; par contre on oublie régulièrement de faire remarquer que, si la totalité des langues connues disposent d'un répertoire d'interrogatifs aptes à exprimer un questionnement portant sur un constituant nominal, ainsi que d'un répertoire d'anaphoriques aptes à représenter le référent d'une expression nominale présente dans le contexte, il est par contre très exceptionnel que les descriptions de langues signalent l'existence de verbes interrogatifs ou des verbes anaphoriques. De telles choses ne sont signalées dans aucune langue ayant fait l'objet d'un nombre important de descriptions détaillées. Autrement dit, il serait peut-être imprudent d'affirmer que le verbe se distingue universellement du nom par le fait qu'il ne se prête pas directement au questionnement et à l'anaphore, mais ce critère est incontestablement beaucoup plus général que tous les autres critères qui ont pu être proposés comme constituant des manifestations typiques du contraste entre nom et verbe.

Creissels distingue quatre défektivités caractéristiques du verbe concernant :

(i) la deixis, il n'existe pas de deixis spatiale sur le verbe (on ne peut pas le montrer sous peine de le faire devenir argument, comme l'exemple en nahuatl nous a permis de le mettre en évidence); (ii) en conséquence, on ne peut pas l'anaphoriser, forme étendue de la deixis; (iii) dans un domaine encore dérivé de la deixis, il n'existe pas de pro-verbalisation (défektivité du verbe faire, cf. infra); (iv) et enfin, autre défektivité fondamentale : il n'existe pas d'interrogation sur le verbe.

Afin d'illustrer ces idées, prenons un exemple de phrase tout à fait neutre au niveau sémantique, puisque nous avons remplacé les mots du français par des mots inexistant dans notre langue :

« Zigueclub borchoume les zagrounettes »

Dans cet énoncé, il est possible d'interroger (c'est qui ?) et d'anaphoriser (c'est celui-là) « Zigueclub ». De même pour « les zagrounettes », on aura un moyen d'interroger (C'est quoi ?) et de montrer (c'est ça !). À l'inverse, lorsque l'on se penche sur « borchoume », qui fonctionne ici comme verbe, on se montre incapable d'interroger ou d'anaphoriser celui-ci. On pourrait dire « Il fait quoi ? », mais on

utiliserait alors un pro-verbe qui lui-même est défectif pour remplacer des verbes dits abstraits (ou d'états mentaux) du type « aimer ». Ainsi, on ne pourrait pas concevoir une phrase avec anaphore comme « Pierre aime Marie et Paul en fait autant ». Creissels exécute ce faux pro-verbe dans l'extrait suivant :

À ce propos, il convient de rejeter la proposition qui a parfois été faite de considérer qu'en français par exemple, *faire* serait un « pro-verbe » qui fonctionnerait par rapport aux autres verbes comme les pronoms par rapport aux expressions nominales ordinaires. En effet, si cette proposition était justifiée, ce qui vient d'être dit n'aurait aucun sens. Mais en réalité, en français (et ceci semble très général), ce n'est pas *faire* à lui seul qui joue relativement aux autres verbes un rôle que l'on pourrait comparer à celui des « pronoms » : ce sont *faire QUOI ?* et *LE faire* (ou *faire ÇA*) qui servent respectivement à questionner et à marquer une anaphore à propos d'un verbe, comme dans les ex.(89) et (90).

(89) **QU'**est-ce qu'il fait ? Il travaille

(90) Il s'est fâché, et il ne **L'**a certainement pas fait sans raison

Il y a bien là le verbe *faire*, mais ce verbe n'a lui-même rien d'interrogatif ou d'anaphorique : il n'est là que comme support syntaxique de l'interrogatif *quoi ?* ou de l'anaphorique *ça* (représentable en fonction d'objet par l'indice *le*), qui sont quant à eux de nature nominale.

#### II.10.4. Quelle solution pour le verbe ?

Nous venons de voir que la deixis et l'interrogation ne sont pas possibles sur le verbe au niveau grammatical. Quelle peut bien être alors la deixis du verbe ? Il semble que l'intonation –ou focalisation– peut fonctionner pour le nom comme pour le verbe. La deixis intonative et la question intonative sont en effet possibles sur le nom comme sur le verbe. Voici un exemple illustrant ceci en anglais :

Buy Bobby a **DONKEY** ? Accentuation focale sur l'objet

Buy **BOBBY** a donkey ? Accentuation focale sur l'agent

**BUY** Bobby a donkey ? Accentuation focale sur le verbe

À l'inverse, la deixis syntaxique (anaphore et pronominalisation) et la question syntaxique ne sont donc possibles que sur le nom. Lorsque l'on effectue une extraction syntaxique, la deixis peut être grammaticalisée, mais il n'est pas possible d'extraire syntaxiquement le verbe directement. En voici un exemple français :

- C'est à **BOBBY** qu'on achète ce bourricot. [syntaxique]
- C'est ce **BOURRICOT** qu'on achète à Bobby. [syntaxique]
- Ce bourricot, on l'**ACHÈTE** à Bobby. [intonatif]

C'est en fait la prosodie qui seule permet en dernier ressort la deixis du verbe (voir Figure II.34). L'extraction syntaxique avec « c'est SN qu- ... » est une

grammaticalisation de la deixis prosodique, comme l'interrogation l'est avec le formant morphosyntaxique « est-ce que... ? ». Il existe donc trois ancrages pour l'apprentissage du langage : (i) l'ancrage interrogatif intonatif ou formantisé sur l'objet-nom, (ii) l'ancrage démonstratif intonatif ou formantisé sur l'objet-nom, et (iii) l'ancrage intonatif interrogatif ou démonstratif sur l'objet-nom et sur l'action-verbe. L'ancrage intonatif semble universel et premier, les formants morphosyntaxiques étant des constructions particulières aux langues (cf. « Qu'est-ce que c'est ? C'est quoi ? », italien « Che cosa è ? », portugais « O que é isto ? » ; de même pour les extractions). Et bien entendu, s'il peut être aussi bien interrogatif que démonstratif, on se souviendra qu'on s'accorde généralement parmi les grammairiens et logiciens pour dire que toute interrogation implique une monstration (« c'est qui ? » ayant l'implicature « c'est quelqu'un »).

En clair, sans ce premier système, l'enfant ne pourrait pas acquérir le verbe avant d'avoir acquis une syntaxe, dont on ne connaîtrait donc pas le *bootstrapping*.

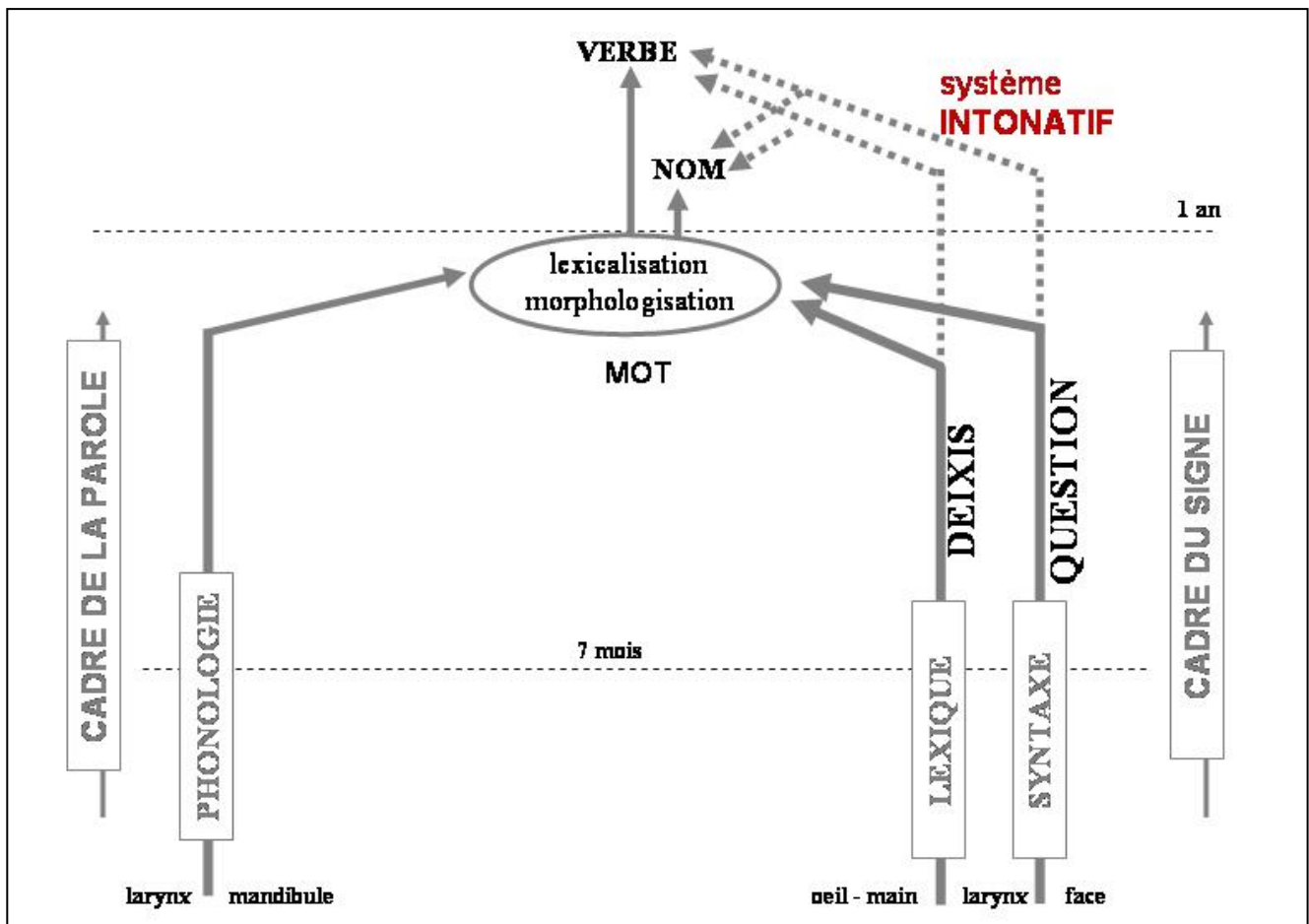


FIGURE II.34 : La deixis syntaxique et la question syntaxique sont possibles sur le nom. Lors d'une extraction syntaxique, la deixis peut être grammaticalisée, mais il n'est pas possible d'extraire syntaxiquement le verbe. Le système intonatif seul permet la deixis sur le verbe (en pointillé sur notre figure), tout comme sur le nom bien entendu.

Carey (2004) insiste sur cette notion de *bootstrapping*, qu'elle considère comme une faculté uniquement humaine, au même titre que l'analyse causale ou la faculté de raisonner sur les états mentaux des autres. L'apprentissage requiert la création de nouvelles ressources de représentations, souvent plus puissantes que celles présentes en sorties manifestes. Le bootstrapping constitue alors le processus sous-jacent et fondamental de la création de tels concepts nouveaux.

## II.11. Sociogenèse et ontogenèse du processus de grammaticalisation

### II.11.1. Sociogenèse de « that » (ça) à « that » (que)

Nous avons explicité en ce qui concernait la mise en place du cadre du signe, l'importance des grands systèmes cérébraux pragmatico-sémantiques que sont les systèmes When/Then, Where/There et What/That dans le développement de l'enfant. Chez ce dernier, ce système When-Where-What se met en place progressivement par l'œil puis par l'index. Si nous nous intéressons maintenant plus en détail à ce système d'un point de vue linguistique, nous affirmerons que les radicaux Wh- et Th-, issus des racines indo-européennes \*Kw- et \*T- de l'interrogation et de la monstration, constituent de véritables racines-outils pour la grammaire dans le développement de la structure prédicat-argument (Français « X voit que... », anglais « X sees that... »). Nous allons alors tenter de démontrer la puissance de tels déictiques dans le processus de grammaticalisation ou sociogenèse linguistique. C'est, à l'échelle temporelle du changement linguistique, ce que l'ontogenèse est à l'échelle de l'individu et la phylogénèse à celle de l'espèce. Nous empruntons pour le processus de grammaticalisation ce terme de sociogenesis à Tomasello (2003b). Il tient compte du fait maintenant reconnu que les enfants contribuent peu à ce processus, mais la créativité et la tolérance sociales beaucoup (*The American Heritage Stedman's Medical Dictionary* [2nd Edition 2004, Houghton Mifflin Company] donne la définition générale : « Sociogenesis : The origin of social behavior that derives from past interpersonal experiences »).

### **II.11.1.1. La puissance des déictiques dans le processus de grammaticalisation**

Diessel (2003a), dans une étude particulièrement approfondie sur la relation entre démonstratifs et interrogatifs dans un échantillon représentatif de 100 langues, confirme ce parallélisme saisissant entre ces deux racines. Avec comme hypothèse forte, le fait que démonstratifs et interrogatifs initient tous deux une recherche d'information guidée par leurs traits sémantiques et syntaxiques, fonctionnant pour focaliser l'attention de l'auditeur sur un objet d'intérêt, visant ainsi à établir une attention conjointe. Ils se démarquent par là très nettement de tous les autres marqueurs grammaticaux, qui visent uniquement à organiser le flux d'information au cœur du discours. De plus, à la différence des marqueurs grammaticaux véritables, il n'existe pas d'évidence que les démonstratifs se soient développés à partir d'une source lexicale non-déictique, leur origine étant de plus la plupart du temps inconnue.

L'existence d'un lien très fort entre ces deux entités monstratives, que sont le démonstratif et l'interrogatif, à fonction déictique ne semble plus à être démontrée.

Demonstratives are deictic terms, which in their basic use function to focus the hearer's attention on elements in the surrounding situation. Interrogatives are question words, which instruct the hearer to search for a specific piece of information in his or her knowledge store. (2003a:636)

Ces deux expressions sont par ailleurs communément marquées par les traits sémantiques comme la personne, la chose, le lieu, la direction, le temps, la manière et la quantité. Diessel (2003a) rappelle alors que ces traits sont essentiellement identiques aux catégories ontologiques proposées par Jackendoff (1983) pour constituer les primitives d'une structure conceptuelle. Outre ces catégories et le fait qu'ils sont à la frontière de plusieurs classes de mots, ils incluent également souvent les mêmes morphèmes par dérivation et portent l'accent (stress), un trait prosodique crucial. Diessel considère que les démonstratifs et les interrogatifs sont utilisés dans un premier temps pour des fonctions internes au langage, constituant ainsi une première marche basique dans le processus de grammaticalisation, bien avant d'assumer pleinement une fonction fondamentalement distincte des marqueurs grammaticaux véritables. Nous venons de voir que les démonstratifs jouent spécifiquement un rôle prépondérant dans l'utilisation du langage et de la grammaire. Ils offrent la possibilité de focaliser l'attention de l'auditeur sur un élément d'intérêt de la situation de parole, et sont fréquemment combinés avec le geste de pointage



correspondant à la référence du discours. De plus, ils fournissent une source historique à la matérialisation des articles définis, des pronoms relatifs, des pronoms de 3<sup>ème</sup> personne, des complémentisateurs, des connecteurs de phrase *inter alia* (Diessel, 2003b). Diessel (1999a,b) explique la ligne de grammaticalisation des démonstratifs en invoquant le fait que l'utilisation exophorique de ceux-ci est à la base des autres types d'utilisations : (i) anaphorique, qui aboutit par exemple au pronom de 3<sup>ème</sup> personne ; (ii) déictique discursive, qui aboutit par exemple aux connecteurs de phrase ; et enfin (iii) « recognitional », qui aboutit par exemple aux déterminants.

Afin d'asseoir sur la longue durée la puissance de ces déictiques dans le processus grammatical, nous partons de l'existence de ces racines stables linguistiquement, les racines de la monstration et de l'interrogation. Nous pouvons citer les racines indo-européennes \*kw- et \*t- (\*kwo, \*kwei / \*to, \*tei, etc.). Selon Diessel (2003a,b), les démonstratifs dans les langues indo-européennes se sont développés à partir de racines déictiques qui faisaient déjà parties de la catégorie démonstrative au préalable (\*to-, \*so-, etc.), tandis que la grande majorité des interrogatifs dérive de mots questions reconstruits (\*kwi-s pour *who*, \*kwi-d pour *what*). Diessel (1999b, 2003b) rappelle à ce propos qu'il est difficile d'affirmer que ces expressions grammaticales soient dérivées d'une source de type symbolique, dans la mesure où les démonstratifs sont généralement si anciens que leurs racines ne sont pas analysables d'un point de vue étymologique.

Nous pouvons ainsi simplement établir les séries de correspondances pour les racines anglaises wh- et th- (*What/That, When/Then, etc.*), allemandes w- et d- (*was/das(s)*), ou françaises qu- et -t (*quoi, que / (ce)t*). Notons que le couple latin bien connu pour corréliser ces deux racines t-/qu- se retrouve à travers les langues romanes pour certains aboutissants de *tam...quam, tot...quot, tum...cum, talis...qualis* (ce dernier ayant laissé en français *tel quel*). Ces racines sont solidement ancrées dans les langues et vont être utilisées par la suite dans les processus de grammaticalisation. Les démonstratifs et les interrogatifs ont fréquemment été renforcés par d'autres lexèmes comme « ecce » en latin vulgaire. Comme illustration, nous reprendrons l'exemple exposé par Diessel (2003a) : nous évoluons d'un démonstratif « ille » en latin à « ecce ille » en latin vulgaire, aboutissant en vieux français à « cest, cel », dont notre « ce ». L'élément de

renforcement tend ainsi à simplement entretenir la fonction grammaticale d'une ancienne forme. En aucun cas, selon Diessel (2003a,b) cet élément de renforcement ne donne naissance à un nouveau type de marqueur grammatical. Il est d'autre part remarquable que les démonstratifs peuvent également provenir de verbes déictiques. C'est le cas précisément du français qui avec un verbe adjoit d'une particule démonstrative a donné un démonstratif à part entière : « vois-ci » a évolué vers un « voici » présentatif démonstratif (comme « voilà » d'emplois plus variés en français actuel).

Les racines de la monstration vont être utilisées pour donner différents types grammaticaux : (1) un démonstratif peut donner un article (anglais *the*, allemand *das*, espagnol *el*, français *le/l dans l'homme*, roumain *l* dans *om-u-l*, etc.) ; (2) un démonstratif peut donner un pronom (anaphore), par exemple *le* (comme pronom) en français, *der* (comme relatif) en allemand, etc. ; et (3) un démonstratif peut donner un complément(is)eur, comme le *that* anglais ou le *dass* allemand. (Pour une revue de la grammaticalisation des démonstratifs, nous renvoyons à l'étude très détaillée de Diessel, 1999b, 2003b). Nous pouvons remarquer à ce propos que dans environ un quart des langues du monde, les pronoms démonstratifs et les déterminants démonstratifs prennent des formes différentes. Par exemple le français utilise les démonstratifs *celle* et *celui* comme des pronoms indépendants et *ce*, *cette* en tant que déterminants (Diessel, 2003b). A l'inverse dans des langues telles que l'anglais, les mêmes démonstratifs sont utilisés en tant que pronom ou déterminant (*this*, *that* et leurs formes plurielles *these*, *those*). De la même façon, l'espagnol, contrairement au français, utilise les mêmes formes démonstratives (*ese*, *este*, *aquel*) comme adjectifs ou comme pronoms (Jungbluth, 1999). Nous renvoyons ici le lecteur à l'étude menée par Diessel (1999a) examinant les propriétés morphosyntaxiques des démonstratifs et leur développement en marqueurs grammaticaux, basée sur un échantillon représentatif de 85 langues. L'auteur montre ainsi que les démonstratifs surviennent dans différents contextes syntaxiques bien distincts et que les langues diffèrent dans l'emploi de mêmes catégories grammaticales ou de catégories grammaticales différentes pour les démonstratifs. L'auteur montre par cette étude que la façon dont va se grammaticaliser un démonstratif est crucialement déterminée par le contexte syntaxique dans lequel il survient.

Les racines de l'interrogation vont permettre de créer également différents types grammaticaux : (1) un interrogatif peut donner un pronom relatif (anglais *who*, français *qui*, etc.) ; et (2) un interrogatif peut donner un complément(is)eur (français *quoi* donnant le clitique *que*, etc.).

Nous pouvons illustrer ces différentes fonctions grammaticales à l'aide d'un simple exemple tiré de la langue allemande, où nous trouvons, pour une seule et même forme basique, en position 1 une conjonction de subordination, en position 2 un article et en position 3 un pronom relatif :

« Ich denke, das das Haus das leer ist leer bleibt »  
    1     2                  3

Dans cette théorie, nous utilisons le même opérateur *das(s)*, issu du mécanisme de monstration, afin de réaliser une prédication.

« That's that Babybot that said that you're naughty »  
                                   1     2                  3      4

Nous observons que le même outil *that* va à la fois servir de présentatif (1), de démonstratif (2), de relatif (3), ou encore de complémentiseur (4), illustrant bien ainsi les différentes fonctions qu'il permet et qui doivent être acquises par l'enfant.

Nous citerons comme autre illustration, l'exemple du *buang*, une langue austronésienne parlée dans le sud du district de Morobe de Lae. Cette langue n'est pas du tout reliée à l'anglais et elle nous intéresse tout particulièrement pour le processus de grammaticalisation du mot *ken*. En effet, de la même façon que le *that* anglais, le *ken* du *buang* peut prendre la fonction de démonstratif, de relatif, voire de complément(is)eur. Dans une conversation personnelle, G. Sankoff explique que *ken*, qui est une particule déictique, peut être utilisée comme un démonstratif postposé (*this*) qui peut être combiné avec *-egi* et *-gone* comme *kenegi* et *kenegone* pour signifier 'here' et 'there' respectivement. Dans une étude d'ensemble, Heine et Kuteva (2002), à propos du fait que les adverbes peuvent suivre une trajectoire de grammaticalisation à part, reprennent cet exemple *ken* du *buang* qui, d'un adjectif locatif (*here*) peut donner naissance à un marqueur démonstratif, lequel peut encore se développer pour constituer un relativiseur. Voici les exemples en question tirés de Heine et Kuteva (2002, p. 383, citant Sankoff 1979: 35-6) que nous avons retrouvés dans Sankoff et Brown (1976) :

Buang (Austronesian, Austric)

(a) Ke mdo ken (adverbe de lieu)

I live here

'I live here'

(b) Ke mdo byaŋ ken (démonstratif postposé)

I live house this

'I live in this house'

(c) Ke mdo byaŋ ken gu le vkev (relativiseur)

I live house that you saw yesterday

'I live in the house that you saw yesterday'

### **II.11.1.2. La récursivité : une clé pour expliquer la différence entre communication animale et humaine ?**

Ces deux versants fondamentaux pour la grammaticalisation, monstration et interrogation, vont également permettre de réaliser une opération cruciale du langage : la récursivité. En partant d'un outil de monstration ou d'interrogation, nous allons aboutir à un outil syntaxique puissant permettant d'emboîter à l'infini des phrases dans d'autres phrases : le phénomène dit de récursivité. La récursivité semble être un phénomène caractéristique du langage humain, puisque les systèmes de communication animale semblent manquer d'une expressivité aussi riche que celle qu'offre la capacité à humaine à la récursivité. Nous nous appuyons sur l'étude menée par Hauser et al. (2002), qui admettent que le langage n'est pas totalement déterminé génétiquement et argumentent l'existence de deux conceptions de la faculté de langage, une conception large, et une seconde plus restreinte, dite étroite (Figure II.35).

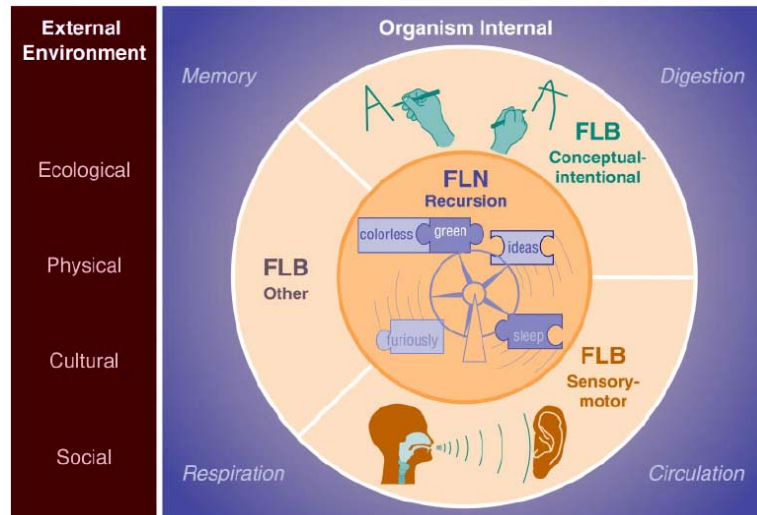


FIGURE II.35 : Représentation schématique des facteurs externes et internes à l'organisme reliés à la faculté de langage. FLB inclut le système sensori-moteur, le système conceptuel et intentionnel et d'autres systèmes possibles ; FLN inclut les computations grammaticales noyaux que les auteurs suggèrent limitées à la récursivité (D'après Hauser et al., 2002).

#### II.11.1.2.a) La faculté de langage au sens large (FLB)

La première faculté nommée FLB (*Faculty of Language Broad*) combine au moins deux systèmes internes à l'organisme, un système sensori-moteur et un système conceptuel-intentionnel.

Le système sensori-moteur est un pré-requis au langage, et tout particulièrement pour la perception de la parole. De la même façon que la littérature sur les capacités de l'enfant a démontré qu'un très jeune bébé était capable d'apparier correctement un visage à une vocalisation correspondante, les singes sont également capables d'intégrer l'information perçue de façon multimodale.

Le système conceptuel-intentionnel constitue de la même façon une faculté cognitive essentielle au langage, notamment la faculté à catégoriser les éléments de l'environnement, démontrant un niveau d'abstraction développé. Bovet et Vauclair (1998) ont réussi à démontrer par leurs expériences menées sur 4 babouins adultes nés en laboratoire (*Papio Anubis*) que ceux-ci présentaient la faculté de catégoriser des objets (nourriture/non-nourriture, que celle-ci soit physiquement présente ou présentée sous forme de photographies) au niveau fonctionnel (est-ce alimentaire ou non-alimentaire ?), avec la faculté de généraliser ensuite cette faculté à de nouveaux objets, au niveau perceptif (paire d'objets identiques/paire d'objets différents), mais également au niveau conceptuel (jugement sur l'appartenance à une catégorie). Les

babouins sont également capables de catégoriser ces objets si ceux-ci sont présentés sous forme de photographies, même si les auteurs soulignent le fait que le transfert n'est pas immédiat ni aussi parfait.

Une seconde illustration de ce système conceptuel/intentionnel vient de l'étude dirigée par De Waal et al. (2005) sur des singes capucins (*Cebus apella*). Ceux-ci montrent que l'attitude de ces petits singes non anthropoïdes face au miroir est différente de l'attitude adoptée face à d'autres singes. Les auteurs ont comparé la réaction des singes capucins lorsqu'ils voient un autre singe connu ou étranger, séparé par une cloison transparente et lorsqu'ils se trouvent face à un miroir. Ils observèrent que face à des congénères familiers, qu'ils soient mâles ou femelles, les singes capucins ont peu de réactions. Les femelles sont moins anxieuses face au miroir que face à une femelle étrangère. Les mâles, quant à eux, réagissent de façon plus ambiguë face aux mâles inconnus, tout comme face au miroir. Toutefois certaines attitudes étaient réservées uniquement lorsque les singes se trouvaient face au miroir, des comportements comme les cris, le repli, etc. Les auteurs concluent de cette étude qu'il serait possible que ces singes capucins possèdent un degré intermédiaire de conscience en reconnaissance/non-reconnaissance face au miroir. Ils sont ainsi capables d'établir une différence entre un congénère lambda et l'image réfléchi. Cette notion de reconnaissance est très intéressante dans la mesure où elle constitue une première étape, classiquement testée en primatologie et en développement de l'enfant, vers une conscience de soi, ou vers une représentation de son propre corps.

#### **II.11.1.2.b) La faculté de langage au sens étroit (FLN)**

A l'intérieur du système langagier au sens large (FLB), qui comprend donc des capacités de communication partagées avec le monde animal, Hauser, Chomsky et Fitch (suivant en cela Chomsky) distinguent un autre système computationnel, la FLN (*Faculty of Language Narrow*), un système proprement linguistique, indépendant des autres facultés (comme celles que nous venons d'évoquer). Une propriété noyau de cette FLN réside dans la récursivité, ce système partant d'un ensemble fini d'éléments et produisant une gamme potentiellement infinie d'expressions discrètes, telles que « Marie voit que Pierre voit que Jean voit que Paul triche aux cartes » par exemple. Chomsky prétend que c'est dans cette capacité de récursivité qu'il faut

chercher le point de rupture expliquant l'absence d'une réelle compétence langagière chez le singe. Ainsi le langage se verrait caractérisé par cette faculté humaine unique.

By this hypothesis [hypothesis 3], FLB contains a wide variety of cognitive and perceptual mechanisms shared with other species, but only those mechanisms underlying FLN –particularly its capacity for discrete infinity- are uniquely human. This hypothesis suggests that all peripheral components of FLB are shared with other animals, in more or less the same form as they exist in humans, with differences of quantity rather than kind. What is unique to our species is quite specific to FLN, and includes its internal operations as well as its interface with the other organism-internal systems of FLB. (Hauser et al., 2002:1573).

Cette capacité d'emboîtement récursif de phrases a sans doute émergé assez récemment dans le développement proprement humain et elle semble résulter d'une capacité de computation adaptative permettant de communiquer une variété sans fin de pensées.

Chez l'animal, plusieurs tentatives ont été réalisées pour démontrer l'existence ou non de cette faculté de récursivité. Nous citerons *inter alia* Fitch et Hauser (2004) chez les singes tamarins, qui possèdent une capacité de traitement de séquences acoustiques régulières, mais ne sont en revanche incapables de traiter une structure récursive.

Hauser, Chomsky et Fitch (2002) concluent de leur étude que la complexité sociale se répercute directement sur les capacités cognitives –les représentations de relation sociale faisant appel à une organisation hiérarchique, ouverte, régies par des règles, et amodale– et que ce fait constitue un précurseur dans l'évolution de la cognition humaine avec langage, ou plus simplement un précurseur de la faculté de langage étroit.

Bien que cette théorie avancée par Hauser, Chomsky et Fitch (2002) paraisse fondamentalement juste et que nous adhérions à l'idée que la récursivité soit un outil syntaxique complexe réservé à l'usage linguistique humain, certains auteurs remettent en question certains de leurs arguments, que nous proposons de résumer ci-après.

En réponse à cet article de Hauser et al. (2002), Jackendoff et Pinker (2005) proposent de discuter l'idée que cette faculté de langage étroit consiste seulement en la récursivité et que cela ne puisse pas être considéré comme une adaptation à la communication. Ils argumentent dans ce but que cette caractérisation est problématique à différents niveaux, notamment la dichotomisation des capacités

cognitives en caractère unique ou identique vis-à-vis des capacités non-linguistiques ou non-humaines, laissant alors de côté des capacités qui auraient pu être présentes dès le départ et qui auraient subies une modification durant l'évolution humaine. Selon Jackendoff et Pinker, cette théorie dichotomique ne laisse pas la place à la notion d'adaptation, et qu'en omettant cette alternative, Chomsky peut alors argumenter que rien ne distingue l'utilisation du langage pour la communication de l'utilisation du langage pour la parole intérieure par exemple. A ce sujet Jackendoff et Pinker répliquent l'argument suivant :

Moreover, the existence of phonological rules that ease articulation, and of syntactic processes with pragmatic communicative functions such as topic and focus, points strongly to language being an adaptation to social communication rather than to internal reasoning. (2005:213)

Jackendoff et Pinker rejoignent Hauser et al. sur l'idée que la vraie récursivité caractérisée par une structure hiérarchique et un enchevêtrement illimité, semble ne pas exister dans tout autre système de communication animal. Selon les auteurs, chez l'humain, il existe une démonstration non ambiguë de récursivité dans le domaine de la cognition visuelle



FIGURE II.36 : Récursivité dans le groupement visuel (D'après Jackendoff et Pinker, 2005)

L'exemple ci-dessus (Figure II.36) qu'ils proposent, montre qu'il peut exister une « infinité discrète » en perception visuelle, comprenant des structures hiérarchiques de profondeur illimitée, l'organisation de ce groupement visuel étant alors gouverné par des principes de Gestalt. Les auteurs veulent ainsi montrer par là que la récursivité en soi ne fait pas partie du FLN tel que l'entendent Hauser et al. (2002). Jackendoff et Pinker argumentent alors que la principale raison pour laquelle la syntaxe récursive ait évolué se trouve dans l'expression de structures récursives



dans la cognition. Ainsi la distinction FLB/FLN échouerait à expliquer pourquoi les humains possèdent le langage et non pas les animaux. La récursivité syntaxique pourrait de ce fait constituer une nouvelle combinaison réalisée à partir de capacités nouvellement re-réglées puisées dans la cognition. Ils remettent d'autre part en question la dichotomie humain/non-humain, qui échoue à établir une distinction claire entre une similarité due à des fonctions analogues ayant évolué indépendamment d'une similarité due à un héritage d'un ancêtre commun récent. Enfin, ils dénoncent la dichotomie syntaxe/lexique qui selon eux ne prendrait pas en compte l'ensemble des phénomènes linguistiques productifs qui ne peuvent pas être analysés en termes de syntaxe étroite, et qui, de ce fait, isole de façon erronée la récursivité et l'établit en tant qu'unique facteur développemental dans l'évolution du langage.

### **II.11.2. Ontogenèse de « that » (ça) à « that » (que)**

Tomasello (2003b) note que la grammaire est souvent associée à l'idée d'une adaptation biologique, unique à notre espèce. Il existerait alors une grammaire universelle qui serait biologiquement déterminée, constituée de similarités de contenu parmi les structures grammaticales des langues du monde. Toutefois, l'exemple des langues austronésiennes qu'il cite, montre que certaines langues peuvent varier de façon remarquable, en ne possédant pas, par exemple, de catégories ou constructions que l'on retrouve régulièrement dans les langues européennes. Une solution à ce dilemme est d'envisager alors la grammaticalisation comme un processus opérant sur le temps historique. L'argument invoqué repose sur l'idée que les processus de grammaticalisation et de syntactisation peuvent subir de forts changements à l'intérieur de la structure linguistique lors de périodes temporelles relativement courtes.

En ce qui concerne la mise en place de la grammaire précocement chez l'enfant, Tomasello (2003a,b) argue qu'elle pourrait être caractérisée par un inventaire de constructions verbales dites « en île », qui définissent ensuite les premières catégories syntaxiques, basées sur le lexique. Tomasello adopte une position théorique qui vise à ne pas séparer la forme de la grammaire.

L'étude menée par Lieven et al. (1997) semble également abonder dans ce sens, avec l'idée que l'organisation grammaticale soit basée sur les items. Suite à leur analyse distributionnelle d'énoncés de 12 enfants anglais, suivis de l'âge de 12

mois à 3 ans, ils proposent que les enfants puissent organiser leurs énoncés sur la base de la combinaison de catégories sémantiques sous-jacentes plutôt que sur la base de connaissances syntaxiques. En effet, selon ces chercheurs, la construction de patrons distributionnels autour des items lexicaux spécifiques pourrait être la bonne façon d'envisager les énoncés d'enfants. Les auteurs reprennent l'idée d'une organisation en îlot, comme définie pour les verbes par Tomasello (2003b). Ces îles seraient des segments (mots ou syntagmes) que l'enfant aurait identifié et qu'il pourrait mettre en relation de façon systématique avec ce qui suit ou précède.

Nous avons discuté préalablement des mécanismes de monstration et interrogation afin de démontrer la puissance d'un tel système déictique. Localiser et vocaliser permettent ainsi d'aboutir à la grammaticalisation, une grammaticalisation complète avec utilisation des complémentiseurs et utilisation du phénomène de récursivité. Nous proposons alors de nous intéresser en détail aux différents stades d'acquisition de cette grammaticalisation chez l'enfant. Comment l'enfant passe-t-il de l'acquisition des démonstratifs à l'utilisation de présentatifs, de relativiseurs, de complémentiseurs pour aboutir à la manipulation récursive dans sa langue ?

Diessel et Tomasello (2000) ont observé les premiers énoncés de 7 enfants constitués de compléments de phrases et qui étaient composés de deux types de verbes : le premier type concerne les verbes épistémiques comme « think » et « know », en les utilisant basiquement à la première personne du présent, sans aucune utilisation de complémentiseur, c'est-à-dire sans utilisation de formes telles que « I think that... ». Le second type de verbes correspond à des verbes d'attention tels que « look » ou « see », utilisés presque exclusivement à la forme impérative, sans négation, et sans complémentiseur une fois encore. Les enfants semblent donc se reposer sur une approche basée sur les items, sans toutefois impliquer systématiquement un enchevêtrement syntaxique au départ.

Diessel et Tomasello (2000) se sont également penchés sur les phrases relatives chez des enfants locuteurs de l'anglais dans la période survenant avant 3 ans, 3.5 ans. Précédemment, Lieven et al. (1997) avaient pu observer que l'arrivée des relatives dans le langage se situait à ce même âge. Diessel et Tomasello (2000) remarquent que les premières phrases relatives surviennent dans des constructions présentatives, telles que « Here is the chair that broke » ou « There's a drink I want ». Ces phrases consistent alors en une phrase principale utilisant le verbe to

be, ce type de phrase étant très répandu chez les enfants de cet âge et en une phrase relative, constituée la plupart du temps d'un verbe intransitif véhiculant une information nouvelle sur le topic préalablement introduit dans la phrase présentative. Ce type de construction exprime une seule proposition, le topic étant introduit par une construction introductive commune. Ainsi, les présentatifs semblent être le second stade fonctionnel des démonstratifs (*Here, there, it is...*).

Ce n'est qu'après 3 ans, 3.5 ans que les enfants vont complexifier leurs constructions relatives, où celles-ci seront rattachées à un nom dans une véritable phrase principale, permettant alors de parler d'énoncés composés de phrases relatives subordonnées. Ainsi, à cet âge, les enfants sont capables de tourner une phrase en relative avec un démonstratif, mais il faudra attendre encore un peu avant qu'ils ne soient capables de réaliser des phrases du type « I think that » en utilisant les verbes d'états mentaux. Ils utilisent d'abord le démonstratif comme un présentatif, ils l'utilisent ensuite pour créer une relative pour finalement se servir de celui-ci comme un puissant outil de récursivité, comme complément(is)eur. Diessel (2003b) souligne que, tout comme les pronoms relatifs, les complément(is)eurs sont fréquemment dérivés de démonstratifs pronominaux. Il cite comme exemple le cas des langues germaniques du nord et de l'ouest qui se développent à partir de démonstratifs survenant à l'origine à l'intérieur d'une phrase principale référant à une proposition suivante.

Le premier type de récursivité utilisé par l'enfant –typiquement ce que Chomsky nomme la FLN « Faculty of language narrow »– est communément un outil de comméragé très répandu: « il m'a dit que... et que... et que... », permettant ainsi une récursivité de complémentisateurs apposés grâce à la conjonction de coordination. Ils ne sont pas emboîtés comme dans l'exemple suivant « Marie voit que Pierre voit que Fred voit que Paul bat sa femme ». Ce second type de récursivité utilise l'outil démonstratif, qui va avoir fonction de complémentisateur, comme le *that* de l'anglais, ou de même une racine d'interrogatif comme le *que* du français. Nous arrivons en plus complexe à une articulation des démonstratifs et des interrogatifs dans des corrélations comparables aux *tam...quam, talis...qualis*, etc., latins, déjà cités. Avec ces corrélations, du type : « il y avait **si** longtemps **qu'**il n'avait pas mangé **qu'**il était près de s'évanouir » ou « ils étaient **tant** courageux et **si** nombreux **qu'**ils réussirent », nous sommes parvenus au cœur du raisonnement type « Si X alors Y ».

Nous pouvons remarquer que ces deux types de récursivité nécessitent sans aucun doute l'activation de l'aire de Broca, sans quoi l'extraction ne serait pas correcte. Mais une activité préfrontale semble requise aussi pour le raisonnement logique (« Si X alors Y »), le narratif ou encore la rhétorique (cf. supra).

Selon nous, lorsque l'enfant commence à utiliser des phrases qui sont emboîtées, cela nécessite du cerveau qu'il mette en route un système de perception active, qui expliquerait que saute aux yeux (*pop out*) l'interprétation la plus évidente, comme « sa femme a été tuée par John », qui n'est pas à première audition (ce n'est pas la femme de John), une simple transformation passive de « John a tué sa femme ». En d'autres termes le cerveau fonctionnerait tout comme il le ferait face au fameux cube de Necker, qui donne lieu à une perception bistable (multistable chez Vasarely) avec un traitement enactif fronto-pariétal, différent d'un traitement logique préfrontal d'une figure paradoxale (à la Escher, comme le célèbre crazy crate ou cageot fou du docteur Cochran) (Figure II.37).

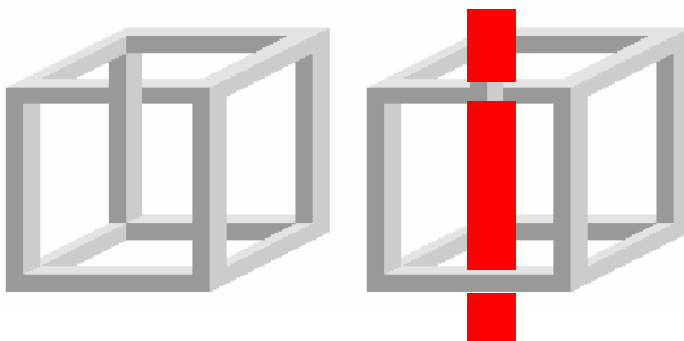


FIGURE II.37 : Le cube de Necker. Sur la figure de droite, nous avons volontairement biaisé (bloqué) la bistabilité du cube afin de montrer l'un des deux états perceptifs produits par cette figure (le plus courant en perspective post-Giotto), laquelle est en réalité paradoxale (à gauche).

Ce cube de Necker est un percept qui présente non pas une ambiguïté, mais une bistabilité. Lorsque l'on observe le cube, apparaissent de façon alternative (en moyenne toutes les 3 secondes, pas moins) les deux interprétations valides, l'observateur basculant soit sur une interprétation, soit sur l'autre.

Comment passe-t-on en langage d'une représentation à une autre ? Comment passons-nous en prosodie syntaxique de « C'est Paul, ça ! » à « Ça, c'est Paul » ? Comment passe-t-on, dans le cadre du mot, de l'envers à *verlan* ? C'est ce qu'ont exploré systématiquement pour les syllabes multistables, Sato et al. (2004, 2006), en montrant que cet effet de transformation verbale relevait bien de la mémoire de travail, et plus exactement d'une boucle phonologique pénétrable par les contraintes de contrôle des gestes articulatoires (*ips* répété donne *psi*, l'inverse n'étant pas vrai,

les gestes des consonnes et de la voyelle étant lancés, coarticulés, en synchronie dans *psi*, pas dans *ips*).

Ces études de perception bistable nous amènent à valider l'existence d'un système qui va justement tenter de stabiliser la perception entre formes concurrentes. Ce système général, C. Abry (Abry et al. 2004) l'a nommé la « stabil-loop » ou stabil-boucle, recrutant au minimum le couple Broca-SMG. Et c'est précisément à l'intérieur de cette stabil-loop que nous allons trouver le *that-path* déjà cité, dont le système est mis en œuvre par cet outil puissant qu'est le geste de pointage déictique. Imaginons l'enfant produisant un flux de parole, comme dans le jargon babbling ou un babillage canonique « par défaut » (sans autre contrôle que l'oscillation de la mandibule, cf. en première partie le « jaw-mama-for-free »), et qui tout à coup veut amener l'attention de sa maman sur un objet précis qu'il désire en le désignant du doigt via le *That-path*. Il va ainsi, via le bras, « chunker » (ou « chopper ») le flux de parole afin que certaines formes soient mémorisées dans la mémoire de travail. Cette mémoire de travail va ensuite avoir pour tâche de stabiliser la forme (choisir une forme entre *amam* ou *mama* par exemple), la stabil-loop intégrant les formes stationnaires les plus naturellement phasées au niveau du contrôle des articulateurs (*mama* dans ce cas précis). Avec le *That-path* intégré dans la *Stabil-loop*, on a d'abord un système déictique attentionnel (un précurseur syntaxique de type *that-grammar*) qui aidera à saisir la sémantique de l'objet pointé via la phonologie du pied du pointer. Mais également par la suite un système de stabilisation d'une seule structure linguistique, parmi d'autres moins stables (ou rangées selon leur degré de stabilité), avant sa mémorisation à long terme. La notion essentielle ici pour nous est qu'à la base de toute grammaticalisation de type *that*, nous trouvons le geste de pointage comme fondement.

### II.11.3. Conclusion

A travers les notions de monstration et interrogation que nous venons d'étudier, nous nous sommes particulièrement intéressée à définir l'acte de communication en lui-même comme un acte sémiotique, qui peut être vu aussi concrètement qu'une action de préhension. Ces actions de préhension, actions dirigées vers un but (goal-directed actions), sont depuis quelques années largement replacées dans les neurosciences cognitives de l'action (cf. notamment par les

recherches de Marc Jeannerod, fondateur de l'Institut des Sciences Cognitives, à Lyon), et elles peuvent aller jusqu'à l'exploration de l'agentivité, de la conscience et du langage (Cf. l'étude pionnière de Castiello et al., 1991). Suivant cette ligne de pensée, notre proposition pour cet acte sémiotique consiste à insister sur l'évidence que la lune est aussi clairement désignée par le doigt qu'une pomme est prise. La pomme n'étant clairement pas le contexte de la prise, mais son objet.

Si l'on revient à l'enfant, lorsqu'il montre l'argument lune, il s'agit d'un acte sémiotique qui peut être aussi palpable que lorsqu'il tire sa mère par son vêtement ; ou bien, agacé par son inattention lorsqu'elle est engagée dans une conversation avec un tiers, et qu'elle tarde à lui répondre, il lui tourne des deux mains la tête vers ce qui l'intéresse ; aussi tangible qu'un contact si l'index vient, lorsque cela est physiquement possible, jusqu'à toucher du doigt le référent, dans « Tu TE le payes ! », avec l'index sur le plexus de l'interlocuteur en parole ou en langue des signes. Mais ce n'est pas le côté « concret » de cet acte sémiotique qui en donne la structure cognitive : ce caractère concret n'est que le résultat d'une emphase dans une communication qui peut très bien fonctionner en esquisse. Les comptines d'élimination en sont une parfaite illusion : le pointer du doigt du « compteur » de comptine n'a pas à toucher chaque joueur (tout au plus le joueur éliminé), mais à l'indiquer, à le désigner suffisamment.

La lune n'est donc pas le « contexte » du doigt, de l'œil ou de la maman... Chacune de ces composantes a une fonction précise : selon le proverbe répété, il n'y a que le sot qui regarde le doigt quand le sage montre la lune.

Cet index pointé est, de plus, grammaticalement invisible : le système attentionnel de la deixis ou monstration opère de façon à ce qu'on le voit de manière angulaire extrêmement précise pour localiser le référent ; mais nous sommes incapables de voir en même temps la présence d'une bague sur ce doigt. C'est une *grammatical blindness* que l'on retrouve y compris dans un texte écrit, exemplairement illustré dans la difficulté particulière à détecter les lettres des mots-outils (ou pointeurs en quelque sorte) par rapport aux autres mots (exemple cité par Rey, 2002, tiré de l'expérience réalisée par Healy, 1976, Drewnowski et Healy, 1977, cf. encore Healy, 1980, 1994) :

FINISHED FILES ARE THE RE-  
 SULT OF YEARS OF SCIENTIF-  
 IC STUDY COMBINED WITH THE  
 EXPERIENCE OF YEARS

Si vous connaissez ne serait-ce qu'un peu d'anglais, vous chronométrez le temps que vous mettez à compter le nombre de « F » contenus dans ces quatre lignes...

Si vous trouvez, ce qui est le cas généralement, que vous avez loupé les « F » d'un petit mot pourtant bien répété, vous aurez (i) compris le phénomène de *grammatical blindness* ; (ii) goûté le proverbe chinois énoncé ci-dessus ; (iii) et en prime apprécié le niveau de compétence demandé aux correcteurs d'épreuves, aux journalistes, aux chercheurs, aux écrivains, aux Académiciens... sans oublier surtout aux enfants des écoles ou à l'étudiante qui rédige sa thèse !

Hallé et al. (2005) confirme cette *grammatical blindness* chez le jeune enfant qui, à 6 mois, préfère porter préférentiellement attention aux mots lexicaux plutôt qu'aux mots de fonction. Il semble en effet que les mots de fonction soient acoustiquement moins saillants que les mots lexicaux, dans la mesure où ils tendraient à être plus courts et à posséder des voyelles réduites entre autres. Toutefois, à l'âge de 13 mois, mais pas à 8 mois, les enfants préfèrent écouter des phrases contenant des mots de fonction réels (*the, this* en anglais par exemple) que des phrases contenant des mots de fonctions non sens (*kuh* par exemple). De plus, à l'âge de 11 mois, mais pas à 10 mois, des différences dans l'amplitude des enregistrements de potentiels évoqués (ERP) sont relevées pour les phrases contenant des morphèmes de fonction non modifiés vs. des morphèmes de fonction prosodiquement et segmentalement modifiés. Pris ensemble, ces résultats suggèrent que les enfants traitent les mots de fonction, mais différemment par rapport aux mots lexicaux, et ceci assez tôt dans leur développement. Il semble ainsi qu'autour de 11-13 mois, ils commencent à être sensibles à la grammaticalité des articles. Une explication pertinente à ce phénomène pourrait résider dans l'idée que les enfants ne seraient pas sensibles à un groupe clitique analytiquement bien formé, mais reconnaîtrait plutôt ce groupe clitique comme une forme familière holistique, c'est-à-dire que « le canard » serait alors une variante possible de « canard » avec ou sans « filler » prononcé à la place de l'article.

Notre objet d'étude étant posé et validé pour notre étude réalisée sur le corpus longitudinal, nous proposons d'aller un peu plus loin dans la relation qui unit le geste à la parole chez l'enfant à partir de ses premiers mots. Le pointage communicatif étant envisagé comme le meilleur prédicteur des premières performances linguistiques, nous proposons de nous intéresser à la relation entre geste et parole chez l'enfant et aux répercussions que ce mécanisme peut avoir dans la mise en place de combinaisons de mots et plus largement dans le langage chez l'enfant. Ces considérations ontogénétiques nous amèneront également à nous pencher sur la phylogenèse de la relation qui unit le geste à la parole, en explorant la théorie gestuelle de l'origine du langage, redevenue très en vogue ces dernières années.

## **II.12. La relation geste-parole chez l'enfant dans l'acquisition du langage**

Les jeunes enfants utilisent les gestes dans un but communicatif très tôt dans le développement et bien avant d'être capables de produire le premier mot. Des études pionnières comme celles de Bates et al (1979) indiquaient déjà que le « pointage communicatif » était le meilleur prédicteur des premières performances linguistiques.

Bates et Dick (2002) rapportent un lien de parenté étroit entre geste et parole chez l'enfant. Ils notent (comme Thelen, 1981) que le babillage canonique est fortement lié à l'apparition de la main rythmique. En effet, les enfants produisent typiquement leurs premiers gestes rythmiques entre 9 et 12 mois, comme le « banging », donnant lieu au battement rythmique des bras, une action souvent liée à l'activité rythmique du babillage, les deux survenant assez fréquemment en co-occurrence chez les enfants autour de 7 mois. Entre 8 et 10 mois, l'enfant commence à donner les signes d'une compréhension des mots, un événement marquant selon Bates et Dick, et qui est corrélé avec l'émergence de gestes déictiques, notamment le pointer impératif.

Morford et Goldin-Meadow (Morford et Goldin-Meadow, 1991), s'intéressant aux rôles que jouent les gestes dans les premières étapes de l'apprentissage du langage, constatent que, d'une part, tous les enfants produisent spontanément des gestes en combinaison avec la parole, et que, d'autre part, ils sont tous capables de comprendre les gestes quand ils leur sont présentés en combinaison avec la parole.



Et ceci pas uniquement quand ces gestes apportent de l'information redondante à la parole, mais aussi lorsque les gestes se substituent à la parole. Ils relatent, eux aussi, que les enfants utilisent les gestes pour communiquer, plusieurs mois avant qu'ils ne commencent à parler.

Il existe ainsi deux types de combinaisons geste/parole qui prédominent durant la période de développement où l'enfant produit des énoncés dont la longueur n'est que d'un mot : (i) les gestes qui se combinent avec la parole pour véhiculer une signification redondante avec la signification du mot parlé (pointer un objet, une pomme, et la nommer « po »), et (ii) les gestes qui se combinent avec la parole pour ajouter un nouvel élément connecté au mot prononcé (« gi(ve)me ! » en pointant la pomme).

Les gestes semblent aider les enfants à dériver la signification d'un énoncé adulte. Schaffer et al. (Shaffer et al., 1983), cités par Morford et Goldin-Meadow (1991) dégagent l'idée que les gestes fournissent souvent de l'information qui peut aider l'enfant à interpréter l'instruction de la mère. Par exemple, la mère pointe un nounours et prononce simultanément « Put the hat on Teddy » (mets le chapeau sur Teddy), utilisant de ce fait le geste pour focaliser l'attention de l'enfant sur la localisation de l'objet foyer.

MacNamara (1977), cité par Morford et Goldin-Meadow (1991), a effectué des expériences sur des enfants de 1 an 2 mois à 1 an 8 mois, concernant deux types de gestes : le premier était celui de tendre un objet vers l'enfant comme si on le lui offrait ; et le second celui de pointer. Il faisait varier de plus l'énoncé qui accompagnait chaque geste. Il observa que l'enfant répondait différemment aux deux gestes. En réponse au geste d'offrir, l'enfant se saisissait de l'objet. En réponse au geste de pointer, l'enfant regardait l'objet sans agir sur celui-ci. Ce qui est le plus important, c'est que lorsqu'il existe un conflit entre l'information véhiculée dans le geste et dans l'énoncé, ou lorsque l'énoncé n'est pas interprétable par l'enfant, le geste prend alors la priorité et l'enfant prend ou regarde l'objet uniquement selon le geste.

Selon Morford et Goldin-Meadow (1991), ces études ne fournissent qu'un aperçu incomplet pour savoir si les jeunes enfants sont capables d'intégrer l'information à travers les modalités gestuelles et verbales. C'est pourquoi leur étude a eu pour but d'explorer la compréhension gestuelle et également la production

gestuelle chez les jeunes enfants, en se focalisant particulièrement sur la capacité de l'enfant à intégrer l'information à travers geste et parole sur une population d'enfants qui se situent à l'étape de l'énoncé d'un mot.

### **II.12.1. Les combinaisons geste-parole chez l'enfant**

Dans leur étude, Morford et Goldin-Meadow (1991) divisent 46 enfants en deux groupes. Le premier comprenant ceux qui produisent seulement les combinaisons geste/parole où les gestes véhiculent principalement de l'information redondante à la parole. Le deuxième groupe est constitué d'enfants qui produisent au moins une combinaison geste/parole dans laquelle ces deux modalités véhiculent une information différente et supplémentaire, ces deux modalités prises ensemble formant une phrase interprétable à deux éléments (« mots »).

Il ressort de ces expériences que les enfants, qui sont tous locuteurs à un mot, utilisent des gestes assez fréquemment, particulièrement par rapport à la quantité de parole qu'ils produisent. En moyenne, ils produisent 1.5 gestes par minute (seul ou en combinaison avec la parole), comparé à 2.8 mots par minutes (seul ou en combinaison avec les gestes). Ce taux de production peut se comparer au taux de production gestuelle chez l'adulte. Selon Bekken (1989), cité par Morford et Goldin-Meadow (1991), une mère produit en moyenne 1.1 gestes par minute quand elle parle à sa fille de 18 mois et 1.6 gestes par minute quand elle s'adresse à d'autres adultes. L'étude de Bekken souligne que 24% de la communication des enfants est produite par des gestes seuls sans parole, alors qu'aucun des adultes testés par Bekken ne produit de gestes sans parole. Cette quantité de gestes sans parole semble être typique des enfants aux étapes de commencement de l'apprentissage du langage.

De plus, les enfants produisent essentiellement trois types de gestes : pointer, donner, lever les mains, à la différence évidente des adultes ou d'enfants plus âgés, qui produisent typiquement un groupe plus large de gestes iconiques et métaphoriques, en plus des gestes de pointage.

Si un geste est compris par les enfants, ils peuvent l'interpréter dans le flux de parole qu'ils entendent et par conséquent intègrent l'information véhiculée par le geste dans l'information parole. Si le geste fournit de l'information redondante par rapport à la parole qui l'accompagne, le geste tend alors à accroître la

compréhension de l'enfant. Si le geste fournit une information différente de la parole, alors cette information peut être intégrée avec la parole, le geste pouvant ainsi servir de substitut à un mot.

Church et Goldin-Meadow (1986) ont révélé (nous le verrons bientôt) que les enfants qui produisent des combinaisons geste/parole non-concordantes dans l'explication d'une tâche de conservation (des grandeurs), ont de fortes chances d'être avantagés par les instructions renforçant la conservation, par rapport aux enfants qui produisent des gestes en concordance avec la parole. Selon Morford et Goldin-Meadow (1991), cela peut mettre en évidence le fait que les locuteurs à un mot, dont les gestes véhiculent une information différente de la parole qui l'accompagne, sont sur le point d'étendre leur vocabulaire parlé à deux mots (Goldin-Meadow et Butcher, sous presse ; cf. infra).

Le répertoire gestuel des enfants semble se restreindre principalement à deux gestes (pointage et don), mais ils sont fréquemment produits et sont facilement compris par des enfants de 18 mois. De plus, ces mêmes gestes semblent à cet âge-là former un système intégré avec la parole. Ces résultats vont alors soutenir l'idée qu'à l'intérieur de ce système intégré geste-parole, les gestes sont utilisés pour *bootstrapper* la parole quand cela est nécessaire (Capone et McGregor, 2004).

Selon Iverson et Goldin-Meadow (2005), les gestes produits de façon précoce dans le développement de l'enfant occupent une fonction communicative forte que celui-ci n'est pas encore en mesure de fournir par le langage oral. Le geste de pointage utilisé seul offrirait alors une opportunité aux enfants d'amener l'attention des autres sur un objet de l'environnement, avant de pouvoir correctement l'étiqueter verbalement. Lorsque l'enfant entre dans le langage, la combinaison geste + mot constituerait alors une technique pour communiquer deux parties d'information dans un seul et unique énoncé, avant de pouvoir produire une combinaison de 2 mots autonome. Ainsi outre le fait que les gestes anti-dateraient le développement langagier oral, ils pourraient également prédire les changements à venir dans ce développement (cf. infra).

Les auteurs ont ainsi voulu tester cette hypothèse en examinant la production gestuelle en relation avec le développement syntaxique et lexical dans les premières étapes de développement du langage. Ils ont testé plus spécifiquement sur 10 enfants de langue anglaise suivis longitudinalement de 10 à 24 mois : (i) si

l'utilisation des gestes pour référer à des objets spécifiques entretenait une relation particulière avec l'émergence d'étiquettes verbales pour ces mêmes objets ; et (ii) si la production de combinaisons geste + mot était effectivement reliée à l'émergence des énoncés à 2 mots. Pour cette étude, les auteurs ont alors sélectionné les sessions se déroulant entre le début de la parole à 1 mot (10-14 mois) et l'émergence des combinaisons à 2 mots (17-23 mois). Dans les premières sessions, les enfants se reposaient en grande majorité sur l'utilisation des gestes pour référer aux objets, avec un quart seulement de références produites par la parole. En effet, les premiers gestes apparaissent fréquemment en premier lieu sans accompagnement vocal, ce qu'avait également relevé Masur (1983) quelques années plus tôt dans son étude portant sur 4 enfants anglais.

Toutefois les gestes deviennent de moins en moins importants à travers le temps, et là où 9 enfants sur 10 produisaient une majorité de références gestuelles en sessions initiales, la quasi-totalité n'en produisait pratiquement plus lors des sessions finales. Les auteurs émettent alors l'hypothèse d'une fonction facilitative du geste, qui permettrait qu'un item lexical individuel entre dans le répertoire de l'enfant en premier lieu par le geste, puis que ce même item soit ensuite transféré à la parole. Et il semble en effet que la modalité gestuelle ait un impact clair sur le développement lexical de l'enfant : significativement plus d'items étaient produits initialement avec un geste, et en moyenne les enfants produisaient un geste pour un objet particulier trois mois avant de produire le mot pour cet objet. L'étude des gestes en relation avec l'entité qu'ils désignent permettrait alors de prédire les items lexicaux qui vont apparaître dans le répertoire verbal de l'enfant. Ainsi le lien entre le geste déictique et son référent semble beaucoup plus clair pour l'enfant que la relation qui peut exister entre les mots et leurs référents, comme le soulignent Iverson et Goldin-Meadow :

Because the relation between a deictic gesture and its referent is more transparent than the arbitrary relation between most words and their referents, gesture can provide children with a temporary way to communicate about objects, one that allows them to circumvent difficulties related to producing speech (Acredolo et Goodwyn, 1988; Werner et Kaplan, 1963). Gesture may thus serve as a transitional device in early lexical development. (Iverson et Goldin-Meadow, 2005:369)

Quelques années auparavant, Butcher et Goldin-Meadow (2000) et Goldin-Meadow et Butcher (2003) s'étaient également penchées sur la relation que pouvaient entretenir le geste et la parole chez l'enfant, particulièrement le rôle que ce

type de relation pouvait tenir entre la production de mot seul accompagné d'un geste dès 12 mois et l'évolution vers des combinaisons à 2 mots dès 18 mois, relation qui avait déjà été relevée par l'équipe romaine de Volterra bien des années plus tôt chez des enfants italiens (Caselli, Volterra et Pizzuto, 1984, pour une revue plus large voir Volterra et al., à paraître in *Festschrift Bates*). Ces auteurs ont suggéré à plusieurs reprises que chez les enfants en développement typique, les gestes sont utilisés dans les premières étapes du développement du langage mais également lorsque les énoncés à 2 mots apparaissent. Cette faculté à combiner 2 symboles linguistiques marque une étape importante dans le processus d'apprentissage du langage. De plus, cette faculté semble constituer, pour l'équipe italienne, une étape importante dans l'apprentissage dans la mesure où cette progression observée lors du passage d'énoncés de 1 à 2 symboles se produit indifféremment à la culture et au langage ambiant de l'enfant. Les auteurs vont jusqu'à parler d'un « trait universel d'apprentissage du langage », dans la modalité parlée aussi bien que signée.

Depuis McNeill (1992), nous savons que geste et parole forment chez l'adulte un système unifié, respectant deux principes : la cohérence sémantique et la synchronie temporelle. En effet lorsque les gens parlent, ils produisent une variété de types gestuels, tels que les gestes iconiques, métaphoriques, les *beats*, les déictiques, etc. Chacun de ces types s'accompagne d'un type caractéristique de parole. L'exemple le plus frappant vient des gestes iconiques qui entretiennent une relation transparente ou motivée avec les idées qu'ils véhiculent et s'accompagnent d'énoncés qui représentent des objets et des événements, remplissant une fonction narrative. Ces événements geste-parole présentent également la caractéristique d'être synchrones : McNeill (1992) trouve que 90% des gestes chez les adultes étaient produits lorsque la personne qui fait ces gestes parle. De plus, le geste et le segment linguistique véhiculant la même information semblent être co-temporels. Cette synchronie impliquerait alors que le locuteur véhicule le même sens dans les deux canaux à sa disposition, indiquant de ce fait que le système est unique et intégré.

Butcher et Goldin-Meadow (2000) se sont alors interrogés pour savoir s'il en était de même chez l'enfant, et à partir de quel âge on pouvait trouver des traces de cette unification temporelle. Ils observèrent alors 6 enfants longitudinalement lors de la transition de la parole à 1 mot vers la parole à 2 mots. Au début de la parole à 1

mot, les enfants semblent utiliser des gestes sans forcément qu'ils soient accompagnés par la parole. Toutefois, lorsque les enfants sont dans la fin de la période de production de parole à 1 mot, les gestes étaient principalement utilisés en co-occurrence avec la parole. De la même façon, lors des premières productions geste-parole, le geste ne semblait pas être synchrone avec la parole. Ces découvertes suggèrent alors qu'au début de la période à 1 mot, geste et parole ne semblent pas tout à fait constituer un système unifié. En ce qui concerne le critère de cohérence sémantique, les auteurs notent que les combinaisons geste-parole véhiculant une même information et celles véhiculant une information différente, mais reliée, tendent à augmenter à travers le temps. Pour 5 enfants, les auteurs observent une convergence entre plusieurs événements : (1) la proportion de communications avec geste seul déclinent à travers le temps ; et (2) la proportion de combinaisons synchronisées geste-parole augmente à travers le temps.

Les auteurs proposent donc que l'intégration geste-parole, telle qu'observée chez l'adulte, paraît être absente au début de la parole à 1 mot et se développe plus tard lors de cette même période à 1 mot. Lorsque les enfants commencent à combiner des gestes avec des mots et ils commencent aussi à synchroniser leurs gestes avec la parole. La découverte la plus frappante dans ces résultats pour les deux auteurs réside dans le fait qu'au cours du développement de l'enfant, une compétence motrice permettant à la main et à la bouche de travailler ensemble va se développer, une compétence qui sera temporellement liée aux changements dans le système sémantique.

### **II.12.2. Quand la non-redondance du geste et de la parole annonce l'énoncé à 2 mots**

Dans Goldin-Meadow et Butcher (2003), la question était de considérer si on pouvait considérer que les combinaisons geste-mot non redondants forment un « pont transitionnel » vers la parole à 2 mots, avec pour soutien à cette prédiction le fait que geste et parole fonctionne comme un système unifié. Lors de la période à 1 mot, la plupart des combinaisons sont constituées d'un geste apportant une information redondante de la parole. Mais les locuteurs peuvent également produire des combinaisons où le geste véhicule une information différente de la parole, ce qui permet à l'enfant d'exprimer 2 éléments différents dans un seul acte communicatif.

Dans cette étude, Goldin-Meadow et Butcher s'appuient sur les 6 enfants étudiés dans Butcher et Goldin-Meadow (2000), suivis entre 12 et 27.5 mois. Lors des premières sessions filmées, les gestes ne paraissent pas être pleinement intégrés à la parole car ceux-ci apparaissent sans la parole, ne formant pas un système unifié sémantiquement et temporellement. Cette relation tend à changer lors que les enfants commencent à produire des gestes avec des mots significatifs qui véhiculent une information similaire à celle du geste. En effet, lors de ce changement, les auteurs observent que les combinaisons geste seul chutent et que les combinaisons geste-parole synchrones augmentent à mesure que les gestes se combinent avec des mots véhiculant le même sens. Les combinaisons geste-parole véhiculant une information différente sont toujours précédées chez les 6 enfants étudiés par des combinaisons véhiculant une information similaire. Lors des premières combinaisons geste-parole avec information différente, le geste tend à indiquer un objet jouant un rôle sémantique dans une proposition, tandis que la parole réfère à un autre objet, exprimant ainsi 2 éléments qui entretiennent des relations sémantiques l'un par rapport à l'autre. En ce qui concerne le timing de ces combinaisons, les auteurs relèvent le fait que la vocalisation semble survenir sur le stroke ou pic du geste et que dès lors que le geste et la parole sont intégrés, les combinaisons geste-parole tendent à être synchrones, que l'information véhiculée soit similaire ou différente d'une modalité à l'autre.

A la lumière des résultats de Goldin-Meadow et Butcher (2003) sur les combinaisons geste-parole au début de la parole à 2 mots, les enfants qui produisent des combinaisons geste-parole véhiculant une information différente, possèdent clairement la faculté de véhiculer 2 éléments à l'intérieur même d'un seul acte communicatif, et ils font ceci bien avant de produire une combinaison à 2 mots. Il existe alors une corrélation forte et fiable (.82) entre l'âge de début de ce type de combinaison geste-parole et l'âge de début des combinaisons à 2 mots, comme le montre la figure II.38 ci-dessous tirée de leur article :

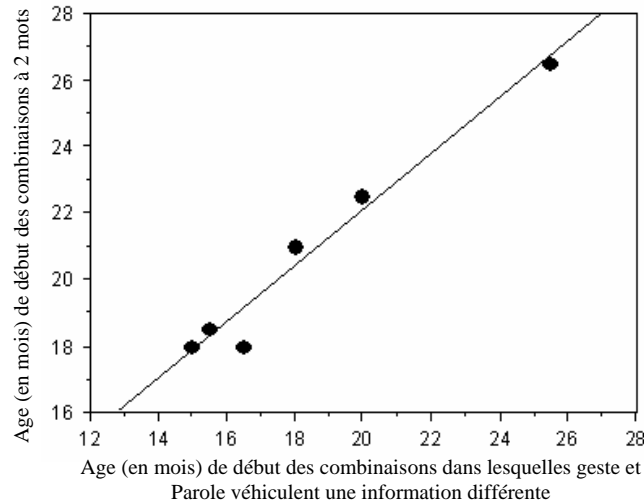


FIGURE II.38 : La figure présente l'âge auquel chaque enfant a produit ses premières combinaisons à 2 mots en fonction de l'âge auquel l'enfant commençait à produire des combinaisons dans lesquelles geste et parole véhiculaient une information différente (D'après Goldin-Meadow et Butcher, 2003).

Les enfants qui étaient les premiers à produire des combinaisons où geste et parole véhiculent une information différente sont également les premiers à produire des combinaisons à 2 mots. De plus, ce qui est très intéressant dans ces résultats réside dans le fait que cette corrélation soit spécifique à ce type de combinaisons geste-parole à information différente, puisqu'il n'y avait pas de corrélation entre l'âge de début du type de combinaison geste-parole véhiculant la même information et l'âge de début des combinaisons à 2 mots, comme nous pouvons le voir sur la figure II.39 suivante :

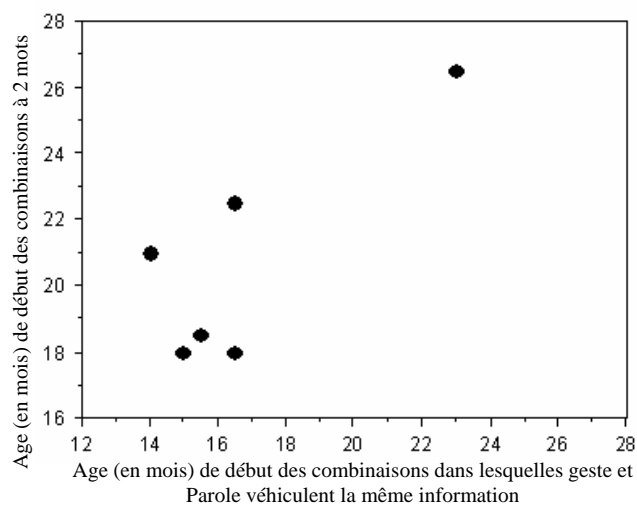


FIGURE II.39 : La figure présente l'âge auquel chaque enfant a produit ses premières combinaisons à 2 mots en fonction de l'âge auquel l'enfant commençait à produire des combinaisons dans lesquelles geste et parole véhiculaient la même information (D'après Goldin-Meadow et Butcher, 2003).



Ce qui est donc surprenant, ce n'est pas simplement le fait que le début des combinaisons dans lesquelles geste et parole véhiculent une information différente précède le début des combinaisons de parole à 2 mots, mais c'est surtout le fait que ce soit réalisé de façon tout à fait prédictible. Ainsi, lors de l'apprentissage précoce du langage, la relation entre geste et parole –particulièrement la relation entre geste-parole avec information différente et parole à 2 mots– paraît être annonciatrice de changements à venir dans le système linguistique de l'enfant.

L'examen de cette relation si révélatrice entre combinaisons geste + mot et combinaisons mot + mot a été reprise récemment par Iverson et Goldin-Meadow (2005). Elles mettent à nouveau en évidence le fait que chez les enfants testés, les premières combinaisons apparaissent plusieurs mois avant les secondes. De plus, tous les enfants produisaient des combinaisons geste + mot *complémentaires* (redondantes) entre 4 et 5 mois avant d'entrer dans les combinaisons à 2 mots (Figure II.40 a), et des combinaisons geste + mot *supplémentaires* (non-redondantes) environ 2 mois avant les combinaisons à 2 mots (Fig. II.40 b). Par ailleurs il existe une corrélation significative (.94) entre l'âge de début des combinaisons geste + mot supplémentaires et l'âge de début des combinaisons à 2 mots, indiquant ici des stades de développement bien distincts. Nous pouvons donc en conclure que c'est la faculté même à combiner deux éléments sémantiques différents (combinaisons supplémentaires) dans un seul acte communicatif qui va prédire le début de la parole à 2 mots.

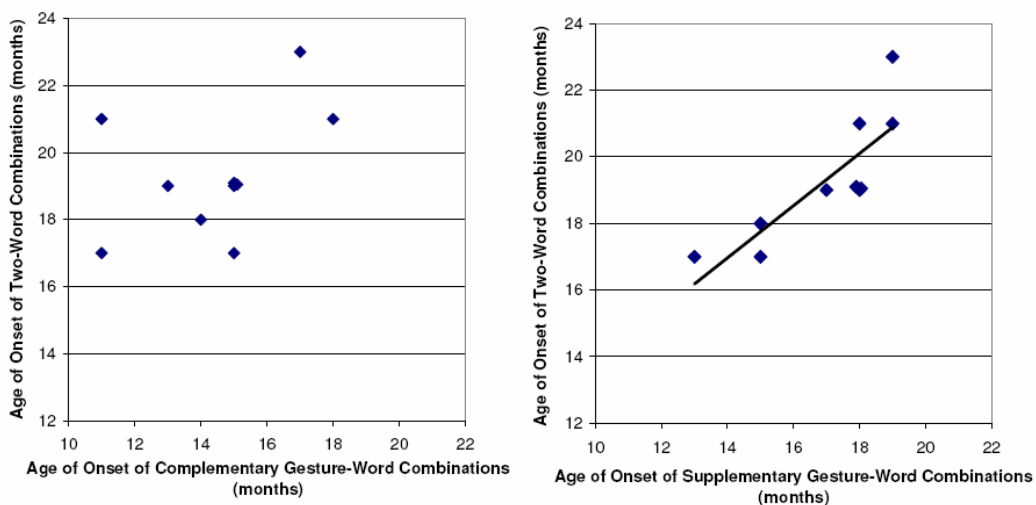


FIGURE II.40 : a) La figure de gauche montre l'absence de corrélation entre l'âge de début des combinaisons geste + mot complémentaires et l'âge du début des combinaisons à deux mots ; b) La figure de droite présente la corrélation entre l'âge de début des combinaisons geste + mot supplémentaires et l'âge du début des combinaisons à deux mots (D'après Iverson et Goldin-Meadow, 2005).

Plusieurs hypothèses sont formulées par Iverson et Goldin-Meadow (2005) pour expliquer le rôle même du geste dans l'apprentissage du langage. Une première explication soulignerait le rôle de celui-ci dans l'interaction avec un autre individu : il serait alors un déclencheur pour indiquer à l'autre que l'enfant est prêt à recevoir le message linguistique qui correspond à la requête du geste. Une autre explication impliquerait l'effet du geste sur l'enfant lui-même : le geste exploiterait des ressources différentes de celles utilisées pour la parole, et de ce fait les significations se prêtant à la représentation visuo-spatiale pourraient être exprimées plus aisément par le geste que par la parole.

Nous venons donc de voir une fois encore que l'âge auquel les enfants produisent leurs premières combinaisons geste-parole véhiculant une information différente est hautement corrélé avec l'âge où ils commencent à produire leurs premiers énoncés à 2 mots, corrélation que l'on ne retrouve pas entre combinaisons où geste et parole apportent une information redondante et début des énoncés à 2 mots. Il s'agit donc bien de la relation entre geste et parole, et non pas du geste *per se*, qui prédit quand l'enfant va débiter ses combinaisons multi-mots.

Özçalışkan et Goldin-Meadow (2005) ont poussé l'étude de ces combinaisons geste-parole à information différente encore plus loin, en émettant l'hypothèse que si ces combinaisons sont des précurseurs des constructions linguistiques, alors on devrait s'attendre à ce que les enfants produisent des combinaisons de type argument+argument ou prédicat+argument parmi les deux modalités avant de produire ces combinaisons dans la parole seule. Leur prédiction repose sur l'idée que l'enfant devrait alors être capable de réaliser des constructions complexes à 2 prédicats dans les combinaisons geste-parole avant la parole seule. Pour répondre à cette question, les auteurs ont testé 40 enfants anglais filmés à la maison, de la période à 1 mot jusqu'à la période à 2 mots, c'est-à-dire ici à 14, 18 et 22 mois. Les auteurs dégagèrent des données 3 catégories d'actes communicatifs : (1) geste seul, (2) parole seule et (3) combinaison geste-parole. La catégorie « combinaison geste-parole » est à son tour subdivisée en 3 types de relation : (i) relation de renforcement

(geste et parole véhiculent la même information) ; (ii) relation de désambiguïation (le geste clarifie le référent d'un mot pronominal, démonstratif ou déictique) ; et (iii) relation supplémentaire (le geste ajoute une information sémantique au message véhiculé par la parole). Et enfin, les auteurs ont focalisé sur les combinaisons geste-parole supplémentaires multi-mots qu'ils catégorisent en trois types : (1) arguments multiples sans prédicat ; (2) un prédicat avec au moins un argument ; et enfin (3) prédicats multiples avec ou sans arguments.

Les résultats font apparaître, qu'avec l'âge, les enfants produisent plus d'actes communicatifs contenant de la parole, et plus de types de mots différents. Cette augmentation est significative non seulement entre 14 et 18 mois, mais aussi entre 18 et 22 mois. La majorité des enfants produit des mots seuls à 14 mois et plus de la moitié commence à produire de la parole à 2 mots dès 18 mois.

En ce qui concerne le geste, avec l'âge, les enfants produisent plus d'actes communicatifs contenant des gestes, plus de marques gestuelles et plus de combinaisons geste-parole. L'augmentation est significative entre les deux périodes d'âge similairement à la parole. A 14 mois, la moitié des enfants produisent des combinaisons geste-parole, tandis qu'à 18 mois, tous (sauf un) produisent ces mêmes combinaisons. Lorsque l'on s'intéresse plus particulièrement aux types de combinaisons produites, on observe que les enfants produisent significativement plus de combinaisons de renforcement que de combinaisons supplémentaires, mais plus de combinaisons supplémentaires que de combinaisons de désambiguïation.

En ce qui concerne le type d'information sémantique véhiculé dans les combinaisons geste-parole supplémentaires, à 14 mois, les enfants produisent très peu de combinaisons argument+argument ou prédicat+argument. A 18 mois, plus de la moitié des enfants réalisent ces constructions, mais ils les produisent avec le geste et la parole, et pas encore dans la parole seule. Les combinaisons de type prédicat+prédicat ne sont pas du tout réalisées à 14 mois, et seuls 3 enfants les produisent à 18 mois. Néanmoins, à 22 mois, plusieurs enfants les réalisent en gestes avec parole, mais pas encore dans la parole seule.

Les enfants ne produisent donc pas d'énoncés à 2 arguments ou 1 argument+1 prédicat dans la parole avant 22 mois. Toutefois nombre d'entre eux les réalisent dès 18 mois dans les combinaisons geste+parole. Peu d'enfants produisent des combinaisons prédicat+prédicat dans la parole à 22 mois, celles-ci étant plutôt

réalisées dans les combinaisons geste+parole. Le fait que ces combinaisons surviennent dans un premier temps sous les deux modalités laisse à penser qu'ils sont encore incapables de les produire dans la parole, bien qu'ils soient déjà capables de comprendre les relations qui lient arguments et prédicats dans leur langue. Les gestes fournissent donc un outil fiable à l'enfant pour étendre son répertoire communicatif, lui permettant de véhiculer de plus en plus d'idées complexes, bien avant de pouvoir le faire en parole seule. Les auteurs se sont alors demandés ce qui pouvait entraîner ce retard dans la parole. Ils argumentent que cela pourrait venir du fait que l'information véhiculée par la main requiert moins de charge cognitive que la voix. Moins de contraintes pèseraient sur la mémoire en utilisant le geste. Ceci semble valable uniquement pour les gestes déictiques : un geste de pointer de l'index possède le bénéfice, comme unité-outil, de ne pas varier de forme selon le référent associé, et il est par conséquent plus facile à réaliser physiquement, mais également plus aisé à être mis en mémoire.

Selon Goldin-Meadow et Butcher (2003), l'étude des premières combinaisons à 2 mots est vraiment pertinente car elles reflètent, dans le développement de l'enfant, sa faculté à exprimer une information propositionnelle à l'intérieur d'un seul et même acte communicatif. Mais leur importance réside également dans le fait que ces combinaisons sont un premier pas vers la syntaxe, dans la mesure où elles gabarissent l'ordre particulier des mots selon le modèle de sa propre langue. Ainsi, cette découverte se trouve être très utile pour traiter une population clinique, puisque les enfants qui montrent un retard de la parole à 2 mots tombent naturellement dans 2 catégories, selon Goldin-Meadow et Butcher (2003) : (i) ceux qui réalisent éventuellement la parole à 2 mots plus tard dans le développement par rapport à l'âge normal (*late bloomers*) ; et (ii) ceux qui continuent à avoir de graves difficultés avec le langage parlé et qui ne sont jamais capables de combiner les mots dans un seul ensemble.

Mais nous allons voir que ce n'est pas le seul bon indicateur de ce changement : le fait que les explications soient concordantes ou discordantes entre geste et parole dans des tâches de type piagétienne, est également annonciateur de changements dans les capacités de l'enfant.

## **II.12.3. Concordance/discordance geste-parole comme index de transition**

### **II.12.3.1. Les gestes dans l'acquisition de la conservation de quantités**

Dès les années 80, Church et Goldin-Meadow (1986) ont tenté de définir le rôle que pouvaient jouer des discordances entre geste et parole chez 28 enfants âgés de 5 à 8 ans, dans des tâches de conservation piagésiennes typiques (quantité de liquide, longueur, nombre). Ces tâches se divisaient en 3 étapes : une égalité initiale, une transformation et une égalité finale. A la fin de ces étapes les auteurs ont étudié les explications produites par les enfants et ont pu les classer en 3 groupes : (i) les enfants conservants (ils trouvent la même égalité avant et après transformation) ; (ii) les enfants conservants partiels (ils trouvent parfois la même égalité, parfois des différences) ; (iii) et les enfants non-conservants (ils répondent que c'est différent aux questions de jugement). Lors des explications, les enfants conservants tendaient à produire des explications équivalentes tandis que les enfants qui échouent à conserver le faisaient rarement. Tous les enfants, sauf un, produisaient lors de leurs explications des gestes en plus des explications verbales. Ces combinaisons geste-parole pouvaient être de nature concordante (véhiculant la même information) ou discordante (véhiculant chacun une information différente). L'analyse montre que les enfants qui produisent une large proportion de réponses avec des discordances geste-parole dans leur explication étaient moins cohérents dans leurs réponses vocales aux questions de conservation, comparé aux enfants qui tendent à être plus concordants. Le rôle même du geste utilisé par les enfants discordants était de nature différente : (i) il pouvait compléter la parole (stratégie de compensation), (ii) il pouvait surpasser la parole ; et (iii) il apportait moins d'information que la parole (par exemple une explication d'équivalence dans la parole, et une explication d'équivalence différente dans le geste).

Church et Goldin-Meadow (1986) en dégagent donc le fait que, d'une part, les enfants dits « concordants » expriment un niveau de raisonnement dans leurs gestes comparable à celui exprimé dans leurs explications verbales ; tandis que les enfants dits « discordants » tendent à exprimer un niveau plus haut de raisonnement dans leurs réponses. Pour ces derniers c'est dans leur composante gestuelle que cette meilleure performance s'exprime, ce qui semble indiquer qu'ils possèdent une

compréhension des éléments centraux de l'explication d'équivalence, mais qu'elle n'est pas encore intégrée dans leurs explications verbales. Les éléments d'information ne semblent ainsi pas encore totalement consolidés dans un seul système explicatif cohérent, et de ce fait, les auteurs prédisent que l'instruction nouvelle donnée par l'expérimentateur sur la tâche devrait être bénéfique à ces enfants discordants, leur permettant de consolider leur connaissance de l'équivalence. C'est ce que les auteurs ont testé sur des enfants concordants et discordants dans une seconde étude, avec une phase de pré-test, suivie d'un entraînement, puis d'une phase post-test. L'entraînement pouvait être soit une condition d'instruction, ou une condition de manipulation. Leurs résultats montrent que pour les deux conditions d'entraînement, les enfants discordants présentaient une plus grande amélioration dans les explications que les enfants concordants, et qu'après ces entraînements, les enfants discordants avaient significativement plus de chances d'ajouter une nouvelle explication d'équivalence à leur répertoire que les concordants.

Suite à ces études, Church et Goldin-Meadow (1986) ont pu argumenter que le geste pouvait être utilisé comme un index des connaissances déjà présentes chez l'enfant, ou comme un index permettant de savoir à quel degré de cohérence ses connaissances sont ancrées. Ainsi, la discordance entre geste et parole constitue un index de la cohérence du système explicatif à la base de la compréhension de l'enfant pour un concept donné. Mais elle fournit également un index de connaissance transitionnelle, par le fait qu'elle indique si l'enfant est prêt ou non à utiliser une instruction pour un concept. En d'autres termes, seuls les enfants discordants sont réellement capables de bénéficier d'une instruction donnée et d'ajouter des explications d'équivalence à leur répertoire, ce qui prouve qu'ils sont déjà prêts et savent déjà beaucoup de choses (même implicitement) sur l'invariance de la longueur, du nombre ou de la quantité de liquide par rapport à l'étape de transformation. Cette connaissance implicite d'équivalence est alors reflétée dans la composante gestuelle et l'expérience via l'entraînement aide à rendre cette connaissance explicite. L'hypothèse générale que posent alors les deux auteurs réside dans l'idée que la discordance entre geste et parole peut être une caractéristique générale du nouvel apprenant, qui va s'attaquer à un concept particulier et qui n'a jusque-là qu'une compréhension incohérente de ce concept,

mais qui se montre prêt à recevoir un entraînement afin de rendre cette connaissance explicite.

### **II.12.3.2. Le geste dans l'acquisition de l'équivalence mathématique**

L'équipe de Goldin-Meadow a pu renforcer cette théorie quelques années plus tard dans une nouvelle étude (Alibali et Goldin-Meadow, 1993), en testant une fois de plus la « disparité » entre geste et parole dans le but d'explorer les processus cognitifs caractérisant la transition d'une compréhension incorrecte vers une compréhension correcte d'un problème. Selon les auteurs, l'instabilité dans laquelle l'enfant va se trouver est le témoin d'un changement imminent dans ses capacités. Cette « disparité » entre geste et parole est vue ici comme fournissant l'évidence que de multiples hypothèses sont considérées pour la résolution d'un seul problème (alors que l'enfant ne cite en général qu'une seule hypothèse dans son explication verbale, qui n'est donc pas révélatrice de toutes les autres hypothèses activées). Ainsi, l'enfant qui produit une discordance active en fait deux hypothèses différentes simultanément, et c'est cette concurrence que les auteurs prennent comme étant la preuve même d'un état de transition. Pour tester cette hypothèse, ils menèrent une étude sur 90 enfants âgés de 10 ans en moyenne, divisés en trois groupes testés chacun dans une condition qui était : (i) soit de résoudre et expliquer une série de 12 problèmes d'addition pendant une session d'entraînement, sans recevoir ni instruction, ni feedback sur ces problèmes ; (ii) soit la même condition, mais avec instruction et feedback après chacun des 12 problèmes ; ou bien (iii) ces 12 problèmes d'addition, suivis d'une généralisation de leurs connaissances à une nouvelle opération constituée de 6 problèmes de multiplication, et avec feedback. L'expérience consistait en un pré-test pour évaluer la compréhension d'équivalence mathématique, suivi d'une des trois conditions, puis du post-test. Les enfants étaient ensuite testés à nouveau deux semaines après. Les auteurs ont alors pu observer que les enfants qui réalisent un état concordant correct en passant par une transition discordante sont capables de généraliser la connaissance gagnée de l'addition à la multiplication. De plus ils maintiennent cette connaissance sur une période de deux semaines, contrairement aux enfants qui ne passent pas par cet état de discordance. La discordance semble donc développer une compréhension plus profonde de l'équivalence mathématique. De plus, l'augmentation de la taille du répertoire associé

à l'entrée dans un état discordant paraît être principalement attribuable à une augmentation dans le nombre de procédures qui sont produites dans le geste, et non pas dans la parole. Ainsi lorsque les enfants progressent dans leur acquisition d'équivalence mathématique, leurs répertoires augmentent à l'entrée dans l'état discordante, puis chute lorsqu'ils sortent de ce même état.

Les résultats d'Alibali et Goldin-Meadow (1993) apportent donc un fort soutien à l'idée d'un modèle d'apprentissage, dans lequel un enfant débute par une hypothèse incorrecte simple, avant d'entrer dans une période de transition nourrie de multiples hypothèses, pour déboucher au final sur une seule hypothèse correcte. Et la discordance entre geste et parole observée est justement causée par les processus qui caractérisent la connaissance transitionnelle.

Goldin-Meadow (2002) apporte également un soutien à l'idée d'un rôle plus direct du geste dans le changement cognitif de l'enfant. D'un côté, si l'on envisage le geste et la parole comme un seul système unifié à la McNeill, alors on s'attend à ce que les gestes réduisent la demande sur les ressources cognitives du locuteur, libérant ainsi la capacité cognitive pour réaliser une autre tâche. Mais à l'opposé, si le geste accompagné de la parole requiert une planification, une exécution et une coordination de deux systèmes cognitifs moteurs séparés, alors on s'attend à ce que les gestes augmentent la charge cognitive du locuteur. Goldin-Meadow (2002) a testé ces prédictions en demandant à des enfants de résoudre un problème de mathématiques, alors qu'il devait simultanément se souvenir d'une liste de mots lors de l'explication de la résolution du problème. Les résultats de l'expérience ont montré que les enfants se souviennent de plus de mots lorsqu'ils font des gestes. Ainsi, le geste paraît en quelque sorte alléger les ressources cognitives du locuteur lors de la tâche d'explication, afin d'optimiser ces capacités pour garder les mots en mémoire.

#### **II.12.4. Relation entre gestes déictiques et mots représentationnels**

Pizzuto et al. (2003) explorent en détail la relation geste-parole chez l'enfant en considérant l'utilisation de gestes, de vocalisations et de mots déictiques vs. représentationnels chez 4 enfants italiens observés longitudinalement de 10-12 mois à 24-25 mois, soit aux deux périodes clés que sont la parole à 1 mot et la parole à 2 mots. Les auteurs pensent que pour savoir précisément si la relation geste-parole



s'unifie à travers le temps, il est alors nécessaire de prendre en compte les types de mots et gestes utilisés dans les premiers énoncés d'enfants. Et spécifiquement, il convient de distinguer les éléments dits représentationnels ou chargés de contenu des éléments dits déictiques. Leurs résultats font apparaître que tous les enfants produisent un geste et des énoncés geste-vocalisation ou geste-mot avant ou en même temps que les énoncés à un mot. Tous produisaient des énoncés geste-vocalisation/mot 4 à 8 mois avant les énoncés mot-mot. Très peu de combinaisons geste-geste ont été trouvées parmi les enfants, elles ont donc été exclues de l'analyse. Les données longitudinales de Pizzuto et al. montrent que les deux modalités semblent chargées de sens et intégrées dans le temps dès les premières étapes du langage, soutenant ici encore la vision de McNeill (1992).

Toutefois, selon ces auteurs, la modalité gestuelle et la modalité vocale ne paraissent pas contribuer de la même façon à l'articulation du sens dans les premiers énoncés à 1 ou 2 éléments, et l'intégration geste-parole en tant que telle semble différer de celle de l'adulte. En effet, les éléments représentationnels et déictiques ne sont pas distribués de manière comparable selon les deux modalités et dans les différents types d'énoncés.

Dans la modalité gestuelle, il existe une prévalence très nette pour les éléments déictiques vs. représentationnels, un patron qui n'est pas observé dans la communication adulte. Les auteurs soulignent alors la pertinence de la deixis gestuelle comme un mécanisme primaire dans le développement précoce du langage. A l'inverse, dans la modalité vocale, une prévalence pour les éléments représentationnels est marquée, particulièrement dans la production d'énoncés à 1 élément, et plus significativement dans les énoncés à 2 éléments codant une information supplémentaire, où les enfants combinent plus fréquemment 2 mots représentationnels plutôt que 2 gestes ou 1 mot représentationnel et 1 geste représentationnel.

Cette étude vient appuyer des études réalisées par les collègues de Pizzuto quelques années plus tôt (Caselli et al., 1984, Caselli et Volterra, 1990) qui ont démontré qu'alors que les enfants possédaient les gestes déictiques dans le répertoire à 16 et 20 mois, la même chose n'était pas vraie pour les mots déictiques. La deixis gestuelle précède la deixis vocale dans le répertoire de la moitié des enfants étudiés par ces auteurs, et les gestes déictiques ne sont pas supplantés plus

tard par les mots déictiques, puisqu'ils continuent à être présents dans les répertoires de tous les enfants à 20 mois. De plus, l'utilisation exclusive d'éléments gestuels de type représentationnel semble constituer un phénomène marginal chez les enfants entendants exposés à un input linguistique vocal. Il existerait donc un rôle spécial pour les gestes déictiques, notamment le pointer de l'index, lors du développement de l'enfant. Ce rôle est encore plus évident dans les énoncés à deux éléments, combinant un geste de pointer avec un mot représentationnel, énoncés composant le type le plus productif utilisé par les enfants.

Cette découverte que la deixis gestuelle précède et supplante la deixis vocale est également citée dans l'étude de Capirci et al. (1996), qui argumentent cette avance en termes d'exigences sur les compétences cognitives et symboliques imposées sur l'enfant par la deixis vocale :

It is of interest to note that, while deictic gestures (especially pointing) were widely employed in children's utterances, deictic words were much less frequent in both complementary utterances and in supplementary combinations of a deictic with a representational word. In line with the observations made by Clark (1978), we interpret the disparity between gestural and vocal deixis as a product of the greater demands that deictic words impose upon children's growing symbolic and cognitive skills relative to deictic gestures. Deictic gestures, which appear very early in children's communicative repertoires, are more directly linked to reality than deictic words: they can be used in the same manner to point to objects, people or locations, without differentiating demonstrative and locative meanings, and the object or event to which they refer can be identified in the physical context of the child's utterance. (1996:669)

Voilà qui indique nettement le rôle prépondérant des gestes déictiques dans la mise en place de l'intégration geste-parole lors des premières étapes du développement du langage. Comparés aux adultes, les enfants semblent bien soumis à des contraintes restreignant l'utilisation d'éléments représentationnels.

## **II.12.5. Le geste comme outil d'apprentissage pour tous ?**

### **II.12.5.1. Sourds vs. entendants**

Le rôle prépondérant du geste chez l'enfant est soutenu par Goldin-Meadow (1999, 2002), qui voit dans cette composante gestuelle une façon de construire la communication *ab ovo*. En effet, des enfants sourds nés de parents entendants, soumis à une méthode d'apprentissage de la langue orale sans exposition à une langue des signes conventionnelle, persistent à utiliser spontanément des gestes « faits maison » pour communiquer. Ce qui est étonnant ici, c'est le fait que les constructions que les enfants utilisent dans leur système gestuel possèdent les propriétés du langage, même façonnées sans bénéfice d'un

input linguistique : c'est ce que l'on peut nommer les propriétés « résilientes » du langage. Ainsi les « mots-gestes » ou « phrases-gestes » générés par ces enfants possèdent tous deux les propriétés trouvées dans les langues naturelles (comme les constructions ergatives) et ils sont de plus utilisés pour des fonctions centrales comme des requêtes, commentaires pour eux-mêmes, ou à des fins métalinguistiques, ou encore pour la narration. Les études menées ont par ailleurs montré que ces systèmes gestuels sont des systèmes créés de façon spontanée et totalement indépendante des gestes de leur mère. Ils semblent prendre comme point de départ les gestes que les gens entendants réalisent pour les transformer ensuite en un système beaucoup plus proche d'un langage.

Dans la situation où les enfants sont exposés à un modèle de langue conventionnel, comme les enfants entendants, ils vont également faire des gestes, même avant de produire des mots, comme nous l'avons vu précédemment. Ces productions gestuelles sont alors vues ici comme reflétant la gamme de sens que l'enfant est capable de véhiculer. Une bonne façon de refléter ce phénomène se situe dans les premières étapes de l'apprentissage du langage, et par le fait que les enfants peuvent exprimer deux éléments sémantiques distincts, un élément dans le geste et un élément dans la parole pour un seul et même système, donnant lieu à des combinaisons dites discordantes qui sont révélatrices des combinaisons futures à deux mots, ce que nous allons voir par la suite plus en détail.

Volterra et al. (in press) présentent également un ensemble de données tirées de différentes études menées au sein de leur laboratoire romain et qui portent sur l'étude des gestes et de la parole chez des enfants sourds (vs entendants). Il apparaît que lors des premières étapes, les enfants sourds et entendants utilisaient seulement les gestes déictiques, tandis que les gestes référentiels, les signes, et les mots n'apparaissent que dans une phase ultérieure, conformément à ce que nous avons cité précédemment. Les deux groupes d'enfants commençaient de plus à combiner deux signes ou deux mots dans un seul énoncé au même âge, autour de 17-18 mois, lorsque leurs vocabulaires de signes ou mots comprenaient environ de 20-40 items distincts. Les enfants entendants et parlants combinent 2 mots référentiels au même stade de développement symbolique que les enfants sourds et signants combinant 2 signes référentiels.

En résumé, lorsque les enfants sont exposés à un input linguistique conventionnel, ils l'utilisent, ceci comme un moyen robuste de construire leur pensée en dehors des bornes du langage conventionnel. Et lorsque les enfants ne sont pas exposés à un input linguistique conventionnel, ils se reposent en quelque sorte sur les gestes afin de combler ce manque, en inventant des systèmes gestuels qui possèdent les caractéristiques des langues naturelles. A partir de modèles gestuels « imparfaits » fournis par les personnes entendantes, ils créent un système linguistique plus élaboré en sortie. Ce processus de transformation d'un modèle imparfait en un modèle élaboré illustre parfaitement, selon Goldin-Meadow (2002) le fait que les enfants ne soient pas des destinataires passifs des systèmes qu'ils utilisent pour communiquer. Ils sont ainsi capables de construire leur propre système de communication, et ils peuvent utiliser leurs mains pour le faire.

Nous pouvons également citer de Capirci et al. (2002) une étude portant sur la communication spontanée d'un enfant italien, entendant, né de parents sourds, bilingue, qui s'est trouvé exposé à la langue parlée et à la langue des signes italienne dès la naissance. Les chercheurs ont mis en évidence le fait que les premières communications de cet enfant consistaient principalement en gestes, découverte cohérente avec les données sur les enfants entendants. L'acquisition de nouveaux mots et de nouveaux signes paraissait plus lente. Elle fut suivie d'une période de croissance rapide, en premier sur les mots entre 19 et 22 mois ; puis dans le vocabulaire des signes autour de 25 mois. L'énoncé à deux mots a émergé en premier dans la production de cet enfant vers 16 mois et son utilisation a augmenté significativement dès 25 mois, âge auquel apparaissent les énoncés à deux signes, lesquels ont dépassé les énoncés à deux mots à 29 mois. En résumé, la production gestuelle et la production verbale de cet enfant bilingue à 16 et 20 mois reste comparable en tous points à celles des enfants monolingues exposés à la parole. La taille du vocabulaire et la productivité verbale et manuelle se rangent dans la même gamme que celles des enfants exposés uniquement à la parole. Il n'existe ainsi pas d'avantage lié à l'utilisation des signes chez cet enfant exposé également à l'input linguistique signé et parlé.

Le lien entre geste et langage paraît si fort et robuste que des enfants sourds qui n'ont pas été exposés à une langue des signes formelle vont créer un système gestuel pour communiquer. Leurs gestes expriment des items lexicaux

simples et des relations sémantiques similaires à celles de formes parlées d'enfants entendants. Leur système prend spontanément les caractéristiques linguistiques, comme le marquage morphologique des noms et verbes (Goldin-Meadow et al., 2004). Comme Goldin-Meadow et Morford (1985) l'observent : « la communication chez les humains est un phénomène inévitable : lorsque nous l'empêchons d'arriver par la bouche, elle émane presque irrésistiblement des doigts. » (p.146).

### **II.12.5.2. Voyants vs. non-voyants**

Comme mis en évidence par Bates dès les années 70 et comme nous venons de le montrer par différentes études, les enfants en période prélinguistique utilisent les gestes, particulièrement les gestes de pointer de l'index, plusieurs mois avant d'utiliser les mots pour référer aux objets. Cette utilisation du geste perdure même après la mise en place des premiers mots. L'étude précédente réalisée par Goldin-Meadow et Butcher (2003), appuyée par une étude antérieure (Butcher et Goldin-Meadow, 2000) a mis en évidence le fait que le timing d'un certain type de combinaisons multimodales prédit avec précision le début des énoncés de parole à deux mots, le geste semblant alors jouer un rôle fondamental dans l'apprentissage du langage. Ce comportement prédictif, même deux fois corroboré (cf. supra), pourrait être simplement un comportement appris en regardant les autres bouger leurs mains lorsqu'ils parlent.

Or Iverson et Goldin-Meadow (1998) ont constaté que les locuteurs aveugles congénitaux réalisent des gestes en dépit d'un manque de modèle visuel, même lors de situations où ils s'adressent à d'autres partenaires aveugles. Cela semble donc indiquer que le système gestuel ne se construit pas sur l'apprentissage d'un modèle de geste fourni par les autres. Les auteurs ont testé cette alternative dans la communication spontanée de 12 enfants aveugles congénitaux âgés de 9 à 18 ans par comparaison avec un groupe de 12 enfants voyants. Il ressort de l'étude que les 12 enfants aveugles font des gestes lorsqu'ils parlent et qu'ils le font à un taux qui n'est pas différent des enfants voyants, en véhiculant la même information et en utilisant une gamme de gestes similaires. Ainsi la production gestuelle ne semble pas exiger une expérience préalable pour se réaliser, et ne constitue donc en aucun cas un processus d'imitation.

Dans une seconde expérience, Iverson et Goldin-Meadow (1998) testent les enfants aveugles afin d'observer leur production gestuelle dans une situation où ils s'adressent à un interlocuteur aveugle également. Là le résultat est une fois de plus probant : les enfants aveugles font autant de gestes, à un taux qui n'est pas différent de celui observé dans une communication entre voyants ou voyant et aveugle. Ces résultats mettent en évidence le caractère extrêmement robuste du geste dans la parole, puisqu'il ne requiert ni modèle, ni observateur. Le geste fait donc partie intégrale du processus de parole lui-même.

C'est à partir de ces mêmes réflexions que Iverson, Tencer, Lany et Goldin-Meadow (2000) ont eu l'idée de d'expérimenter cette prédiction en observant un groupe d'enfants qui sont sensés ne pas réaliser de gestes, des enfants aveugles congénitaux.

La question était donc de savoir si ces enfants aveugles faisaient des gestes lors des premières étapes de développement du langage, et si c'était le cas, si l'utilisation de ces gestes était conforme à celle réalisée par les enfants en développement normal ; enfin comment ces gestes affectaient l'apprentissage du langage chez l'enfant aveugle. Pour cela les auteurs ont recodé des données vidéos de 3 enfants aveugles à partir d'études menées par Andersen et collègues<sup>20</sup>, et de 2 enfants aveugles à partir d'une étude menée par Landau et Gleitman<sup>21</sup>, observés entre 14 et 28 mois, pour établir une comparaison avec 5 enfants voyants (Butcher et Goldin-Meadow, 2000). Les enfants voyants commencent à produire des combinaisons à deux mots de 18 à 22.5 mois, un âge typique. Deux enfants aveugles commencent à les produire dès 18 mois, tandis que les trois autres enfants aveugles ne le faisaient pas à cet âge et commençaient à les produire à 28-30 mois, 29 mois et 23.5 mois.

---

<sup>20</sup> Andersen E.S., Dunlea A., Kekelis L.S. (1984). Blind children's language: Resolving some differences, *Journal of Child Language*, 11, 645-664.

Andersen E.S., Dunlea A., Kekelis L.S. (1993). The impact of input: Language acquisition in the visually impaired, *First Language*, 13, 23-49.

Dunlea A. (1989). *Vision and the emergence of meaning*, Cambridge: Cambridge University Press.

Dunlea A., Andersen E.S. (1992). The emergence process: Conceptual and linguistic influences on morphological development, *First Language*, 12, 95-115.

<sup>21</sup> Landau B et Gleitman L.R. (1985). *Language and experience: Evidence from the blind child*, Cambridge, MA: Harvard University Press.

La production de parole semble ne pas être aussi homogène que chez les enfants voyants : certains enfants aveugles ne produisant que quelques mots ; d'autres plus de mots que les voyants ; ou alors simplement autant.

En ce qui concerne la production de gestes, nous avons vu que la proportion d'actes de communication avec gestes chez les enfants voyants reste haute jusqu'à 18 mois, avant de subir un rapide déclin. Il n'en est pas tout à fait de même chez les enfants aveugles, qui affichent deux patrons de comportement : (i) un des enfants se comporte plutôt comme l'enfant voyant, c'est-à-dire avec une grande proportion de gestes entre 14 et 17 mois, suivie d'un déclin rapide à 26 mois ; (ii) les 4 autres enfants ont une proportion d'actes de communication relativement bas. La vaste majorité des gestes produits par les aveugles étaient des gestes déictiques (73%), tout comme les voyants (82%), pour indiquer un référent ou attirer l'attention sur un référent. Il faut noter que les enfants aveugles tendent à plus utiliser la paume de la main plutôt que l'index pour établir la référence, ce qui peut être expliqué par l'absence de vision qui rend la référence plus imprécise, d'où l'utilisation de la paume, elle-même beaucoup moins précise que l'index. Il semble que les voyants utilisent les gestes plus fréquemment pour référer à des objets hors d'atteinte, tandis que les aveugles les utilisent pour référer à des objets à leur portée, plus proximaux. La communication des enfants aveugles peut alors sembler plus restreinte au contexte immédiat par l'absence de vision, contrairement aux enfants voyants qui disposent de plusieurs options disponibles pour communiquer sur des référents situés plus distalement.

Dans le développement, les enfants voyants changent leur mode de communication, allant d'une préférence pour le geste à une préférence pour la parole entre 16 et 20 mois (voir Figure II.41), âge auquel ils utilisent plus les mots que les gestes pour étiqueter les objets de leur environnement.

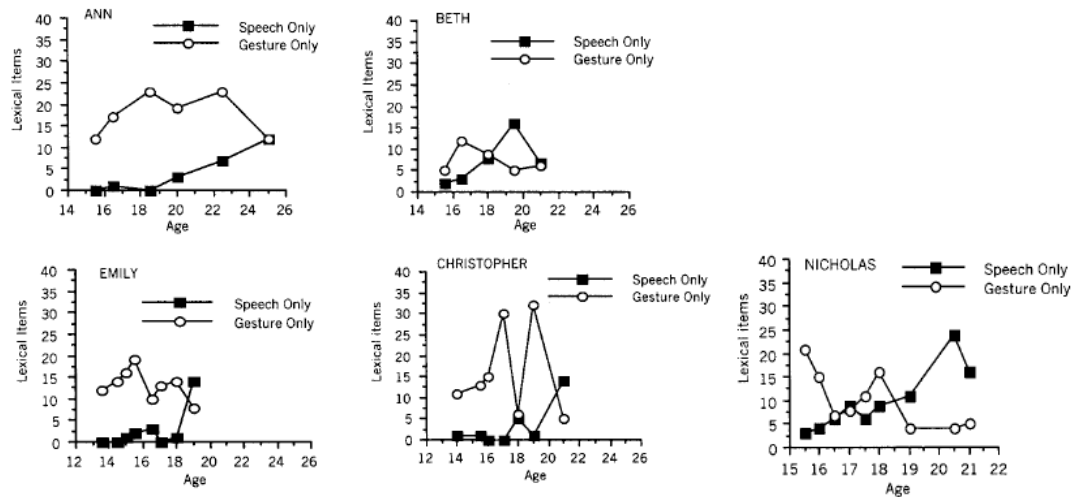


FIGURE II.41 : Nombre d'items lexicaux (en y) dans la parole seule et le geste seul produits par les enfants voyants en fonction de l'âge (en x). (D'après Iverson et al., 2000)

Les figures suivantes tirées de Iverson et al. (2000) permettent de mettre en évidence le fait que lors des premières sessions, les enfants voyants réfèrent aux objets en utilisant plus la modalité gestuelle que vocale, et cette tendance reste constante jusqu'à peu près 18 mois. Lors des dernières sessions, on observe à l'inverse que le nombre d'items présents dans la parole égale, voire surpasse, le nombre d'items dans le geste.

Les auteurs concluent de ces résultats que le geste peut remplir une fonction de « bootstrapping » dans le développement lexical, en procurant à l'enfant une façon de référer aux objets dans l'environnement, sans avoir encore la capacité de prononcer l'étiquette verbale correspondante. Le geste possède donc un rôle indéniable dans les premières étapes d'apprentissage du mot chez l'enfant voyant. Cette constatation se vérifie également pour les enfants aveugles, qui possèdent les mêmes types de gestes que les voyants, ce qui indique que les gestes émergent même lorsque aucun modèle visuel n'est disponible, constituant alors une composante plus que robuste dans l'apprentissage de la communication humaine. Certes les enfants aveugles utilisent les gestes à un degré moindre que les enfants voyants, et les utilisent dans un espace plus restreint. Mais cela peut s'expliquer par le fait que le geste, spécifiquement le geste déictique, est là pour tourner la tête de l'interlocuteur vers un objet d'intérêt tout en opérant un contrôle visuel par alternance du regard, et que ce dernier contrôle manque aux enfants aveugles, les poussant à se reposer davantage sur la parole.



### **II.12.5.3. Geste et parole chez les enfants à développement atypique**

Que pouvons-nous tirer des études menées sur geste et parole chez une population clinique d'enfants atteints de syndromes comme ceux de Williams ou de Down ? L'étude de ces deux pathologies est intéressante dans la mesure où elles semblent être en miroir l'une de l'autre. En effet, les enfants atteints du syndrome de Down (SD) présentent des problèmes dans l'acquisition du langage liées à des questions de morphologie et de syntaxe. Tandis que les enfants atteints du syndrome de Williams (SW) paraissent avoir une commande inhabituelle du langage : leur compréhension semble être plus limitée que le langage expressif, mais ce langage tend à être grammaticalement correct, complexe et fluide au niveau superficiel.

Plusieurs études italiennes relatées par Volterra et al. (sous presse) font apparaître que les enfants atteints de SD possèdent des répertoires action+geste significativement plus grands que les enfants en développement typique, et qu'il semble exister un « avantage gestuel » chez ces enfants. Capone et McGregor (2004) affirment également que la production gestuelle est une « force » pour les enfants atteints du SD. Ces enfants pourraient alors compenser des facultés réduites de production de langage par une plus grande production gestuelle. D'autres études citées apportent l'idée qu'il existe un lien fort entre geste et langage chez ces enfants : les enfants atteints de SD produisent des quantités similaires de gestes et mots, ils combinent ces deux modalités avec une fréquence comparable aux enfants en développement typique. Toutefois, lorsqu'ils les combinent, ils le font principalement de façon redondante sur le plan informatif : ils réalisent des combinaisons équivalentes de deux éléments représentationnels portant sur un même référent, et véhiculant la même signification. Cet échec à produire des combinaisons multimodales les différencie des enfants en développement typique. Les enfants à SD peuvent bénéficier d'un paradigme d'apprentissage exploitant la modalité gestuelle, et ils semblent ainsi utiliser les signes plus souvent que les mots dans le but de combler leurs besoins communicatifs.

Par opposition, les enfants atteints du SW préfèrent avantager leur modalité vocale. Les études rapportent qu'ils présentent un retard dans l'âge où ils

commencent à produire des gestes. De plus, ils n'ont qu'une utilisation limitée des gestes, se focalisant sur des fonctions déclarative ou instrumentale.

### **II.12.6. Conclusion**

Ces revues de la littérature, qu'elles concernent un déficit de la vision, de l'audition, ou des troubles pathologiques, indiquent que quelle que soit la langue d'origine, la modalité d'input ou le profil cognitif, il existe une forte évidence pour une continuité étonnante entre le développement prélinguistique gestuel et linguistique oral. En effet, les compétences symboliques mises en évidence dans la modalité linguistique parlée co-surviennent de façon robuste avec des compétences cognitives et représentationnelles générales, mise en évidence dans la modalité gestuelle, phénomène qui perdure chez l'adulte. Cette proposition est de plus soutenue actuellement par la découverte des neurones miroirs, dont nous avons déjà parlé plus haut. Il semble exister un lien puissant entre les facultés de représentation et les facultés motrices. Un tel lien entraîne une compréhension plus nette des relations entre actions fortement structurées, celles du geste et celles de la parole.

Le développement du geste émerge de façon prédictible. Ainsi, les enfants qui ne parlent pas et ne produisent pas de pointers sont plus susceptibles d'être retardés par rapport aux enfants qui pointent pour communiquer. De la même façon, les enfants qui n'ont pas encore produit d'énoncés à deux mots, mais qui sont dans une étape de combinaisons bimodales, sont sur le point de voir émerger des combinaisons parlées. Capone et McGregor (2004) proposent un calendrier développemental du geste impliquant les phases successives suivantes : (1) 10-13 mois : apparition du pointer, qui prédit les premiers mots ; (2) 12-13 mois : les premiers mots émergent, les gestes remplissent une fonction complémentaire à la parole ; (3) 15-16 mois : préférence gestuelle ou vocale ; (4) 18-20 mois : préférence pour les mots parlés, utilisation de combinaisons geste+parole, augmentation significative des mots, pointers augmentés en combinaison avec les mots parlés ; (5) 2-5 ans : intégration geste-parole ; (6) âge scolaire : les gestes aident à la transition vers l'acquisition de concepts.

## II.13. De l'intégration du langage à travers les gestes dans la cognition humaine

Action influences thought as much as thought motivates action  
(Iverson et Thelen, 1999 : 37)

Notre travail s'est largement intéressé aux origines ontogénétiques des premiers mots, avec l'idée principale que geste de pointer et babillage sont deux pré-requis fondamentaux à la réalisation des premiers mots chez l'enfant.

Il nous semble à ce stade que le temps est venu de nous pencher sur l'origine phylogénétique d'un tel lien geste-parole. Pour cela, nous discuterons de cette fameuse spéculation d'une théorie gestuelle à l'origine du langage récemment remise en selle.

Le geste ainsi associé à la parole forme un noyau dur comme reflet des intentions communicatives d'un locuteur, voire pour l'intégration de sa propre pensée. Iverson et Thelen (1999) suggèrent que les origines développementales de ces deux modules soient ancrées dans les premiers liens main-bouche, étroitement couplées dans l'activité cognitive mutuelle. Elles expliquent que le couplage relativement synchrone observé entre geste et parole chez l'adulte se formerait dès la naissance dans ces actions motrices de la main et de la bouche, qui évolueraient ensuite de façon interactive à mesure que les enfants entrent dans le langage. Elles appuient leur théorie sur des études neurophysiologiques reliant langage et mouvement qui tendent à démontrer que certaines fonctions linguistiques et certaines fonctions motrices peuvent partager des mécanismes communs dans le cerveau. Par exemple les aires motrices semblent s'activer dans les tâches de langage qui n'impliquent pas explicitement la production de parole. D'autre part, certaines études valideraient l'idée que les aires du langage seraient activées lors de tâches motrices où l'usage du langage serait peu approprié (p. ex. utiliser le langage pour guider les mouvements). Erhard et al. (1996) montrent à cet égard que des portions de l'aire de Broca sont autant activées lors de tâches motrices impliquant les mains que lorsqu'un sujet pense simplement à bouger ses mains. Le fait que l'aire de Broca soit impliquée dans certaines activités motrices qui requièrent les mains ou même la zone oro-faciale, pourrait en faire un bon candidat neural pour expliquer ce lien geste-parole si particulier. En soutien à cette idée, nous relevons le fait que des

patients aphasiques de Broca montrent également des interférences entre geste et parole. De plus, la relation geste-parole est préservée même lorsque d'autres activités motrices ont été sévèrement altérées, indiquant alors la possible existence de mécanismes cérébraux communs pour le geste et la parole. Iverson et Thelen envisagent ces deux modules comme deux systèmes de mouvements reliés qui se coordonnent l'un l'autre progressivement, à mesure que la nature de cette coordination change au cours du développement.

Les auteurs proposent un décours développemental précis pour expliquer le couplage du système geste-parole. Tandis qu'à la naissance l'activité de ces deux systèmes est couplée, un contrôle indépendant émerge peu après dénotant une utilisation grandissante et adaptative des mains et de la bouche, fréquemment à travers des activités rythmiques dans les deux modalités, qui vont parfois être coordonnées. Émerge ensuite un couplage non synchrone du geste et de la parole, qui laissera place plus tard, chez l'adulte, à un couplage synchronisé (voir Figure II.42).

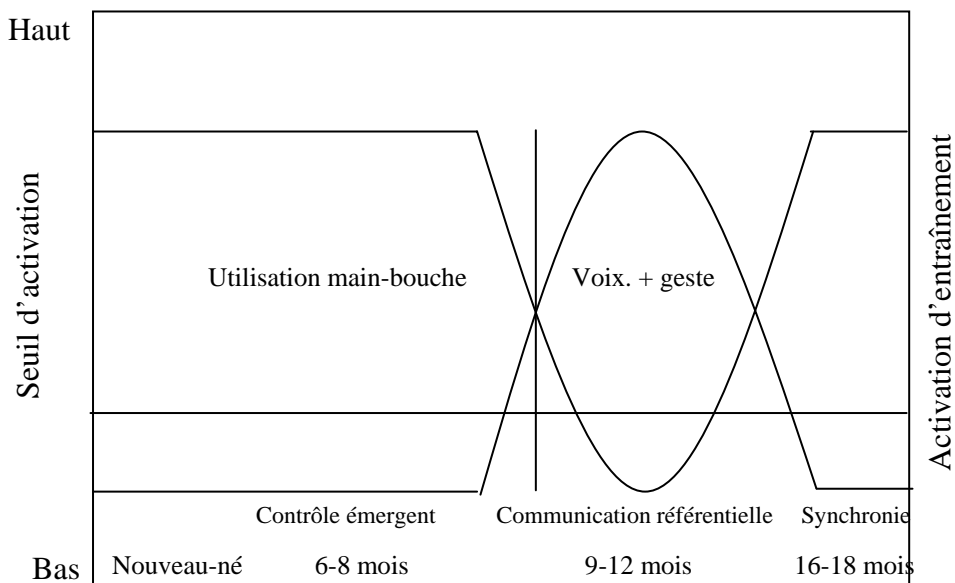


FIGURE II.42 : Niveaux du seuil d'activation et de l'activation d'entraînement dans le système oro-manuel lors des deux premières années de vie chez l'enfant (D'après Iverson et Thelen, 1999).

Sans même parler de la coordination nécessaire pour sucer son pouce in utero, on observe dès la naissance des conséquences comportementales caractéristiques des connexions entre système oral et manuel. C'est le réflexe de Babkin du nouveau-né : quand on presse la paume de ses mains on obtient une

ouverture de la bouche, comme un effet d'anticipation de l'arrivée de la main dans leur bouche. Par la suite, dès que l'enfant est capable de saisir un objet avec la main, il l'amène automatiquement à la bouche afin de l'explorer (orocognition). L'activité rythmique de percussion avec les mains (*banging*) annonce également l'activité rythmique de la mandibule (*babbling*). Notons également encore une fois l'importance du pointer de l'index dès 9 mois lors d'interactions communicatives, un pointer qui co-survient généralement avec une production vocale. Enfin, comme nous venons de le voir, cette relation geste-parole accompagne les premiers mots de l'enfant et annonce indiscutablement l'arrivée de combinaisons à deux mots. Ainsi ce couplage semble précoce chez l'enfant et il est activé progressivement de manière plus différenciée au cours du développement.

Il est alors intéressant, suite aux nombreuses découvertes réalisées dans le développement de l'enfant, de se demander comment de telles capacités linguistiques ont pu évoluer à partir de notre ancêtre commun. Pourquoi le langage chez l'humain et surtout comment s'est-il développé ? De nombreux auteurs ont tenté de décrire ce scénario développemental en replaçant le rôle fondamental du geste dans l'évolution menant vers le langage.

### **II.13.1. Plaidoyers pour une théorie gestualiste de l'origine du langage**

Dans une théorie gestuelle de l'origine du langage, Place (2000) invoque plusieurs principes généraux permettant une reconstruction évolutive : (i) un principe de sélection, où un avantage sélectif s'est opéré progressivement au sein d'un groupe ; (ii) un principe de réplique de la phylogenèse par l'ontogenèse ; (iii) le principe de régression à des adaptations précoces ; et (iv) le principe selon lequel chaque mutation va laisser des traces structurales sur les membres de l'espèce dans laquelle elle se développe.

A contrario d'autres chercheurs, tel Chomsky prônant une seule grande mutation fondamentale, qui aurait permis à certains primates d'utiliser le langage, Place envisage l'évolution linguistique comme une succession de mutations :

Once we begin to view semantics as the handmaiden of pragmatics and syntactics as the handmaiden of semantics, it becomes increasingly difficult to endorse Chomsky's (1957, 1965) belief that, in order to explain the human ability to construct and construe complex sentences, we must postulate an innate language faculty which appeared *deus ex machina* in a single gigantic mutation at the dawn of human prehistory. Mutations there must have been. How else can we explain the fact that

we talk and nonhumans, even with the best human instruction, barely do so? But what we must look for is not just one mutation, but a number of mutations spread out over millions of years, each one building on what has gone before, each one providing a selective advantage to the group in which it occurs which has enabled its members to survive and pass on their genes, when those who lacked that mutation went to the wall. Nor should we expect to find that the selective advantages which have promoted the survival of the groups in which such mutations have occurred have always been advantages conferred by improvements in interpersonal communication. (Place, 2000)

Au cœur de ces mutations, Place met l'accent sur deux d'entre elles, qui semblent fondamentales au processus de communication interpersonnelle et qui contribuèrent au développement du langage, à savoir le pointage référentiel et les changements survenant dans les cavités glottiques et supraglottiques :

Two mutations which may well have been selected in this way are that which made referential pointing possible and that or those which produced the changes in the mouth and larynx which made possible the production of vocal speech.

Ainsi, la communication linguistique humaine se révélerait être en premier lieu un langage gestuel, où les requêtes impératives seraient construites sur la base d'un prédicat qui prendrait la forme d'une action mimée. Cette capacité est déjà présente chez les chimpanzés qui sont tout à fait capables avec un ou plusieurs arguments d'indiquer un objet via le pointage référentiel. De plus, les chimpanzés semblent capables d'apprendre à utiliser le pointage dans le but d'obtenir de la nourriture par leur soigneur. Mieux encore, le bonobo Kanzi a appris à utiliser des gestes référentiels pour indiquer qui-fait-quoi, quel objet est manipulé, mais seulement dans un contexte où l'énoncé sera dirigé vers son soigneur. Cela pose la limite de son apprentissage, puisqu'il ne serait pas en mesure de répondre à un geste de pointage lorsqu'il est produit par quelqu'un d'autre. Le pointage référentiel qui serait dirigé vers un congénère n'aboutirait pas, confirmant l'observation que les grands singes ne pointent pas naturellement dans leur état naturel.

Une évidence neurophysiologique vient également en soutien à cette idée, à savoir que l'aire de Broca dans le cortex inférofrontal gauche humain a son homologue chez le singe nommée F5 où les neurones miroirs, dont nous avons déjà parlé précédemment, déchargent lors des productions de mouvement de main contrôlés visuellement, ainsi que lors de la perception visuelle des mêmes mouvements réalisés par un congénère. A ce propos, Arbib (2005), sur la base de l'hypothèse des neurones miroirs établie par l'équipe de Rizzolatti, explique que cette hypothèse constitue un développement neurologique impliquant que les hominidés aient un prélangage basé primitivement sur les gestes manuels (proto-signes), bien avant d'avoir un prélangage basé principalement sur les gestes vocaux (proto-parole).

En effet, Corballis (2002) souligne que les enregistrements de neurones dans l'aire F5 du cerveau du singe révèlent qu'ils ont un rapport avec les gestes manuels plutôt qu'avec les vocalisations, même si cette région est envisagée comme l'homologue de Broca dans le cerveau humain. Ce *mapping* de perception sur la base de l'exécution semble ainsi fournir plutôt un point de départ manuel pour le langage, tendant à soutenir l'idée que le langage trouve son origine dans les gestes, et non dans les vocalisations.

Place envisage un scénario pour l'évolution du langage, dans lequel il distingue les étapes suivantes : (1) une première étape consistant dans le fait que les chimpanzés et les humains aient un ancêtre commun. Ainsi ils partagent des facultés similaires telles que l'utilisation d'outils/armes, l'imitation des mouvements en apprentissage pour réaliser des manipulations qui impliquent ces outils, la communication d'un individu par le mime d'une action ; (2) la deuxième étape de ce processus implique l'émergence de premières vraies phrases en langue gestuelle. Celle-ci implique l'utilisation des gestes de pointage référentiels permettant de savoir qui réalise l'action, quel objet est manipulé et vers quel individu l'objet doit être transféré. Cette faculté ne se retrouve pas chez les chimpanzés qui sont incapables de répondre à un tel signal ; (3) la troisième étape concerne l'émergence de la vocalisation qui vient s'adjoindre aux gestes. Cette nouvelle capacité permet alors de référer à des objets qui ne sont plus physiquement présents dans la scène, capacité que ne permettait pas le geste de pointage ; (4) cette étape voit l'émergence de la faculté à compter jusqu'à 10 sur les doigts des deux mains de façon iconique, glissant ensuite vers un aspect plus symbolique lorsque le comptage progresse au-delà du nombre permis par les deux mains ; (5) l'étape suivante voit l'apparition de représentations d'objets utilisant des symboles arbitraires (des noms en d'autres termes). Pour être capable d'apprendre la signification de mots aussi facilement que le fait un enfant âgé de 2 ans, une mutation s'est certainement opérée et s'est vue sélectionnée uniquement dans l'espèce humaine. Certes les grands singes, tels que les bonobos, peuvent apprendre une communication symbolique et peuvent créer des phrases dans un « proto-langage » comme le nommerait Bickerton (1990), une faculté tout de même exceptionnelle de leur part. Toutefois ils ne bénéficient pas d'une explosion du vocabulaire comme l'enfant humain. Place ajoute que c'est certainement à ce moment précis que l'aire de Wernicke a dû évoluer, devenant le

noyau central de l'interprétation et de la production de noms ; (6) l'étape finale de ce processus développemental réside dans cette faculté-clé pour le langage qu'est la syntaxe. Cette faculté aurait été favorisée dans la communication linguistique pour libérer l'attention de l'auditeur vis-à-vis du contexte et pour libérer l'attention de l'auditeur sur les gestes du locuteur dans le but de désambiguïser le message. La syntaxe offre ainsi la liberté de parler d'items qui ne font pas particulièrement partie de l'ici-et-maintenant du locuteur. Arbib (2005) argumente ainsi qu'à la différence des langages utilisés par les singes, qui sont restreints à un répertoire spécifique, les langages humains sont à l'inverse productifs, constitués de mots, de marqueurs grammaticaux qui, combinés entre eux, offrent aisément un stock de phrases très large, et ces langages humains ont la capacité de produire de nouveaux mots qui peuvent ajouter à l'étendue du vocabulaire existant.

Ce qui pourrait en partie expliquer pourquoi les humains ont évolué vers cette capacité du langage réside dans l'idée que l'imitation a joué un rôle crucial dans l'acquisition du langage humain. Certes les singes, via les neurones miroirs sont capables d'imitation, mais ils ne le sont pas aussi pleinement que les humains qui, outre le fait de reconnaître la performance de l'autre comme des mouvements familiers et de les répéter ensuite, sont également capables de reconnaître que telle ou telle performance combine des actions nouvelles qui peuvent être proches de variantes d'actions déjà existantes dans leur répertoire. Ainsi l'ancêtre commun du chimpanzé et de l'humain était capable d'imitation « simple » de mouvements de mains, mais le système d'imitation plus complexe ne semble s'être développé qu'avec la lignée des hominidés.

Arbib propose de la même façon un décours possible pour l'évolution du langage : tandis que les pré-hominidés seraient capables de saisie, seraient dotés d'un système miroir et d'un système d'imitation simple pour cette saisie, l'évolution des hominidés s'est vue dotée en plus, toujours pour la saisie, d'un système d'imitation complexe, avec un système de communication basé sur le gestuel, facilitant un répertoire ouvert, ainsi que la proto-parole (une production d'ouvertures et fermetures du conduit vocal entraînant des séquences vocales, sans pour autant impliquer que ces séquences constituent un langage). L'évolution culturelle chez Homo Sapiens s'est vue, quant à elle, dotée d'un langage avec des structures prédicat-arguments reposant sur une syntaxe et une sémantique bien établies.



En opposition à la grammaire universelle de Chomsky avec ses structures pré-établies dans le génome humain, Arbib propose que le cerveau du premier Homo Sapiens était déjà « prêt-pour-le-langage » (*language-ready*), mais qu'une certaine évolution culturelle au sein des sociétés humaines était nécessaire pour qu'on puisse aboutir à un langage humain tel que le nôtre aujourd'hui.

Arbib distingue plusieurs critères différenciant un cerveau « prêt-pour-le-langage » d'un cerveau utilisant le langage. Sept critères sont retenus pour le premier : la nomination, l'intentionnalité, la parité (ce qui compte pour le locuteur doit compter pour l'auditeur), la structuration hiérarchique, l'ordre temporel, et la socialisation. Quatre critères supplémentaires viennent compléter cette liste pour le langage : la symbolisation, l'ensemble syntaxe/sémantique/récurtivité, la possibilité d'aller au-delà de l'ici-et-maintenant et l'apprenabilité.

Une évidence est que les gestes ont un but communicatif. Arbib (2005) émet l'hypothèse que les singes utilisent les mouvements des mains seulement pour les actions pragmatiques, c'est-à-dire des mains utilisées pour interagir physiquement avec des objets ou d'autres créatures. A la base, les singes auraient une communication « involontaire » de différents comportements à travers des gestes oro-faciaux et vocaux. Arbib suggère alors que le système des neurones miroirs (cf. supra) aurait pu rendre possible l'évolution depuis des mouvements de mains dédiés à des actions pragmatiques vers des gestes communicatifs contrôlés volontairement. Selon lui, le langage se trouve compris à l'intérieur de la saisie (cf. son article en collaboration avec Rizzolatti, intitulé *Language within our grasp*) et l'homologie du système miroir chez le singe et chez l'homme procurerait ce chaînon neurobiologique manquant pour parfaire l'hypothèse que la communication gestuelle (basée sur les gestes manuels) précède la parole dans l'évolution du langage.

### **II.13.1.1. La dominance de la main droite dans l'évolution du geste vers la parole**

Corballis (2003) explique que la dominance de l'hémisphère cérébral gauche pour la vocalisation est une caractéristique récurrente dans de nombreuses espèces humaines (grenouilles, oiseaux, mammifères ; pour un patron de latéralisation des vervets opposé à celui des humains, cf. Gil-da-Costa et Hauser, in press). En revanche la forte prédominance de la main droite et la présence de l'asymétrie main

droite/cerveau gauche pour la parole chez l'homme paraissent être un élément unique, qui pourrait être survenu suite à l'association entre gestes manuels et vocalisations dans l'évolution du langage. En effet, l'activité de la main droite (et donc du cerveau gauche) reste prédominante pour les mouvements manuels lorsque des locuteurs sont engagés dans une conversation. Cette dominance du cerveau gauche est également trouvée chez les signeurs atteints de surdité, particulièrement pour le contrôle fonctionnel de la langue des signes.

Même si des chercheurs comme Hopkins ont réussi à démontrer une préférence de la main droite parmi les chimpanzés captifs pour certaines activités telles que la nourriture bimanuelle, le ratio des utilisateurs de main droite et gauche est nettement moins élevé chez le singe (2:1) tandis que chez l'humain il atteint quasiment 9:1.

Meguerditchian et Vauclair (2006) ont étudié les gestes communicatifs manuels et leurs asymétries chez les primates non-humains, dans un but de clarification de l'hypothèse de l'origine gestuelle du langage et de sa latéralisation. De nombreuses études antérieures avaient déjà apporté la preuve d'une continuité entre humain et chimpanzé. Cette preuve constituait alors un soutien à l'hypothèse qu'il pouvait exister un système gestuel de communication latéralisé dans l'hémisphère gauche chez notre ancêtre commun, et que ce système avait pu évoluer pour le langage chez l'homme. Les auteurs ont donc voulu tester cette hypothèse auprès d'un échantillon comprenant 60 babouins (*Papio anubis*). Les babouins possèdent un geste manuel, qui est spécifique à leur espèce, caractéristique pour sa fonction communicative. Ce geste consiste en un frottement de la main rapide et répété sur le sol et est principalement utilisé afin de menacer ou intimider d'autres individus. Les auteurs notent que ce geste est très probablement intentionnel puisqu'il est toujours utilisé en direction d'un partenaire social, que celui-ci soit un congénère ou un humain. Meguerditchian et Vauclair se sont alors intéressés à l'étude de ce geste spécifique dans deux contextes, le premier lors d'interactions agonistiques spontanées, et le second lors d'interactions entre babouins et expérimentateur. Sur les 60 babouins observés, 35 étaient classés comme droitiers, 10 comme gauchers et 15 avec une dominance manuelle ambiguë. Parmi les 45 sujets latéralisés (droitiers ou gauchers), 78% étaient droitiers. Leurs observations rapportent une asymétrie manuelle pour les gestes indiquant très

probablement une dominance cérébrale gauche pour le contrôle des fonctions engagées dans la communication par la main. Le fait que ces observations aient été réalisées aussi de façon spontanée, au sein d'un groupe social de babouins, renforce l'évidence que cette dominance manuelle n'est en aucun cas un artefact qui aurait pu être induit par les humains. De plus ce résultat met en évidence une différence claire dans les patrons de latéralité et dans l'utilisation de gestes dans des situations communicatives vs. non-communicatives. Les auteurs suggèrent alors l'idée suivante :

[...] the communicative function of the hand could imply a different cerebral substrate than that involved in their manipulative functions: a communicatory left-hemisphere system may be involved for the production of gestures. This system would more strongly favor the use of the right hand than bimanual coordinations for object manipulations. (2006:173)

Ils soulignent le fait que leurs résultats peuvent être corroborés par d'autres études, notamment celles en imagerie cérébrale réalisées sur les chimpanzés (Gannon et al., 1998, Hopkins et al., 1998, Cantalupo et Hopkins, 2001, Tagliatela et al., 2006). Elles indiquent que les asymétries des aires homologues du langage, ne sont pas corrélées à la dominance manuelle que l'on observe lors d'actions motrices dans des situations non-communicatives. Elles le sont par contre pour les gestes communicatifs, cette préférence manuelle étant liée avec une asymétrie dans l'aire homologue de Broca. L'étude de Meguerditchian et Vauclair sur les babouins, et les autres études réalisées sur les chimpanzés, tendent donc à attester de cette continuité existante entre asymétrie pour la parole reliée au geste et l'asymétrie des gestes communicatifs chez ces primates non-humains. Et ceci même s'il existe une dominance manuelle droite moindre chez le primate non-humain. Cette découverte semble alors soutenir l'hypothèse de l'origine gestuelle du langage, avec une dominance hémisphérique gauche déjà présente chez notre ancêtre commun pour le contrôle de l'action manuelle, qui aurait servi de substrat ou de pré-requis pour le développement du langage chez l'homme.

En appui à ces travaux comportementaux de Meguerditchian et Vauclair, nous avons cité les preuves neuro-anatomiques récemment publiées par Tagliatela et al. (2006) sur les différences morphologiques existantes entre hémisphères dans le gyrus frontal intérieur (IFG) de 56 chimpanzés captifs, études corrélées avec les observations sur la latéralité manuelle qu'ils utilisaient le plus fréquemment pour les gestes communicatifs. Leur hypothèse de base résidait dans l'idée que la main utilisée peut être reliée au comportement communicatif et que, ainsi, on pourrait

observer une asymétrie dans IFG, ce comportement ne se répliquant pas dans le cadre de gestes non-communicatifs (par exemple une atteinte simple). De la même façon que Meguerditchian et Vauclair, les auteurs classèrent les chimpanzés selon leur dominance manuelle : sur les 56 chimpanzés, 33 étaient droitiers pour les gestes communicatifs, tandis que 23 ne montraient aucune préférence, ou une préférence pour la main gauche. De la même façon, 24 chimpanzés étaient classés comme droitiers pour la tâche d'atteinte simple et 32 comme gaucher/aucune préférence manuelle. De leurs observations et analyses ressort un résultat attendu : les chimpanzés qui montrent une préférence significative pour l'utilisation de la main droite dans le cadre de gestes manuels ont une activation plus grande et plus à gauche dans IFG, comparé à la partie droite. Aucune différence n'était relevée dans le *planum temporale* (PT) et dans les aires motrices manuelles. A l'inverse, lors de la tâche d'atteinte, la préférence manuelle était associée à une asymétrie dans les aires motrices manuelles mais ni dans IFG, ni dans PT. Les auteurs concluent alors que les asymétries dans IFG étaient en rapport spécifique avec les actions manuelles communicatives, tandis que la latéralisation structurelle observée dans les aires motrices manuelles était reliée à un comportement similaire mais qui ne posséderait pas la caractéristique d'être communicatif. Ce résultat peut être mis en parallèle avec l'étude sur les sujets humains menée par Løevenbruck et al. (2005) déjà citée précédemment, dans la mesure où chez les humains, IFG gauche est activé lors de la prononciation de phrases déictiques. Les résultats de cette étude dirigée par Tagliatela et al. (2006) ajoute un crédit supplémentaire à la proposition que la latéralisation gestuelle puisse avoir joué un rôle significatif dans l'évolution vers le langage humain, avec l'idée que l'ancêtre commun ait possédé ce système de communication gestuelle latéralisé, qui aurait ensuite évolué chez l'homme afin de favoriser la production de la parole.

Ces réflexions sur la latéralisation gestuelle et la dominance manuelle nous amènent à simplement mentionner les observations sur la préférence manuelle dans les pointers des 6 enfants français du corpus que nous avons étudié (cf. infra).

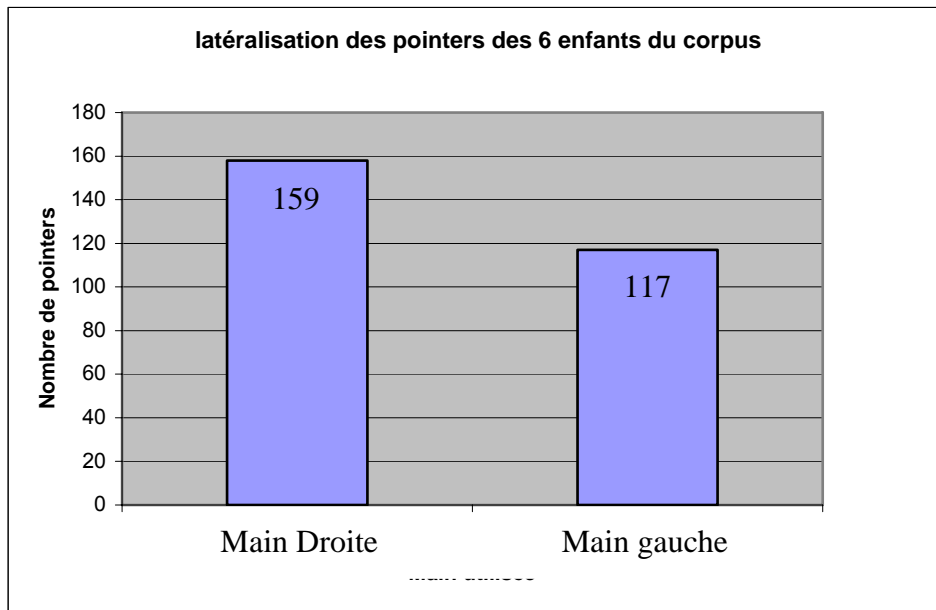


FIGURE II.43 : Nombre de pointers réalisés au total avec la main droite et la main gauche pour l'ensemble des 6 enfants français du corpus (cf. 3<sup>ème</sup> partie).

Chez nos 6 enfants, il apparaît que trois d'entre eux ont clairement une préférence manuelle droite (Anatole : 10 pointers sur 12 ; Célia : 22 sur 31 ; et Tom : 67 sur 100) ; les trois autres n'ayant pas vraiment une préférence manuelle gauche, (Jules : 40 sur 75 ; Nicolas : 12 sur 21 ; et Lise : 21 sur 37). Ce n'est que sur le total des pointers relevés parmi tous les enfants, que nous avons 159 main droite (la présence de Tom grand pointeur, droitier, n'y est pas pour rien), contre 117 main gauche (Figure II.43).

### II.13.1.2. La théorie gestualiste selon Corballis

Corballis considérant que le langage a évolué à partir des gestes manuels, pose que différents éléments vocaux sont venus se surajouter progressivement. Le fait que les vocalisations soient graduellement incorporées au système gestuel a eu pour conséquence d'amener à la latéralisation des gestes manuels eux-même, avec une préférence pour la main droite, suite à l'émergence de ce « centre moteur de la parole » situé dans l'hémisphère gauche.

Corballis suggère que nous avons hérité en tant que primates d'un contrôle intentionnel assez puissant impliquant le visage et les membres antérieurs. Toutefois notre héritage en matière de contrôle vocal est plutôt faible. En effet les cris des primates les plus proches de l'humain, le bonobo et le chimpanzé, sont réalisés

largement sous le contrôle de l'émotion et correspondent davantage à des rires ou des pleurs que des sons réellement articulés. Pourquoi alors l'homme a-t-il surimposé à ses gestes la parole ? Corballis a lui aussi tenté de répondre à cette question incontournable (pour tous les gestualistes) en proposant son propre scénario évolutif.

L'émergence de la bipédie il y aurait 6 millions d'années aurait eu pour conséquence d'instaurer un langage basé sur les gestes de façon prédominante. Toutefois, il y a 2 millions d'années, de nombreux changements ont entraîné une utilisation accrue des mains pour d'autres activités que la communication, notamment la fabrication d'outils. Corballis avance, d'autre part, le rôle de l'œil chez l'humain pour la communication. En effet, tandis que la couleur sombre de la sclera chez les primates non-humains semble être une adaptation afin de cacher la direction du regard aux autres primates ou prédateurs, l'œil humain quant à lui –notre sclera est blanche avec un œil allongé horizontalement– semble avoir évolué pour augmenter la communication plutôt que de la cacher. Cet élément cité par Corballis nous semble pertinent seulement dans la mesure où nous estimons que l'œil serait la première partie du corps révélatrice d'une attention dirigée chez l'enfant.

Critiquant le point précis de la bipédie dans la théorie gestuelle exposée par Corballis, Tomasello (2002) estime que ce dernier focalise sur la libération des mains, alors que cette réflexion ne peut en aucun cas expliquer clairement la différence socio-cognitive qualitative entre symboles référentiels humains (pointage déclaratif) et signaux impératifs de régulation d'interaction des chimpanzés (*mands*). En effet, nous pensons que cette théorie « corballisienne » ne met pas assez en avant le fait que, au contraire des chimpanzés, les enfants humains pointent référentiellement vers des objets ou événements extérieurs, et ils font ceci dans le but de partager l'attention. Ces caractéristiques humaines nous semblent être fondamentales si l'on veut comparer la communication humaine et non-humaine chez les primates ou les carnivores. Ce que souligne également Tomasello :

These are both important features of human linguistic communication and lay the groundwork for the quintessential human use of language –idle conversation about external topics– and so are plausibly responsible for a major transition (2002:791)

Pour Corballis, une des clés pour expliquer comment nous sommes passés des gestes au langage repose en partie dans le développement de l'aire de Broca, recrutée pour les gestes chez le singe et siège de la parole chez l'homme. Cette aire

étant impliquée depuis longue date dans les gestes manuels, son élargissement peut avoir reflété l'incorporation de la syntaxe dans la communication gestuelle. Le scénario proposé par Corballis implique que le langage ait longtemps requis la combinaison de gestes manuels, faciaux et vocaux, et que ce n'est que récemment que la parole dite vocale est devenue dominante. Corballis argumente que c'est à travers l'association de l'aspect manuel avec l'aspect vocal que la dominance de la main droite est née. Ainsi, avec l'élévation du genre homo, la pensée récursive aurait émergé basée sur les adaptations aux problèmes sociaux complexes. La récursivité sous-tend une syntaxe complexe. Les vocalisations auraient ensuite pris le pouvoir sur les gestes.

En effet, l'aire de Broca est asymétrique chez la grande majorité des humains et cela de façon assez frappante, avec un côté gauche jouant seul un rôle dans la syntaxe. La région homologue du côté droit semble être impliquée plutôt dans ce qui est nommé « syntaxe musicale » ou rythme. L'aire de Broca peut avoir été le siège de l'interaction entre le manuel et vocal, comme le souligne Corballis :

In any event, to revert to the matter at hand, as it were, righthandedness may well have evolved from the synchronization of manual and facial gestures with a lateralized system of vocal production (2003:205)

La dominance du côté droit pour les gestes se retrouve dans la communication : les droitiers tendent à faire plus de gestes avec la main droite lorsqu'ils parlent, tandis que les gauchers montrent un patron beaucoup plus flou et utilisent les deux mains. De plus, il semble que 90% de la population humaine testée montre un mouvement plus grand du côté droit de la bouche lorsqu'ils parlent. Il existerait alors une asymétrie des gestes manuels et faciaux induite par une asymétrie présente dans le contrôle de la vocalisation.

Afin de mieux appréhender cette dominance manuelle chez l'adulte, il paraît nécessaire de s'intéresser également aux origines ontogénétiques de celle-ci. Bates et Dick (2002) relatent le fait qu'une décentration vers la main droite existe dès le début de la vie. Elle semble plus grande pour les gestes porteurs de contenu communicatif et/ou symbolique que pour les gestes non-communicatifs et non-symboliques. Dans une étude rapportée par Bates et al. (1986), codant les gestes déictiques, symboliques et les mouvements de main non symboliques à 13, 20 et 28 mois, les auteurs notent que cette décentration sur la main droite est significativement plus grande pour les gestes de pointer (80% avec la main droite), et

significativement plus grande aussi pour les actions symboliques plutôt que non-symboliques. Cette distinction est ici importante et il convient de noter que la latéralisation liée au langage et au geste est plus évidente lorsque l'on discute de gestes familiers et porteurs de sens. Ainsi, les enfants à développement normal montrent une tendance pour la main droite de façon significative pour l'activité manuelle. Des déficits dans la production gestuelle sont plus communs lors d'un dommage à l'hémisphère gauche, et ils tendent à être corrélés à la présence et la sévérité des symptômes aphasiques. Bates et Dick (2002) supposent que le geste, la parole et le langage sont « étroitement couplés neurologiquement et développementalement ». Il existe un chevauchement dans le contrôle neural pour la parole et le geste, de telle façon que l'activation neuronale propagée d'une région à l'autre peut être à la base de leur co-occurrence. La synchronie geste-parole serait donc à la base originare des premières activités main-bouche.

Nous avons poussé l'examen du rapport entre geste et parole au-delà de la première année chez l'enfant, afin de dégager le caractère précurseur des gestes de pointers dans la faculté de langage ultérieure. Notre dessein était d'arriver à donner une explication rationnelle du lien entre ces deux modules dans une perspective ontogénétique. Cette perspective nous a ensuite mené à considérer la phylogénèse afin de trouver une racine plus profonde chez notre ancêtre commun.

En vue de la partie suivante, nous désirons apporter dès maintenant des éléments de réponse au fait que lors du visionnage de notre corpus, nous nous sommes aperçue que le geste de pointer n'était pas systématiquement synchrone avec la vocalisation affiliée, loin de là (cf. aussi Pizzuto et al., 2005). Nous aurions donc bien deux modules qui coopèrent pour arriver au premier mot, mais apparemment ils ne le font pas obligatoirement et naturellement —disons pas par principe— en synchronie. Afin de tenter d'apporter une ébauche d'explication à ce phénomène observé chez l'enfant, nous avons dû reconsidérer les formes abouties de cette coopération dans les langues de l'adulte, à la lumière des théories et des expériences menées sur cette coordination du geste et la parole.

## **II.14. La coordination geste-parole chez l'adulte**

Lorsque nous communiquons un message à un destinataire, nous utilisons évidemment plusieurs modalités, naturellement la parole, souvent accompagnée de



gestes de la main. Plus spécifiquement, qu'en est-il de ces composantes dans une situation de communication face-à-face où il s'établit une relation triadique entre un objet, un destinataire et un destinataire qui tente d'apporter l'attention de l'autre sur l'objet d'intérêt par les moyens dont il est doté : la voix et le geste. Comment s'établit alors le dialogue (*crosstalk*) entre ces deux modalités ?

Les locuteurs bénéficient d'un ensemble riche de formes gestuelles pour exprimer différents types de contenu. Nous pouvons rappeler la classification établie par McNeill et al. (1990), où les auteurs distinguent principalement deux grandes catégories de gestes : les emblèmes et les gestes co-verbaux. Les premiers ont la capacité à remplacer la parole et constituent de ce fait des gestes conventionnels variant selon les cultures. Ils fonctionnent ainsi indépendamment et sont pleinement porteurs de sens. Toutefois ces gestes ne constituent pas en eux-mêmes un système linguistique à part entière, dans la mesure où on ne peut combiner plusieurs emblèmes, et qu'ils ne sont pas soumis à des règles grammaticales. Les gestes co-verbaux, quant à eux, accompagnent la parole et ne peuvent s'en substituer. Ils sont par nature inextricablement liés au contexte et ne peuvent être interprétés qu'à l'intérieur de celui-ci. Parmi ces gestes co-verbaux, McNeill distingue : (1) les gestes iconiques, dépendant du contenu sémantique apporté par la parole et qui peuvent véhiculer une information complémentaire ou supplémentaire à la parole à propos d'une entité concrète ; (2) les gestes métaphoriques, qui se distinguent des gestes iconiques par le fait qu'ils décrivent une entité plutôt abstraite ; (3) les battements ou gestes « bâtons » (*beat gestures*) consistant en des mouvements rythmiques de main ou doigt liés à la parole et dont la fonction discursive est principalement pragmatique ; et (4) les gestes déictiques qui ont pour fonction d'indiquer un objet de référence dans une scène.

Le pointer de l'index constitue un élément incontournable de la communication humaine. Celui-ci co-survient fréquemment avec la parole, se caractérisant par un bras et un index tendus dans la direction de l'objet d'intérêt. Nous pointons aussi bien des choses que nous connaissons que des choses que nous ne connaissons pas. Pointer peut être une façon de déclarer, faire une remarque, ou poser une question. Toutefois, l'intention sous-jacente toujours présente dans ce geste de pointer de l'index réside dans le but d'attirer l'attention de quelqu'un d'autre vers un objet ou un événement du monde (Masataka, 2003), ce monde étant le monde

possible instancié par la situation de communication, ce qui rend compte de la deixis in absentia ou *Deixis am Phantasma* (différente de la *demonstratio ad oculos*) pour reprendre les termes du grand Viennois précurseur du schéma de la communication, Karl Bühler.

Intéressons-nous dans un premier temps à établir ce qui est connu sur la coordination main-bouche chez l'adulte, et sur le comportement de ces deux modules, afin de tenter de nous donner des pistes et de mieux comprendre dans quelle mesure une simple synchronisation entre ces deux modules chez l'enfant n'est pas forcément au rendez-vous des observables dans leur coordination.

### **II.14.1. Quels modèles théoriques pour expliquer la relation geste-parole ?**

McNeill (2000) a tenté d'expliquer les différents types de traitement de l'information modulaire des productions gestes-parole. Il reprend trois modèles fondamentaux pour l'étude de cette relation : (i) le modèle de Krauss et al. (2000) ; (ii) le modèle de De Ruiter (2000) ; et (iii) son propre modèle.

Selon McNeill, les stratégies de ces trois approches ne sont pas forcément différentes, mais elles se démarquent par le locus où le module gestuel, dans le module de parole, est introduit.

#### **II.14.1.1. Modèle de Krauss, Chen et Gottesman (2000) : lien préconceptualiseur**

Krauss et al. utilisent selon McNeill (2000) un « lien préconceptualiseur », c'est-à-dire que le module gestuel est lié à la structure de traitement de l'information de façon pré-conceptuelle dans la mémoire de travail. En d'autres termes, des structures pré-existantes —avant même les processus de planification ou de conceptualisation— seraient déjà disponibles dans la mémoire. Cette relation pré-existante entre gestualité et information verbale serait ensuite mobilisée lorsque le locuteur cherche un mot. Voyons plus en détail ce modèle KCG (Krauss, Chen et Gottesman).

Krauss et al. proposent leur modèle afin d'éclairer le traitement par lequel les gestes lexicaux sont produits. Ils étudient cette catégorie particulière de gestes, appelés « gestes représentatifs », ou encore « gestes illustratifs ». Cette catégorie

n'est pas aussi clairement définie que les catégories concernant les gestes symboliques, déictiques ou moteurs. Pour Krauss et al., ces gestes surviennent uniquement avec la parole. Ils varient dans leur durée, ne sont pas répétitifs, ils sont complexes et de forme elle aussi variée. Plusieurs d'entre eux sont porteurs d'une relation significative avec le contenu sémantique de la parole qui les accompagne. Plus globalement, cette catégorie a été désignée pour regrouper les gestes qui ne sont ni symboliques, ni déictiques, ni moteurs. Cependant, il faut noter que cette classe est tout à fait cohérente et concerne des mouvements générés dans un processus uniforme, qui va jouer un rôle dans la production de parole.

Dans leur modèle, les chercheurs supposent que les gestes lexicaux et la parole impliquent deux systèmes distincts qui vont dialoguer (*crosstalk*) ensemble.

Lors de la production de parole, la conceptualisation implique d'avoir recours, selon les modèles courants, à des connaissances déclaratives et procédurales dans le but de construire une intention communicative. A la sortie de cette étape de conceptualisation, on obtient une structure conceptuelle contenant des spécifications sémantiques. Lorsqu'on arrive à l'étape de formulation, ce message préverbal va subir des transformations : d'un côté le codeur grammatical cartographie le concept qui doit être lexicalisé dans un lemme, dans le lexique mental. Des significations vont être liées au contenu du message. D'un autre côté, l'information syntaxique liée au message va être utilisée, afin que la structure conceptuelle soit transformée en structure de surface. On accède ensuite aux formes de mots qui sont stockées dans la mémoire lexicale et on peut ainsi planifier la prosodie de l'énoncé, partie d'un codage phonologique qui transforme cette structure de surface en une planification phonétique. A la fin de l'étape articulatoire, on obtient une parole manifeste, qui est contrôlée par le locuteur lui-même, en utilisant un feedback à but correctif.

En ce qui concerne la production de gestes, Krauss et al. proposent une explication qui place les origines du geste dans les représentations liées à la mémoire de travail, qui vont être exprimées par la parole. Dans leur modèle, ils proposent plusieurs hypothèses sur la mémoire et les représentations spatiales : (i) la mémoire possède différents gabarits pour représenter les connaissances, et ce qui est contenu en mémoire va être codé sous ces différents gabarits ; (ii) lorsqu'un concept est activé sous un gabarit donné dans la mémoire, il va également activer d'autres concepts sous d'autres gabarits ; (iii) ce sont ces différents gabarits qui vont

déterminer les différences entre concepts. Pour établir une représentation mentale complète d'un concept, il peut être nécessaire de prendre en compte les inputs des autres gabarits ; et (iv) certaines représentations dans un format peuvent être traduites dans une forme représentative d'un autre gabarit.

Leur modèle repose sur l'idée qu'il sélectionne des traits spatiaux/dynamiques et que cela va transformer l'information stockée dans les formats spatial et dynamique en un ensemble de spécifications spatiales/dynamiques. Ce sont ces spécifications qui vont précisément être traduites par un planificateur d'actions dans un programme moteur qui fournit au système moteur un ensemble d'instructions pour effectuer le geste lexical. A la sortie du système moteur, on obtient alors un mouvement gestuel, contrôlé de façon kinesthésique. Les auteurs notent néanmoins que les traits de l'item lexical recherché peuvent, ou non, faire partie de l'intention communication du locuteur.

Selon Krauss et al., un des rôles majeurs de ces gestes lexicaux, est celui de faciliter la récupération lexicale. Le traitement consiste à ce que la sortie du système de production de geste pénètre dans l'encodeur phonologique via un contrôle kinésique, afin de favoriser la récupération de la forme lexicale.

Au moment où la durée du geste est étroitement liée à l'articulation de l'affilié lexical, un mécanisme va informer le système moteur du moment où il doit cesser le geste. Le fait d'entendre la production de l'affilié lexical permet d'instaurer le signal qui va réclamer la fin du geste<sup>22</sup>.

Dans la relation entre modules geste et parole, Krauss et al. se posent plus dans une perspective d'un modèle interactif, rejetant le modèle d'autonomie des modules. Ils soutiennent leur vision par deux évidences convergentes : la première vient des études concernant les relations temporelles entre geste et parole, particulièrement celle de Morrel-Samuels et Krauss (1992). Dans cette étude, les chercheurs étudient la coordination temporelle entre le geste et la parole qui partage son sens. Selon eux, les gestes ne dérivent pas des images, comme le propose McNeill, mais plutôt d'un système de représentations motrices de concepts, dont plusieurs d'entre elles viennent à être matérialisées dans la parole. L'origine commune entre geste et parole se situerait alors au niveau pré-sémantique de

---

<sup>22</sup> « Essentially, we are proposing that hearing the lexical affiliate being articulated serves as the signal to terminate the gesture . » Krauss et al. (p. 270, 2000)

l'intention de communication du locuteur. Les résultats de leur étude montrent que les gestes sont synchronisés avec la parole et qu'ils sont initiés avant ou en même temps (jamais après) que le début de l'affilié lexical. De plus ils observent une corrélation entre la familiarité de l'affilié lexical et la grandeur de l'asynchronie : plus l'affilié est familier, plus l'écart de l'asynchronie va se réduire. De la même façon, plus l'affilié lexical va être familier, et plus la durée du geste associé sera courte. Les auteurs concluent de cette découverte que la familiarité de l'affilié lexical peut prédire l'asynchronie entre geste et parole et la durée du geste. Ces évidence soutiennent l'idée que la production du geste est liée aux traits de l'accès lexical.

Morrel-Samuels et Krauss (1992) se posent nettement dans une vision interactive pour la relation geste-parole, et ils mettent en avant le fait que la relation entre durée du geste et familiarité de l'affilié lexical met à mal la vision balistique de la production geste/parole telle qu'elle est soutenue par Levelt et al. (1985). En effet, afin de planifier un geste de durée suffisante, le locuteur devrait connaître à l'avance combien de temps l'accès lexical va prendre. Si on se place d'un point de vue balistique, il faudrait alors supposer que les gestes dont la durée est longue sont associés aux mots non-familiers, ou alors que le locuteur lui-même a la capacité de prédire la quantité de temps nécessaire pour l'accès lexical et de s'en accommoder.

Leur seconde évidence vient des effets du geste sur la parole. La parole n'est en fait pas aussi fluide lorsqu'elle est accompagnée de gestes, lorsqu'il s'agit de parole à contenu spatial. Cette observation semble donc nourrir l'idée que les gestes facilitent d'une certaine façon l'accès lexical.

L'approche proposée ici par Krauss et al. est donc un modèle de traits pour expliquer les origines du geste. C'est une alternative aux modèles dits « imagistiques » proposés par De Ruiter ou encore McNeill. Selon Krauss et al., ces modèles présentent de nombreux problèmes pour expliquer la production gestuelle. On ne doit pas, par exemple, négliger d'expliquer que les images sont représentées à la base par un ensemble de mouvements de main, et qu'implicitement certains mécanismes doivent rendre abstraits certains éléments d'une image afin de les traduire ensuite sous forme d'instructions destinées au système moteur. La fonction de cette opération se retrouve ainsi dans le module de génération de « sketch » de De Ruiter (voir plus bas). Le second problème de ce type de modèle est le fait que pour certains gestes, l'« isomorphie physique » est manquante dans le contenu

conceptuel de la parole qui co-survient (par exemple les gestes métaphoriques). Les modèles de type « imagistique » semblent ne pas être, selon les auteurs, les modèles appropriés pour expliquer la production de tels gestes.

Pour résumer brièvement le point de vue de la théorie KCG, retenons que c'est l'utilisation même du geste qui va aider à la récupération lexicale. Selon McNeill, afin d'atteindre le but de récupération de l'item lexical, le modèle KCG implique qu'il doit avoir eu accès à l'information contextuelle. Néanmoins, ce contenu est exclu par une conception modulaire.

#### **II.14.1.2. Modèle de De Ruiter : lien conceptualiseur**

En gardant toujours à l'esprit la comparaison des modèle selon McNeill, il place l'étude de De Ruiter et la relation geste-parole dans une optique « lien conceptualiseur », où la composante du geste est liée au conceptualiseur à la structure du traitement de l'information, au niveau de la génération du message.

De Ruiter (2000) propose un modèle ou architecture de traitement général pour la production de gestes, visant à poser les bases pour étudier ensuite les processus et représentations impliquées dans les gestes et la parole et qu'il appelle le « Sketch Model » et qui serait, selon lui, une extension du modèle proposé par Levelt (1989) pour la production de parole. Toutefois, ce modèle semble se différencier de celui de Levelt en plusieurs points, principalement parce que De Ruiter considère et explique les différents types gestuels par un modèle de traitement de l'information modulaire. De plus, le Sketch modèle implique que les gestes soient produits en trois étapes : (1) la sélection de l'information qui doit être exprimée dans le geste, (2) la génération d'un programme moteur et (3) l'exécution de celui-ci.

Selon De Ruiter, les gens n'ont pas d'affiliés lexicaux mais plutôt des affiliés « conceptuels ». En citant McNeill (1992), il argumente que les gestes sont synchronisés avec et significativement reliés à une information discursive de haut niveau. Cette explication paraît de plus pertinente afin d'expliquer l'occurrence de gestes occasionnels reliés apparemment à un seul mot, comme dans le cas d'un pointage déictique accompagné d'un mot. L'idée étant selon De Ruiter la suivante :

All content words have an underlying conceptual representation, but not all conceptual representations have a corresponding content word. (2000:291)<sup>23</sup>

De plus, De Ruiter pense que l'encodeur grammatical tel que mis en valeur dans le modèle de Krauss par exemple ou Levelt, est un candidat improbable pour expliquer l'initiation du geste, dans la mesure où l'input de formulation est un « message préverbal » qui se trouve sous la forme d'une représentation propositionnelle, qui par conséquent n'a pas accès à l'information « imagistique » dans la mémoire de travail. De Ruiter suppose que c'est le rôle du conceptualiseur d'initier le geste, puisqu'il a accès à la mémoire de travail, et peut ainsi accéder à la connaissance propositionnelle pour la génération de messages préverbaux et à l'information spatio-temporelle pour la génération des gestes. Le conceptualiseur va ensuite envoyer une représentation appelée « sketch » aux modules de traitement suivants. Cette notion est à la clé de la réflexion de De Ruiter pour ce modèle et c'est pour cette raison qu'il a nommé celui-ci le modèle « sketch ». Une fois qu'un sketch contenant une ou plusieurs représentations est créé, il est directement envoyé au planificateur de geste, qui va construire un programme moteur résultant dans la réalisation du geste.

Selon De Ruiter, le modèle « sketch » est un bon candidat pour expliquer la synchronie geste-parole, bien qu'il soulève la difficulté à définir en tant que telle la notion de synchronisation, et qu'il reste difficile et ambigu d'affecter le bon affilié à un geste donné. Toujours selon l'auteur, la synchronie entre geste et parole peut être définie par deux intervalles de temps souvent difficiles à délimiter. Toutefois, on retrouve l'évidence d'un début de geste précédant le début de la parole qui co-survient, ceci par une durée de moins d'une seconde.

Cela nous amène à considérer la notion de phase dans le geste. Kendon (1980) avait déjà dégagé les principales composantes d'une unité gestuelle, qu'il décompose de la façon suivante : une phase préparatoire du geste suivie d'un « stroke », dont la tenue précède le retour. Le « stroke » constitue la phase gestuelle porteuse de sens et peut donc être considérée comme la partie la plus cruciale d'une unité gestuelle. De Ruiter s'est également intéressé en détail à la phase de tenue gestuelle (*gestural hold*), et à laquelle il ajoute deux éléments supplémentaires distincts par rapport à Kendon (1980) dans son modèle : (i) la phase précédant la

---

<sup>23</sup> « Tous les mots de contenu ont une représentation conceptuelle fondamentale, mais pas toutes les représentations conceptuelles ont un mot de contenu correspondant »

prise ou pré-stroke, où la main retourne vers sa position initiale et attend que la parole soit produite avant de réaliser le « stroke » du geste (le « stroke » étant également ici considéré comme la partie significative du geste) ; et (ii) la phase suivant la tenue ou « post-stroke », où la main reste inerte après que le « stroke » ait été complètement réalisé et jusqu'à ce que la parole ait été complètement produite. La phase de « pré-stroke » peut ainsi être utilisée afin de réaliser une relation de cohérence entre parole et geste, tandis que la phase de « post-stroke » va donner au locuteur l'opportunité de prolonger le geste avec la parole si ce premier a été trop rapide.

Dans la partie « pré-stroke », De Ruiters fait l'hypothèse que le Sketch va être envoyé à ce moment là au Planificateur de geste avant la fin de la construction du message préverbal. Ce temps permet ainsi au Planificateur de geste de préparer le programme moteur et de préparer la main à la phase de « stroke », en lui faisant reprendre sa place initiale. Une fois le message envoyé au Formulateur, un signal retour va être renvoyé au Conceptualisateur, qui va envoyer le reste du programme moteur aux unités motrices. Dans la partie « post-stroke », une fois que le message est terminé, le Conceptualisateur va ensuite envoyer un signal de retrait au Planificateur de geste.

Ainsi, selon le modèle évoqué par De Ruiters, la parole et le geste sont traités indépendamment et en parallèle, sauf dans le conceptualisateur, et les gestes ne peuvent pas être interrompus pendant la phase de préparation du geste, quand, par exemple, il y a un problème pour générer la parole.

Par comparaison avec la théorie « growth-point » (GP) de McNeill (1992), le modèle Sketch suppose que geste et parole prennent leur base à partir de la même représentation, soit le GP pour McNeill ou l'intention communicative pour De Ruiters. De la même façon, ces deux modèles s'accordent pour dire qu'ils sont planifiés par les mêmes processus. Néanmoins, De Ruiters soulève le fait que McNeill n'explique pas comment les GPs se développent dans les gestes et la parole, et peut par conséquent présenter un problème de circularité. De Ruiters conclut que la théorie GP ne donne pas d'explication sur la façon dont le système geste et parole réalise la synchronie observée par l'auteur.

Par comparaison avec le modèle de Krauss et al. (2000), qui s'inspire comme De Ruiters de la théorie de Levelt, De Ruiters a modifié de façon conséquente les



fonctions du conceptualiseur. Pour Krauss et al., les gestes sont générés par le conceptualiseur, par le processus séparé appelé sélecteur de trait spatial/dynamique, que nous avons évoqué plus tôt. Ces auteurs expliquent ainsi la synchronisation par le fait que le contrôle auditif peut terminer le geste en percevant l'affilié lexical (parlé). A l'inverse, dans le modèle sketch, le contrôle peut utiliser la parole et l'output de l'encodeur phonologique (« parole intérieure ») pour terminer les gestes. De plus, ce modèle de Krauss n'est pas en mesure d'expliquer tous les phénomènes de synchronisation, particulièrement les phénomènes de prise pré-stroke ou post-stroke, cités par De Ruitter dont nous avons parlé précédemment, dans la mesure où selon eux, les gestes sont terminés une fois que l'affilié lexical correspondant a été produit, et que c'est spécifiquement la phase de pré-stroke qui indique que le geste peut également être initié en synchronisation avec la parole.

En résumé, selon McNeill, le modèle de De Ruitter exclut le contexte, ne permettant pas, lors du traitement de l'information de récupérer ce que le contexte devrait avoir été, un contexte où l'instrument et la direction constituaient le contenu véhiculé.

#### **II.14.1.3. Modèle de Cassel et Prevost : lien post-conceptualiseur**

Citons pour mémoire que McNeill mentionne également une étude de Cassel et Prevost (1996), qui se basent sur un « lien post-conceptualiseur », où la composante du geste est liée à la structure de traitement de l'information à l'étape équivalente du formulateur, appelé également « planificateur de phrases ». Ce planificateur de phrase a pour fonction de trouver le rhème et de choisir un geste pour le représenter, en mettant l'accent sur le verbe qui correspond au geste.

McNeill conclut sur ces trois modèles en mettant en évidence le fait que tous ces modèles excluent le contexte, ayant pour conséquence d'agrandir le fossé qui sépare réalisations écologiques et réalisations guidées par un modèle. McNeill suggère que parole et geste sont deux réalisations ou manipulations du contexte.

Ainsi, après avoir mis en évidence ces différents modules, discutés par McNeill, voyons quelles solutions compte apporter ce dernier avec sa théorie « Growth Point » (GP).

#### **II.14.1.4. Modèle de McNeill : la théorie Growth Point (GP)**

Iverson (2006) explique dans son plaidoyer pour les travaux de McNeill, que ce dernier focalise sur la question de comment le geste participe activement dans le processus de production de parole et dans la pensée. Alors que Morrel-Samuels et Krauss cités plus tôt proposent une explication où le geste facilite l'accès lexical, la théorie de McNeill, également soutenue par Goldin-Meadow, propose que les gestes facilitent la pensée. Geste et parole seraient ainsi des manifestations co-expressives d'un système intégré.

Sa théorie repose sur plusieurs axes fondamentaux, avec l'idée que chaque idée qui va naître chez un individu constitue un « growth point » (GP), qui constitue une opposition dialectique entre deux modes de structuration du sens : d'un côté un mode spatial, analogique, holistique et imagé, et d'un autre côté un mode séquentiel, digital, combinatoire et linguistique.

Le GP a, selon McNeill, la capacité de prédire les points de synchronisation parole-geste. Il part du principe que les énoncés produits par les individus contiennent du contenu en dehors de leur propre structure. Cette faculté de prédiction va induire en retour la validité du GP, dans la mesure où il peut expliquer l'occurrence des événements geste-parole à un haut degré de précision temporelle. Nous pouvons entrevoir ici le problème de circularité que soulevait plus haut De Ruiter.

McNeill considère que geste et parole sont systématiquement organisés en relation l'un avec l'autre, et que le geste a du sens. Ils forment des combinaisons significatives, souvent non redondantes, avec les segments de parole avec lesquels ils sont synchronisés. Le terme qu'emploie McNeill pour caractériser ce type de relation est « co-expressif ». Le geste et sa parole co-expressive synchronisée exprime la même unité idée sous-jacente, mais n'expriment pas forcément des aspects identiques de celle-ci.

McNeill utilise la notion de « growth point » (GP) pour une unité analytique – une unité psychologique minimale suivant Vygotsky– qui combine imagerie et contenu catégoriel linguistique. Ces GPs sont inférés à partir de la totalité des événements communicatifs en focalisant plus particulièrement sur la synchronie geste-parole et la co-expressivité.

Si nous résumons en d'autres termes cette théorie, pour McNeill, gestes, parole, pensée et expérience ne forment qu'un seul et unique système.

Iverson (2006) ajoute, au sujet de la théorie Growth Point, que l'identification des contrastes de sens qui sont à la base d'un GP exploite l'évidence fournie par le contenu sémantique et la construction grammaticale de la parole co-expressive, mais exploite également l'événement de synchronisation entre geste et parole, la forme du mouvement du geste co-expressif de la parole, le timing de la préparation du geste, par exemple.

Selon McNeill les points de synchronisation peuvent survenir à différents moments de la tenue gestuelle : (i) lors du stroke, où l'image matérialisée par le geste va se synchroniser automatiquement avec les catégories linguistiques qui font partie du GP ; (ii) lors de la tenue pré-stroke : si dans un seul énoncé, deux contextes se chevauchent, le contexte incarné dans le point de croissance, va pouvoir être retardé alors que l'autre contexte va être incarné sous une forme lexicale ; (iii) dans le début de la préparation : au moment même où la prochaine unité GP va commencer à se matérialiser. Cette synchronisation entre phase préparatoire et phrase de stroke gestuel semble soutenir son modèle GP.

McNeill insiste sur le fait que chaque énoncé contient un contenu hors de sa propre structure, et que cet autre contenu fait le lien entre l'énoncé et le contexte, par la pensée. C'est pour cette raison entre autres que McNeill rejette absolument la thèse de la modularité.

McNeill et Duncan (2000) résume la notion de GP en mettant en évidence qu'il ne s'agit ni d'un mot, ni d'une image, mais plutôt d'une imagerie globale et de catégories linguistiques en même temps, et il explique son origine en cherchant à modéliser le processus même de fabrication de l'énoncé :

Its essential feature is a dialectic of these forms of thinking, and it gives rise to speech and gesture through their collaboration (...) Speech-gesture synchrony is therefore explained genetically, as an inevitable consequence of how the idea unit itself took form and its resistance to interruption during unpacking. (2000:148)

Depuis McNeill (1992), l'idée que le geste, avec le langage, aide à constituer la pensée s'est véritablement répandue. Le modèle que nous venons de voir repose sur l'hypothèse de mise en paquets d'information (IPH : *Information Packaging Hypothesis*). Pour résumer, le geste reflète la représentation mentale imagée activée au moment de la parole, et il est impliqué dans la planification conceptuelle du

message à verbaliser, dans le fait que les gestes aident les locuteurs à mettre en paquet l'information spatiale dans des unités appropriées pour la verbalisation. Le geste joue donc un rôle prépondérant dans le processus de production de parole car il joue un rôle dans le processus de conceptualisation. Par conséquent, selon cette vision, le geste joue un rôle dans la pensée.

C'est cette théorie qu'ont voulu tester Alibali, Kita et Young (2000) en focalisant sur le rôle fonctionnel du geste lors du processus de production de parole. Ils confrontent cette prédiction faite par McNeill à une autre hypothèse répandue, celle de la récupération lexicale (LRH : *Lexical Retrieval Hypothesis*). En quelques mots, cette hypothèse considère que le geste soit impliqué dans la génération des formes de surface des énoncés, il aurait donc un rôle actif dans l'accès lexical, spécialement pour les mots à contenu spatial. Ils joueraient alors un rôle direct dans le processus de parole. Alibali et ses collègues ont donc voulu tester ces deux hypothèses afin d'éclaircir l'implication réelle du geste dans la génération de formes de surface de l'énoncé, et dans la planification conceptuelle des messages à travers deux tâches expérimentales, une tâche de conservation (explicative) et une tâche de description chez 22 enfants. Les prédictions sont les suivantes : selon LRH, le geste est impliqué dans la génération de formes de surfaces des énoncés, donc la production gestuelle lors de ces deux tâches devrait être comparable. Selon IPH, le geste a un rôle dans la planification conceptuelle, donc la production gestuelle dans les deux tâches devrait être différente : la tâche d'explication induirait plus de gestes représentationnels pour explorer les dimensions des objets, et plus de gestes non redondants afin de considérer les multiples dimensions liées aux objets.

Les résultats indiquent que les formes de surface des énoncés parmi les tâches sont comparables. Les enfants produisent plus de gestes substantifs et non-redondants dans la tâche d'explication, ce qui soutient la vision que les gestes spontanés sont impliqués dans la planification conceptuelle des énoncés, et ce qui est donc cohérent avec IPH et incohérent avec LRG. Les résultats laissent ouverte la possibilité que les gestes puissent être impliqués dans la récupération lexicale, mais ils montrent tout de même que la récupération lexicale n'est pas le seul endroit dans le processus de parole où le geste est impliqué. Les découvertes suggèrent fortement que les gestes substantifs sont générés lorsque l'information spatiale est « packagée » dans une seule unité informative qui doit être verbalisée. Cela est

également compatible avec l'idée que le geste joue un rôle facilitatoire dans ce processus et qu'il puisse jouer alors un rôle dans la pensée et la mémoire.

Alibali et ses collègues interprètent leurs résultats en invoquant le rôle du geste dans la production de parole, un rôle allant au-delà de la récupération lexicale. L'action même de réaliser un geste peut aider un locuteur à organiser l'information spatiale dans le but de la verbaliser. Le geste jouant ainsi un rôle dans la conceptualisation du message qui doit être verbalisé. De plus, le geste peut ne pas seulement jouer un rôle dans la production de parole, mais plus largement dans d'autres activités cognitives plus générales.

A la question de savoir de quelle façon le geste est impliqué dans la récupération lexicale, nous citerons les travaux réalisés par Krauss (1998). Cet auteur estime que si les gestes lexicaux aident le processus d'accès lexical, on en trouverait beaucoup plus, accompagnés de de plus de pauses à des non-jonctions de phrases grammaticales, dans la parole spontanée que dans la parole mémorisée. Il a pu démontrer que la durée d'un geste lexical est étroitement liée à la quantité de temps prise par le locuteur pour accéder à son affilié lexical. Krauss montre également que le fait d'empêcher les locuteurs de faire des gestes rend l'accès lexical plus difficile, avec une parole ralentie, spécifiquement lorsque le contenu est spatial, des pauses plus nombreuses, des mots incomplets, répétés, des phrases recommencées, ainsi que de nombreuses pauses. Cela soutient l'hypothèse que les gestes lexicaux facilitent cet accès lexical. Ce qui reste moins clair, c'est à quel point du processus le geste accomplit-il cela. Il existe, selon Krauss, trois points du processus de production de parole où les gestes peuvent jouer un rôle : (1) dans la conceptualisation, le geste peut aider à formuler le concept qui va être exprimé dans la parole, (2) le codage grammatical où l'information contenue dans le geste peut aider à transformer le concept en lemme dans le lexique et enfin (3) le codage phonologique où le geste peut aider à récupérer la forme du mot ou lexème. Ces trois points constituent trois facettes du processus qui peuvent affecter la parole en rendant lente ou dysfluente.

#### **II.14.1.5. Modèle de Kita et Özyürek: l'hypothèse d'interface**

Kita et Özyürek (2003) se sont plus récemment attachés à comprendre les processus cognitifs qui sont à la base même des gestes réalisés de façon spontanée

lors de production de parole. Ils mettent en évidence les différences reliés aux différents modèles présentés dans la littérature sur des questions telles que comment les gestes sont-ils reliés sur un plan informatif au contenu de la parole concurrente, ou à quel niveau du processus de production de parole, le contenu du geste est-il déterminé. Afin de mieux comprendre l'enjeu de telles questions, les auteurs confrontent alors trois hypothèses dans une étude comparative de production gestuelle réalisée par des locuteurs japonais, turques et anglais. Cette étude va également nous permettre de resituer entre elles les études citées auparavant et d'en dégager plus clairement les atouts et désavantages.

#### **II.14.1.5.a) Hypothèse d'imagerie libre (HIL)**

La première hypothèse qu'il teste est l'hypothèse d'imagerie libre (*Free Imagery Hypothesis*), soutenue par des auteurs tels que De Ruiter ou Krauss et al. Selon celle-ci, les gestes sont générés à partir de l'imagerie de la mémoire de travail, et leur contenu se construit sur la base de la mémoire, des événements à long terme ou d'autres processus de pensée. Les gestes sont donc générés « pré-linguistiquement », indépendamment du potentiel de représentation du langage. Krauss et al. considèrent que les gestes sont générés à partir de l'imagerie spatiale dans la mémoire de travail, activée au moment de la parole, tandis que De Ruiter propose que les gestes représentationnels soient générés par le processus qui génère aussi la parole, à savoir le conceptualisateur, qui produit un message pré-verbal qui va être nourri dans le module de formation linguistique. Toutefois leurs modèles sont similaires dans le fait que les gestes sont générés avant que le processus de formation linguistique ne se produise. Cette hypothèse HIL prédit alors que l'information codée dans un geste n'est pas influencée par la manière dont l'information pourrait être verbalement exprimée.

#### **II.14.1.5.b) Hypothèse lexico-sémantique (HLS)**

La seconde hypothèse est l'hypothèse lexicale sémantique (*Lexical Semantic Hypothesis*), soutenue par des auteurs tels que Butterworth et Hadar (1989). Elle consiste à affirmer que les gestes sont générés à partir de la sémantique des items lexicaux dans la parole qui accompagne le geste. Butterworth et Hadar soutiennent qu'un item lexical génère un geste iconique à travers un ou plusieurs de ses traits sémantiques qui peuvent être interprétés dans l'espace. Les gestes iconiques sont générés à partir du produit de l'étape computationnelle dans la production de parole,

après qu'un item lexical abstrait ait été sélectionné à partir d'un lexique qui serait alors organisé sémantiquement. La prédiction de cette hypothèse serait que les gestes dits représentationnels ne codent pas ce qui n'est pas codé dans la parole concurrente.

#### **II.14.1.5.c) Hypothèse d'interface (HI)**

Cette dernière hypothèse (*Interface Hypothesis*) est celle présentée par Kita et Özyürek (2003). Elle propose que les gestes soient originaires d'une représentation d'interface entre parole et pensée spatiale. Cette représentation est en soi une représentation motrice, c'est-à-dire qu'elle code l'information sur l'action, et est organisée dans le but de production de parole. Selon leur vision, les gestes ne coderaient pas seulement les propriétés spatio-motrices (et par conséquent non-linguistiques) du référent, mais coderaient également la structure de l'information sur le référent de façon relativement compatible avec des possibilités de codage linguistique. En d'autres termes, pour parler, l'information qui doit être exprimée, doit être travaillée pour le dessein de la parole, impliquant l'idée que le fameux « *thinking for speaking* » de Slobin (1987, 1996) soit nécessaire. L'information complexe et riche doit ainsi être organisée en plus petits paquets d'information, afin que chaque paquet conserve une complexité informative appropriée pour la verbalisation dans une seule unité de traitement pour la production de parole. Cet argument repose fondamentalement sur la théorie *Growth Point* de McNeill, dans la mesure où la planification de l'énoncé implique un jeu entre pensée imagée et pensée linguistique. De ce point de vue le produit de la pensée imagée se manifesterait par le geste, et le produit de la pensée linguistique se manifesterait par la parole co-expressive. Les gestes seraient donc impliqués dans le processus de mise en paquets (*packaging*) de l'imagerie spatio-motrice en unités d'information disponibles pour la production de parole. Cette hypothèse prédit alors qu'un geste soit formé à partir des possibilités de formulation linguistique (ce qui va à l'encontre de HIL), et en même temps que ces gestes vont pouvoir coder l'information spatio-motrice non exprimée dans la parole (ce qui va à l'encontre de HSL).

Dans le but de tester ces différentes hypothèses, les auteurs les ont testées lors d'une comparaison inter-langues (anglais américain, turc et japonais), par l'analyse des gestes produits dans des narrations obtenues à partir d'un même stimulus, à savoir le récit de l'histoire d'une aventure de Titi et Grosminet fait à une

personne qui ne l'aurait pas vu. Les prédictions annoncées pour chacune des hypothèses sont les suivantes : H1 prédit que les gestes vont mettre en relief l'influence des possibilités de formulation linguistique et vont régulièrement coder certains détails spatiaux non exprimés verbalement dans la parole, puisque les gestes sont générés à partir de représentations imagées des référents événements. H1L ne prédit aucune différence inter-langue dans le contenu gestuel, les gestes codant de façon constante les détails spatiaux non exprimés verbalement. Et H3L prédit que les gestes vont refléter des différences dans les possibilités de codage linguistique dans les trois langues, mais que les gestes ne codent pas régulièrement les détails spatiaux non exprimés verbalement.

Les résultats font apparaître l'existence d'une variation entre les différentes langues dans l'utilisation des gestes iconiques. Les gestes représentent un événement spatial de façon similaire à la parole, mais les gestes incluent en plus des détails spatiaux qui ne peuvent pas être exprimés dans la parole concurrente. Le japonais et le turc ont typiquement besoin de deux verbes pour exprimer la manière et la trajectoire contrairement à l'anglais.

Ces résultats permettent aux auteurs d'affiner leur modèle afin de spécifier comment le contenu d'un geste représentationnel est déterminé. Le modèle de Kita et Özyürek est construit sur celui de Levelt (1989), et comme tel il comporte une distinction fondamentale entre processus de planification au niveau conceptuel (le conceptualisateur) et processus de formulation de parole (le formulateur) pour aboutir à une étape finale d'articulation. Selon Levelt, le conceptualisateur transforme l'intention communicative en une représentation propositionnelle appelée « message préverbal », qui entre (*feed*) le formulateur. Ce dernier récupère alors les items lexicaux sur la base de spécifications conceptuelles du message pré-verbal et spécifie ensuite la réalisation syntaxique, morphologique et phonologique d'un énoncé. Plus précisément, le formulateur transforme le message préverbal de deux façons. D'une part le codeur grammatical transforme le contexte qui doit être lexicalisé en un lemme dans le lexique mental, dont les significations sont liées au contenu du message préverbal. Ainsi en utilisant l'information syntaxique du lemme, la structure conceptuelle est transformée en structure de surface. D'autre part, en accédant aux formes de mots stockées dans une mémoire lexicale et en construisant un plan approprié pour la prosodie de l'énoncé, le codeur phonologique transforme



cette structure de surface en un « plan phonologique » ou ensemble d'instructions pour le système articulatoire. L'output de l'étape articulatoire étant la parole, avec un locuteur qui la contrôle et qui l'utilise comme une source de feedback correctif.

Dans le modèle de Kita et Özyürek, le conceptualisateur de Levelt est coupé en deux parties : une partie qui planifie la communication et génère une « intention communicative » (la macro-planification de Levelt), et qui détermine quelles modalités d'expression vont être impliquées ; et une autre partie qui génère le message (la micro-planification de Levelt) et qui consiste à formuler une proposition qui doit être verbalement formulée en prenant en compte le but communicatif de l'énoncé et le contexte discursif.

Dans cette perspective, le contenu d'un geste devra donc être déterminé par : (i) l'intention communicative générée dans le planificateur de communication ; (ii) les schémas d'actions sélectionnées sur la base de traits, ou imaginés en espace réel ; et (iii) un feedback on-line du formulateur via le générateur de message. Ces trois contributions vont déterminer conjointement le contenu gestuel, ce contenu gestuel n'étant alors pas pleinement spécifié dans les mécanismes dédiés à la communication, comme pour le conceptualisateur de Levelt, mais spécifié dans un mécanisme plus général de génération de l'action. De plus leur modèle présente la caractéristique d'échanger l'information on-line entre le générateur de message et le générateur d'action d'une part, et le formulateur et le générateur de message d'autre part. Cela permet ainsi au contenu gestuel d'être formé « en ligne » à partir des possibilités de formulation linguistique.

Ce modèle s'inspire fortement de la théorie du *Growth Point* proposée par McNeill, élaborée par McNeill et Duncan (2000), dans le sens où geste et parole sont co-expressifs et leurs plans sont formés par la dialectique entre expression linguistique et représentation spatio-motrices. Selon Kita et Özyürek, les gestes sont générés à partir d'un mécanisme de génération d'actions génériques utilisées à des fins communicatives ; une idée en contraste avec De Ruiter (2000), McNeill (1992) ou McNeill et Duncan (2000), qui proposent que le geste soit généré par un mécanisme dédié uniquement à la communication.

De la même façon, De Ruiter (2000) proposait que le contenu gestuel soit pleinement spécifié dans le conceptualisateur (au sens de Levelt), basé sur l'intention communicative. En contraste, Krauss et al. (2000) proposaient que

l'intention communicative ne joue pas de rôle dans la détermination du contenu gestuel dans la plupart des gestes. Kita et Özyürek proposent une solution alternative : l'intention communicative spécifierait seulement de façon grossière le domaine d'information qui doit être exprimé, et l'information spatio-motrice effectivement réalisée, venue du générateur d'action génériques, pourra inclure l'information qui fait partie de l'intention communicative.

Enfin, ce modèle diffère de celui de Levelt (1989) dans le fait qu'il existe dans le modèle de Kita et Özyürek un feedback direct à partir du formulateur au niveau de la planification conceptuelle de la parole.

#### **II.14.1.6. Geste + parole : un bénéfice pour le locuteur mais également pour l'interlocuteur**

Une vision alternative de cet ensemble geste-parole envisage le bénéfice destiné à l'interlocuteur plutôt qu'au locuteur. Les gestes sont alors considérés comme un soutien aux activités conjointes communicatives et informatives pour les auditeurs. Le locuteur et l'interlocuteur sont engagés dans un projet conjoint de communication et les gestes font partie intégrante de l'utilisation du langage. Il est également fréquent que les interlocuteurs aient à se rattraper sur des déclencheurs gestuels lors de situations particulières, comme dans le fameux « effet cocktail-party » où l'environnement bruyant rend la parole ambiguë et nous force bien souvent à nous focaliser sur les lèvres ou sur les mouvements gestuels de notre interlocuteur afin de comprendre le message.

Lors de ces situations, il est un geste qui paraît encore plus intrinsèquement lié à la parole, le geste déictique, souvent accompagné d'une expression déictique linguistique (*this, that, here, there...*). Généralement le geste de pointage de l'index sert alors à déplacer l'attention de l'auditeur vers une région proche du référent qui intéresse le locuteur. Si tous deux ont leur attention est focalisée sur une région similaire, alors la résolution de la référence est, sinon automatique, du moins grandement facilitée.

Par ailleurs, dans l'hypothèse formulée par Louwerse et Bangerter (in press), en plus que les gestes déictiques aident l'auditeur à identifier la cible indirectement, en guidant leur regard vers sa région, soit en établissant un focus conjoint d'attention entre le locuteur et l'interlocuteur, leur hypothèse prédit que ces gestes déictiques

sont substituables à certaines expressions linguistiques spatiales. Leurs résultats montrent que les participants à leur expérience passent plus de temps sur une cible correcte, quand le pointage est présent, ou lorsque la description de la localisation est vocalisée, mais pas lorsque les deux sont combinés : ce qui soutient cette hypothèse de substitution ou de vicariance.

## **II.14. La coordination temporelle geste-parole**

### **II.14.1. Levelt... ou la voix attend le doigt**

L'étude pionnière sur ce sujet est celle de Levelt et al. (1985), première étude à poser ce problème dans un cadre expérimental avec suivi du mouvement. Ils s'intéressent à la synchronisation du geste et de la parole, et particulièrement à l'interdépendance temporelle entre ces deux modules lors de la planification motrice et lors de l'exécution motrice. Ils opposent deux conceptions théoriques : la conception 'interactive' et la conception 'balistique', en démontrant que cette dernière semble être la plus plausible pour expliquer ce phénomène. Levelt et al. étudient la relation de synchronisation entre la parole et les gestes, spécifiquement une classe particulière de gestes directement reliés à la parole, les gestes déictiques. Levelt et al. soulignent le caractère prioritaire de l'utilisation de cette classe de gestes déictiques, dans la mesure où ils sont très dépendants du message linguistique qui co-survient, et sont mesurables clairement, grâce à un 'apex' temporel très marqué (le maximum d'extension de l'index au bout du bras).

Dans cette étude, les auteurs ont mené 4 expériences distinctes mais qui traitaient toutes de la coordination du geste de pointage avec les énoncés déictiques. Nous ne verrons ici que les 2 premières expériences, qui sont les plus pertinentes pour notre étude. Les expériences consistaient à obtenir des gestes de pointers de l'index et/ou des réponses verbales à la présentation de stimuli qui se situaient à différentes localisations. Ils ont relevé pour cette étude trois mesures principales pour la performance du geste : (1) le temps (temps de réaction ou TR) d'initiation du mouvement (le moment où, après présentation du stimulus, le doigt se lève), (2) le temps (TR) de l'apex (le moment où le doigt est porté par le bras à son extension maximale) et (3) la durée d'exécution (ou MT *Movement Time* : différence entre le temps d'apex et le temps d'initiation).

Dans leur première expérience, ils observent une parfaite synchronisation entre geste et parole dans le champ ipsilatéral, mais cette tendance ne se retrouve pas dans le champ contralatéral. Dans la condition on-line (le sujet réagit immédiatement après l'allumage de la diode), le temps d'exécution du geste déictique n'est pas synchronisé avec l'énoncé linguistique déictique. Levelt et al. soulignent alors le fait que ces résultats peuvent être dus à la possibilité que le locuteur essaie d'ajuster le mot déictique au geste de pointage de l'index.

L'expérience 2 nous intéresse plus particulièrement pour notre travail. Il s'agit de comparer les conditions parole seule, geste seul, et geste/parole ensemble, afin de mettre en évidence les effets d'une modalité monstrative sur une autre. Le dispositif expérimental consistait en 2 LEDs, une proche et une éloignée, auxquelles les sujets devaient faire référence par la voix, par le geste ou par les deux simultanément en utilisant les expressions néerlandaises « dit lampje » (*this light*) et « dat lampje » (*that light*). Dans la condition geste seul, les expérimentateurs ont demandé aux 12 sujets de pointer uniquement avec la main droite.

Dans la condition on-line de cette expérience 2, on observe les mêmes résultats que dans l'expérience précédente, c'est-à-dire pas de synchronisation geste/parole, la parole arrivant plus tard que l'apex du geste.

Les résultats de l'expérience montrent dans un premier temps qu'il existe un degré de synchronisation entre geste et parole, le début de la parole survenant plus tardivement pour une cible lointaine et contralatérale plutôt que proche et ipsilatérale. On observe qu'en tâche simple, le temps de réaction (TR) vocal est plus rapide que le TR manuel, mais qu'en tâche double, le TR vocal est très retardé par rapport au TR manuel, qui lui n'a pas subi de forts changements.

De ce fait, on ne peut donc pas affirmer que les deux systèmes de production sont indépendants l'un de l'autre. Les auteurs observent également que dans la condition où geste et parole sont réalisés ensemble, l'initiation du mouvement est significativement retardée en présence de la parole, de quelques 14ms. Ce résultat permet aux auteurs de suggérer que ce retard dans l'initiation du mouvement peut être la conséquence de la planification de la parole par le sujet. Ce qui permet d'argumenter en faveur de la théorie balistique, c'est le fait que le mouvement entier de pointage soit quelque peu retardé mais la durée de ce geste n'est pas en soi affecté par la présence de la parole. Le mouvement de pointage est en effet initié

légèrement plus tôt en l'absence de la parole ou dans une situation où les cibles sont indiquées par la même expression verbale. En d'autres termes, la préparation de la parole n'affecte que légèrement la préparation du geste lors de la phase de planification, mais n'a plus aucune conséquence dans la phase d'exécution du geste. Les auteurs concluent donc que « la parole et le geste déictique sont interactifs dans la phase de planification, mais bien proche du [modèle] balistique dans la phase d'exécution »<sup>24</sup>. Les auteurs envisagent donc ce phénomène comme une adaptation de la parole au geste plutôt que l'inverse. Ainsi la voix s'adapte au geste mais le geste ne semble pas s'adapter à la voix.

Nous utiliserons une bonne partie de ces résultats, notamment les durées de mouvement (MT), pour situer les performances de nos enfants par rapport à ces données sur l'adulte.

## II.14.2. Holender

Ces résultats de Levelt et al. sont similaires à ceux trouvés par Holender (1980) quelques années plus tôt. Même si les deux études ont toutes deux vocation à étudier la relation entre geste et parole, les méthodes expérimentales sont différentes et elles ne se prêtent pas à une comparaison directe. En effet, Holender a étudié l'interférence entre une réponse vocale et une réponse manuelle pour un même stimulus, en demandant à ses sujets de réaliser une tâche de nomination (d'une lettre L, N, R ou S) et une tâche de pression de touche en réponse à un stimulus qui était une de ces lettres – tâche bien différente de Levelt et al. qui utilisent des productions manuelle et vocale déictiques. Sa première expérience nous montre que lors de tâches simples, le TR vocal est toujours plus rapide que le TR manuel, alors que dans la tâche double (nomination + pression), le TR vocal est plus lent que le TR manuel de façon constante, sans que le temps du geste ne soit altéré. La parole semble également être retardée et attendre le geste, comme montré par Levelt et al (1985). Holender en déduit que la réponse vocale ne peut pas être libérée avant un délai minimum après l'initiation de la réponse manuelle. Dans la mesure où dans la tâche simple, la parole est plus rapide, Holender suggère que lors de la tâche double, les sujets doivent alors retarder leur réponse vocale afin de la synchroniser au mieux avec la réponse manuelle.

---

<sup>24</sup> « ... *Speech and deictic gesture are interactive in the planning phase, but well-nigh ballistic in the execution phase.*” (Levelt et al., 1985, p.162)

Ces résultats sont confirmés dans une deuxième expérience réalisée par Holender qui demandait aux sujets de réaliser soit une double performance, soit une synchronisation des réponses dans la tâche double. Pour le groupe soumis à la condition de réponse rapide, pour chaque réponse dans la tâche double comme pour les réponses dans les tâches simples correspondantes, on obtient le même patron que celui trouvé dans l'expérience précédente : la réponse vocale est significativement plus lente que la réponse manuelle, la parole étant en retard d'en moyenne 83ms sur le geste. L'auteur ne trouve pas par contre de différences significatives entre le groupe soumis à la tâche de double performance et le groupe soumis à la tâche de synchronisation des réponses.

Les résultats montrent clairement, de la même façon que Levelt, que la parole attend le geste et que celle-ci est fortement retardée en double tâche, à cause de la charge de traitement augmentée, alors que l'on n'observe presque aucun effet sur le mouvement manuel.

Holender conclut sur ses données en proposant que les réponses de nomination et de pression de touche sont réalisées par des traitements indépendants : lorsque ces deux modalités sont utilisées en coordination, ces traitements sont en compétition dans le but d'établir un espace de ressources commun aux deux<sup>25</sup>.

### **II.14.3. Feyereisen**

Une autre étude s'est attaquée à la question de la relation temporelle geste/parole, toujours dans un paradigme de type double tâche : c'est une étude de Feyereisen (1997), qui a tenté de répliquer, d'une certaine manière, et d'étendre les résultats obtenus par Levelt et al. (1985).

La première expérience de Feyereisen consistait: soit (i) en une tâche simple, pointer ou vocaliser [ti] ou [ta] ; soit (ii) en une double tâche, où le sujet devait appuyer sur une touche pour activer la présentation du stimulus, puis la relâcher pour effectuer un mouvement de pointage à la présentation du stimulus, une croix apparaissant d'un côté ou de l'autre sur un écran d'ordinateur, exécutant ce geste tout en vocalisant soit [ti], soit [ta], pour le stimulus gauche et droit respectivement.

---

<sup>25</sup> « ... when used together , these processors compete for a common processing capacity pool... » (Holender, 1980, p. 431)

Les sujets ont été divisés en deux groupes, le premier pointait avec la main droite, et le second avec la main gauche. Une des motivations pour cette expérience était de se demander s'il existait des différences de latéralité dans la relation geste/parole. Dans cette expérience, Feyereisen n'a malheureusement pas pu répliquer l'effet de l'avantage du côté ipsilatéral (réponses plus rapides de la main droite pour le champ droit et idem pour la gauche) trouvé chez Levelt et al. (1985). Cependant, Feyereisen a réussi à répliquer l'effet trouvé chez Levelt et al. (1985) en ce qui concerne les temps d'initiation de réponses gestuelle et vocale retardés dans la double tâche. En effet, dans son expérience le geste de pointage est en avance en moyenne de 183ms sur la parole. Toutefois, Feyereisen n'a observé aucun effet d'asymétrie manuelle en relation avec la spécialisation hémisphérique dans le traitement du langage

Dans leur article, Levelt et al. (1985) se demandaient si les résultats auraient été similaires s'ils avaient utilisé une autre classe de gestes, comme par exemple les gestes iconiques, que l'on retrouve fréquemment en accompagnement de la parole dans le discours. Feyereisen a testé cette hypothèse dans une deuxième expérience où il remplace simplement le geste de pointage de l'index par un geste illustratif à l'apparition du stimulus. Les sujets devaient alors associer 4 formes de main à 4 symboles-stimuli (extension oblique de la main, opposition pouce-index, geste de quelqu'un prenant un cube de 5 cm, présentation du dos de la main, nommés respectivement verbalement par « barre », « boule », « bloc », « blanc »). Dans cette expérience, où plusieurs modifications ont été faites, Feyereisen trouve de façon inattendue et en contradiction avec sa première expérience, un effet significatif de latéralité de la main utilisée, l'activité de la main droite retardant la réponse vocale contrairement à la main gauche.

Feyereisen a donc réussi dans un premier temps à répliquer l'expérience 2 de Levelt et al. (1985), c'est-à-dire qu'il trouve des temps d'initiation oral et gestuel augmentés dans la tâche double par rapport aux conditions de tâche simple. Toutefois, dans sa deuxième expérience, Feyereisen a échoué à étendre cet effet à d'autres classes de gestes.

Cette étude soutient une fois de plus l'idée d'une compétition existant entre les deux modules voix/doigt lors de la phase de planification de la réponse. Feyereisen explique que les gestes précèdent de façon typique les mots ou syntagmes qu'ils

illustrent, et essaie de donner une piste de réflexion sur cette asynchronie geste/parole, dans le sens où la taille du lexique mental est plus grande que le répertoire des gestes. Il s'appuie sur les données de Morrel-Samuels et Krauss (1992), qui ont trouvé que l'asynchronie est reliée positivement à la familiarité du mot et que la durée du geste est reliée à ce facteur en sens inverse. Selon eux, l'idée importante est que les gestes peuvent faciliter l'accès au lexique. Selon Feyereisen, cette idée paraît tout de même contradictoire avec le fait que la performance du geste dans son expérience en double tâche ne fait pas chuter mais au contraire augmenter le temps de nomination. Il s'agirait alors d'un effet inverse à celui proposé par Morrel-Samuels et Krauss (1992) : les gestes ne faciliteraient pas l'accès lexical mais les énoncés de mots non familiers retardés devraient laisser une plus grande liberté à la production de gestes.

#### **II.14.4. Quand la main pointe la voyelle sur le visage**

Une autre étude nous paraît pertinente pour la compréhension de la relation geste-parole chez l'adulte, celle de Attina et al. (2004), réalisée au sein du laboratoire de l'Institut de la Communication Parlée, décrivant pour la première fois la coordination temporelle entre les différents articulateurs –lèvres et main– du Langage Parlé Complété (LPC) :

Le Langage Parlé Complété (LPC) est un augment naturel de la lecture labiale. Il est composé de clés digitales réalisées à l'aide de la main placée à différentes positions particulières sur le côté du visage afin de désambiguïser des syllabes de type CV. (2004:197)

Le LPC est formé de 2 paramètres, le positionnement de la main sur le visage ou autour et la forme de la main ou configuration des doigts. Tandis que le premier paramètre code les voyelles, le second établit une distinction entre consonnes. Attina et al. précisent que la main en elle-même n'est pas suffisante pour identifier les phonèmes de la langue, et l'information visible qui peut être lue sur les lèvres demeure une composante essentielle. Cependant combiner toutes ces informations permet d'identifier une syllabe de type consonne-voyelle. De nombreuses études se sont penchées sur le LPC en tant que tel, mais peu d'entre elles se sont intéressées plus particulièrement à la question de la relation entre visage et main d'un point de vue du contrôle moteur. C'est précisément ce qu'ont cherché à étudier Attina et al. Leur but était de préciser comment le geste de la main avec le visage co-produit l'information consonantique et vocalique. Pour répondre à cette question, ils ont



étudié l'organisation temporelle des augments manuels en relation avec l'organisation temporelle propre des lèvres et de l'acoustique de la parole.

Les résultats de leurs expériences ont montré que la mise en configuration de la main était achevée au début acoustique de la consonne, et que la cible labiale était atteinte largement après que la formation de la main ne soit terminée. Leurs résultats confirment une synchronisation de la cible manuelle vocalique sur le début acoustique de la consonne. Plus précisément nous pouvons résumer le patron de temporel anticipatoire de ces résultats en trois points. (1) Le mouvement de déplacement de la main vers sa position commence en anticipant de plus de 200 ms sur le début acoustique consonantique de la syllabe CV. Cela signifie que le début du geste se situe pendant la syllabe précédente, c'est-à-dire pendant la voyelle précédente. (2) La cible de la main a été atteinte autour du début acoustique de la consonne et anticipant là aussi largement sur la cible labiale vocalique. Enfin (3) la cible manuelle a été atteinte en moyenne 172-256 ms avant la cible labiale vocalique (voir figure II.44).

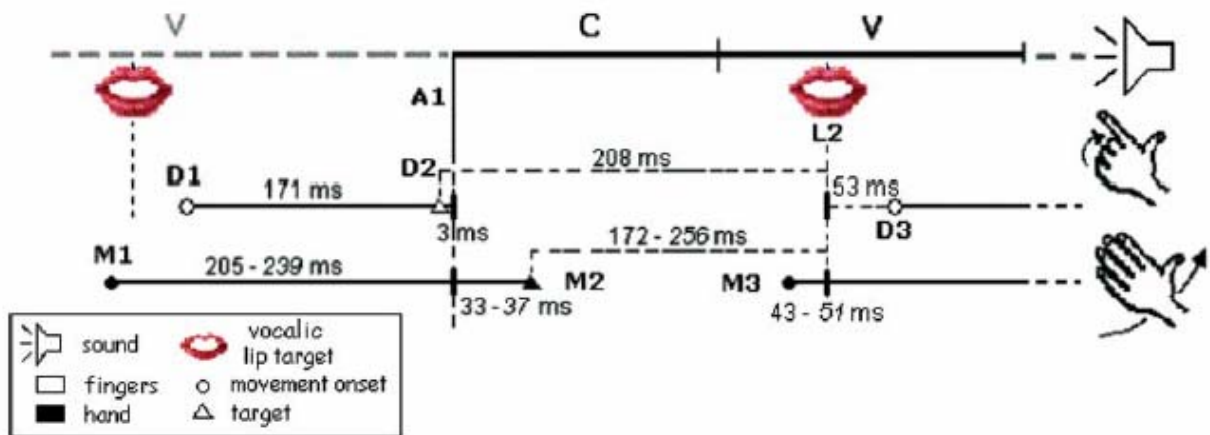


FIGURE II.44 : Patron temporel de la coordination entre son, lèvres, formation de la configuration des doigts de la main et placement de la main, pour la production du Langage Parlé Complété français (les valeurs en italiques sont celles des séquences avec consonnes). (D'après Attina et al., 2004, p. 208)

Afin de confirmer cette découverte, Attina et al. ont enregistré trois autres codeuses de LPC. Leurs résultats montrent que chez ces trois autres sujets, la main est largement en avance sur les lèvres et la synchronisation du placement de la main sur le début de la réalisation acoustique consonantique pour des séquences CV a été confirmée.

L'anticipation de la main LPC sur les lèvres (alors que la main, par construction de ce code inventé par Cornett en 1965, devrait désambiguïser les lèvres) optimise un contrôle moteur coordonnant des gestes compatibles, ici les contrôles de constriction des consonnes de la parole avec les positions (contacts) de la main pour les voyelles. Il n'y a sans doute pas simplement un phénomène général d'anticipation, mais un patron de coordination répondant aux exigences du cerveau amené à contrôler la main et la bouche dans une structure déjà linguistiquement organisée dès le contrôle de la parole.

### **II.14.5. Coordination oui, mais synchronisation... à voir**

Suite à l'étude de Morrel-Samuels et Krauss (1992), qui montrait que pour l'anglais, l'initiation du geste précède l'initiation de la parole affiliée, d'autres chercheurs, et particulièrement De Ruiter et Wilkins (1998), se sont posé la question de savoir si les gestes anticipent également la parole dans d'autres cultures linguistiques. Ils ont alors étudié cette relation dans une approche comparativiste entre le néerlandais et l'arrernte (une langue aborigène d'Australie, Wilkins, 1999). Ils cherchent à découvrir si la synchronisation des gestes et de la parole est la même pour tout locuteur humain. Pour cela, De Ruiter et Wilkins ont choisi ces deux langues et cultures gestuelles très différentes. L'arrernte est une langue à marque de cas avec un ordre des mots libre, alors qu'en néerlandais c'est l'ordre des mots qui signale les différentes fonctions grammaticales. De plus, en Arrernte, les gestes jouent apparemment un rôle plus important qu'en néerlandais, et l'excursion des gestes dans la langue des signes de cette culture utilise un espace plus grand que celui du néerlandais. Ces différences poussent les auteurs à croire qu'ils pourraient relever des différences dans la synchronisation entre geste et parole affiliée.

Leur étude a consisté en un visionnage de narrations vidéos de 4 locuteurs de chaque communauté, décrivant des situations ou événements dans leur environnement particulier respectif. Les expérimentateurs se sont plus profondément intéressés aux gestes qui fonctionnent pour coder des orientations spatiales réelles, relevant ainsi environ 15 excursions manuelles pour chaque locuteur, dont les auteurs ont décrit chaque composante. Pour chaque geste, avec sa phase préparatoire et sa détente ou « stroke », De Ruiter et Wilkins ont repéré la zone de la parole affiliée au geste: en d'autres termes la réalisation du mot ou syntagme relié

pragmatiquement et sémantiquement au geste. Ils ont recueilli ainsi 100 gestes pour l'arrernte et 80 gestes pour le néerlandais.

Leurs résultats montrent que dans 94 % des cas, les gestes précèdent la parole pour les locuteurs du néerlandais, et dans 88% du temps, il en est de même pour les locuteurs arrernte (voir figure II.45 ci-dessous).

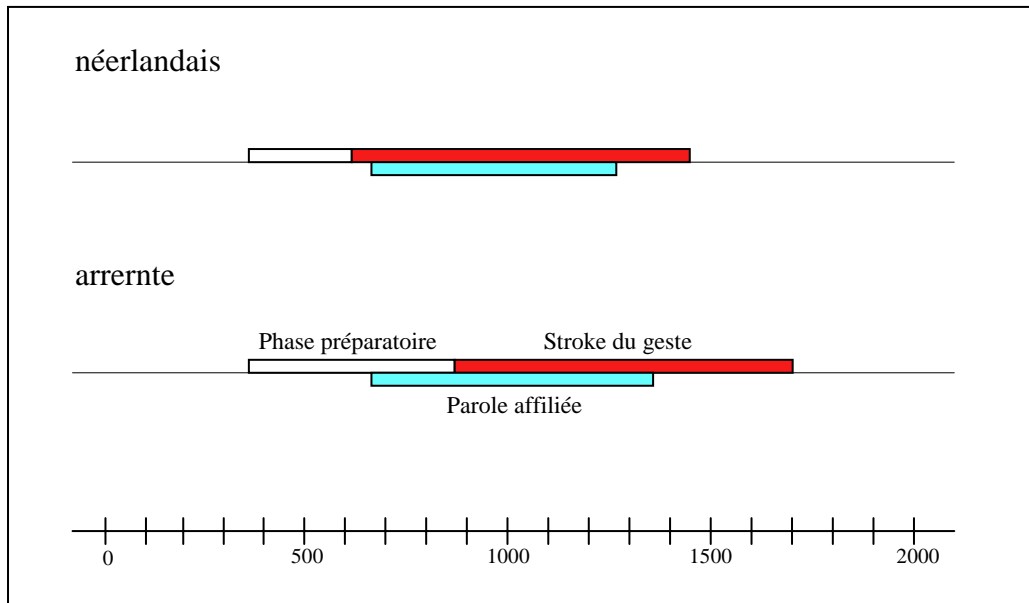


FIGURE II.45 : Comparaison du geste et de la parole affiliée dans le temps en néerlandais et arrernte. (D'après De Ruiter et Wilkins, 1998)

Mais les choses changent si nous considérons que la partie centrale du geste est le *stroke*. Nous pouvons alors remarquer que, si l'exemple du néerlandais peut être en faveur d'une synchronisation de l'initiation du *stroke* du geste avec celle de la parole, il n'existe pas de patron de synchronisation geste/parole, commun aux deux langues. L'exemple de l'arrernte nous permet clairement de mettre en évidence que cette stratégie peut être différente : ici nous observons que la parole affiliée précède le *stroke* du geste.

En arrernte, les locuteurs vont utiliser un espace gestuel plus grand, et les gestes vont avoir des mouvements de préparation plus longs, de l'ordre de 803ms comparé à 559ms en néerlandais. Pour expliquer ce temps de préparation plus long, les chercheurs tentent de trouver l'explication dans la théorie de Levelt et al. (1985), et expliquent que le geste et la parole sont initiés en synchronie, et qu'ensuite va

s'établir un dialogue (*crosstalk*) entre les deux modalités. Ainsi la conséquence devrait être que les gestes en arrernte soient retardés par rapport à la parole affiliée.

A la lumière de ces différentes études, il semble bien que la relation entre geste et parole ne soit pas une relation basée uniquement sur le phénomène de synchronisation. De Ruitter (2000) avait déjà prédit qu'étudier la synchronisation était un problème mal posé :

It should be pointed out that the issue of temporal synchronization is a nebulous one. (De Ruitter, 2000, p. 297)

Si nous avons vu que la plupart des cas, notamment en double tâche, le geste a tendance à précéder le début de la parole, l'exemple de l'arrernte nous montre également que la parole peut survenir avant le début du stroke du geste. Il semblerait donc qu'il ne faille pas parler de synchronie mais plutôt d'*harmonie* selon notre propre terme, entre geste et parole. L'harmonie entre ces deux modules réside dans le fait que le geste puisse survenir avant, pendant ou après la parole, mais toujours en coordination... Avec le défi de découvrir quel peut bien être le rapport quantitatif pour une telle coordination.

#### **II.14.6. Pour une synergie de la voix et du doigt : considérations théoriques pour une synchronisation non obligatoire**

La littérature relative à la coordination du geste et de la parole dans la communication parlée est, comme nous venons de le montrer, assez dense et assez complexe. C'est un sujet qui se prête à de nombreuses spéculations théoriques, ainsi qu'à de nombreuses études empiriques afin d'en comprendre le fonctionnement. Nous nous sommes intéressée également à cette question de la coordination geste-parole suite à notre étude réalisée sur le corpus d'enfants français. Alors que notre travail consistait à prédire et soutenir comment pointer de l'index et vocalisations de babillage se mettent en relation pour arriver aux premiers mots chez l'enfant, la question de la coordination, voire de la synchronisation, nous est naturellement « venue sur le tapis ». Comme nous l'avons décrit précédemment, de nombreux chercheurs ont tenté d'étudier la relation entre geste et parole chez l'enfant (Butterworth, 2003 ; Bates et Dick, 2002, Volterra et al., 2004 pour ne citer qu'eux), et des travaux comme ceux de Goldin-Meadow (2003) portant sur la résilience du langage ont pu montrer que le stade du geste non congruent avec un mot précédait

les énoncés à 2 mots lors de l'acquisition du langage chez l'enfant. Pour résumer, l'enfant débute par une phase de congruence (l'enfant pointe son papa et dit « papa »), suivie d'une phase non-congruente (l'enfant pointe la chaussure de son papa et dit « papa »), qui laisse place ensuite à une phase de congruence augmentée à 2 mots (l'enfant pointe la chaussure de son papa et dit « chaussure papa »). Parmi les gestes utilisés par l'enfant, notons une fois encore la prédominance du geste de pointer de l'index, qui offre un paradigme remarquable pour l'étude de l'acquisition du langage, comme noté par Tomasello et al. (2005), qui argumente que ce geste accompagné d'un partage de l'attention-intention donne certains bénéfices « gratuits » à l'enfant, tels que la syntaxe et l'accès au lexique ou l'imitation pour s'ajuster à sa langue maternelle.

Toutefois la notion de congruence signifie-t-elle pour autant que geste et parole se phasent en synchronie ? C'est une question ouverte qui fait débat. En effet, la question de la synchronisation, pour établir la synergie entre la voix et le doigt, reste un champ d'investigation qui donne lieu à des résultats encore très contradictoires selon la langue ambiante de l'enfant. Nous citerons par exemple les travaux de l'équipe britannique de Goldin-Meadow qui tendent à prouver que les enfants anglais synchronisent progressivement avec l'âge geste et parole, qui s'opposent en quelque sorte aux travaux de l'équipe romaine menée par Pizzuto (Pizzuto et al., 2005), qui ne trouvent pas cette tendance à la synchronisation entre geste et parole chez tous les enfants italiens. Sur notre propre corpus d'enfants français, nous notons également une tendance à des vocalisations qui peuvent survenir avant, pendant, au moment de l'apex, ou après le geste de pointer de l'index. Ainsi à la question de savoir si congruence rime avec synchronisation, nous répondrons simplement « pas théoriquement », en considérant l'influence des différents systèmes linguistiques.

Pour clarifier notre prise de position, nous avançons un premier point qui consiste à argumenter qu'au stade du geste non congruent avec le mot, il n'existe aucune raison intrinsèque pour que le geste pointant vers les chaussures soit synchrone avec l'émission vocale « papa ? » (pour montrer les chaussures qui appartiennent à papa ou le demander). Bien entendu aucune raison théorique ne pourrait justifier qu'ensuite dans l'énoncé de 2 mots qui suivra, on s'attende à une

quelconque synchronicité (?) par quel principe (?) entre « chaussure » et « papa ? » dans « chaussure papa ? » !

Notre second argument serait de savoir pourquoi cette synchronisation n'apparaîtrait pas dans le stade antérieur, le stade où le geste est congruent avec le mot, soit lorsque l'enfant pointe son papa et dit « papa ? ». Afin de répondre à cette question, nous apportons deux pièces d'information. Premièrement, chez l'enfant, il semble qu'il existe des sites du mot —la notion de site est emprunté à Lowenstamm ici— qui vont recevoir des démonstratifs (des articles, etc.) propres à la langue ambiante. A l'âge adulte, ces sites accueillent ensuite des démonstratifs déterminatifs et/ou locatifs (exemple : « Cet homme-ci » où « cet » est un déterminant et « ci » une adposition locative). De fait, chez l'enfant ces deux fonctions déterminative et locative sont « lumpées ». Le site du « ça » (déterminatif) et le site du « là » (locatif) n'en forment plus qu'un. Plusieurs arguments linguistiques et philosophiques seraient à développer contre la proposition avancée par certains philosophes, linguistes et cognitivistes (résumé dans Hurford, 2003) qu'il n'y a jamais de locatif dans le démonstratif. La meilleure preuve empirique est sans doute la possibilité pour les langues de faire évoluer un locatif en démonstratif (« ici » devenant « cet », Diessel, 1999b). Deuxièmement, dans les langues du monde décrites, il existe deux sites possibles pour la détermination ou la localisation : un site en début de mot et un site en fin de mot. Les langues européennes de familles différentes en sont une bonne illustration. Certaines langues peuvent privilégier un site plutôt qu'un autre, par exemple « l'homme » en français versus « omul » en roumain (cf. pour les considérations typologiques dans les langues européennes, Riemsdijk, 1999) ; ou encore « **(the)** house » en anglais versus « huset » en suédois. Certaines langues peuvent utiliser les deux sites : comme français « cet homme-ci » ou allemand « das Haus da ». Cela signifie donc que : a) les sites peuvent ne pas être remplis, rester vides (« house » en anglais) ; b) un seul des deux sites peut l'être (« l'homme ») ; ou c) les deux sites peuvent l'être (« cet homme-ci »). A partir de ces deux informations, nous pouvons proposer que le pointer chez l'enfant n'a théoriquement pas à être plus en phase avec le début qu'avec la fin du mot.

Nous proposons donc, dans cette perspective, d'étudier la synchronisation voix-geste dans une tâche de désignation avec focus (insistance de détermination). Chez l'adulte, le contrôle expérimental des tâches volontaires est possible dans la

mesure où nous pouvons mesurer les productions verbales et gestuelles finement pour étudier ce problème de synchronisation. En effet, chez le sujet adulte, il nous est possible de manipuler volontairement la position du focus vocal (prosodique) sur l'un des deux sites : « CET homme-ci » vs. « cet homme-CI ». Dans le but d'éviter la dispersion évoquée précédemment dans les études sur le geste et la parole, nous avons pris le parti scientifique de limiter le problème de la coordination du doigt et de la voix à (1) une situation de synergie où le doigt et la voix convergent vers le même but d'ostension, le pointer en l'occurrence, ceci pour l'enfant comme pour l'adulte ; et (2) une situation où il n'existe que deux possibilités de focaliser ou situer l'ostension, avec un contrôle volontaire possible expérimentalement chez l'adulte. Une expérience pilote a été menée avec Coriandre Vilain à l'Optotrak, faisant apparaître plusieurs problèmes pratiques qui sont maintenant résolus et l'expérience finale est en cours de réalisation.

## **II.15. En guise de transition...**

A la lecture de cette longue seconde partie, nécessaire afin d'ancrer les mécanismes et étapes charnières qu'implique le développement du cadre du signe, et suite à la première partie, concernant le cadre de la parole, nous voici à présent dotée de deux cadres de contrôle bien identifiés au sein de notre dispositif de travail : d'un côté les mécanismes nécessaires à l'enfant pour ancrer la phonologie dans le langage, et de l'autre les mécanismes nécessaires afin d'ancrer la sémantique. Ces deux cadres ou flux se développant parallèlement au cours de la première année de vie de l'enfant, avant de dialoguer pour donner naissance au mot.

Toutefois, il ne suffira pas de nous contenter uniquement de ce que ces revues de littératures nous apportent (même si leurs contenus nous sont indispensables pour ancrer tout particulièrement une sémantique neurale) si nous voulons expliquer ce qui se développe et ce qui évolue à l'intérieur de ces deux cadres. Nous perdrons de vue notre enjeu principal, qui est de chercher à les lier à un moment donné du développement. Le pivot central de notre étude repose bien sur ce lien, entre phonologie et sémantique, entre vocalisations et gestes de pointage, et dans la partie suivante, nous allons tenter de démontrer quelle opération unit ces deux versants, et ceci dans les termes des contraintes motrices qui mûrent chez l'enfant, pour que celui-ci réussisse à mettre en place ses premiers mots. Nous

vous proposons donc maintenant de découvrir les ressorts cachés et les promesses langagières de ce mécanisme qui permet l'intégration et le dialogue entre le cadre de la parole et le cadre du signe.



**TROISIEME PARTIE :**

**QUAND LE CADRE DE LA PAROLE EST  
ENCHÂSSÉ DANS LE CADRE DU SIGNE**

### **III.1. Une hypothèse de travail pour trois questions princeps: le mot, le pied et la syllabe**

#### **III.1.1. Quels critères pour les premiers mots ?**

De nombreuses études ont été attentives à l'arrivée des premiers mots chez l'enfant, mais peu d'entre elles proposent une classification claire et pertinente pour caractériser une vocalisation en tant que mot. Nous retiendrons les critères de Vihman et McCune (1994) qui offrent une méthodologie reposant sur les différences développementales de l'utilisation du mot par l'enfant (à partir d'un corpus de 10 enfants âgés de 9 mois à 1,6 an, filmés mensuellement à la maison, en interaction ou jeu libres avec leurs mères). Pour leur analyse, les chercheurs se sont intéressés aux sessions concernant les âges de 1,5 an et 1,6 an, qu'ils ont transcrits phonétiquement. Vihman et McCune ont listé un inventaire exhaustif de formes vocales, en notant le contexte d'utilisation. Une quantification préliminaire a porté sur l'acceptabilité d'un candidat comme mot potentiel. Pour identifier un mot en tant que tel, les auteurs ont eu recours à des critères impliquant le contexte, la forme phonétique de la vocalisation, et la relation que cette vocalisation entretient avec les autres vocalisations.

Les critères basés sur le contexte impliquent : (1) le contexte déterminatif, qui s'applique seulement aux mots ayant un sens spécifique, facilement identifiable à l'intérieur d'un contexte et qui incluent des noms plus concrets, ou plusieurs mots de type relationnel. Ce critère ne s'applique pas à une réponse imitative à un stimulus purement verbal ; (2) l'identification maternelle : la mère identifie au moins un exemplaire vocal comme la marque du mot supposé ; (3) l'utilisation multiple : l'enfant utilise le mot plus d'une fois ; enfin (4) la présence d'épisodes multiples : y a-t-il plus d'un épisode par utilisation ?

Le deuxième ensemble de critères implique la forme de la vocalisation : (5) comme le lien complexe : par lequel l'enfant forme un lien avec plus de deux segments de la forme adulte ; (6) le lien exact, où l'enfant produit au moins un exemplaire que même une oreille non entraînée devrait reconnaître comme un exemplaire du mot supposé ; et (7) le lien prosodique, plus particulièrement (a) par rapport au modèle, quand le lien est réglé sur la cible adulte ; et (b) par rapport aux

marques, c'est-à-dire lorsque l'enfant utilise un effet vocal spécial de façon répétée, dans des contextes pragmatiquement plausibles pour le même mot supposé.

Le troisième et dernier ensemble de critères concerne la relation du mot supposé avec les autres vocalisations : (8) les marques (ou effet vocaux spéciaux répétés) imitées, s'il existe au moins un exemplaire imité ; (9) l'invariance, tous les exemplaires du mot montrant la même forme phonologique ; et enfin (10) les utilisations appropriées, où les utilisations surviennent toutes dans des contextes qui suggèrent probablement le même mot.

Ces critères d'identification réunis permettent alors de décider le statut de chaque vocalisation sélectionnée. Une vocalisation a été considérée comme un mot quand elle remplissait au moins quatre de ces critères. Au-delà de la démonstration de l'efficacité et de la robustesse de cette méthode, Vihman et McCune ont eu également la volonté de montrer qu'il existe une séparation entre les mots liés au contexte, référentiels, utilisés par certains enfants, et l'usage plus large de ces mots qui émerge plus tard chez l'enfant. Ces réflexions sur le moment où « un mot devient un mot » nous amène également à nous questionner sur le moment précis où l'enfant « fixe » une vocalisation étiquetée en tant que mot dans sa mémoire.

### **III.1.2. Quand peut-on parler de permanence du mot chez l'enfant ?**

L'évolution de l'utilisation des mots chez l'enfant suit une trajectoire développementale assez bien connue à présent. Alors qu'autour d'un an, au tout début du langage expressif, le taux d'acquisition des mots est lent (quelques mots par mois), les chercheurs observent vers la fin de la seconde année, une explosion de la croissance du vocabulaire, aboutissant à un lexique productif de presque 50-100 mots. Après cette explosion, les premières combinaisons de mots émergent, au moment même où débute l'utilisation du langage pour référer à des choses qui ne sont pas physiquement présentes. L'explosion lexicale a souvent été interprétée dans la littérature par le fait que des compétences représentationnelles se développaient au même moment, avec l'idée que les mots réfèrent aux choses, qu'il existe des contraintes linguistiques, que l'enfant établit un mapping plus rapide entre l'entité et le mot et qu'il montre un intérêt plus prononcé dans la catégorisation des choses dans le monde. Toutefois, Dapretto et Bjork (2000) soulignent que ces

explications échouent à expliquer que la compréhension excède de loin la production de mots dans les premières étapes d'acquisition du langage. Ce fossé entre perception et production s'explique en partie par le fait que la compréhension réclame la reconnaissance des sons-mots et leur catégorisation, tandis que la production exige la récupération du patron des sons associés au sens donné. Dans les expériences que Dapretto et Bjork ont menées sur 30 enfants suivis de 14 à 24 mois, les chercheurs ont fourni une évidence empirique solide sur les changements significatifs qui opèrent dans les processus de récupération de mots dans la seconde année de vie de l'enfant, au moment même où des développements cruciaux sont observés dans les vocabulaires productifs des enfants. De plus, la présence de déclencheurs visuels faciliterait selon eux la récupération du mot par l'enfant dans les premières étapes d'acquisition lexicale, une évidence plutôt importante lorsque l'on sait que des facultés pauvres de récupération de mots dans les étapes initiales du développement lexical peuvent limiter ultérieurement le nombre total de mots dans les vocabulaires productifs des jeunes enfants.

De plus, ces mêmes chercheurs font état de travaux sur les bases neurales de l'acquisition du langage, qui montrent qu'il existe une réorganisation à l'intérieur du substrat neural du traitement du langage au moment où la plupart des enfants sont au cœur de l'explosion du vocabulaire. Ainsi, les réponses cérébrales discriminant les mots compris des mots inconnus étaient bilatéralement et largement distribuées à 13-17 mois, tandis qu'elles étaient limitées aux régions temporales et pariétales de l'hémisphère gauche à 20 mois. Le changement qui intervient dans les patrons d'activation corticale impliqués dans le traitement du langage peut alors suggérer qu'un système latéralisé, émergeant plus tard serait alors impliqué dans le développement des fonctions linguistiques plus avancées. Ces recherches indiquent alors que, dans l'apprentissage du langage, une période sensible peut se montrer fondamentale pour le déclenchement d'un patron latéralisé, classiquement optimal pour les spécialisations fonctionnelles du traitement du langage. Dapretto et Bjork interprètent ces découvertes dans le sens où le développement de fonctions linguistiques complexes émergerait de l'interaction dynamique entre un organisme doté d'un système neural très bien adapté au traitement de l'information linguistique, et un environnement riche et structuré qui fournirait le type d'input requis pour le développement de représentations linguistiques matures.

### III.1.3. Des premiers mots similaires pour tous ?

Dans son ouvrage *Comment la parole vient aux enfants*, Boysson-Bardies (1996) a étudié les différentes stratégies utilisées par 9 enfants dont 7 français dans la production de leurs premiers mots. Comme il y a peu de rapports aussi détaillés pour le français, nous en avons repris ci-dessous les éléments intéressants pour les patrons des premiers mots de notre langue.

#### III.1.3.1. Première stratégie : a minima

La première stratégie est celle qu'elle nomme la stratégie a minima et concerne Emilie, Sean et Timmy. Emilie, lors du babillage, évite les longues productions intonées, et 61% de ces productions sont monosyllabiques (contre 40% pour les autres enfants). Elle tend également à privilégier les consonnes occlusives (60% de ces consonnes contre 49% pour les autres enfants). Lors de ses premiers mots, elle consacre ses vocalisations à l'information articulatoire et privilégie les syllabes et consonnes produites lors du babillage. Elle réduit le plus souvent ses bisyllabes en monosyllabes et garde la syllabe qui commence par l'occlusive. Elle conserve également les voyelles pertinentes, ce qui lui évite les cas d'homophonie. Ses premiers mots sont du type [ka] pour canard, [ke] pour clés, [po] pour chapeau, [papa] pour papa. Au stade de 25-30 mots, elle commence à augmenter les productions bisyllabiques en introduisant une voyelle d'appoint (exemple : à boire [abɔ̃a] ou brosse [abɔ̃]) ou en redoublant la syllabe (babar donne [baba], canard donne [kaka]). A ce stade, 40% des occurrences sont bisyllabiques, 8% ont plus de deux syllabes. Emilie tend à accentuer les tendances phonétiques déjà présentes dans son babillage. C'est une stratégie plutôt simple, efficace, à économie articulatoire. Cela lui permet de développer rapidement un vocabulaire en évitant d'avoir trop d'homophones, même si son répertoire est assez pauvre et que les mots se distinguent par leur voyelle. Cette stratégie a minima est une stratégie souvent retrouvée chez les enfants anglophones des Etats-Unis, comme le signale M. Vihman.

Sean utilise également cette stratégie. Il possède 77% de monosyllabes dans son répertoire, et lorsqu'il essaie de produire des bisyllabes, il les réduit à 65% en monosyllabes. Il évite le système de reduplication, et ses premières productions monosyllabiques sont, contrairement à Emilie, quasi-homophones. Ces mots sont

construits autour de deux schémas : syllabe CV avec occlusive labiale [b, p] + voyelle ou syllabe CV avec occlusive dentale [t, d] + voyelle. Il utilise en revanche un répertoire vocalique très varié. Sean va ensuite développer son répertoire avec des mots qui se terminent par une consonne, toujours monosyllabiques, et toujours vélaires du type block [ba:k]. Lorsqu'il atteint un inventaire de 25 mots, les formes sont simples avec un geste articulatoire minimal, proches des CV initiales. Sean constitue un bon exemple du style référentiel ou analytique, ses premiers mots renvoyant à des objets, animaux, personne.

Enfin Timmy, un enfant américain étudié par Vihman, constituerait ce que Boysson-Bardies appellerait un « schéma idéal », un schéma rigide qui sert pour tous les mots adultes. La base de son vocabulaire repose sur 2 types de CV où C = [b, g] et V = [a]. Les mots commençant par [b] correspondent aux mots adultes commençant par la même voyelle, et les mots commençant par [k] correspondent aux mots adultes qui contiennent une vélaire (en initiale ou finale de mot). Ainsi Timmy possède à 16 mois un inventaire de 15 mots, [ba] ou [ga] homophones. A 17 mois, de timides variations vocaliques surviennent, ainsi que des dissyllabes (surtout de redoublements), avec un schéma vocalique rigide : la première voyelle est [a] et la seconde [i].

### **III.1.3.2. Deuxième stratégie : « les charmes de la conversation »**

Simon ne produit à 18 mois que quelques mots mais pourtant il réalise de longues phrases, utilisant de façon prédominante des bisyllabes. Il porte tout particulièrement attention à la forme prosodique des discours adultes et son entrée dans la parole repose en priorité sur l'intonation ou le rythme. Il choisit pour cela un contour d'intonation qui correspond à un contour de groupe de mots et le remplit par des syllabes variées et bien articulées. Il fait également varier ses contours, soit descendant, soit montant, avec une majorité de montants, ce qui correspond au patron français. Tout au long de son observation, Simon a gardé cette stratégie intonative, un style fréquent chez les enfants français selon Boysson-Bardies. En effet, le découpage en mots n'est pas franchement favorisé par l'organisation prosodique du français, qui lui préfère un regroupement plus large, en proposition par exemple. Le contour d'intonation du français implique un allongement des syllabes terminales et un contour montant que l'on retrouve chez Simon.

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

Marie, tout comme Simon, expérimente largement les monologues, en produisant beaucoup de polysyllabes. A 13 mois, elle possède 5 mots, une majorité de bisyllabes sans schéma phonétique particulier. A 17 mois, elle possède 15 mots, la plupart du temps noyés dans des phrases de 4 à 7 syllabes. Les formes CVCV sont prédominantes dans son répertoire. Elle utilise les voyelles principales du français et privilégie les occlusives (sauf les vélares), ainsi que les mots commençant par des nasales [m, n] ou des fricatives [v]. A plus de 17 mois, elle possède 25 mots dont certains commencent par la latérale [l], et elle commence à produire beaucoup plus de phrases de 5 à 6 syllabes qui peuvent être interprétées. Son répertoire est plus varié dans le sens où elle nomme les objets et utilise déjà des verbes, adjectifs, pronom, expressions et le pronom « moi ».

A 10 mois, Léo possède 7 mots, il est plutôt précoce. Il privilégie les mots dissyllabiques, en particulier ceux avec [l]. Ses productions sont longues, dont 73% ont plus de deux syllabes. A 14 mois, il a 15 mots et à 17 mois, il possède 25 mots, variés, qui sont enchâssés dans une expression. La sémantique de ses premiers mots est multiple, avec des noms concrets, des expressions, verbes, adjectifs...

Cette stratégie utilisée par ces 3 enfants donne un certain poids à l'intonation et au rythme de l'énoncé, dénotant une certaine expressivité. Le choix des mots est très varié au niveau sémantique, parfois aux dépens de la fidélité segmentale. Ce style purement expressif a pour conséquence de retarder le développement lexical, mais va en revanche favoriser l'insertion sociale de ces enfants. Le développement du lexique est plus varié ici que dans la stratégie à minima.

### **III.1.3.3. Troisième stratégie : une alternative**

A l'intérieur de cette stratégie, les enfants vont prendre comme unité de base le mot et ils vont chercher à reproduire les syllabes. Dans leurs productions, des omissions, des redoublements syllabiques ou des schémas préférentiels vont être privilégiés, sans qu'une technique ne soit forcément mise en avant particulièrement.

Charles produit ses premiers mots à 12 mois, généralement monosyllabiques mais variés. Ses 10 premiers mots sont destinés à des besoins vitaux. A 14-15 mois, son vocabulaire est constitué de 15 à 20 mots, pour la plupart monosyllabiques et dissyllabiques, toujours assez variés au niveau phonétique (occlusives labiale,

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

dentale, nasale, fricatives et latérales). Il tend parfois à omettre la consonne initiale mais respecte au mieux la stratégie syllabique et phonétique du mot cible.

Noël possède un inventaire de 5 mots, puis 15 mots et enfin 25 mots à 13, 16.5 et 17.5 mois respectivement. Ses productions sont importantes car Noël répète plusieurs fois les mots. Il a tendance à favoriser les disyllabes, opère quelques réductions de syllabes et son répertoire est essentiellement occlusif.

Henri est un enfant qui a très peu babillé, c'est un enfant plutôt silencieux, mais vif et qui a marché très tôt. Dès 10 mois, il indique par une production de type « hein » et une intonation montante, qu'il désire qu'on lui nomme les objets. Vers 16-17 mois, il commence à peine à produire 2 ou 3 mots, qu'il prononce avec prudence. Jusqu'à 18 mois, son répertoire est limité globalement à papa, maman et non. Ce qui est en revanche tout à fait surprenant chez cet enfant silencieux jusque là est le fait qu'à 20 mois, il se mette subitement à ressortir tout le vocabulaire qu'il avait accumulé pendant ce temps, soit 30 mots pour les animaux et objets et 15 mots pour définir les relations ! A 22 mois, sa production de mots explose littéralement avec par exemple pour une seule séance de 30mn plus de 150 mots prononcés, ainsi que de vraies phrases de deux ou trois mots. Cet enfant est tout à fait atypique dans la stratégie qu'il a adoptée. Il possède un système de règles complexes qui gouverne la structure de ses mots. D'un côté il est capable de produire tous les phonèmes, mais de l'autre il exige que ceux-ci soient à certaines places dans le mot. Henri a élaboré en quelque sorte un système de substitutions régulières. A 30 mois, cet enfant produit des mots rares et sophistiqués, ainsi que des phrases syntaxiquement élaborées. Boysson-Bardies conclut sur cet enfant en supposant qu'il soit rentré immédiatement dans le stade phonologique, et qu'il s'est refusé alors aux productions approximatives. Il aurait ainsi sauté la période concernant les 50 premiers mots.

Pour conclure sur cette étude, nous pouvons remarquer l'utilisation de stratégies pleinement variées pour produire les premiers mots chez l'enfant. Ainsi, l'enfant pourra suivre un patron très régulier et où l'économie articulatoire prédominera. D'autres favoriseront les contours intonatifs et le rythme de la langue maternelle et enfin certains adopteront des stratégies tout à fait à part, à l'image d'Henri qui voit son vocabulaire exploser lorsqu'il est prêt à articuler les sons qu'il désire produire.



### III.1.4. Définir le mot ou en percer le contrôle ? Trois questions pour...

Les linguistes dans leurs ouvrages, leurs manuels, leurs dictionnaires, se sont attachés à définir le mot, faisant comme Louis Agassiz qui critiqua Darwin pour n'avoir jamais donné la définition de l'espèce, alors qu'il travaillait sur la variation des espèces... La définition de l'homme comme un animal bipède s'est déjà attirée les sarcasmes depuis les philosophes Grecs, jusqu'à Pinker, qui l'expédie en deux mots dans son *Instinct du langage* comme un exercice de lexicographie plutôt qu'un effort de connaissance biologique. Découvrir le système comportemental et neural qui contrôle la production du mot chez l'enfant est pour nous la véritable priorité. De même la définition de la syllabe nous a semblé un vain exercice de linguistique, à côté de la proposition de MacNeilage d'en connaître la génératrice neuro-motrice.

Nous devons à Edy Veneziano (d'une brève mais dense conversation avec C. Abry, le 10 décembre 2005 à l'occasion du colloque *Emergence of Language Abilities* à Lyon), la plus « étonnante » des trois questions suivantes. Pourquoi les premiers mots sont-ils d'un mot (*Mean Length Utterance=1*)? Alors que l'énoncé d'input de la langue ambiante donné à l'enfant est réputé être de plus d'un mot de long ? Nous avons déjà de notre côté posé deux autres questions. Pourquoi les premiers mots sont-ils massivement de (une)deux syllabes ? Et pourquoi après tout des syllabes ? Nous pensons que la réponse à cette dernière question peut être donnée, à la MacNeilage, par le contrôle du rythme mandibulaire. Mais quelle solution pour la longueur du premier mot, voire du premier énoncé (ce que nous prétendons hardiment proposer du même coup) ?

Notre hypothèse de travail ou cadre conceptuel (*framework*) repose sur l'idée qu'il existerait un rendez-vous développemental entre ce que nous nommons le cadre de la parole et le cadre du signe. Le cadre de la parole (*Speech Frame*) a été précédemment illustré comme celui de la théorie *Frame/Content* de MacNeilage (1998). Le cadre du signe (*Sign Frame*) est caractérisé pour nous par le contrôle visuo-moteur du « bras porteur », menant au fameux « pointer impératif » de l'index autour de neuf mois. Quand le babillage canonique, présent dès les six mois, s'intègre-t-il avec le pointer, à partir de neuf mois ? Plus fonctionnellement parlant : quand peut-on intégrer rythme de la parole et détente du geste ? Sachant quelle est la métrique de l'unité intégrable, disons la syllabe, quelle sera la métrique de l'unité

intégrante ? Les deux sont-elles compatibles ? Nous allons énoncer sur ces points des prédictions mesurables en termes de signaux pour un tel rendez-vous.

Dans cette partie, nous présentons des données qui semblent soutenir notre propre manière de comprendre la relation incontournable entre phonologie et sémantique. Notre but ultime est de lier le contrôle rythmique du flux du « babillage-syllabe » (à six-sept mois), avec le contrôle sémiotique de la deixis dans un pointer discret (à neuf mois), ceci dans le but d'intégrer le babillage dans un gabarit « pointer-pied ». Ce qui est finalement le gabarit des premiers mots (vers douze-treize mois).

La proposition que la syllabe du babillage devrait entrer dans le gabarit du mot n'est pas nouvelle : la dominance du cadre (*frame*) pour MacNeilage et Davis (Davis et al., 2002), le *schème vocal moteur* pour Vihman (McCune et Vihman, 2001), sont des propositions soutenues par des données en faveur du filtrage articulatoire des premiers mots. Plusieurs autres propositions théoriques concernent les premiers mots. D'autres ont l'avantage sur le cadre présenté ici de traiter les comportements apparaissant avant les premières étapes du babillage canonique (inter alia : Koopmans-van Beinum et Van der Stelt, 1986, Locke, 1997, Oller, 2000), avec ou sans emphase sur les gestes... et les *grunts* ou grognements (voir McCune et al., 2003). Nous les prenons comme des acquis nécessaires et valables : qui dirait que les premiers pas des jeunes enfants commencent à partir de compétences en locomotion zéro ? Depuis la page dans *Nature* d'Iverson et Goldin-Meadow (1998), la gestuelle dans le développement du langage a été illustrée par ces auteurs et d'autres, parmi lesquels Butterworth (2003), Bates et al. (voir Bates et Dick, 2002, pour une revue tenant compte du développement cérébral). Iverson et Bates ont collaboré sur ce point avec l'équipe de Rome (Volterra et al., 2004). Tomasello et al. ont étudié aussi bien les gestes chez les enfants que chez les grands singes (pour une revue récente et une proposition évolutionnaire différente sur le pointer, voir Leavens, 2004).

Mais les liens entre l'unité sémantique et son expression resteraient selon nous de l'ordre du miracle, si l'on utilisait une version de la phonologie gabaritique du mot, simplement pour reformuler d'une autre façon le fait que le mot lemmatique possède une manifestation morphophonologique plutôt stable. A notre connaissance, il n'existe aucune proposition d'un contrôle pour cette unité, comparable au contrôle

cyclique de la syllabe, qui soit donc incarné dans un système de contrôle neural d'un articulateur : nous disons ici le bras à la place de la mandibule. C'est donc un nouveau cadre (*frame*) pour lequel s'ouvre la recherche de ses corrélats neuraux. Nous pouvons citer que dans les (méta)analyses majeures des tâches de mots en imagerie cérébrale, la définition de cette unité *mot* n'est même pas posée. Même pour les études de production de mots, quand elles fractionnent et parcellent la chronométrie dans le comportement pour une tâche typique de dénomination d'image (*picture naming*), Indefrey et Levelt (2004) ne traitent jamais de la question : pourquoi un total de TR de 600ms à partir de l'apparition de l'image jusqu'à « l'initiation articulatoire » (en fait le début du signal acoustique) ? Le temps consommé par les processeurs n'est pas conçu comme un temps pour le contrôle d'un articulateur, avec des contraintes corporelles internalisées dans une habileté du système neural. Quoi qu'il en soit sans que la production de parole imaginée devienne plus rapide que la parole à haute voix (mesuré en premier par Landauer, 1962), les deux étant « incarnées cérébralement ».

Ainsi nous voyons notre proposition comme une extension nécessaire de la théorie *Frame/Content*, avec le cadre de la parole finalement niché (*embedded*) dans le cadre du signe, soit un cadre pour deux cadres. Comme pour l'intégration de la syllabe phonologique abstraite dans le contrôle du rythme mandibulaire, l'unité abstraite qui sied le mieux pour réaliser l'interface entre le contrôle du cadre de la « parole-syllabe » et le cadre du « signe-mot », est également une unité métrique célèbre. Le pied (*foot*) est prosodiquement contrôlé, accentué, intonné, focalisé, etc. Sa formulation la plus fameuse dans une approche développementaliste, publiée par Allen et Hawkins (1979), est dans la formule trochaïque proposée par Gerken (1994) comme un gabarit métrique *Strong(Weak)* [fort(faible)]. Quoiqu'il en soit, que l'on suive ou non cette formulation, la présence d'une unité de pied est hautement disputée dans la littérature développementale. Mais cette unité semble inévitable...Même pour ses opposants principaux argumentant à partir de données françaises. Cf. dernièrement Hilaire-Debove et Demuth (2005), qui obtiennent des réponses de réduplication dissyllabique aux expériences de troncation de mots avec leurs plus jeunes sujets (environ 2 ans). En fait Demuth et Johnson (2003) ont utilisé le corpus de la fille de Deville datant de 1891. Par conséquent la durée des mots monosyllabiques CV vs. CVCV ne pouvait évidemment pas être mesurée, pour tester

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

s'il y avait une tendance ou non vers l'isochronie. En dépit de leur remarque qu'il y aurait eu besoin bien entendu de mesures acoustiques, mais que l'allongement compensatoire des voyelles, ou la durée des voyelles dites tendues (*tense*) qui seraient bimoraïques, n'ont pas été rapportés pour les enfants français, cette clause reste de pur style puisque que les auteurs qu'ils citent n'ont pas, à notre connaissance, traité spécifiquement de ces questions en français (contrairement au japonais). Par conséquent l'hypothèse du pied dans le développement ne peut pas être actuellement rejetée, et la « subminimalité » de tels mots CV peut apparaître comme un mirage (laissant de côté la question sur le français adulte qui ferait ou non partie des langues « sans pied », *footless languages*). En ce qui concerne le statut de cette unité en psycholinguistique, nous prendrons hardiment la responsabilité de rapporter une conclusion d'Ann Cutler (pour l'isochronie cf. Cutler, 1980), qu'il s'agit probablement d'une unité de contrôle en production, sinon de perception (discussion privée avec C. Abry, *InterSpeech-ICSLP*, Jeju, Corée, 2004).

Il n'en reste pas moins que cette unité est particulièrement appropriée pour prendre en compte les une-deux syllabes typiques des premiers mots. Le pied est aussi capable d'intégrer chez les adultes 1-3...syllabes. Le pied permet en outre le contrôle de l'isochronie : par exemple, en anglais, que vous prononciez *muse*, *music*, *musical*, *musically*, vous garderez la même fenêtre temporelle, dans laquelle vous pourrez introduire ces 1, 2, 3, voire 4 syllabes (Figure III.1)

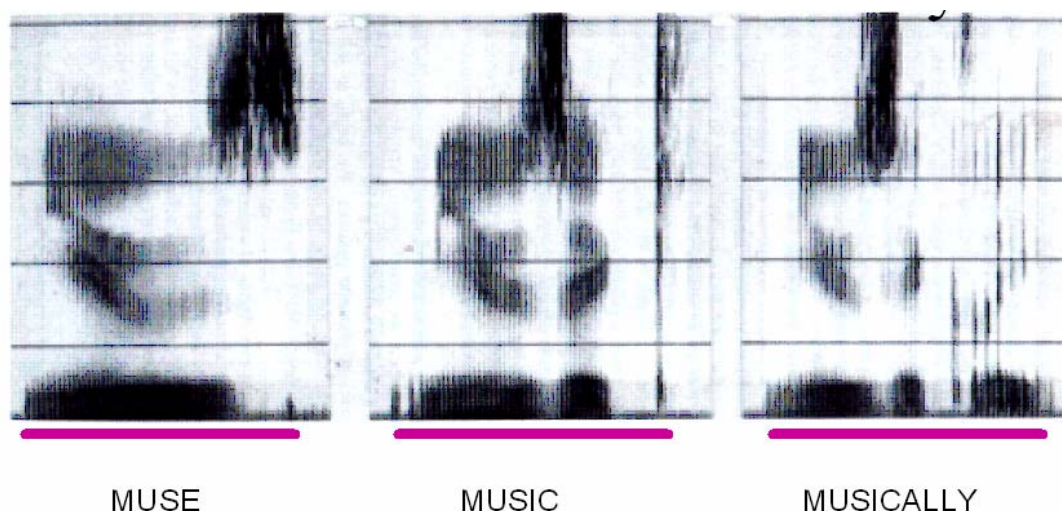


FIGURE III.1: Sous chaque production, soit pour chaque pied (foot), on peut mesurer la même longueur de barre, démontrant une tendance à l'isochronie remarquable (D'après Lindblom, 1991)

Afin de démontrer l'importance de cette notion qu'est le pied, ainsi que l'isochronie qui le régit, nous allons vous démontrer que toute unité linguistique n'est

pas tombée « surnaturellement » de son crayon sur le papier d'un théoricien de la grammaire ou de la poétique. Cette unité peut expliquer ce qui paraîtrait autrement inexplicable. Lors d'une conférence à Barcelone en août 2003, Wendy Sandler (Sandler, in press) spécialiste de la phonologie de la langue des signes, a montré les vidéos d'un sujet féminin signant en ASL. La première vidéo montre cette personne en train de signer FAINT « s'évanouir » : les deux poignets fermés s'élèvent au niveau du front et sont ensuite projetés vers le bas, les mains s'ouvrant comme lorsqu'on laisse tomber quelque chose... ou quelqu'un. Il faut savoir qu'en ASL, FAINT est composé de MIND+DROP. La vidéo monte les gestes du signe MIND (le doigt de la main droite va pointer sur la tête) et ceux du signe DROP (les deux poignets fermés s'élèvent au niveau de la ceinture et sont ensuite projetés vers le bas, mains ouvertes). Entre ces deux gestes, nous observons une fusion, un *blending* qui n'est pas que sémantique, mais aussi phonologique, fusion résultant du processus de lexicalisation ou grammaticalisation. Nous avons voulu mesurer la durée de chaque geste (Wendy nous ayant aimablement envoyé ses signaux) : la chute ou phase de downstroke de FAINT a la même durée (à une image près) que le downstroke de DROP. Nous observons donc une forte tendance à l'isochronie entre les deux mouvements de chute, bien que FAINT, avec les mains qui partent du niveau du front, réclame une plus grande amplitude que DROP, qui part de la ceinture.

Comment expliquer ce phénomène bien repéré (sinon mesuré) par Wendy Sandler. Celle-ci invoque ce qu'elle nomme une « syllable conspiracy » ou conspiration (un terme courant des linguistes chomskyens) en faveur de la syllabe : tout signe serait d'une syllabe et la fusion de deux signes serait encore une syllabe. Sémantiquement nous avons bien la formule suivante : 1 signe A + 1 signe B = 1 signe C. En laissant de côté l'idée de la conspiration (qui fait penser à l'anthropomorphisme de l'horreur du vide), en faveur de la réalité du contrôle moteur pour l'isochronie (bien connue pour votre paraphe en grande ou petite taille tendant à être de même durée en dépit de la variation d'amplitude), il reste à décider si l'unité métrique qui intègre phonologiquement la phonologie de chacun des deux signes est la syllabe ou le pied. (Rappelons qu'il n'existe pas de consensus pour la langue des signes sur l'existence d'unités comme la syllabe ou le phonème, sachant que les descriptions sont plutôt en traits ou chérèmes, de position, configuration et

mouvement des mains.) En termes phonologiques la réduction de 2 syllabes en une se traite plutôt par chute d'un des deux noyaux vocaliques (Mittèrand prononcé aussi en disyllabe Mitt'rand ; cf. encore la dièrèse ou non à Lyon, soit « L Lyon », ou Lyon en synérèse). Et même si le processus est phonétiquement continu, la phonologie n'offre pas de processus de raccourcissement comme elle le prévoit pour le pied (cf. l'hexamètre homérique et la poésie anglo-saxonne). Phonologiquement l'unité métrique « papier » qui correspond donc le mieux à une unité de contrôle neuro-motrice est précisément ce pied. Tout comme nous arrivons à mettre une seule syllabe longue (muse) et plusieurs syllabes raccourcies (musically) dans une même fenêtre de contrôle du timing, le pied du mot phonologique anglais, nous arriverions à caser les phases de downstroke de DROP et FAINT dans un même pied phonologique du signe de l'ASL.

Comme pour la syllabe et le mot, de nombreuses études se sont également penchées sur les concepts de la syntaxe, ou de la grammaire, et de la sémantique, toutes mettant en évidence cette boîte noire qui fait que l'on pose une étiquette sur une chose du monde dans lequel nous vivons sans pour autant comprendre comment cela est possible, à l'image de cette réflexion désabusée de Tomasello concernant le symbole en philosophie ou en linguistique :

About « the many approaches to linguistic symbols that exist on the intellectual scene, [s]uffice it to say that criteria such as arbitrariness and spatial-temporal displacement are decidedly unhelpful when looking at actual communicative processes. After all, Pavlov's dog associated the arbitrary sound of a bell with food that was not at the time perceptually present. And the expressions « duality of patterning » and « stands for » (as in the locution « symbols stands for their referents ») simply put a new name on the phenomenon without providing any further insights.» (2003:1994)

Avec notre proposition d'une phonologie du signe et du mot ancrée sur le geste de pointer déictique-sémiotique —qui est un captureur, un outil « harpon » sémantique (*semantic grabber*) pour le célèbre *fast mapping* lexical, et un précurseur syntaxique pour les connecteurs de type *that*, et qui est en même temps un « *chopper* » de disyllabes dans le débit du babillage, avec un corps ou gabarit (*template*) phonologique, enfin un « *chunker* » pour la mémoire verbale de travail lexical— nous faisons appel à une unité de contrôle principale formant le pied, en quelque sorte un « pointer-pied ».

De fait, nous nous sommes placés théoriquement dans un système d'attention/intention partagée à l'intérieur duquel le geste de pointer de l'index va être

un geste de contrôle phonologique via le pied. Nous proposons donc, en d'autres termes, de nous intéresser à cette phonologie du pointer, toujours dans l'optique précise d'accéder à une phonologie de contrôle du mot.

### **III.2. Pourquoi un geste *discret* si voyant pour le mot ?**

The practice of pointing with the index finger as a way of establishing reference to objects in the common stimulus environment of speaker and listener is a linguistic universal which by common consent plays an essential role in the acquisition of word-meanings.

(Place, 2000)

Nous posons ici l'hypothèse que les gestes brachio-manuels, particulièrement les pointers bras/index, peuvent être considérés comme un « cadre » dans lequel va s'ancrer le signe, ou proto-signe.

Dans la littérature portant sur le geste, de nombreuses études ont pu montrer que les gestes déictiques constituaient les gestes les mieux reconnus par l'enfant. Rondal et al. (1997) se sont intéressés dans leur ouvrage à la fois à la langue des signes et aux aspects para-verbaux de la communication. Dans la communication, il y a des postures corporelles qui mobilisent le bras, la main, la tête et le reste du corps. L'ensemble des modifications posturales et des gestes qui accompagnent la communication comportent des signaux qui servent à accentuer, ponctuer, nuancer tout ou une partie d'un énoncé, distinguables selon plusieurs catégories, dont une qui nous intéresse plus particulièrement : les gestes déictiques. Ces gestes comprennent le signalement d'une orientation, d'une direction, d'une position en relation avec ce dont on parle, par exemple désigner du doigt un objet référé ou indiquer de la main une direction à suivre.

D'autres études ont permis de mettre en relief que ce sont les gestes les mieux reconnus par les adultes. Et ce sont également les premiers gestes reconnus par les enfants. Cette idée confirme notre fil directeur qui veut que la monstration soit un système très important pour l'acquisition du langage, et que celui-ci s'ancre sur elle pour se développer.

Il nous paraît également important d'apporter quelques points d'information sur le geste de pointer en tant que tel. Nous considérons ce geste comme étant

typiquement un geste discret, et nous l'isolons volontairement des gestes dits rythmiques. L'étude menée par Petitto et al. (Petitto et al., 2001, 2004) s'était déjà intéressé à la fréquence de gestes chez les enfants, mais selon nous, c'est une erreur de vouloir mélanger tous les gestes, et mélanger particulièrement des activités manuelles rythmiques avec des gestes de pointers discrets. Dans la littérature, les gestes discrets ont souvent été envisagés comme faisant partie, ou comme étant une composante des gestes rythmiques. En d'autres termes un geste rythmique serait composé de gestes discrets. Afin de remettre cette notion en question, nous en appelons à une étude menée par Schaal et al. (2004), faisant suite à d'autres expériences qui avaient déjà mis en relief que les mouvements rythmiques et discrets pouvaient relever de deux régimes de contrôle différents. Dans ce sens, Schaal et al. ont comparé l'activité cérébrale lors de mouvements de poignet rythmiques et discrets à l'aide de la neuro-imagerie fonctionnelle. La condition rythmique requérait des oscillations continues du poignet à une fréquence confortable choisie par le sujet. La condition de mouvement discret impliquait un mouvement de flexion-extension modérément rapide. Les résultats de cette expérience indiquent que le mouvement rythmique est activé seulement dans les aires cérébrales contralatérales au poignet en mouvement, c'est-à-dire l'hémisphère gauche pour la main droite, avec en particulier le cortex sensori-moteur primaire et le prémoteur, l'aire motrice supplémentaire et le cortex cingulaire. Une seule activation ipsilatérale a été trouvée dans le cervelet. Toutes ces aires impliquées dans la tâche rythmique ne sont pas inattendues dans la mesure où il s'agit des aires motrices, souvent impliquées dans des mouvements simples. En contraste, le mouvement discret a présenté des activations tout à fait différentes dans des régions du cerveau plutôt diverses. Les auteurs notent une activité exclusive au mouvement discret dans l'hémisphère contralatéral dans la partie rostrale du cortex prémoteur dorsal, dans l'aire de Broca, le cortex pariétal (BA7, BA40), la partie antérieure de la zone cingulaire rostrale, l'aire BA38 sur le gyrus temporal supérieur et de façon générale, l'hémisphère cérébral ipsilatéral et le cervelet bilatéral. Toutes ces aires activées lors du geste discret sont plutôt liées à la planification du mouvement, nous ne sommes plus dans le cas d'un geste qui ne requérait que les aires motrices primaires. Le cortex prémoteur dorsal, l'aire de Broca et le précunéus dans le pariétal sont de plus des aires spécifiquement activées pour le geste discret.



En résumé, le mouvement rythmique active seulement un petit nombre d'aires motrices primaires unilatérales, tandis que le mouvement discret active une variété d'aires motrices non primaires contralatérales, avec en supplément une très forte activité bilatérale dans le cervelet et plus largement dans le cerveau. Le mouvement rythmique active significativement moins d'aires cérébrales que le mouvement discret, en conséquence de ces résultats, il paraît plus clair désormais que le mouvement de bras rythmique n'est pas composé de strokes discrets, et qu'il faille alors considérer ces deux actions comme indépendantes l'une de l'autre, même si l'articulation utilisée (ici le poignet) est identique. Cette découverte incite alors Schaal et al. à réfuter l'hypothèse que le mouvement rythmique est généré avec l'aide du système de mouvement discret. Ils ajoutent enfin que le circuit permettant le mouvement rythmique semble inclus dans le circuit du mouvement discret.

### **III.3. Retour sur notre hypothèse « étonnante »...**

Cette proposition d'intégration de la syllabe de babillage rythmique dans un « pointer-pied » discret va être testée en analysant le ratio entre les distributions des durées de cycles de babillage et les temps de mouvement de détente (strokes) des gestes de pointer chez les bébés. Si le premier mot peut contenir deux syllabes, ce ratio doit être un ratio harmonique de 2:1. Nous allons réaliser une quantification des régimes ou modes comportementaux (modes préférés) de la parole et du geste. Le premier sens évoqué par la notion de mode fait référence à une fréquence ou résonance en robotique, ici biocybernétique, d'un système contrôle-actuateurs. Seules quelques expériences pionnières ont essayé d'établir la fréquence préférentielle du système mandibulaire (depuis Sorokin et al., 1980). Quant au couplage des systèmes du bras et de la mandibule, après plusieurs tâches de « speech-tapping » purement rythmiques, est encore en phase d'exploration. Ce qui paraît clair à partir des résultats présents est que la durée moyenne décroît avec l'âge, pour le pointer, et aussi pour les cycles de mandibule: les deux sont plus rapides chez l'adulte. Par conséquent, cela n'est pas dû au facteur masse-taille (croissance) dans le calcul des fréquences propres du système mandibulaire et du système brachial. Il est plus probable que cela tienne à la maturation neuronale, par exemple la myélinisation, mais pas seulement (cf. la synaptogenèse et la maturation des systèmes neuro-transmetteurs). Le second sens mis en avant par la notion de

mode est la dominance statistique dans la biométrie comportementale, le plus grand nombre d'occurrences : c'est celui que nous allons traiter ici.

Nous allons évaluer les implications d'une proposition à trois étapes. Premièrement (1a) si le rythme du babillage canonique correspond à un mode de 3Hz, (1b) alors un premier mot à deux syllabes —s'il est « chunké » dans le flux du babillage canonique (avant l'acquisition d'une métrique spécifique, réglée en fonction de la langue et de sa réorganisation intra-mot, trochaïque...iambique)— va tendre vers une durée d'environ 600-700 ms. Deuxièmement, (2a) si le premier mot est contrôlé par un gabarit qui est le pied, (2b) alors la même tendance vers l'isochronie doit être observée dans la durée d'un premier mot monosyllabique. Troisièmement, (3a) si le pied du premier mot est gabarisé par le geste de pointer, (3b) alors la durée d'un stroke de pointer sera également d'environ 600-700 ms.

Nous utiliserons des tendances statistiques comme une première approche-test : à savoir le pic d'occurrences, c'est-à-dire le mode dans les valeurs de distribution de la durée (ou de la fréquence) du cycle de babillage, et le mode pour les durées des mouvements de détente (stroke) dans le pointer. Le ratio des deux modes devrait tendre vers 2:1. Nous allons d'autre part analyser longitudinalement tous les bébés du corps sur lequel nous avons travaillé afin de mettre en évidence le ratio pointer/syllabe de chacun. Nous nous intéresserons particulièrement à un bébé qui est dans le profil de pointer général, mais qui présente une réduction marquée dans la durée de ses détentes vers le début de ses premiers mots. Cette relation pourrait être en effet problématique pour son ratio pointer/syllabe.

### **III.4. Analyses du corpus**

#### **III.4.1. Le mode de distribution des durées de détente du pointer : tendance globale**

Nos mesures ont été réalisées à partir d'un corpus vidéo de six enfants français, filmés à la maison toutes les deux semaines, entre 6 et 18 mois, au total environ 57 heures (corpus enregistré par Stefanie Brosda, en 1999-2000). Comme les enfants commençaient à produire leur babillage canonique plus tôt ou plus tard que 7 mois, certains étaient filmés plus que d'autres (de 18 à 23 sessions), leur suivi prenant fin autour de 17 mois. Ce corpus s'est trouvé disponible pour une étude de

pointers spontanés, dans la mesure où il n'y avait aucune procédure d'élicitation. Comme attendu, l'éthologie du bébé domestique est aussi difficile que la primatologie non-humaine en pleine jungle ! Au bout du compte 276 événements de pointer avec vocalisation ont pu être mesurés, ce qui vaut bien 207 cris d'agression enregistrés pour 37 chimpanzés au cours d'une enquête de terrain sur 5 mois (Slocombe et Zuberbühler, 2005).

La présence de gestes de pointer a été observée de 7,20 mois à 17,18 mois. Les premières occurrences sont apparues chez les enfants entre sept et dix mois. Généralement, le début survenait autour de 9 mois, une date classique pour l'émergence du fameux « pointer impératif ». Après avoir situé les 276 événements, ils ont été mesurés à partir du début du premier mouvement détectable visuellement sur l'écran vidéo, dans une inspection image par image (40 ms) ; idem pour la fin du temps de mouvement (Movement Time ou MT). Nous avons eu comme tout le monde des problèmes méthodologiques pour détecter l'initiation des mouvements de parole et de geste, sur la seule inspection des enregistrements vidéos (cela peut même s'aggraver quand la synchronicité est en jeu, une question contestable et contestée, dont nous parlerons ultérieurement). Ce problème sera seulement résolu avec progrès quand nous disposerons de nouveaux enregistrements avec suivi de mouvement (movement tracking), pour d'autres bébés. Pour le pointer, notre choix de la détente ou stroke, soit le temps de mouvement (MT), est commun dans la littérature ; mais c'est bien sûr un biais de préférence du mouvement, à vrai dire sans fondement théorique (les phases de tenue sont sans aucun doute tout aussi importantes). Pour les mesures de syllabes CV et de mots, nous nous sommes servis essentiellement de l'éditeur audio PRAAT et de nos propres logiciels maison. Notons que le fait que nous-mêmes et d'autres opérateurs obtiennent parfois (voir Fig. 5) les « mêmes » valeurs pour les syllabes (cycles), à savoir S1 initiale et S2 suivante, n'est pas une garantie contre l'absence d'un biais de mesure dans les événements initiaux et/ou finaux.

### **Enfants vs. Adultes**

La figure III.2 offre les distributions cumulatives brutes, obtenues avec les durées des cycles syllabiques et des détente de pointers (strokes) en x, et le rang de ces valeurs de durée en y (rangs normalisés). Les deux résultats adultes sont donnés à titre de référence seulement. Les données de pointage adulte proviennent

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

de Levelt et al. (1985), une étude pionnière sur le pointage du doigt et de la voix (chez des sujets néerlandophones), avec 38 valeurs moyennes de temps de mouvement (MT) disponibles (moyenne 417.24ms ; écart-type :  $\pm 87.19$ ms). Les durées de syllabes correspondent à 18 valeurs moyennes extraites à partir de l'analyse d'un passage continu en anglais dans une étude extensive réalisée par Campbell (1992), ayant pour objectif la synthèse de parole (moyenne 172.33ms,  $\pm 108.65$ ms). Ce ratio adulte Pointage/Syllabe indicatif est 2.42, un ratio peu éloigné des valeurs de nos enfants (voir plus loin la Table III.7).

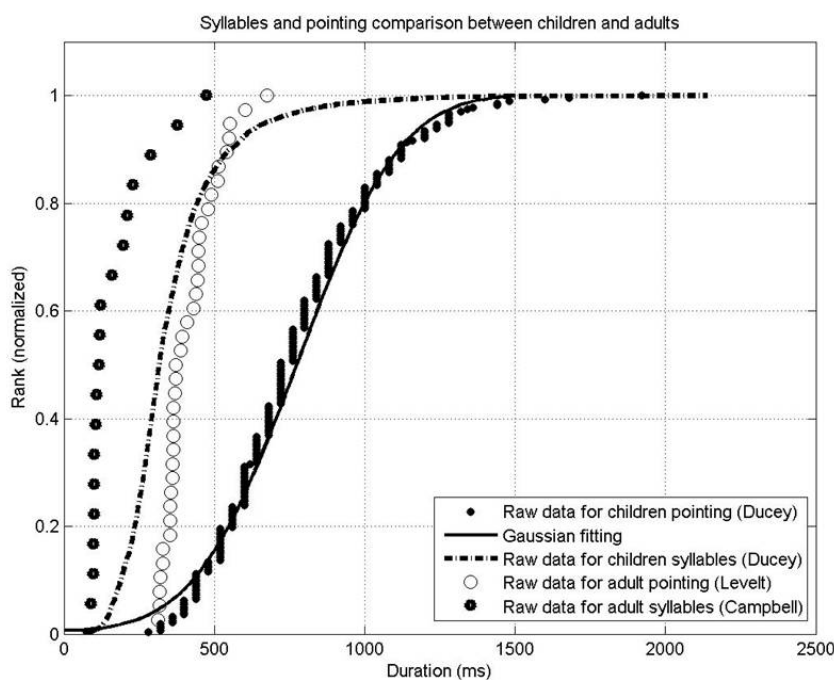


FIGURE III.2: Distributions des fréquences cumulées (durée-rang) pour les 4153 syllabes (courbe de tirets-points suivant les données brutes) et 276 pointers (points noirs, avec un fitting approximé par une courbe gaussienne, la plus à droite) produits par 6 enfants français entre 6 et 18 mois. Comme simple indication pour comparaison avec les adultes : les cercles fins correspondent aux 38 valeurs moyennes de temps de mouvement (MT) pour les données de pointer (deixis néerlandaise) de Levelt et al. (1982) ; et les cercles épais (les plus à gauche) correspondent aux 18 valeurs moyennes de durée de syllabes (analyse d'un passage continu en anglais) mesurées par Campbell (1992).

Les valeurs des enfants s'étendent de 280 ms à 1920 ms (un premier stroke de pointer très lent, voir Figure III.5). La durée moyenne est de 776 ms avec  $\pm 267$  ms d'écart-type ; l'approximation gaussienne est réalisée sur les données brutes. Nous avons donné la comparaison avec les adultes en utilisant les moyennes des 38 expériences pour lesquelles le temps de mouvement était disponible, ainsi que nous venons de le dire, et ce sont celles qui ont été exécutées avec la main droite dans Levelt et al. (1985). Elles dépendent du nombre de choix dans les déictiques

(this/that lamp, en hollandais dit/dat lampje), de la direction et de la distance des LEDs-cibles (champ gauche/droit ; proche/éloigné). Ces valeurs moyennes de temps de mouvement vont de 313 ms (un déictique, pour la direction proche droite choisie entre 2 LEDs, réponse « on-line » instantanée) jusqu'à 675 ms (deux déictiques, pour une direction gauche éloignée parmi un choix sur 4 LEDs, et une réponse « offline » : après que la LED se soit allumée, le sujet attendait la question de l'expérimentateur « Which light ? », avant de démarrer). Dans l'ensemble et comme attendu, les pointers spontanés des enfants sont plus lents que ceux des adultes, mettant en évidence un facteur maturationnel. La distribution de chaque enfant (normalisée en rang, dès lors que le nombre des observations collectées pour chacun d'entre eux était différent) comparée à la courbe cumulée générale, montre que trois d'entre eux (Figure III.3) sont plus proches de ce comportement général (Anatole: 763ms±255; Jules: 769ms±222; et Célia: 826ms±261), tandis que les autres (Figure III.4) sont soit plus rapides (Tom: 681ms±239) ou plus lents (Nicolas: 926ms±291; et Lise: 927ms±313), ce que nous verrons plus en détail ci-dessous.

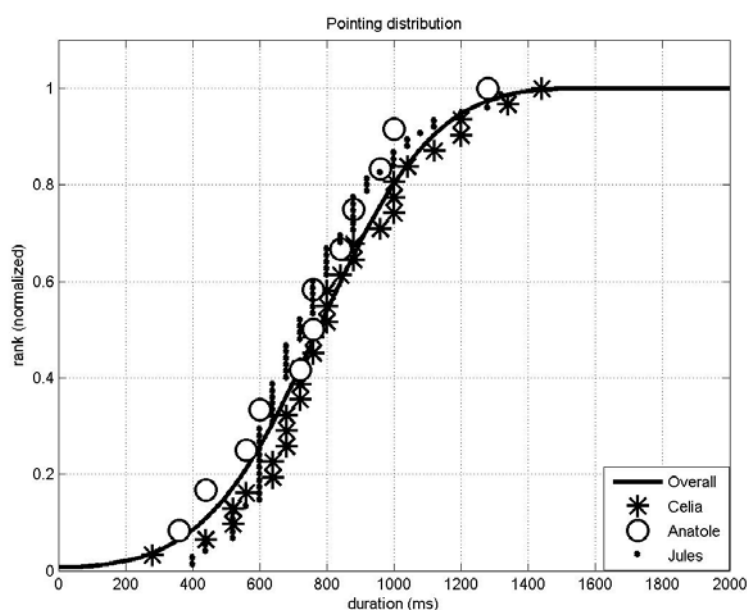


FIGURE III.3: Distributions normalisées des détentés de pointer pour Célia, Anatole et Jules. Leurs données sont plutôt proches de la courbe de distribution générale obtenue pour les 6 enfants (cf. Figure III.2).

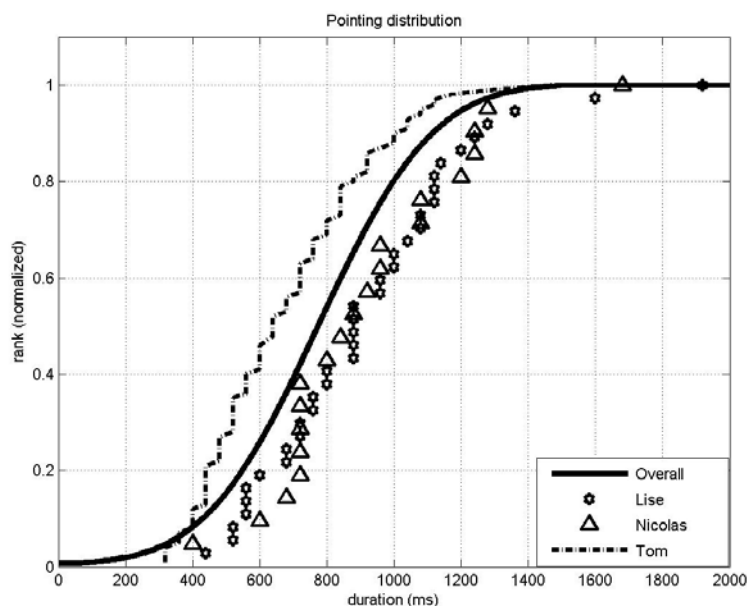


FIGURE III.4: Distributions normalisées des détentes de pointer pour Lise, Nicolas et Tom. Leurs données sont plutôt plus rapides (Tom, à gauche) ou plus lentes (Lise et Nicolas, à droite) que la tendance générale obtenue à partir des 6 enfants (cf. Figure III.2).

De 9 à 17 mois (mises à part les deux rares occurrences à 7-8 mois), on observe une tendance décroissante claire dans les moyennes et les distributions, allant d'environ 1s à environ 600 ms à 15-17 mois, en passant par un palier autour de 800 ms de 10 à 14 mois (Figure III.5). A nouveau, la maturation correspond à des gestes plus rapides.

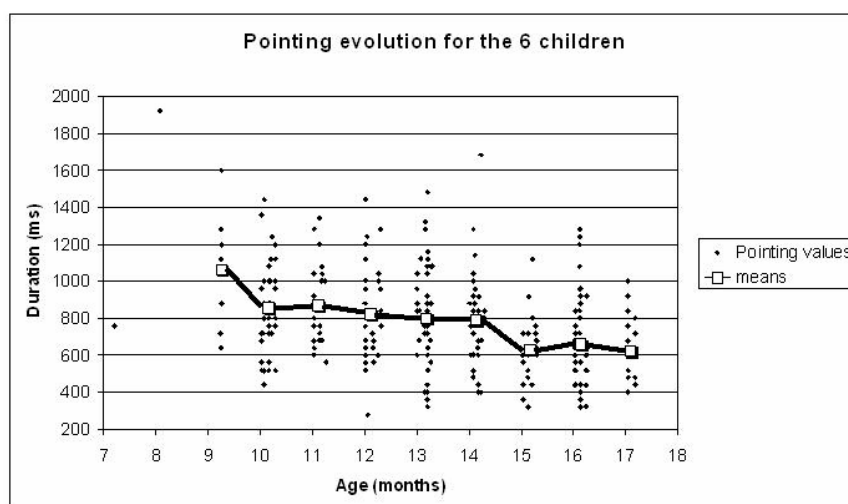


FIGURE III.5: Evolution des durées de détentes de pointer (diamants) pour les 6 enfants français entre 6 et 18 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis deux premiers points isolés (aucun à 6 mois, et 2 diamants à 7-8 mois), le pointer débute véritablement à 9 mois, et décroît en durée –après une période entre 10-14 mois où il est à environ 800 ms – vers 600 ms à 15 mois.

### III.4.2. Le mode de distribution des durées de détente du pointer : analyses individuelles

Après avoir exposé cette tendance globale dans les durées de pointer, intéressons-nous plus particulièrement à chaque enfant du corpus, afin de mettre en évidence leur propre distribution, ainsi que l'évolution dans la durée à travers les mois pour chacun d'eux.

#### III.4.2.1. Anatole

Anatole a réalisé une douzaine de pointers accompagnés de vocalisations lors des enregistrements. La première occurrence relevée se situe à 7 mois 20 jours mais reste très isolée, ses pointers survenant plutôt entre 12 et 15 mois (voir Figure III.6). Nous observons une tendance décroissante dans la durée des pointers chez Anatole sur la figure III.6. Les valeurs débutent autour de 800 ms à 7 mois pour tomber entre 600 et 400 ms à 15 mois. Les moyennes de la durée des pointers parmi les mois sont représentées sur la figure III.7 : de 7 à 12 mois, la moyenne tourne autour de 920 ms, puis descend à 880 ms à 14 mois, avant de chuter à 536 ms à 15 mois. On observe donc une décroissance dans la valeur des durées chez Anatole.

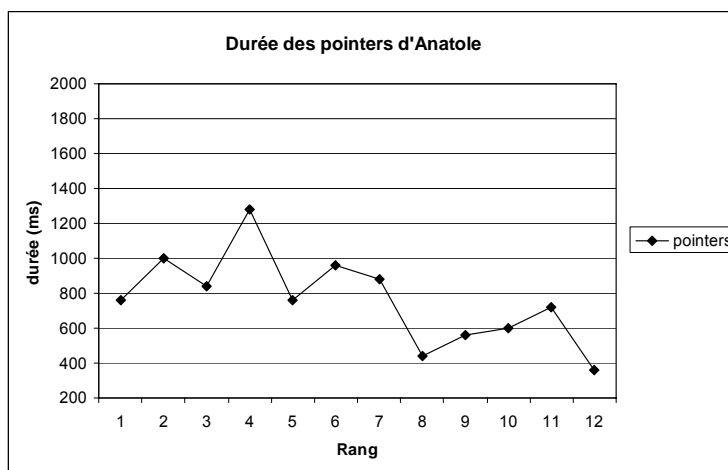


FIGURE III.6 : Evolution de la durée (ms) pour les 12 pointers d'Anatole, classés par rang d'apparition.

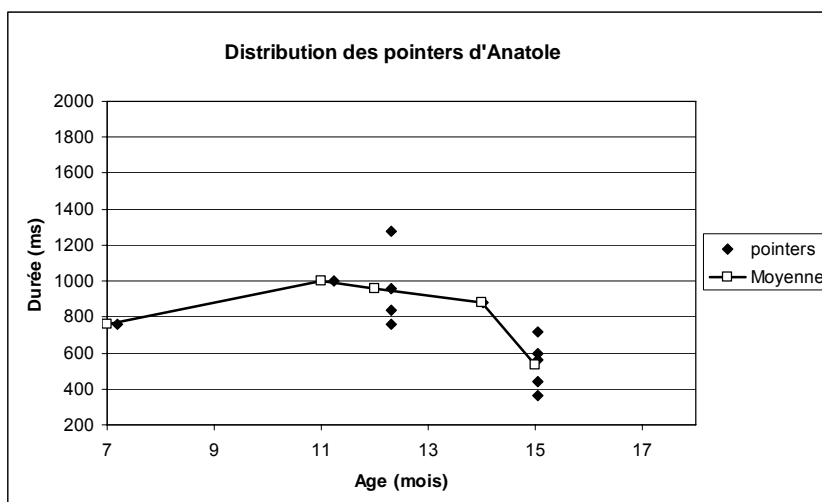


FIGURE III.7 : Evolution des durées des détentes de pointer (diamants) pour Anatole entre 7 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis un premier point isolé (1 diamant à 7), le pointer débute véritablement à 12 mois, et décroît en durée –après une période entre 12 et 14 mois où il est d'environ 950 ms– vers 530 ms à 15 mois.

Anatole est un des enfants qui se rapproche le plus du comportement général observé, par rapport à la courbe cumulée générale (Figure III.3). Ses pointers durent en moyenne 763ms avec un écart-type de  $\pm 255$ ms. Nous pouvons noter que sur 12 pointers, 10 d'entre eux sont réalisés avec la main droite, tandis que les 2 autres sont réalisés avec la main gauche (cf. plus haut).

### III.4.2.2. Célia

Célia totalise un nombre de 31 événements de pointers (Figure III.8). Ses premiers pointers débutent à 9 mois 24 jours et vont jusqu'à 15 mois 29 jours. On note l'existence de deux occurrences à 9 mois, suivie d'une « explosion » dans l'utilisation de ce geste à 10 mois, la plupart de ses pointers se concentrant surtout entre 10 et 12 mois. On note encore quelques rares occurrences à 13 et 15 mois.

La tendance décroissante dans les valeurs de pointers observée chez Anatole n'est pas aussi nette chez Célia (Figure III.9). Elle débute avec une moyenne relativement rapide à 9 mois (680ms), puis les pointers sont plus lents à 10 et 11 mois (883ms et 900ms respectivement). On note un étonnant changement dans ces valeurs à 12 mois, puisque la moyenne chute abruptement à 540ms, pour remonter à 940ms à 13 mois. Enfin, à 15 mois, la moyenne de la durée des pointers se stabilise autour de 680ms. Nous reparlerons en détail de cette chute dans les valeurs à 12 mois de Célia plus tard dans notre discussion.



## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

Les valeurs de pointers de Célia suivent, tout comme Anatole, le comportement général observé sur la courbe cumulée générale (Figure III.3). Ses pointers durent en moyenne 826 ms, avec écart-type de  $\pm 261$ ms. Sur les 31 pointers relevés, notons que 22 d'entre eux ont été réalisés avec la main droite, et les 9 restant avec la main gauche (cf. plus haut).

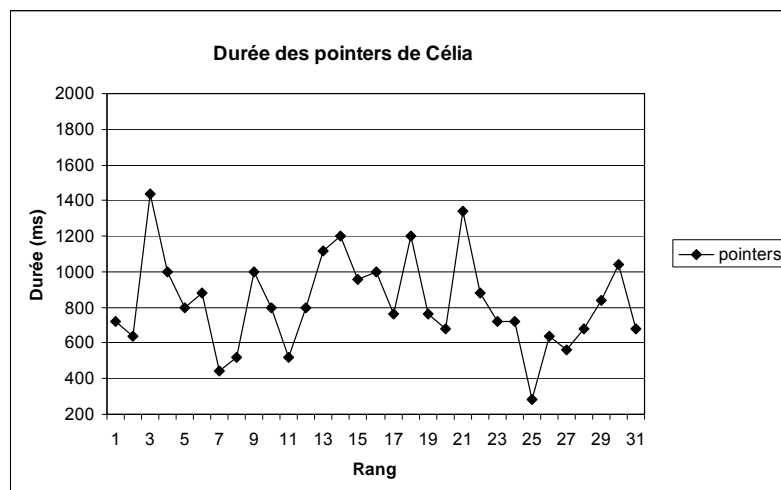


FIGURE III.8: Evolution de la durée (ms) pour les 31 pointers de Célia, classés par rang d'apparition.

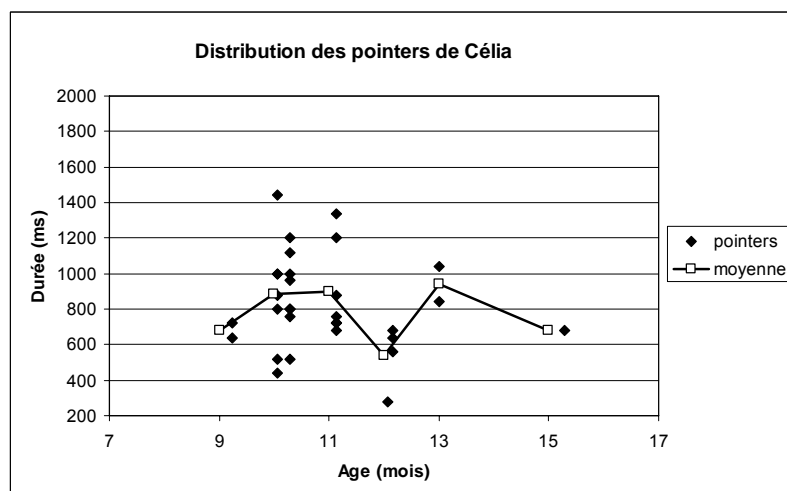


FIGURE III.9 : Evolution des durées de détentes de pointer (diamants) pour Célia entre 9 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Deux pointers commencent à 9 mois ; on note une majorité de productions entre 10 mois et 12 mois.

### III.4.2.3. Jules

Jules a réalisé 75 pointers, soit beaucoup plus que les deux enfants précédents. Etant donné qu'aucune procédure d'élicitation de pointers n'a été recherchée, nous observons cette grande variation naturelle dans la réalisation

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

spontanée de pointers parmi les enfants. Les premiers pointers accompagnés de vocalisation de Jules sont apparus à l'âge de 10 mois 6 jours et se sont étendus jusqu'à l'âge de 16 mois 26 jours (Figure III.10). On peut noter à ce stade de l'analyse une plage de variation dans l'apparition de ces premiers pointers, qui pour Jules débutent plus tard que pour Anatole et Célia par exemple. Chez Jules, les pointers relevés sont très concentrés dans cette période entre 9 et 16 mois, le nombre d'occurrences augmentent vite sans rupture comme observé chez Anatole pour qui quelques pointers apparaissent à 7 mois avant d'observer une explosion à 12 mois.

Si l'on observe la tendance des moyennes de durées de pointers dans le temps (Figure III.11), on note une évolution progressive vers des pointers plus rapides. Ainsi, de 10 à 13 mois, les valeurs sont assez proches (930ms, 857ms, 827ms, 844ms respectivement pour 10, 11, 12 et 13 mois), avec une moyenne globale d'environ 865ms. Puis nous observons un palier décroissant dès 14 mois, où la vitesse des pointers s'accélère (662ms, 667ms, 628ms à 14, 15 et 16 mois respectivement) jusqu'à 16 mois ; la moyenne des durées entre 14 et 16 mois étant de 652ms, soit quasiment 200ms de gagnées par rapport aux mois précédents.

Jules se situe, comme Célia et Anatole, dans le profil général de comportement, suivant la courbe cumulée générale (Figure III.3). La moyenne pour la durée des pointers de Jules s'élève à 769ms (768.53), avec un écart-type de  $\pm 222$ ms. Contrairement à Célia et Anatole qui semblent plus utiliser la main droite pour pointer, Jules utilise les deux mains indifféremment : ainsi, sur 75 pointers, 35 sont réalisés avec la main droite et 40 avec la main gauche (cf. plus haut).

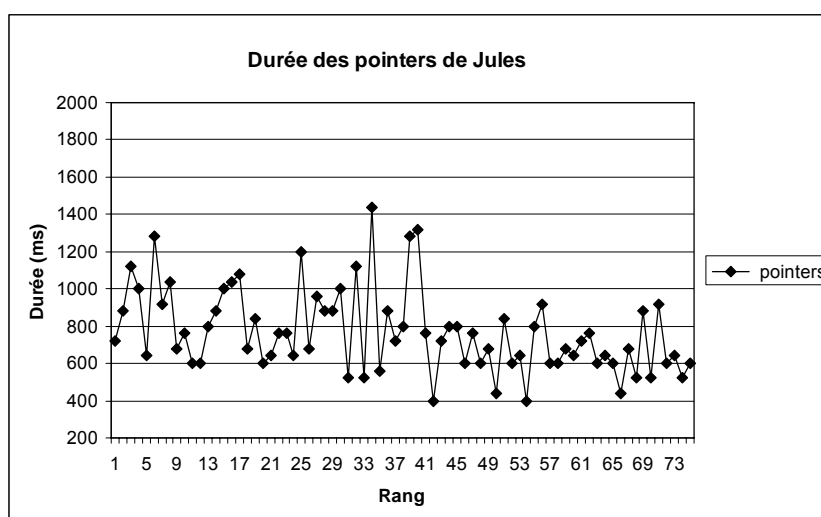


FIGURE III.10: Evolution de la durée (ms) pour les 75 pointers de Jules, classés par rang d'apparition.

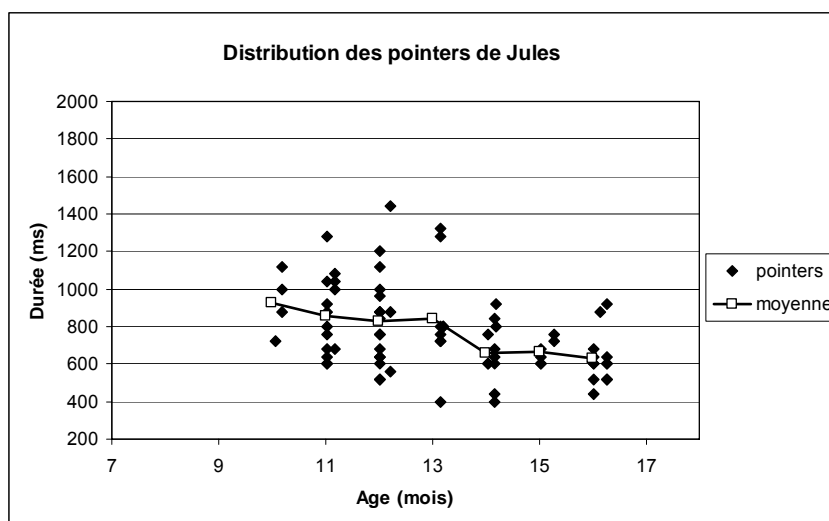


FIGURE III.11 : Evolution des durées des détente de pointer (diamants) pour Jules entre 10 et 16 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Les occurrences sont assez concentrées dans cette période d'observation. On constate sur la figure une tendance à la baisse des durées des pointers à 14 mois.

#### III.4.2.4. Nicolas

Nicolas a réalisé un total de 21 pointers. Ses premiers pointers avec vocalisation sont apparus à l'âge de 10 mois 23 jours, et se sont étendus jusqu'à l'âge de 16 mois 12 jours. On peut remarquer sur la figure III.12 que Nicolas réalise en fait son premier pointer à 10 mois 23 jours, mais qu'il s'agit d'une occurrence isolée, de même qu'à 12 mois. La production de pointers commence à être plus fournie plutôt dès 13 mois.

En ce qui concerne l'évolution des durées de pointers à travers l'âge (Figure III.13), on note chez Nicolas une évolution peu régulière. A 10 mois, le pointer est relativement lent (1240 ms), il chute ensuite à 12 mois autour de 720 ms. Toutefois, la durée des pointers diminue à nouveau autour de 960ms entre 13 et 14 mois (880ms et 1040ms respectivement), pour redescendre ensuite et se stabiliser autour de 915ms environ à 15 et 16 mois (920ms et 913ms respectivement). Si l'on prend ces résultats de façon globale, on note tout de même une légère décroissance dans la durée de ses pointers.

Contrairement à Anatole, Célia et Jules, Nicolas ne possède pas un comportement de pointer qui suit la courbe de tendance générale (voir figure III.4). Son propre comportement apparaît comme plus lent que le comportement général,

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

avec une moyenne générale de durées des pointers de 926ms, et un écart-type de  $\pm 291$ ms. Nicolas semble ne pas avoir également de dominance manuelle trop marquée, puisque 12 des 21 pointers sont réalisés avec la main gauche, et les 9 restant avec la main droite (cf. plus haut).

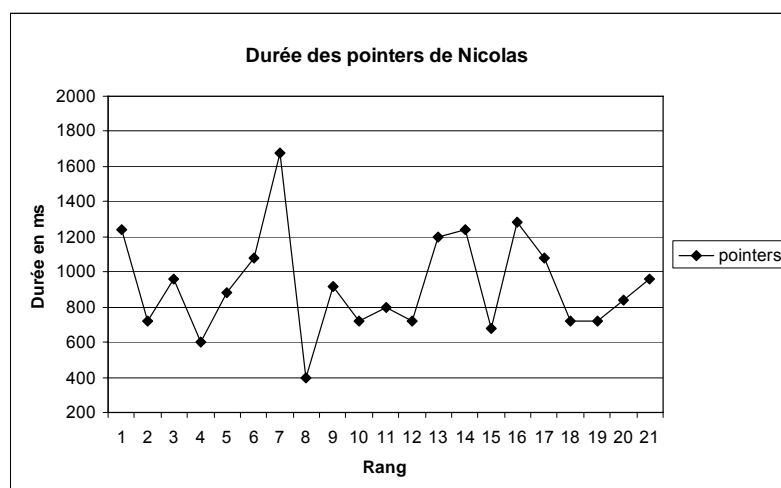


FIGURE III.12 : Evolution de la durée (ms) pour les 21 pointers de Nicolas, classés par rang d'apparition.

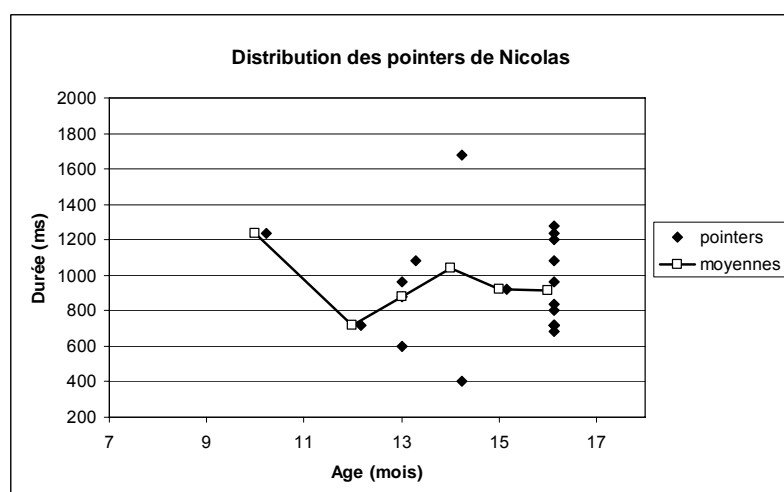


FIGURE III.13 : Evolution des durées des détentes de pointer (diamants) pour Nicolas entre 10 et 16 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis deux premiers points isolés (1 diamant à 10 et 1 diamant à 12 mois), les pointers débutent à 13 mois, la majorité étant produite à 16 mois.

### III.4.2.5. Lise

Parmi les sessions concernant Lise, on dénombre un total de 37 pointers. Nous relevons les premiers pointers à l'âge de 8 mois 8 jours, et les derniers pointers à l'âge de 15 mois 21 jours. Sur la Figure III.14, nous pouvons remarquer une

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

première occurrence très isolée à l'âge de 8 mois, les productions devenant plus conséquentes dès 9 mois.

Lorsqu'on s'intéresse à la distribution des durées de pointers à travers le temps, et particulièrement les moyennes relevées pour chaque mois (Figure III.15), on distingue très clairement une décroissance des valeurs, et ceci très nettement entre 8 et 10 mois, où la chute est très brutale (1920ms puis 1216ms puis 797,5ms à 8, 9 et 10 mois respectivement). A partir de 12 mois, les valeurs s'équilibrent autour d'une moyenne globale d'environ 900 ms (970ms, 864ms, 993ms, 787ms à 12, 13, 14 et 15 mois respectivement). Au moment où Lise commence à produire le plus des gestes de pointers accompagnés de vocalisations, on observe donc cette diminution des durées.

Il est intéressant d'analyser les données de Lise car elle suit de façon très étroite le comportement de l'enfant précédent, Nicolas. Celle-ci ne présente pas un comportement proche de celui de la courbe cumulée générale (voir Figure III.4), et fait donc partie du groupe des trois enfants classés comme plus différents. Son comportement est proche de celui de Nicolas dans la mesure où ses pointers sont également plus lents que la moyenne générale, à savoir 927ms (une valeur très proche de Nicolas !), avec un écart-type de  $\pm 313$ ms. Lors de la réalisation de ses 37 pointers, Lise a utilisé 21 fois la main gauche, et 16 fois la main droite (cf. plus haut). Comme Nicolas, elle semble utiliser plus la main gauche, mais dans des proportions bien faibles par rapport à la main droite.

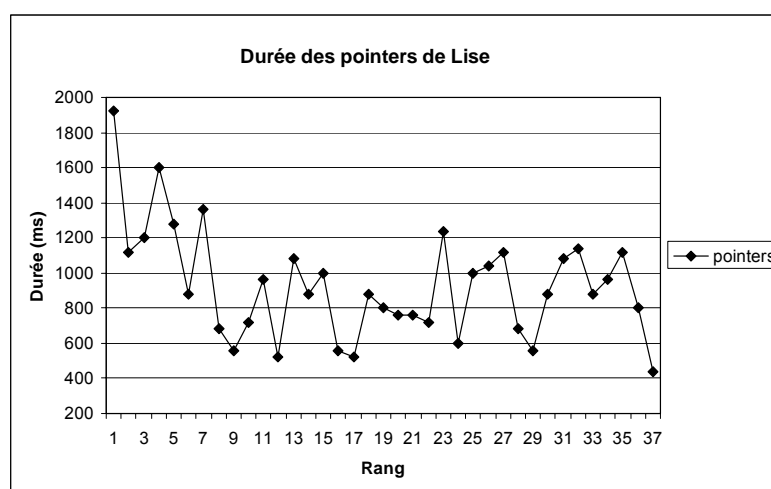


FIGURE III.14: Evolution de la durée (ms) pour les 37 pointers de Lise, classés par rang d'apparition.

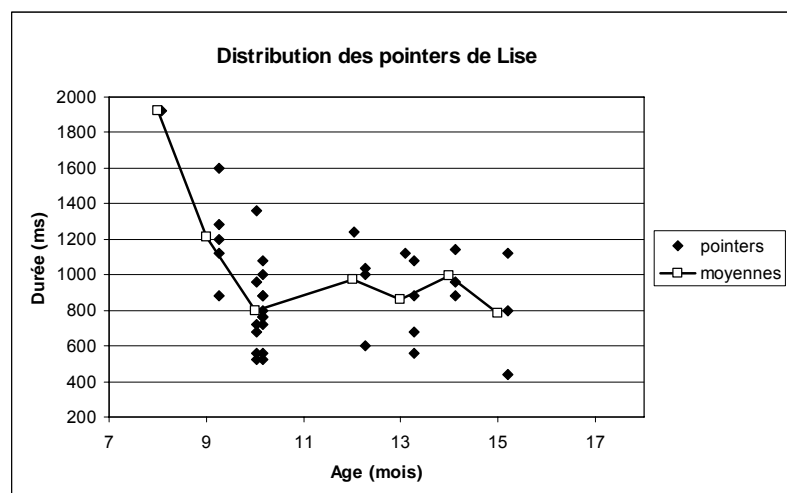


FIGURE III.15 : Evolution des durées de détente de pointer (diamants) pour Lise entre 8 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis un premier point isolé (1 diamant à 8 mois), les pointers débutent véritablement à 9 mois, avec une explosion notable à 10 mois. La durée décroît fortement de 8 à 10 mois, pour se stabiliser autour de 900ms.

### III.4.2.6. Tom

Tom a réalisé une centaine de pointers : c'est l'enfant qui en a totalisé le plus grand nombre de tous. Ses pointers ont commencé à 10 mois 21 jours, et se sont prolongés jusqu'à l'âge de 17 mois 18 jours. Tom réalise peu d'occurrences les premiers mois (on note 1 production à 10 mois et seulement 2 productions à 11 mois). C'est sensiblement à partir de 13 mois, et de façon écrasante, que la réalisation de pointer-parole « explose » (voir Figure III.16).

La tendance décroissante des valeurs de durées de pointer à travers l'âge semble également se confirmer une fois encore avec Tom (voir Figure III.17). Entre 10 et 14 mois, les valeurs de moyennes sont sensiblement les mêmes, autour de 768ms (720ms, 780ms, 760ms, 812.5ms à 10, 11, 13 et 14 mois respectivement), puis on observe, dès 15 mois jusqu'à 17 mois, une décroissance amorcée de ces valeurs, puisque la moyenne globale entre 15 et 17 mois tombe à 577ms (547ms, 562ms et 622ms à 15, 16 et 17 mois respectivement), soit une différence de quasiment 200ms entre ces deux périodes, avec des pointers très rapides dans les derniers mois !

Lorsqu'on observe les données de Tom par rapport à la courbe cumulative générale (Figure III.4), on s'aperçoit, de la même façon que pour Lise et Nicolas, que

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

le comportement de Tom est assez atypique, et ne rentre pas dans le profil le plus proche de la courbe générale. Mais contrairement à Lise et Nicolas, qui ont des durées de pointers moyennes plus lentes que la normale, Tom se distingue des autres enfants par le fait (i) qu'il produit beaucoup plus de pointers certes, mais (ii) surtout car sa durée moyenne de pointers parmi toutes ses réalisations est nettement plus rapide que tous les autres enfants, soit à 681ms, avec un écart-type de  $\pm 239$ ms. On note chez Tom une dominance dans l'utilisation de la main pour pointer (cf. plus haut), en effet, il a utilisé bien plus fréquemment la main droite (67 fois sur 100) que la main gauche (33 fois).

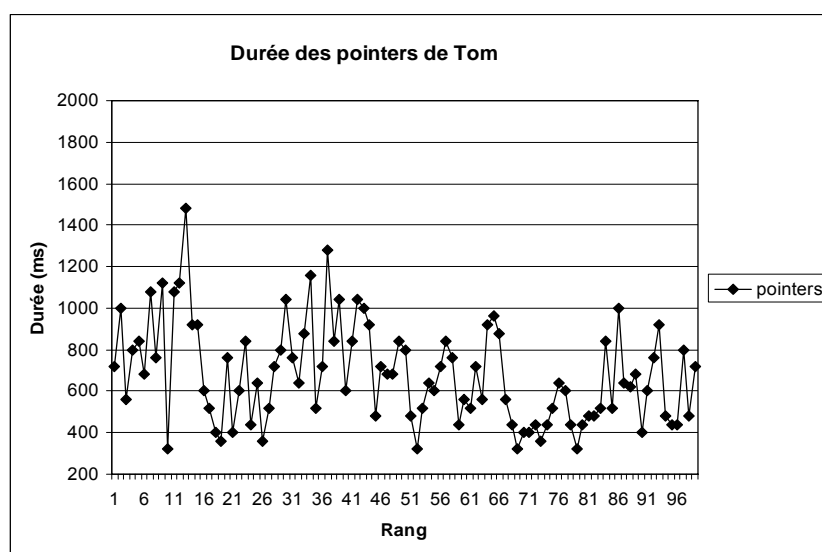


FIGURE III.16 : Evolution de la durée (ms) pour les 100 pointers de Tom, classés par rang d'apparition.

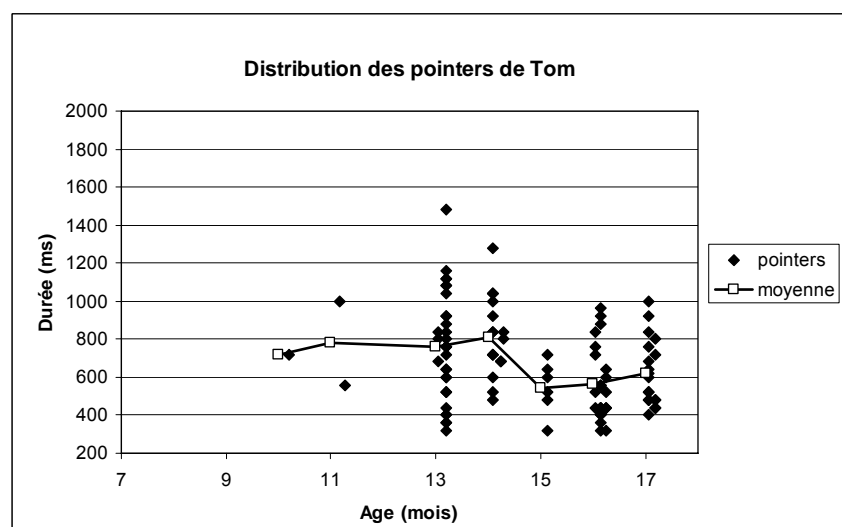


FIGURE III.17 : Evolution des durées des détentes de pointer (diamants) pour Tom entre 10 et 17 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement

visuel par mois). Hormis quelques premiers points isolés (1 diamant à 10 mois et 2 diamants à 11 mois), les pointers débutent véritablement à 13 mois, avec une explosion notable à ce moment-là. La durée stable jusqu'à 14 mois (autour de 800ms) décroît par la suite de 15 à 17 mois autour de 600ms, soit presque 200ms de gagnées.

### **III.4.2.7. Conclusion**

En conclusion de ces analyses par individus, nous avons pu remarquer que la moitié des enfants semble suivre le patron global de décroissance, tandis que d'autres semblent s'en écarter, en adoptant des stratégies différentes, c'est-à-dire en jouant sur des pointers plus lents, ou des pointers plus rapides. L'âge d'apparition des pointers avec vocalisation varie également selon les enfants, débutant pour certains autour de 7-8 mois et pour d'autres vers 9-10 mois. Nous observons pour quasiment presque tous les enfants une décroissance dans les valeurs moyennes de durées de pointers à travers le temps, ce qui peut témoigner d'une certaine maturation.

### **III.4.3. Le mode de distribution du cycle du babillage : apports théoriques**

Il existe une importante variation sur les mesures de durées de syllabes faites par différents chercheurs, ce qui est dû à l'évidence à un trop grand nombre de facteurs.

Nous pouvons ainsi constater l'influence claire des patrons linguistiques de l'anglais américain, du français et du gallois dans l'étude suivante. Cette étude de Vihman et al. (in press) permet de s'interroger sur une acquisition comparable ou non du rythme pour des enfants issus d'environnements linguistiques variés. Ont été étudiés les patrons rythmiques des productions d'enfants acquérant trois langues dont l'accent et le rythme sont distincts, l'anglais américain, le français et le gallois. A deux points de développement importants : le premier au moment où l'enfant possède 4 mots, et le second où l'enfant possède 25 mots. Les facteurs rythmiques contribuent chez l'enfant à ajuster les modèles de durées typiques d'une séquence de type CVCV à ceux de la langue adulte. Ils permettent de mieux comprendre l'évolution des durées segmentales dans le développement linguistique de l'enfant.

Bien qu'on observe un allongement de la syllabe finale général, les patrons de durée de ces trois langues chez l'adulte sont distincts. Le patron accentuel anglais



pour des dissyllabes est trochaïque, soit fort-faible. Cette langue est caractérisée par des facteurs perceptifs de « raccourcissement » qui affectent le noyau vocalique. Le français accentue la syllabe finale, principalement par un effet d'allongement, donnant des dissyllabes uniformément iambiques. Enfin, le gallois tend à avoir des dissyllabes de type trochaïque, tandis que dans des unités plus grandes, comme la phrase, le gallois opte soit pour un patron iambique, soit pour un trochaïque. L'accent gallois se caractérise également par une intensité plus grande sur la syllabe initiale. L'anglais est traditionnellement donné comme une langue de type accentuel (*stress-timed*), à l'inverse du français qui sera de type syllabique (*syllable-timed*), le gallois se plaçant entre ces deux catégories.

Partant de ces connaissances sur l'adulte, Vihman et al. ont alors voulu tester les effets des différences d'input linguistique sur l'apprentissage rythmique, concernant le patron accentuel du mot ou de la phrase, et ce au niveau de l'alternance CV. Ils ont pour cela analysé les productions dissyllabiques des enfants, dans le babillage et dans les premiers mots, et les productions dissyllabiques de leurs mères aux mêmes points de référence.

Les auteurs observent que pour la production du dissyllabe [babi], par exemple, trouvée dans les trois langues<sup>26</sup> pour les locuteurs adultes :

- l'anglais est marqué par une première voyelle longue ;
- le français est marqué par une seconde voyelle longue ;
- le gallois est marqué par une consonne médiane (MidC=Mid Consonant) longue.

Les auteurs se sont ensuite penchés sur les réalisations dissyllabiques des enfants, observant en premier lieu qu'elles sont largement dominées par la présence d'une consonne médiane occlusive. Ce qui les a conduits à centrer leur étude uniquement sur les VCV comportant une consonne de ce mode d'articulation. Vihman et al. relèvent qu'à l'étape de 4 mots, la majorité des enfants, comme les adultes, montrent une proportion de V2 plus longues que V1, donnant un effet d'allongement sur la syllabe finale. A l'étape de 25 mots, les tendances sont très différentes, permettant d'affirmer que les patrons sont plus proches des patrons adultes cibles. En effet, les enfants américains voient leur V1 s'allonger et leur MidC se raccourcir. Les enfants français ont une V1 qui devient plus courte et une V2 qui

---

<sup>26</sup> En anglais [ba:bi], en français [babi:] et en gallois [bab:i].

s'allonge. Et les enfants gallois ont une V1 qui se raccourcit tandis que V2 et MidC s'allongent. Il semble donc alors qu'à l'étape de 4 mots, on ne pourrait pas assigner aisément un enfant issu de ces trois groupes dans son groupe linguistique ambiant, uniquement sur la base de son patron rythmique. Une explication serait que les premiers mots issus du babillage sont fortement contraints par les possibilités articulatoires propres aux enfants, et que ces contraintes biomécaniques sont communes aux enfants apprenants de langues différentes, donnant des patrons segmentaux communs. De plus, l'influence des patrons rythmiques globaux de la langue adulte semble être très faible, dans ces vocalisations largement dominées par le babillage.

Ce n'est que plus tardivement qu'on observe un progrès dans l'atteinte du patron adulte. Il faut toutefois remarquer que les enfants américains ont, à l'étape de 25 mots, une tendance à la variabilité, en étant moins proches des patrons adultes, à l'inverse des enfants français et gallois, qui tendent à se conformer étroitement au patron adulte de leur langue. Il semble ainsi plus aisé pour les enfants français, ainsi que pour les gallois, d'établir un *mapping* direct du rythme d'une cible adulte à une forme qui existe dans son répertoire articulatoire, dans le cas étudié ici des dissyllabes. Cet avantage semble lié aux formes dissyllabiques françaises qui sont toutes iambiques, tandis que l'anglais, et le gallois dans une certaine mesure, doivent faire face à un mélange de modèles trochaïques et iambiques.

Ainsi, une langue accentuelle comme l'anglais possède de façon inhérente une grande variabilité rythmique, qui fournit donc un patron relativement plus difficile à intégrer pour les enfants, comparé à une langue syllabique comme le français. Les patrons rythmiques adultes vont ainsi réclamer de la part de l'enfant une maîtrise plus pointue de la complexité phonotactique, relativement grande, et encore plus particulièrement pour les enfants américains apprenant l'anglais. Cette maîtrise du patron rythmique semble ainsi se développer parallèlement à la mise en place des processus d'apprentissage lexical.

Konopczynski (1993) s'est également posé la question de l'acquisition de la structuration rythmique, dans différentes langues européennes. Pour étudier les patrons rythmiques, l'unité choisie a été la syllabe, dans la mesure où la durée syllabique semble être un bon corrélât du rythme en général. Konopczynski observe pour les enfants français, que vers l'âge de 2 ans, lors du babillage tardif, les

« vocoïdes » (vocalisation sans consonne) vont chuter en nombre, alors qu'ils ne possèdent pas encore une structuration rythmique bien établie. Par contre, lors du proto-langage, les durées des « syllaboïdes » vont subir un changement. En effet, après avoir été isochrones, la durée des syllabes va petit à petit changer selon la position. Suivant en cela le patron français, la syllabe non-finale va se raccourcir au profit d'un allongement de la syllabe finale, un phénomène dont nous reparlerons ci-après sur un cas précis.

Par comparaison, pour l'acquisition de l'anglais, Konopczynski argumente, d'après les travaux d'autres auteurs, qu'au début l'isochronie existe en anglais comme en français, et que l'allongement final se met en place, mais de façon moins régulière que pour le français. Quant à la prééminence de l'accent, elle ne serait repérée auditivement qu'à partir de l'âge de 2 ans. L'enfant anglais doit alors faire face à une complexité accentuelle imprévisible qui va l'obliger à quasiment apprendre chaque accent pour chaque mot. En effet il semble que les mères anglaises utilisent un fort accent d'emphase lorsqu'elles s'adressent à leur bébé, accent qui ne serait pas situé là où il devrait l'être dans la production normale. Cet argument pourrait ici expliquer pourquoi, chez Vihman, les enfants anglais ont encore vers 16-17 mois un patron rythmique assez peu proche de celui de l'adulte.

Konopczynski a également étudié la structure accentuelle d'enfants portugais, une langue à accent lexical à fonction contrastive, qui place généralement l'accent sur l'avant-dernière syllabe (bien qu'il soit possible d'accentuer les autres syllabes). L'auteur souligne le fait qu'à 9 mois, on observe une quasi-synchronie des syllabes, et qu'à 12 mois, un début d'allongement est repérable. Il semble exister une certaine précocité dans la mise en place de la structure accentuelle chez les enfants étudiés par l'auteur. L'espagnol montre à peu près le même patron, avec un emplacement aléatoire de l'accent, jusqu'à la maîtrise à 2 ans des règles accentuelles, avec un accent sur la pénultième dans la majorité des cas.

L'exemple du hongrois est un exemple intéressant cité par l'auteur, dans la mesure où l'ordre des mots est libre, et où l'accent se pose sur la syllabe initiale du mot. Les études de Kassai<sup>27</sup> citées par Konopczynski montrent que l'enfant peut

---

<sup>27</sup> Kassai I. (1988). Prosodic development: stressing procedures of a Hungarian child. *Budapest: 6<sup>th</sup> International Phonology meeting*, Discussion Papers 1, 32-34. Kassai I. (1991). The emergence of intonation and stress in Hungarian: a case study. *Actes du XIIIe ICPHS*. Aix-en-Provence, 1, 328-332.

placer l'accent sur la syllabe qu'il désire, tout en marquant néanmoins une préférence pour la première ou la dernière qui est allongée. On observe donc que même dans une langue où l'accent se place en initiale de mot, l'enfant va passer par une étape de transition où il pratique l'allongement final. L'enfant a de plus tendance à placer plus d'accents sur ses productions que ne l'exige la règle d'accentuation hongroise, où normalement seule l'information nouvelle doit être accentuée. Puis, progressivement, l'enfant va venir placer l'accent en initiale, conformément au patron adulte.

En conclusion des études menées par Konopczynski, il ressort nettement que, malgré les différentes stratégies d'accentuation inhérentes aux différentes langues étudiées, les enfants semblent toujours passer par une étape d'isochronie au début, suivie ensuite de l'apparition d'un allongement final de la syllabe. L'allongement final constitue une transition vers la mise en place de la structuration rythmique, accentuelle et temporelle, liée aux contraintes de la langue de l'enfant. Konopczynski (1990) a posé l'hypothèse d'une « horloge neurale interne », dont le fonctionnement se verrait ensuite bouleversé par les contraintes liées à la langue propre de l'enfant.

Van der Stelt (1976) a conduit une étude visant à étudier les bases du développement du rythme dans deux groupes spécifiques d'enfants : un premier groupe d'enfants sourds dès la naissance ; et un deuxième groupe d'enfants devenus sourds entre 3 et 26 mois, suite à une méningite. Tous ces enfants sont scolarisés dès 4 ans, avec une méthode visant à leur apprendre à parler en leur donnant des indices articulatoires. Ont été comparés les mouvements diadochokinétiques des lèvres et de la langue, dans une tâche où les enfants devaient répéter de plus en plus vite des énoncés du type [tututututu]. Les résultats, non attendus par l'auteur, sont ceux d'une hypothèse nulle, puisque les enfants qui sont devenus sourds avant 9 mois ne diffèrent pas dans l'ensemble des enfants sourds de naissance. Les résultats individuels montrent par ailleurs que c'est la période pendant laquelle l'enfant a été entraîné à un modèle rythmique qui va être décisive: plus sa mémoire kinesthésique a été entraînée, meilleurs sont ses résultats, quel que soit le moment où la surdité est apparue chez l'enfant.

Nous venons de voir de façon générale les grandes tendances liées à la mise en place du rythme propre à la langue chez l'enfant, avec des données comparatives sur diverses langues utilisant des systèmes accentuels variés. Il sera maintenant

question de l'émergence du rythme chez l'enfant d'un point de vue moteur, en nous s'intéressant particulièrement à la mise en place du contrôle mandibulaire dans le contrôle du rythme propre à la langue de l'enfant.

### ***III.4.3.1. Le babillage canonique comme une activité rythmique***

Le babillage est conçu par de nombreux chercheurs comme faisant partie d'un processus global de développement des comportements rythmiques de l'enfant. Thelen (1981) dans son étude sur le processus développemental du comportement rythmique chez l'enfant, a pu relever une tendance générale des mouvements à former des patrons répétitifs. Elle a observé particulièrement que l'activité des battements de main culminait autour de 6-7 mois chez l'enfant, soit à la même période que le début du babillage. Selon elle, ces stéréotypies rythmiques se retrouvent dans d'autres espèces, et semblent constituer des patrons moteurs très organisés qui vont être à la base de compétences motrices ultérieures, coordonnées et plus complexes.

De la même façon, Kent (1992) défend l'idée que la structure rythmique ne soit pas uniquement dédiée au comportement vocal, mais fasse partie de comportements moteurs plus généraux, dans la mesure où ces comportements surviennent dans la même gamme d'âge que le comportement rythmique vocal. Les activités rythmiques des jambes, du tronc, des mains, des doigts surviennent immédiatement avant ou en même temps que le début du babillage.

Ejiri (1998) a mené une étude sur 28 enfants japonais entre 5 et 9 mois, afin de relever quels types de comportements rythmiques émergent lorsque l'enfant entre dans le babillage, et afin de déterminer à quelle fréquence ces activités rythmiques surviennent dans des conditions naturelles. Deux hochets, un audible et un autre inaudible, ont été donnés à l'enfant afin de mesurer son activité rythmique. L'auteur observe qu'à l'étape du début du babillage, les enfants secouent plus significativement les hochets, avec un nombre moyen de secousses rythmiques atteignant son pic autour du début du babillage. Plus précisément, les activités de main rythmique sont survenues plus fréquemment aux étapes du début du babillage, et un mois après, plutôt que durant les mois précédant son émergence, et ceci indifféremment à l'audibilité du hochet (cf. aussi pour le « banging », Masataka, 2001).

Lorsque l'on s'intéresse à l'étude du babillage, on est tenté dans un premier temps, de se poser la question de la mise en place du contrôle du rythme mandibulaire chez l'enfant. Quelle est donc sa fréquence ? Sa fréquence est-elle différente selon la langue maternelle de l'enfant ? Et surtout quelle est l'évolution de ce contrôle du rythme mandibulaire depuis son apparition ? Nous allons tenter de répondre à ces questions en nous appuyant sur les données disponibles dans la littérature, et sur celles recueillies par nous dans notre laboratoire.

### ***III.4.3.2. Quelle fréquence pour le rythme du babillage ?***

Bickley et al. (1986) ont mené une étude acoustique afin d'étudier l'apparition de « proto-syllabes » dans les vocalisations d'une enfant. Leur méthode a consisté à examiner l'enveloppe du signal audio du babillage, soit la courbe d'intensité, pour établir la présence ou l'absence de régularités rythmiques, en prenant comme base les périodicités de basse fréquence, qui reflètent les mouvements articulatoires alternant des gestes d'ouverture et de fermeture du conduit vocal. La fonction d'auto-corrélation (AC) et la transformée discrète de Fourier (DFT), obtenues à partir de l'enveloppe, ont été testées et comparées aux mesures manuelles (cf. Figure III.18 ci-dessous). On observe une bonne correspondance globale entre les deux ensembles de mesures.

En ce qui concerne l'apparition d'un comportement rythmique caractérisé, les auteurs notent que de 21 à 25 semaines, soit environ vers 5 mois, moins de 30% des énoncés contiennent une variation rythmique. Ce n'est que vers la semaine 31, autour de 7 mois, qu'une augmentation dans ce pourcentage d'énoncés rythmiques est significative chez l'enfant. Pour l'enfant étudiée dans cet article, le début du babillage canonique a été noté auditivement dans la semaine 33. Ce qui amène Bickley et al. (1986) à conclure que le début effectif du rythme mesuré, précède le repérage audio du début du babillage d'au moins deux semaines.

Cette figure III.18 de Bickley et al. (1986) nous a permis de récupérer des données (toujours précieuses) de fréquence du babillage. Nous avons mesuré pour cela graphiquement les valeurs pour chaque point.

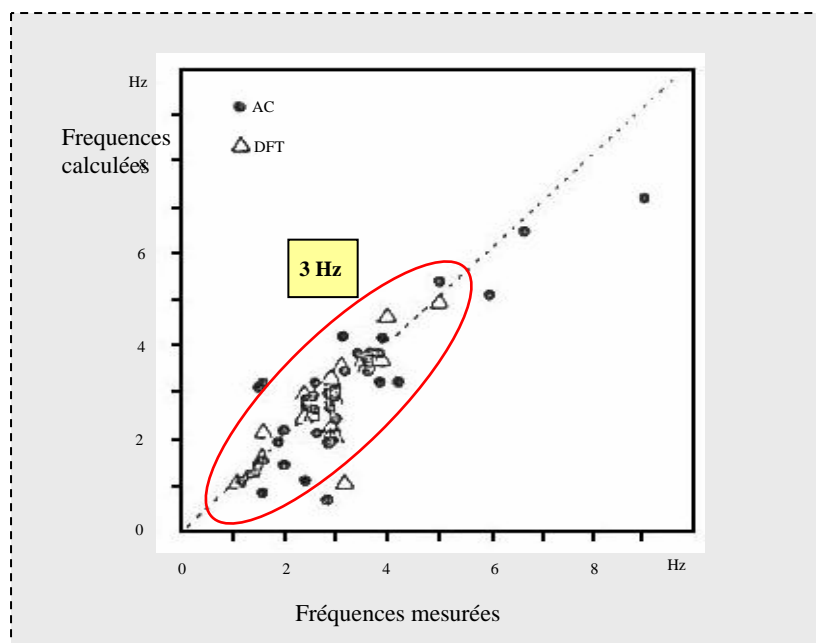


FIGURE III.18 : Correspondance entre fréquences du babillage mesurées manuellement et calculées automatiquement. L'abscisse représente les fréquences mesurées à la main à partir de la forme de l'onde ; l'ordonnée les fréquences calculées (cf. texte) par les méthodes d'autocorrélation (AC : cercles) et de Fourier (DFT : triangles). L'ellipse ajoutée par nos soins montre la concentration des données autour de 3 Hz (D'après Bickley et al., 1986).

D'après notre lecture des mesures de Bickley et al. (1986), nous pouvons mettre en évidence la fréquence rythmique moyenne du babillage pour cet enfant. Les deux méthodes ont donné une dispersion des fréquences autour de 3Hz. En effet, la fréquence moyenne est de 2.94Hz pour la méthode de mesure à la main, et de 2.8Hz pour les méthodes automatiques (AC et DFT confondues). Cela signifie que l'enfant produit en moyenne un rythme de 2.87 cycles par seconde, autrement dit qu'elle réalise près de trois syllabes à la seconde.

Koopmans van Beinum (1993) s'est également intéressée aux durées de syllabes moyennes chez plusieurs enfants, à différentes étapes développementales. Lors des étapes 2 et 3 –correspondant à une phase arrêt/début de phonation dans une seule unité respiratoire vers 6 semaines, et à des mouvements articulatoires en combinaison avec des mouvements phonatoires acquis plus tôt– on observe une certaine variabilité inter-enfants avec des gammes de valeurs allant de 400 à 575ms (moyenne 475ms) pour l'étape 2 ; et allant de 455 à 625 ms (moyenne 500ms) pour l'étape 3. La durée de la syllabe semble ne se stabiliser qu'au cours de l'étape 5, qui correspond à l'émergence du babillage canonique. En effet, la variabilité est plus

restreinte, avec une gamme de valeurs allant de 400 à 450ms parmi les 4 enfants de cette étude, avec une moyenne générale de 425ms. Koopmans van Beinum indique également que le débit d'articulation (*articulation rate*) est de 2.36 syllabes/seconde en moyenne. Ce résultat apparaît comme convergeant clairement avec les résultats que nous avons extraits de l'étude de Bickley et al. (1986).

L'étude de den Os (1990), portant sur l'évolution des propriétés temporelles dans la parole d'un enfant hollandais suivi longitudinalement de 1 à 3 ans, a montré une évolution du nombre moyen de syllabes par mot. Lorsque l'enfant passe de l'étape d'un mot à l'étape de trois mots, l'auteur n'observe pas d'augmentation significative dans le nombre de syllabes par énoncé. La majorité des énoncés à un mot est constituée de 2 syllabes. Ce n'est que vers l'âge de 2 ans et 6 mois qu'on peut observer une augmentation du nombre de syllabes par mot. Den Os s'est alors plus particulièrement penché sur la durée moyenne des syllabes chez cet enfant, et lorsqu'il la convertit en débit d'articulation, il trouve que cet enfant néerlandophone possède un débit d'environ 3 syllabes/seconde.

Boysson-Bardies et al. (1981) ont étudié l'organisation temporelle du babillage *tardif* chez un enfant français, suivi toutes les semaines pendant 2 mois entre 1;6 ans et 1;7 ans. Les auteurs ont analysé le débit d'articulation et son évolution dans la période dite « pivot » entre le babillage et le langage avec sens. Puis la relation entre la synchronisation segmentale et les contours intonatifs qui les ordonnent. Ce babillage tardif est considéré comme « des séquences syllabiques articulées, qui sont groupées sous un contour intonatif ou pitch, globalement similaire à celui des séquences linguistiques »<sup>28</sup>. A cette étape, un adulte ne peut pas établir une correspondance entre des séquences de babillage tardif et des mots ou expressions. Les données qu'ils ont obtenues leur permettent de soutenir la proposition que la structure temporelle du babillage tardif va pouvoir remplir le rôle d'indicateur de la transition du babillage vers un comportement plutôt de type articulatoire, dirigé nettement vers des cibles lexicales. Le principal contraste observé entre les deux périodes réside dans les durées syllabiques des séquences, et leur évolution. A 1;6 ans, les durées moyennes sont très stables et indépendantes de la longueur de la séquence. Par contraste, à 1;7 ans, on observe certains changements, les durées

---

<sup>28</sup> «*Late babbling can be characterized as articulated syllabic sequences that are grouped under an intonational or pitch contour roughly similar to that of sequences of language*» (1981:525).



syllabiques sont plus longues lorsque que l'enfant produit des séquences bi- et trisyllabiques. Cette augmentation n'affecte pas la durée des séquences de 4 à 5 syllabes. A cet âge, on note donc un allongement de la durée syllabique, que l'auteur traduit par un essai d'adaptation de ses événements articulatoires pour atteindre une cible correspondant aux réalisations du lexique adulte. Autrement dit, nous proposons que l'enfant suive déjà le patron rythmique du français consistant à allonger la durée de la syllabe finale.

D'après les données de durées syllabiques données par Boysson-Bardies et al. dans leur article, nous avons voulu établir précisément les débits d'articulation de cet enfant aux deux âges étudiés. Nous avons alors calculé qu'à 1;6 an, le sujet a une durée de syllabe moyenne de 218.20ms, ce qui porte à un débit de 4.6 syllabes/seconde. A 1;7 ans, la moyenne de la durée de ses syllabes est de 244.09 ms, soit 4.1Hz. Si on prend toutes les données obtenues à ces âges, la durée moyenne syllabique est de 203.29 ms, soit 4.3Hz de moyenne générale. Nous sommes ici dans du babillage tardif mêlé aux premiers mots.

Pour résumer, nous prendrons la proposition de Thelen (1981) comme point d'ancrage pour le babillage canonique : le babillage canonique s'inscrit dans la même périodicité que les autres activités rythmiques. A notre connaissance, le premier enregistrement des mouvements d'activité rythmique des bébés, autour de ce stade (5), a été réalisé sur leurs mains avec le système OPTOTRAK, pour tester l'hypothèse du rythme du langage amodal de Petitto (Petitto et al., 2001). Petitto rapporte que le rythme de bébés signant (entendant) possède très tôt (aussi tôt que le babillage canonique à 6 mois pour les enfants exposés à la parole) une composante de basse fréquence (autour de 1Hz) (voir Figure III.19) qui ne se rencontrerait pas dans l'activité manuelle des bébés exposés à la parole (Petitto et al., 2001 ; détails dans Petitto et al., 2004, pour les 3+3 bébés enregistrés à 6, 10 et 12 mois). Cette affirmation reste controversée, n'ayant jusqu'à présent pas été répliquée dans d'autres études (les résultats contraires de Meier et al. sont discutés dans Petitto et al., 2004).

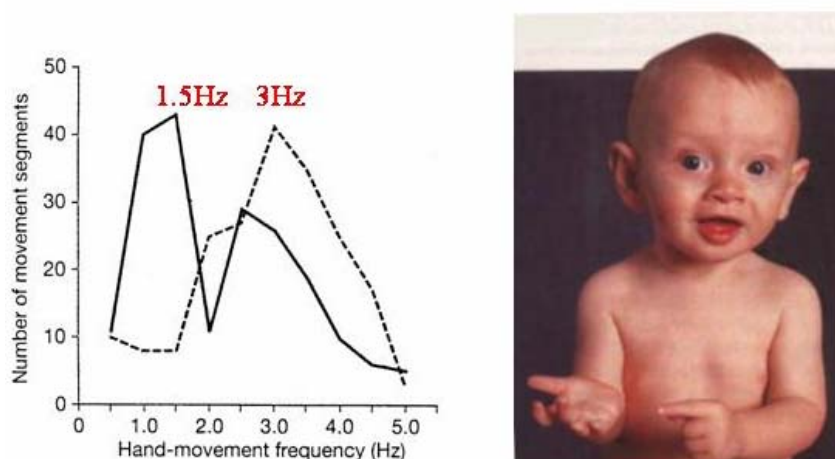


FIGURE III.19 : Distribution des fréquences du mouvement de la main dans l'activité rythmique de bébés entendants exposés à la langue des signes (ligne pleine) et pour les bébés exposés à la parole (ligne pointillée), tous âges confondus de 6 à 12 mois. Pour chaque groupe, 200 segments de mouvements ont été aléatoirement sélectionnés. Seuls les bébés exposés aux signes ont une distribution bimodale des fréquences de mouvements : le premier mode (pic de gauche) tombe autour de 1.5 Hz et le second mode (pic de droite) tombe autour de 2.5 Hz. Par contraste, les fréquences de mouvements de main des bébés exposés à la parole sont unimodales, autour de 3 Hz. Une comparaison des deux groupes a révélé que le second mode des fréquences de mouvement produites par les bébés exposés aux signes, vers 2.5 Hz, n'était pas significativement différent du 3Hz des bébés exposés à la parole (D'après Petitto et al., 2001).

De plus, cette basse fréquence apparaît très tôt (6 mois) et pourrait être interprétée comme un pointer répétitif, un geste très fréquent dans la langue des signes, spécialement dans les bases de données d'enfants sourds: les bébés signant pourraient être plutôt précoces dans les premiers signes (cf. p. ex. Bonvillian et al., 1985).

Dans le but d'ancrer nos prédictions sur un comportement non controversé, nous ne prendrons en compte que les distributions des activités rythmiques de la main obtenues dans la même gamme de fréquence pour les bébés signant ou non. Le pic de distribution à 3Hz, pour les 3 bébés non exposés aux signes, et le mode à 2.5Hz, pour les 3 bébés exposés aux signes, ne sont pas significativement différents. Par conséquent, en nous en tenant à la proposition d'une étape rythmique générale selon Thelen (1981), et en soutenant plus quantitativement (grâce aux données de main communes aux deux groupes d'enfants de Petitto et al.) la proposition d'un contrôle isofréquentiel pour les mouvements rythmiques de la main et de la mandibule dans le babillage canonique, autour de cette période de rythmicité généralisée, nous pouvons avancer que les deux vont fonctionner autour de 3Hz.

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

Rappelons que les soutiens empiriques pour cette proposition d'un cycle de babillage à environ 3Hz, comme le sont ces mains de six bébés, sont synthétisés par Koopmans-van Beinum et Van der Stelt (1993) avec les différents résultats que nous avons exposés précédemment.

“Initially [up to 4-5 months], large differences exists between the mean values of syllable[-sized units] duration, but as soon as the stage of canonical babbling is reached [at about 6-7 months], the mean durations are more or less equal for the four [Dutch] children, with a mean value of 424 ms, or an articulation rate of 2.36 syll/s. However, the relative slowness of the production system at this age may be illustrated by the fact that this value is still about twice as long as the mean syllable duration given by De Boysson-Bardies et al. (1981), for a French child at the age of one and a half. Den Os (1990) studying temporal properties in the speech of one Dutch child between one and three years of age, reports a mean articulation rate of about 3 syll/s. Koopmans-van Beinum (1992) obtained articulation rate values of 6.44 syll/s for read aloud speech and 6.71 syll/s for conversational speech for a Dutch professional adult speaker.” (Koopmans-van Beinum and Van der Stelt, 1993: 76).

Pris ensemble, ces résultats montrent une tendance maturationnelle vers des cycles de « babillage/syllabe » plus rapides.

En ce qui concerne nos données, des mesures répétitives de trains de syllabes (CVCVCV... ; incluant les CVCVs syllabiques ou S1/S2, voir ci-après) ont donné des durées de cycles syllabiques variables autour d'une moyenne 355ms, soit 2.8Hz.

Lorsque l'on compare ces fréquences (*articulation rate*) disponibles pour l'enfant, nous remarquons que le rythme de la parole de celui-ci est quasiment deux fois plus lent que celui de l'adulte. En effet, nous trouvons des fréquences de 2.7Hz, 2.87Hz, 2.36Hz et 3Hz pour l'enfant (jusqu'à 3.46Hz de moyenne pour Lise dans notre corpus), tandis que nous trouvons, dans la littérature des fréquences de 5 à 6.5Hz pour l'adulte, soit un rapport du double. Pour l'évolution plus tardive de cette fréquence chez l'enfant, nous pouvons citer encore Kowal et al. (1975), qui ont étudié le tempo de parole de 4 à 12 ans chez l'enfant. Dans ces données on observe qu'à l'âge de 4 ans, le débit d'articulation est à 2 syllabes/seconde, tandis qu'à 12 ans, il augmente à 4 syllabes/seconde. Cette étude a le mérite de nous montrer que cette évolution est assez lente et nécessite la maturation de nombreux mécanismes afin d'atteindre la cible adulte normale.

Même si chez l'adulte, le rythme de production syllabique se situe autour de 6Hz, soit deux fois plus rapide que chez l'enfant, une étude menée par Ackermann et Hertrich (2003) nous semble particulièrement appropriée pour mettre en évidence le pas de base dans le contrôle du rythme syllabique. Ils démontrent en effet que les activités rythmiques de patients souffrant de pathologie cérébelleuse retombent à

3Hz, cette même fréquence que nous avons trouvée chez l'enfant. Les sujets testés dans cette étude souffrent d'un désordre cérébelleux, qui peut avoir pour conséquence de compromettre le contrôle des muscles du conduit vocal : il s'agit de dysarthrie ataxique. Cette pathologie se caractérise principalement par un débit de parole ralenti, des productions de consonnes et voyelles distordues et une modulation prosodique des énoncés altérée ; mais on observe également une tendance vers des durées de syllabes isochrones dans les énoncés. Elle touche le cervelet, zone connue comme étant une « horloge interne », responsable du contrôle du timing entre les différents domaines moteurs.

Ackermann et Hertrich (2003) ont voulu en outre tester chez ces patients le timing de la syllabe dans une série d'études acoustiques et kinématiques, en utilisant la technique d'imagerie par résonance magnétique fonctionnelle. Avec des tâches diadochokinétiques orales, c'est-à-dire une répétition de mouvements moteurs alternés, ici la répétition de syllabes identiques offrant une alternance d'ouvertures et de fermetures du conduit vocal. Ces tâches ont l'avantage de fournir une estimation brute du débit de parole, et de faire apparaître la limite supérieure pour le tempo de la parole. Leurs résultats tendent à démontrer que les patients cérébelleux tombent à un débit de 3Hz lors de cette tâche de répétition de syllabes. De la même façon, dans la maladie de Friedreich (désordre ataxique plus sévère), lors des mêmes tâches, les patients atteignent un plateau à environ 3Hz. Ces découvertes revêtent une importance fondamentale par le fait qu'elles soutiennent l'idée que le cervelet contribue à des étapes pré-articulatoires dans le contrôle moteur de la parole et qu'il semble organiser la structure rythmique syllabique de la parole interne. Pour notre dessein, cette découverte tend à prouver le côté robuste de ce mode de fréquence dans les activités rythmiques, un mode que l'on retrouve aussi bien dans le bras que dans la mâchoire chez l'enfant, et dans la production d'adultes avec cette pathologie, nous révélant peut-être quel est le rythme de base du langage.

A travers ces diverses études, nous avons pu rendre compte de la fréquence propre du contrôle cyclique d'oscillations mandibulaires trouvé chez l'enfant et qui varie en gros de 2.5 à 3.5 Hz, comme chez les enfants de notre corpus. Ces résultats nous amènent à nous poser d'autres questions sur le développement du rythme chez l'enfant, et notamment la question de l'adaptation de l'enfant au rythme propre de sa langue maternelle. Nous avons évoqué ci-dessus les patrons de structuration

rythmique de différentes langues, et il serait alors tout à fait justifié de se demander maintenant comment évolue la mise en place du rythme propre à la langue française dans le développement de l'enfant. Son évolution suit-elle celle de la mise en place du contrôle du rythme mandibulaire? Quel est l'âge critique pour son émergence ?

### **III.4.3.3. Adaptation de l'enfant au rythme propre à la langue maternelle**

Dans la lignée des études portant sur le rythme, nous pouvons citer les travaux de Konopczynski (1990, 1998). Elle définit le rythme comme « principalement déterminé par la durée relative des syllabes, l'organisation des pauses et des accents<sup>29</sup> ». Chez l'adulte, nous avons vu que le français est principalement considéré comme une langue dite syllabique (*syllable-timed*), avec des syllabes principalement ouvertes, de durée égale. Konopczynski considère cette langue comme plutôt « *trailer-timed* », dans la mesure où la proéminence —caractérisée par une durée plus longue— se porte sur la fin de chaque groupe.

Son étude a porté sur 12 bébés français, suivis hebdomadairement de l'âge de 9 mois à 12 mois, puis suivis une fois par mois dans les 12 mois suivants, c'est-à-dire jusqu'à 24 mois. Ses résultats montrent lorsque que l'enfant entre dans ce qu'elle nomme le « proto-langage », soit la période du babillage canonique, l'organisation syllabique est constituée majoritairement de syllabes canoniques d'une durée courte isochrones. Plus tard la durée de la syllabe va de plus en plus dépendre de la place qu'elle occupe au sein de l'énoncé : les syllabes en position non-finale (NFS) vont progressivement être de plus en plus courtes ; Les syllabes en position finale (FS) semblent avoir une durée assez instable pendant une longue période, avant de devenir deux fois plus longues que les NFS, tendant alors à se conformer au patron français adulte (cf. Figure III.22).

---

<sup>29</sup> "... [the basic rhythm of a language is] mainly determined by the relative duration of syllables, the organization of pauses and accentuation." (1998:178)

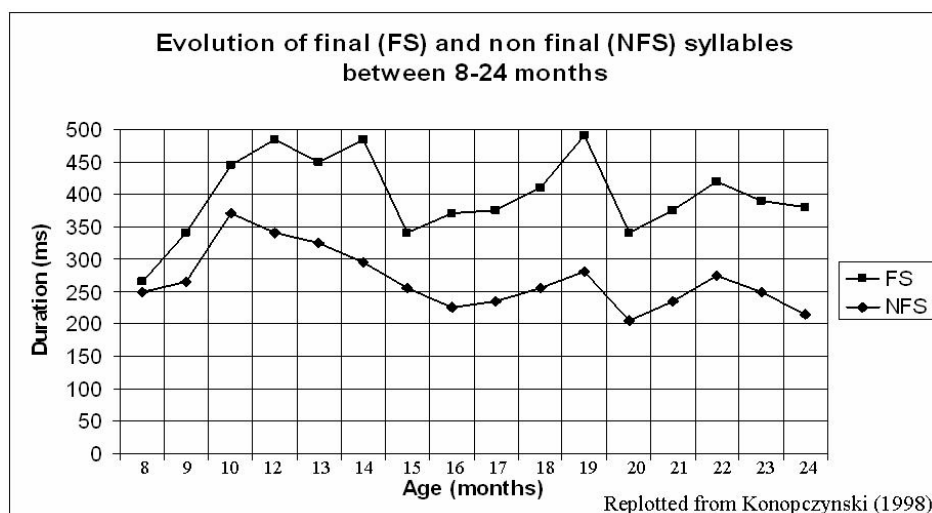


FIGURE III.22 : Evolution des syllabes finales (FS) et des syllabes non-finales (NFS) chez un sujet entre 8 et 24 mois (données replottées d'après Konopczynski 1998) (N.B. : le mois 11 est manquant dans ses données)

Konopczynski explique la grande variabilité de la durée de la syllabe finale par le fait que l'enfant doit s'habituer à produire ce patron et que cela implique des essais et des erreurs. L'allongement de la syllabe en position finale semble être l'indicateur majeur de l'émergence de la structuration temporelle chez l'enfant. Ainsi, l'acquisition de la structure syllabique typique du rythme de la langue française semble être réalisée au milieu de la seconde année de vie. Notons que l'enfant français a commencé à maîtriser ses structures rythmiques, dès 12-13 mois, au moment même où les premiers mots vont apparaître dans ses productions.

#### **III.4.3.4. Le cas de la naissance de la métrique française de Célia : pointers, syllabes et mots**

Nous avons vu que Célia était une enfant qui se situait dans le profil moyen du groupe pour le pointer (Figure III.3). Mais, contrairement aux autres enfants, on n'observe pas chez ce sujet une décroissance en durée qui se produise de façon monotone. Elle présente un soudain changement de détente entre 11 (900ms) et 13 mois (940ms), à savoir à 12 mois (540ms), au moment où l'apparition de ses premiers mots devient corroborée (par au moins 1 des 6 critères de Vihman et McCune, 1994). Quel est alors le profil longitudinal de ses durées de syllabes ? Et plus spécifiquement, qu'arrive-t-il autour de ses premiers mots dans son régime métrique en tant qu'enfant exposée au français ?

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

Sur la figure III. 23, les premières (S1) et secondes (S2) syllabes, débutent toutes deux en isochronie autour de 400 ms (2.5Hz) à 6-8 mois; tombent à 200 ms (S1) et 300 ms (S2) à 10-11 mois, soit à la fréquence moyenne de 3Hz. Ceci juste avant l'apparition claire des premiers mots à 12 mois, où la seconde syllabe saute clairement à une durée de 600 ms et oscille entre 500 et 600-700 ms pour les 13-15 mois restant du suivi (qui finit à 2.5 Hz à nouveau). Ce profil est quelque peu lissé quand le ratio S2/S1 à l'intérieur de chaque item mesuré est tracé (Figure III.24), mettant à nouveau en évidence le saut des 12 mois, quand la seconde syllabe devient 2,5 fois plus longue que la première, avec par la suite une décroissance de ce ratio vers 13-15 mois.

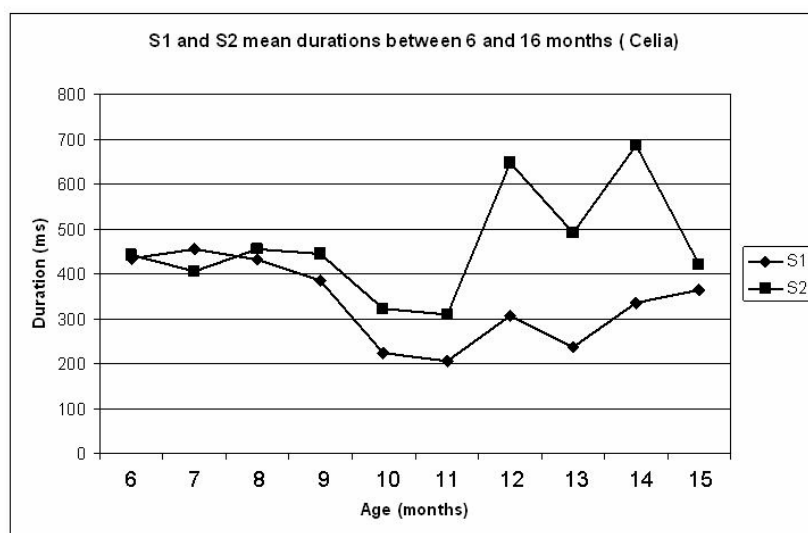


FIGURE III.23 : Evolution de la durée moyenne des cycles des premières (S1 : diamants) et secondes (S2 : carrés) syllabes dans les performances de Célia entre 6 à 16 mois. Notons l'augmentation dans S2 autour de l'émergence de ses premiers mots à 12 mois.

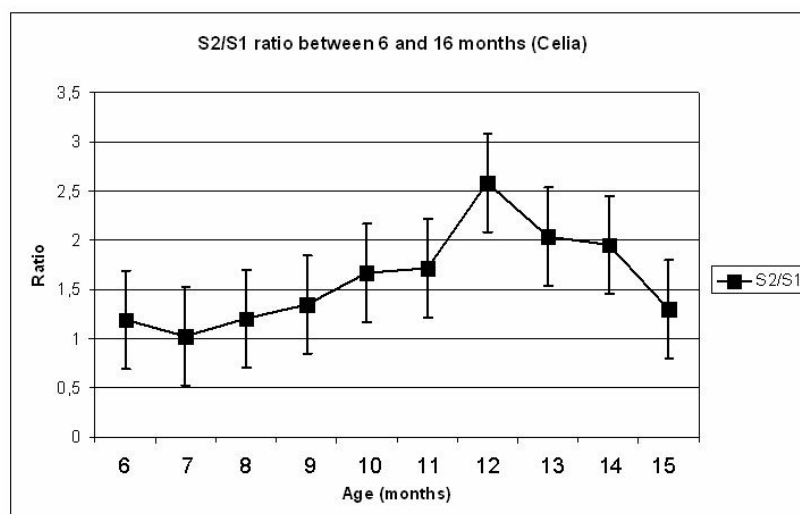


FIGURE III.24: Evolution du ratio S2/S1 (seconde syllabe sur la première) pour Célia. Ce ratio met en évidence le changement métrique à la période des premiers mots vers un an (cf. Figure III.23).

Ce comportement est aussi illustré par le bébé français déjà cité (Konopczynski, 1998). Ce sujet présente globalement les mêmes changements dans le patron des syllabes. Les premières et secondes syllabes commencent en isochronie à 8 mois (à environ 4 Hz) et augmentant en durée autour des premiers mots, la seconde syllabe atteignant un maximum à 12 mois, suivi par une chute à 15 mois. Elles oscillent ensuite en parallèle, gardant leurs valeurs séparées, leur durée moyenne globale étant d'environ 300 ms (3.3 Hz) à 2 ans. Prises ensemble, ces observations montrent que le patron iambique français est acquis, dans le babillage, autour des premiers mots.

Qu'en est-il alors des premiers mots et du pointer dans nos données ? Une simple addition devrait montrer, au moins pour Célia (voir Fig.5), qu'une séquence CVCV réalisée à 12 mois avec S1+S2 autour de 900 ms ne pourrait pas être contenue dans son pied moyen de 540 ms... Mais cela ne se passe pas ainsi. Les mots CVCV de Célia correspondant à ces gestes de pointer durent entre 450-650 ms (avec une moyenne de 527ms). En conséquence, elle n'a aucun problème de cadre-pied/contenu-syllabe, son pointer de mot maximum à 12 mois étant de 680 ms. Elle va ensuite produire à nouveau des exemplaires plus longs de pointers. Le fait est que lorsque les syllabes de babillage décroissent, elles augmentent ensuite avec l'émergence de ses premiers mots –mettant en évidence l'acquisition d'un patron hyper-français S1S2 (où  $S2/S1=2.5$ , une tendance qui est maintenue dans les mots de Célia à 12-13 mois). Ces mots ne sont pas simplement produits en coupant dans le flux de babillage les première et seconde syllabes... Ainsi, il y a une production autonome des mots parallèle à cette restructuration dans le babillage. Ce résultat ne signifie pas pour autant que de tels changements dans la métrique (prosodique) soient indépendants.

### **III.4.4. Le mode de distribution des durées de cycles de la syllabe : tendance globale**

Dans le but de tester plus précisément le ratio entre pointage et syllabe chez les enfants, nous avons réalisé nos propres mesures de cycles du babillage sur les enfants de notre corpus. Pour cela, nous avons prélevé des échantillons de productions de babillage parmi les 6 enfants du corpus pour obtenir les cycles,



## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

uniquement durant les sessions filmées où les enfants produisaient des gestes de pointers.

Nous ainsi pu collecter 4153 productions de cycles syllabiques parmi les 6 enfants, entre 7;20 et 17;18, qui ont été mesurées avec l'éditeur PRAAT et un script associé. La figure III.25 présente la distribution cumulée brute, obtenue avec la durée des syllabes en x et le rang de ces durées en y.

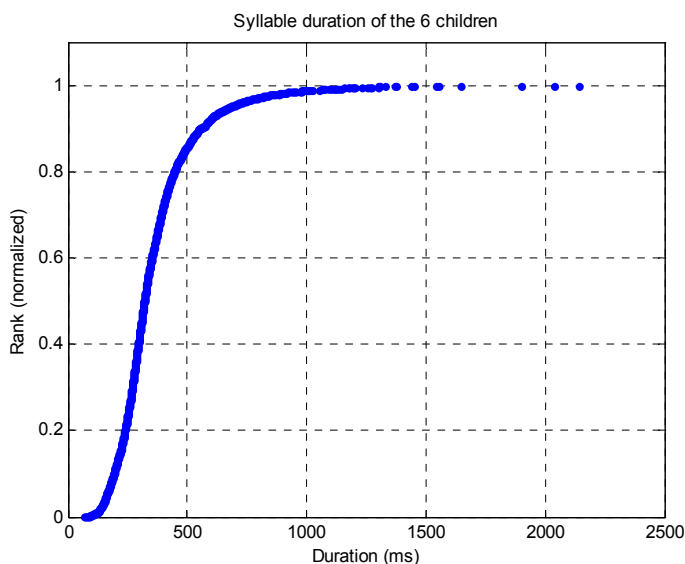


FIGURE III.25: Distribution des fréquences cumulées (durée-rang). Les données brutes correspondent à 4153 syllabes produites par les 6 enfants français entre 6 et 18 mois.

Le cycle moyen de la syllabe est de 355.19ms ( $\pm 176.60$ ms), soit un rythme de babillage à 2.8Hz. Les présentations en rang normalisé de chaque enfant (dès lors que le nombre d'observations collectées pour chacun était différent, voir TABLE III.7) versus la distribution brute générale montrent que 3 des 6 enfants (Figure III.26) sont plutôt proches du comportement général observé (Tom: 360.94ms $\pm$ 147.50 ; Nicolas: 337.72ms $\pm$ 174.40 ; et Anatole: 356.40ms $\pm$ 113.99). Tandis que les trois autres (Figure III.27) ont des scores qui tendent à être plus rapides (Lise: 289.35ms $\pm$ 87.89; et Jules: 314.05ms $\pm$ 105.38) ou plus lents (Celia: 404.59ms $\pm$ 231.85) que le comportement général observé. Enfin, ces bébés français babillent entre 2.5Hz et 3.5Hz, des valeurs proches des nombreuses données citées précédemment pour le babillage, et du rythme de la main mesuré pour les 6 enfants, exposés aux signes ou non, dans Petitto et al. (2004).

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

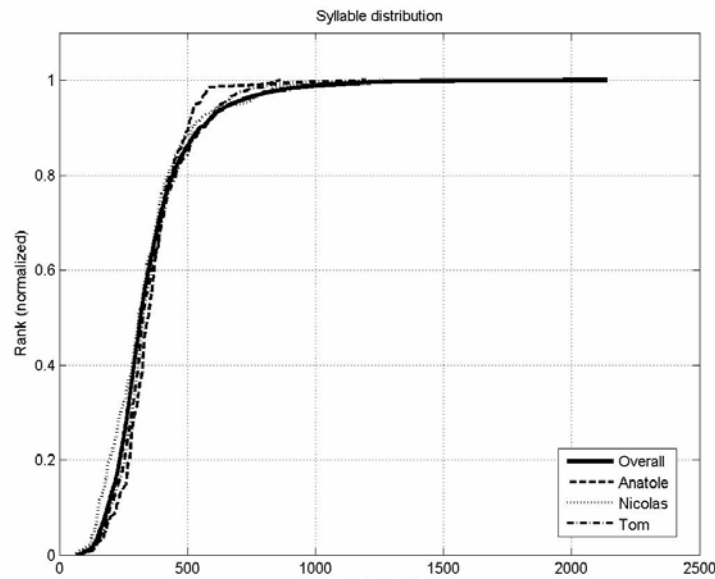


FIGURE III.26: Distributions normalisées des cycles de syllabes pour Anatole, Nicolas et Tom. Leurs données sont plutôt proches de la distribution générale observée pour les 6 enfants (ligne pleine).

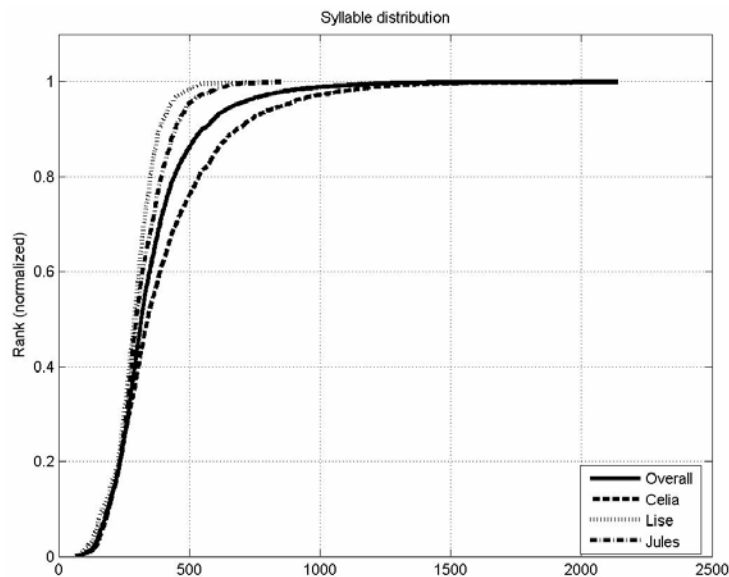


FIGURE III.27: Distributions normalisées des cycles de syllabes pour Célia, Lise et Jules. Ils divergent dans leurs plus hautes valeurs de la distribution générale observée pour les 6 enfants (ligne pleine).

Nous pouvons également noter que les bébés les plus rapides (Tom) ou plus lents (Lise et Nicolas) pour les données de pointage ne correspondent pas aux babilleurs les plus rapides (Lise et Jules) ou plus lents (Célia). Ce qui nous permet déjà d'introduire une idée qui sera développée ultérieurement, selon laquelle il existe une relation de phase assez lâche entre pointage et babillage, suffisante en principe pour permettre une compatibilité entre deux systèmes de porteuses très différents.

Mais tout d'abord, intéressons-nous aux mesures de durées de syllabes réalisées sur les 6 enfants du corpus dans une étude de cas individuels.

### III.4.5. Le mode de distribution des durées de cycles de la syllabe : analyses individuelles

Pour les 6 enfants du corpus, un échantillon de productions de cycles de syllabes a été prélevé, en relevant ces passages uniquement lors des sessions où des gestes de pointers étaient apparus, ceci afin d'obtenir la fréquence de la syllabe dans la période où l'enfant utilise le geste de pointer.

#### III.4.5.1. Anatole

Nous avons relevé pour Anatole un échantillon de 210 syllabes (voire nombre de syllabes par mois, Table III.1) s'étalant dans une période allant de 7 à 15 mois. Anatole fait partie des enfants qui ont un comportement individuel se rapprochant du comportement général observé parmi les 6 enfants.

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
7	48	371.6
11	27	305.91
12	35	372.86
14	69	338.09
15	31	398.96

TABLE III.1 : Nombre de syllabes relevées chez Anatole. Les durées moyennes sont indiquées par mois.

Les 210 syllabes d'Anatole s'étendent sur une plage allant de 103.91ms à 862.40ms, avec une moyenne générale se situant à 356.40ms, soit une fréquence de 2.81Hz et un écart-type de  $\pm 113.99$ ms. Lorsque l'on observe l'évolution de ces durées à travers le temps, elles sont plutôt stables (figure III.28).

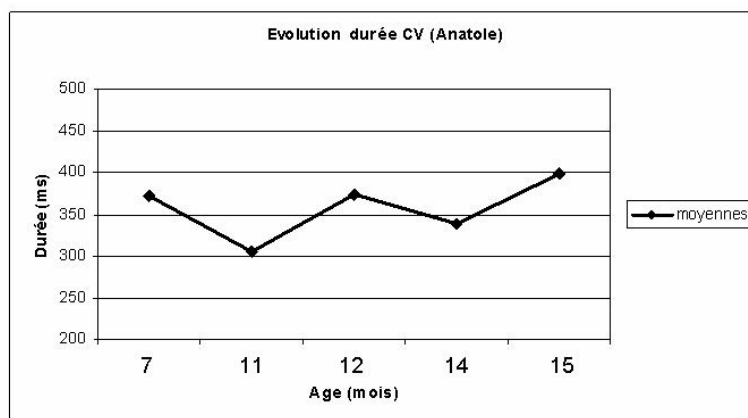


FIGURE III.28 : Evolution de la durée moyenne des cycles syllabiques par mois chez Anatole.

### III.4.5.2. Célia

Célia se situe dans un profil qui diffère du comportement général observé, spécialement pour les valeurs hautes relevées, Célia semble présenter un patron plus lent de durées pour ces valeurs. Nous avons pu bénéficier pour ce sujet des mesures de cycles syllabiques effectuées dans le cadre d'une étude menée par Romain Trollat et Anne Vilain d'une part, et Claire Lalevée d'autre part, ce qui explique le nombre assez important de mesures relevées, soit 1539 syllabes. Il convient tout de même de noter qu'un nombre important de mesures avait été faites pour Célia parmi toutes les sessions filmées par ces chercheurs, et que nous avons restreint délibérément ce nombre de données aux sessions qui nous concernent particulièrement, celles où le geste de pointer apparaît. Ces 1539 syllabes s'étalent donc pour notre étude entre 9 et 15 mois (Table III.2).

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
9	48	441.54
10	335	314.70
11	232	305.36
12	593	473.20
13	91	440.53
15	240	435.43

TABLE III.2 : Nombre de syllabes relevées chez Célia. Les durées moyennes sont indiquées par mois.

Les valeurs de durées de production de cycles de syllabes s'étalent de 81.12ms à 2138.55ms, avec une moyenne de 404.59ms, soit une fréquence de

2.47Hz et un écart-type de  $\pm 231.85\text{ms}$ . L'évolution de la durée des syllabes (Figure III.29) de Célia à travers le temps fait apparaître une chute des valeurs entre 9 et 11 mois, ce qui correspond à un babillage de plus en plus rapide, atteignant les 3Hz à 11 mois, juste avant que l'apparition du patron iambique français avec les premiers mots (cf. supra), nous donne une augmentation des valeurs moyennes dès 12 mois, valeurs qui se stabilise autour de 13-15 mois.

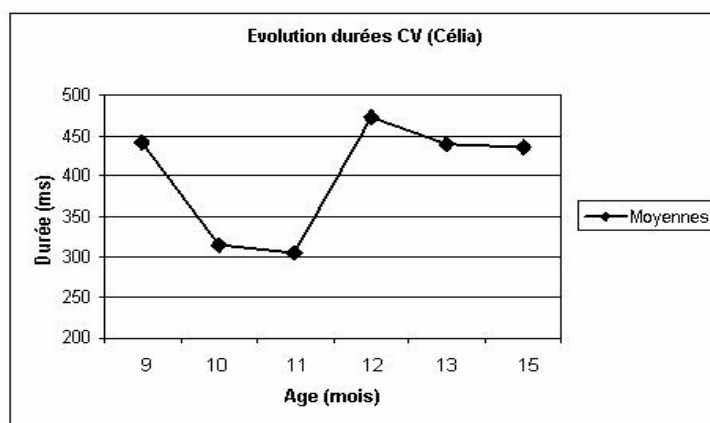


FIGURE III.29 : Evolution de la durée moyenne des cycles syllabiques par mois chez Célia.

### III.4.5.3. Jules

Nous avons pu collecter pour ce sujet un total de 1087 syllabes, de l'âge de 10 mois jusqu'à 16 mois (Table III.3).

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
10	116	372.19
11	133	315.14
12	130	368.92
13	128	313.48
14	302	275.59
15	81	322.47
16	197	298.75

TABLE III.3 : Nombre de syllabes relevées chez Jules. Les durées moyennes sont indiquées par mois.

Les 1087 syllabes réalisées par Jules ont des valeurs qui s'étendent de 80.57ms à 722.29ms, avec une durée moyenne générale de 314.05ms, soit une fréquence de 3.18Hz, et un écart-type de  $\pm 105.38\text{ms}$ . La figure III.30 nous montre ci-dessous un patron d'évolution des durées de syllabes avec une petite tendance globale à la décroissance, avec environ 70ms de moins entre 10 et 16 mois.

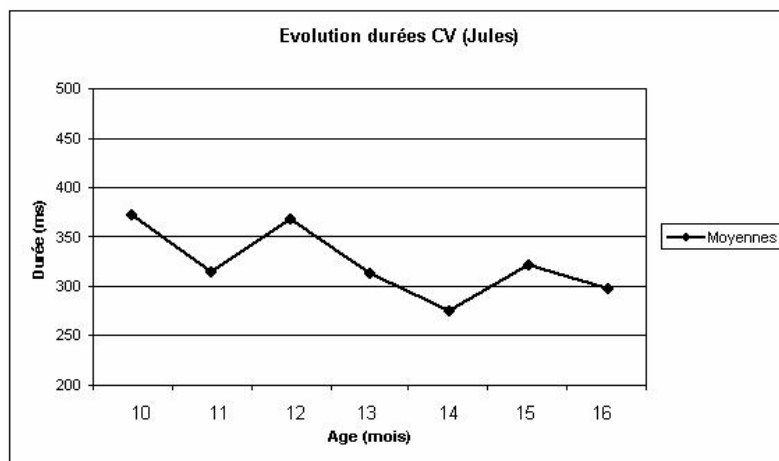


FIGURE III.30 : Evolution de la durée moyenne des cycles syllabiques par mois chez Jules.

#### III.4.5.4. Nicolas

Nous avons pu relever parmi les sessions où Nicolas pointe, qui s'étaient entre 10 et 16 mois, un total de 235 syllabes dans ses productions babillées (Table III.4).

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
10	44	395.34
12	8	651.36
13	94	285.65
14	18	322.84
15	15	333.19
16	56	341.04

TABLE III.4 : Nombre de syllabes relevées chez Nicolas. Les durées moyennes sont indiquées par mois.

Les 235 syllabes réalisées par Nicolas s'étendent de 66.94ms à 1270.40ms, avec une moyenne des durées de 337.72ms, soit une fréquence de 2.96Hz, avec un écart-type de  $\pm 174.40$ ms. L'évolution des durées de ces syllabes à travers le temps (Figure III.31) nous montre, tout comme pour Jules, un changement net du rythme du babillage de 10 à 12 mois. Entre 13 et 16 mois, on observe une légère remontée des valeurs, mais qui restent toutefois nettement inférieures à celle mesurée à 10 mois.

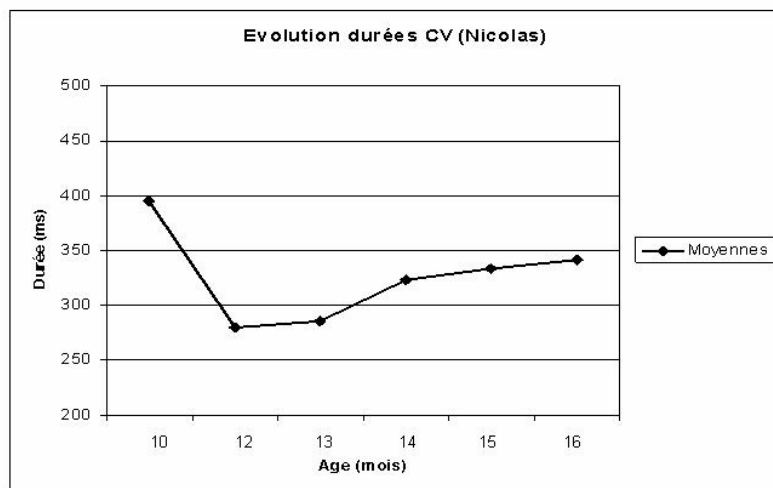


FIGURE III.31: Evolution de la durée moyenne des cycles syllabiques par mois chez Nicolas.

### III.4.5.5. Lise

Comme Jules, Lise est une enfant qui diffère du comportement général, avec un babillage plus rapide que la moyenne. Nous avons collecté un échantillon de 441 syllabes parmi les sessions où les gestes de pointers surviennent, soit entre 8 et 15 mois (Table III.5).

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
8	41	302.95
9	35	310.86
10	156	311.98
12	44	298.88
13	37	290.04
14	72	262.72
15	56	229.22

TABLE III.5 : Nombre de syllabes relevées chez Lise. Les durées moyennes sont indiquées par mois.

Les valeurs de ces 441 syllabes s'étendent pour Lise de 80.57ms à 722.29ms, avec une moyenne des durées de 289.35ms, soit une fréquence de 3.46Hz, et un écart-type de  $\pm 87.89$ ms. Sur la figure III.32, nous pouvons observer que les données décroissent doucement à travers le temps, partant d'environ 300ms à 8 mois pour arriver en dessous de 250ms à 15mois.

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

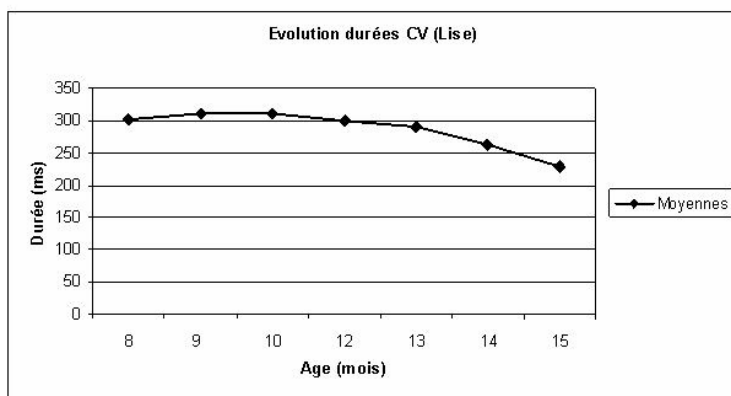


FIGURE III.32 : Evolution de la durée moyenne des cycles syllabiques par mois chez Lise.

### III.4.5.6.Tom

Pour Tom, nous avons recueilli un total de 641 syllabes parmi les sessions de pointer sur une période allant de 10 à 17 mois (tableau III.6).

Age (mois)	Nombre de syllabes	Durée moyenne des syllabes (ms)
10	46	471.81
11	81	331.35
13	100	343.82
14	152	358.64
15	16	331.97
16	106	350.10
17	140	367.88

TABLE III.6 : Nombre de syllabes relevées chez Tom. Les durées moyennes sont indiquées par mois.

Ces 641 syllabes vont de 24.04ms à 604.43ms, avec une moyenne de 360.94ms, soit une fréquence de 2.77Hz, avec un écart-type de  $\pm 147.50$ ms. La figure III.33 ci-dessous donne l'évolution des durées des syllabes pour Tom. Nous observons une chute brutale des valeurs entre 10 et 11 mois d'environ 140ms, puis une phase stable par la suite jusqu'à 17mois.



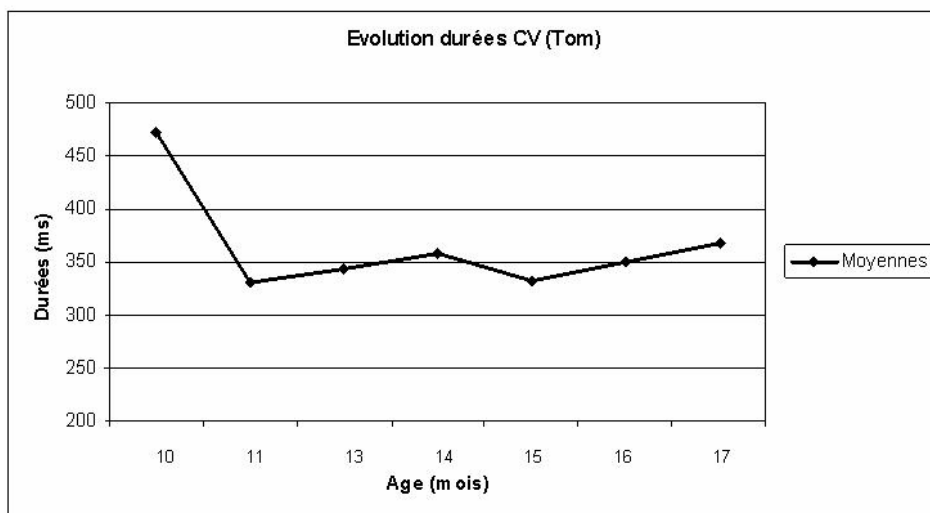


FIGURE III.33 : Evolution de la durée moyenne des cycles syllabiques par mois chez Tom.

### III.4.5.7. Conclusion

A la lumière de ces résultats sur les durées de babillage des enfants de notre corpus, nous pouvons observer que les valeurs sont conformes en moyenne à celles trouvées dans la littérature, se situant à des fréquences comprises entre 2.47Hz et 3.46Hz. En ce qui concerne l'évolution des valeurs de durées de syllabes à travers le temps, nous ne retrouvons pas de patron régulier de décroissance parmi tous les enfants. Pour 3 d'entre eux (Célia, Nicolas et Tom) on a au début des séances de pointer, une accélération du babillage, qui correspond peut-être à une période avant les premiers mots (c'est clairement le cas de Célia). Pour les autres il n'y a pas de tendance nette à une décroissance.

### III.4.6. Le ratio pointer/babillage pour les enfants du corpus

Nous avons précédemment mis en valeur pour chaque individu du corpus les valeurs de durées de pointers et de durées de syllabes, ainsi que leur évolution dans le temps. Afin d'affiner notre prédiction sur le ratio harmonique entre pointer et syllabe, il nous semble à présent opportun d'effectuer une analyse par individu pour mettre en évidence le ratio entre ces deux modalités, afin de montrer si oui ou non cette tendance générale se retrouve pour chaque enfant.

#### III.4.6.1. Anatole

Les résultats d'Anatole présentés précédemment ont fait apparaître une durée moyenne de pointers de 763.33ms, ainsi qu'une durée moyenne de syllabes de

356.40ms. Si nous calculons le ratio Pointer/Syllabe pour ce sujet, nous obtenons une valeur de 2.14, soit un résultat très proche du ratio annoncé. La figure III.34 suivante nous permet de comparer les courbes de distributions de syllabes et de pointers. Nous avons ajouté sur cette figure deux autres courbes, théoriques : la première (Syll\*2 sur la figure) est obtenue en doublant les valeurs des syllabes de la distribution des syllabes observées ; la seconde (Syll\*3 sur la figure) est obtenue en triplant les valeurs de ces mêmes syllabes observées. Ces deux courbes permettent de visualiser avec précision si l'on obtient une approximation plutôt proche de la prédiction 2 syllabes pour 1 geste par rapport à la distribution des pointers observés.

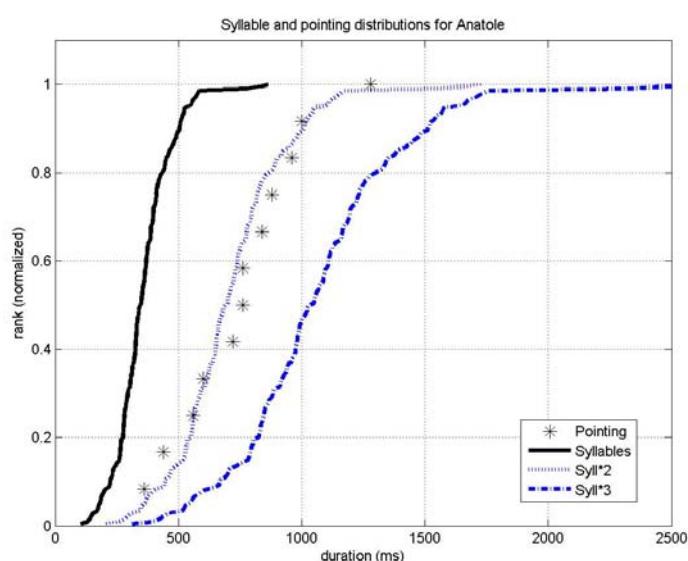


FIGURE III.34: Distributions cumulées brutes d'Anatole entre 7 et 15 mois pour ses 210 syllabes (ligne pleine) et ses 12 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées.

Dans le cas d'Anatole, nous venons de dire que le ratio trouvé est proche de celui prédit, et la figure III.34 nous permet ainsi de démontrer que la courbe de pointers obtenue permet pleinement d'intégrer 2 syllabes dans un geste, puisqu'elle suit de façon étonnante, malgré le peu de pointers produit par cet enfant, la courbe théorique de Syll\*2 (mais ne permet pas d'intégrer des productions trisyllabiques).

### III.4.6.2. Célia

Nous avons relevé pour Célia une durée moyenne de pointers de 826.45ms et une durée moyenne de syllabes de 404.59ms. Le ratio Pointer/Syllabe vaut 2.04, soit une valeur encore plus proche du ratio harmonique prédit.

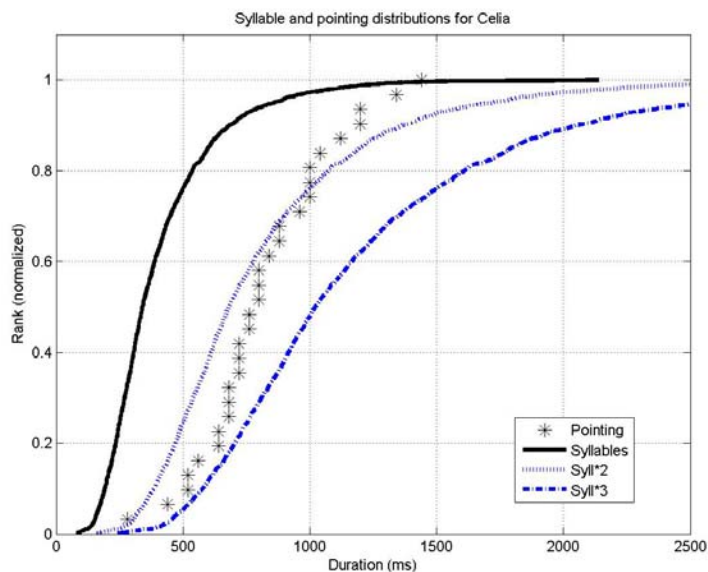


FIGURE III.35 : Distributions cumulées brutes de Célia entre 9 et 15 mois pour ses 1539 syllabes (ligne pleine) et ses 31 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées.

Sur la figure III.35, nous pouvons observer une fois encore que la courbe de distribution des pointers présente dans l'ensemble des valeurs supérieures à la courbe théorique de Syll\*2, démontrant ainsi que dans son geste de pointer, Célia est capable d'intégrer 2 syllabes (sauf pour ses valeurs les plus longues de syllabes, à l'extrême supérieur de la courbe ; noter encore que ses pointers peuvent contenir presque 3 de ses syllabes les plus courtes).

### III.4.6.3. Jules

La durée moyenne de pointers de Jules est de 768.53ms et sa durée moyenne de syllabes est de 314.05ms, ce qui nous donne un ratio Pointer/Syllabe de 2.45, une valeur encore une fois assez proche du ratio prévu.

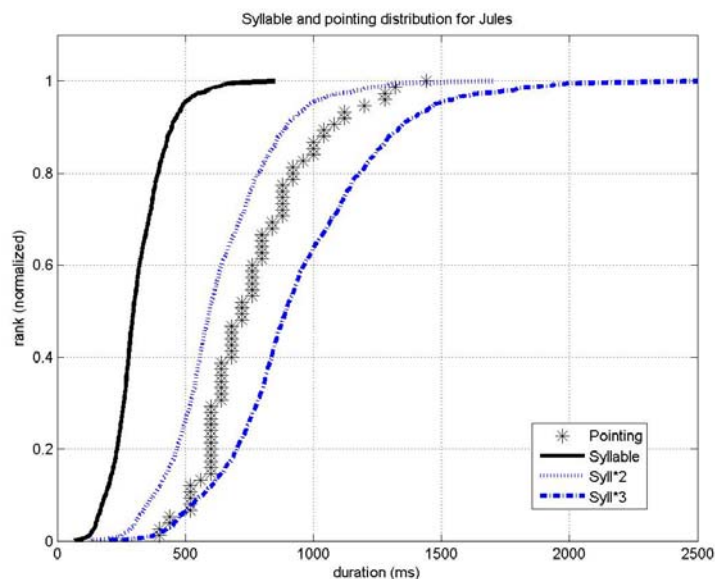


FIGURE III.36 : Distributions cumulées brutes de Jules entre 10 et 16 mois pour ses 1087 syllabes (ligne pleine) et ses 75 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées.

Sur la figure III.36, nous pouvons remarquer que la courbe de pointers de Jules permet de « couvrir » les valeurs de la courbe théorique obtenue en doublant les valeurs de la distribution des syllabes, démontrant qu'un geste de pointer de Jules peut intégrer aisément 2 syllabes (mais pas 3). Ce qui confirme notre prédiction.

#### III.4.6.4. Nicolas

Pour Nicolas, nous avons obtenu une durée moyenne de pointers de 925.71ms, ainsi qu'une durée moyenne de syllabes de 337.72ms. Le calcul du ratio Pointer/Syllabe nous donne pour ce sujet un résultat de 2.74, une valeur qui approche d'un rapport harmonique de 3.

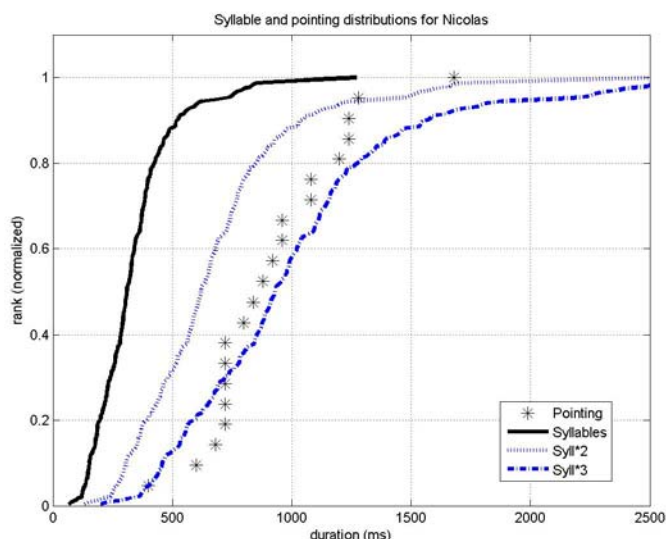


FIGURE III.37: Distributions cumulées brutes de Nicolas entre 10 et 16 mois pour ses 235 syllabes (ligne pleine) et ses 21 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées.

La figure III.37 permet d'observer que la courbe de pointers obtenue pour Nicolas permet d'intégrer, par rapport aux courbes théoriques Syll\*2 et Syll\*3, jusqu'à 3 syllabes dans des valeurs rapides (jusqu'à un pointer d'à peu près 750ms).

### III.4.6.5. Lise

Lise a une durée moyenne de pointer de 927.03ms et une durée moyenne de syllabes de 289.65ms, soit un ratio Pointer/Syllabe de 3.20. Lise est, encore plus clairement que nous venons de le voir pour Nicolas, dans un rapport harmonique de 3, capable donc d'intégrer 3 syllabes dans un seul geste de pointer.

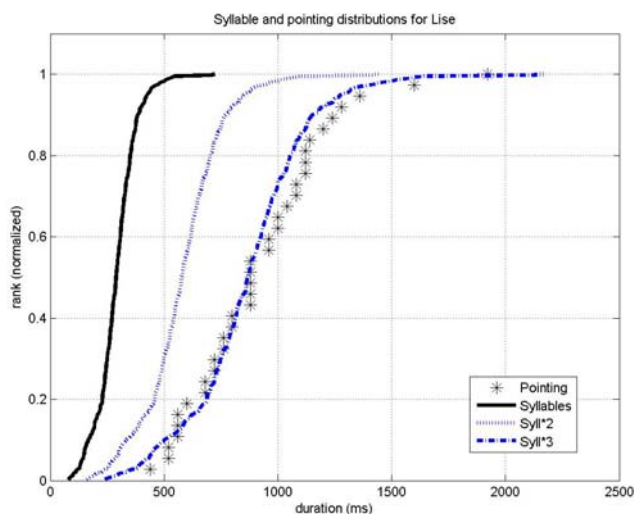


FIGURE III.38 : Distributions cumulées brutes de Lise entre 8 et 15 mois pour ses 441 syllabes (ligne pleine) et ses 37 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées.

### III.4.6.6. Tom

La durée moyenne des pointers pour Tom se situe à 681.01ms, tandis que la durée moyenne de ses syllabes se situe à 360.94ms. Le ratio Pointer/Syllabe obtenu pour ce sujet est donc de 1.89, une valeur encore une fois proche du ratio de 2.

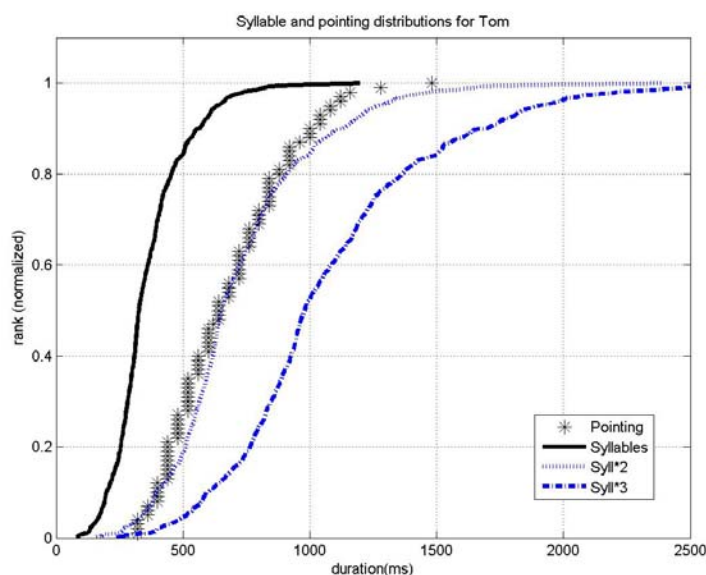


FIGURE III.39: Distributions cumulées brutes de Tom entre 10 et 17 mois pour ses 641 syllabes (ligne pleine) et ses 100 pointers (étoiles). 2 courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observée.

La figure III.39 montre très clairement que la courbe de durées de pointer de Tom suit la courbe théorique obtenue en doublant les valeurs de la courbe de distribution de syllabes (Syll\*2 ; sauf pour les valeurs supérieures à 1000ms à l'extrême de la courbe). Mais de façon générale, Tom est tout à fait capable d'intégrer 2 syllabes dans son geste de pointer.

### III.4.7. Le ratio pointer/babillage : notre « hypothèse étonnante »

Reprenons les résultats de pointers et de syllabes obtenus chez les 6 enfants du corpus, ainsi que les différents ratios obtenus pour chacun d'eux (dans le tableau III.8 ci-dessous):

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

Tab.1: Pointing / Syllable results for the 6 children of the corpus					
Subjects	SYLLABLE		POINTING		Pointing/syllable ratio
	Syllabic cycle events	Mean syllable duration (ms)	Pointing events	Mean pointing duration (ms)	
Anatole	210	356.40	12	763.33	2.14
Célia	1539	404.59	31	826.45	2.04
Jules	1087	314.05	75	768.53	2.45
Nicolas	235	337.72	21	925.71	2.74
Lise	441	289.35	37	927.03	3.20
Tom	641	360.94	100	681.01	1.89
	4153	355.19	276	775.87	2.18

TABLE III.7 : Durées moyennes des pointers et des syllabes obtenues pour 6 enfants français suivis par quinzaine entre 6 et 18 mois, avec leurs ratios Pointer/Syllabe. Le nombre d'événements de pointers spontanés avec vocalisation, observables pour tout le corpus est de 276 (de 12 à 100 par enfant). Le nombre de cycles de babillage mesurés par échantillonnage autour de ces événements de pointer est de 4153 au total (de 210 à 1539). Le rapport des moyennes des durées mesurées sur tous les événements (pointers 775.87 : syllabes 355.19), donne un ratio global Pointer/Syllabe de 2.18, soit tout proche de la valeur harmonique prédite de 2:1. On peut donc toujours enchâsser deux cycles de syllabes mandibulaires dans une détente de pointer.

Si l'on prend la valeur moyenne des durées des détentes (strokes) sur un total de 276 pointers avec vocalisation, observables pour les 6 sujets du corpus, nous obtenons 775.87ms (1.29Hz, avec des moyennes par sujet allant de 1.08Hz à 1.47Hz). La valeur moyenne des durées des 4153 syllabes mesurées pour ces mêmes 6 sujets, est de 355.19 ms, soit un rythme moyen de 2.82Hz (avec des sujets entre 2.47Hz et 3.46Hz). Le ratio Pointer/Syllabe général obtenu pour ces deux moyennes est de 2.18 (775.87ms : 355.19 ms), à savoir 2 syllabes pour 1 stroke de pointer. C'est le ratio harmonique prédit de 2 :1. Si on revient sur chaque enfant, nous observons bien que 4 d'entre eux sont plus proches de cette valeur 2:1 (Tom: 1.89; Célia: 2.04; Anatole: 2.14; Jules: 2.45) que les 2 restant, proches de 3:1 (Nicolas: 2.74; Lise: 3.20). Quoi qu'il en soit, et c'est valable pour tous ces enfants, tous sont capables d'intégrer au moins une-deux syllabes sous un stroke de pointer.

Ce résultat démontre que les deux systèmes peuvent fonctionner en harmonie. Afin d'obtenir la potentialité de ces deux systèmes, l'approche globale que nous avons utilisée se voulait neutre par rapport à la sémantique des mots supposés. De plus, nous ne considérons aucune relation pointer-syllabe intra-événement. Ce que nous mettons en évidence doit donc être considéré comme une tendance

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

corrélacionnelle, pas plus ou pas moins qu'un ratio général dans la métrique corps/tête. Les questions d'isométrie, d'allométrie et d'hétérochronie, bien connues des chercheurs essayant de relier les gènes du plan corporel (Hox) au comportement linguistique (et social) (voir la question de la croissance du corps/cerveau ou l'allométrie du conduit vocal et la controverse sur les universaux de Néandertal dans les systèmes vocaliques, Heim et al., 2002), avec ou sans une approche évo-dévo, sont trop vagues pour notre question comportementale présente. Nous établissons seulement qu'avec la distribution du cycle de babillage d'un enfant entre 6 et 18 mois, on peut statistiquement prédire la durée de son *stroke* de pointer, une durée qui sera dans l'ensemble égale ou supérieure à la durée de deux cycles de babillage ou syllabes. C'est ce que nous montrent les distributions générales ci-dessous ; et ce que nous confirmerons les données individuelles des 6 enfants présentées plus bas.

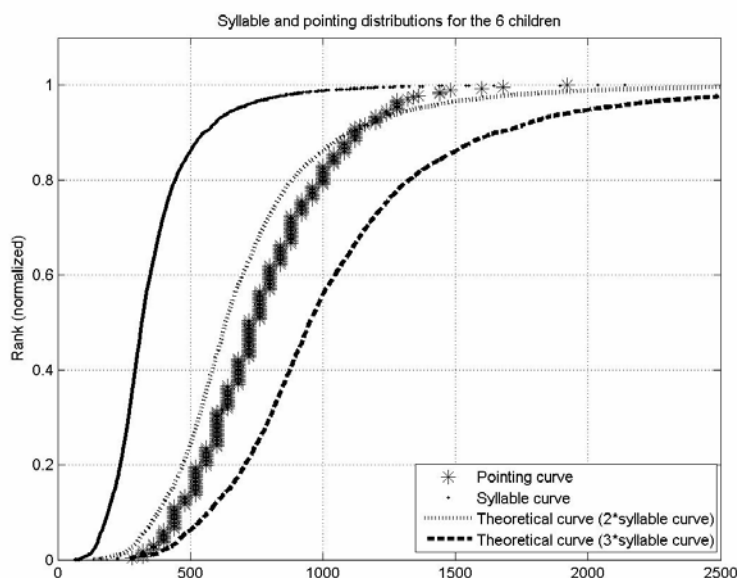


FIGURE III.40 : Les deux distributions cumulées brutes des 6 enfants français entre 6 et 18 mois, pour leur 276 pointers (étoiles) et leurs 4153 syllabes (distribution la plus à gauche). Les deux courbes théoriquement prédites, obtenues en doublant (petits pointillés) et triplant (tiretés) les valeurs de la distribution des syllabes observées, suggèrent que, dans l'ensemble, la durée du *stroke* ou détente de pointer peut être prédite à partir du cycle de babillage observé, comme un cadre (*frame*) ou gabarit (*template*) pour 2 syllabes.

La figure III.40 ci-dessus illustre simplement le succès statistique de notre « étonnante hypothèse ». Lorsque l'on double les valeurs de la distribution des syllabes observées, nous obtenons une approximation plus proche de cette prédiction pour la distribution du pointer observée que si l'on multiplie ces valeurs par le nombre entier supérieur 3. De plus, nous pouvons noter que la distribution du pointage observée est souvent avec des durées de *stroke* correspondantes plus longues que la courbe théorique de la syllabe doublée, ce qui signifie qu'un pointer



peut couvrir suffisamment deux cycles de syllabe mandibulaire (une courbe de prédiction qui échoue seulement pour de très longs strokes, rares exemplaires). Ceci confirme notre prédiction que les premiers mots pourraient contenir deux syllabes dans un pointer. Par exemple, si on a une durée de stroke de pointer de 700ms, la durée de la syllabe correspondante dans la courbe des syllabes est d'environ 300ms. Ainsi, le geste de pointage couvre facilement 2 syllabes. Si le stroke de pointer dure 1000ms, la durée de la syllabe correspondante est d'environ 450ms : ainsi nous pouvons une fois encore mettre facilement 2 syllabes dans un geste de pointer. Et plus intéressant encore est le phénomène inverse, si le pied a un effet de contrôle de l'isochronie. En d'autres termes, nous défendons l'idée que le Pointer-Pied contraint le flux de parole des enfants dans un patron syllabique harmonique de 2:1.

Ainsi, à la question « Pourquoi des mots de 2 syllabes ? », nous proposons finalement que ce ratio harmonique trouvé ici pour les enfants français puisse fournir un patron bisyllabique pour les premiers mots. Dans cette optique, nous nous sommes demandé si nos 6 enfants français montraient une tendance à réaliser plus de productions bisyllabiques dans les événements relevés pour étudier le cycle du babillage. La figure III.41 suivante montre, pour chaque enfant, des histogrammes représentant le nombre de syllabes par énoncé répétitif (c'est-à-dire des énoncés de 2 syllabes ou plus).

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

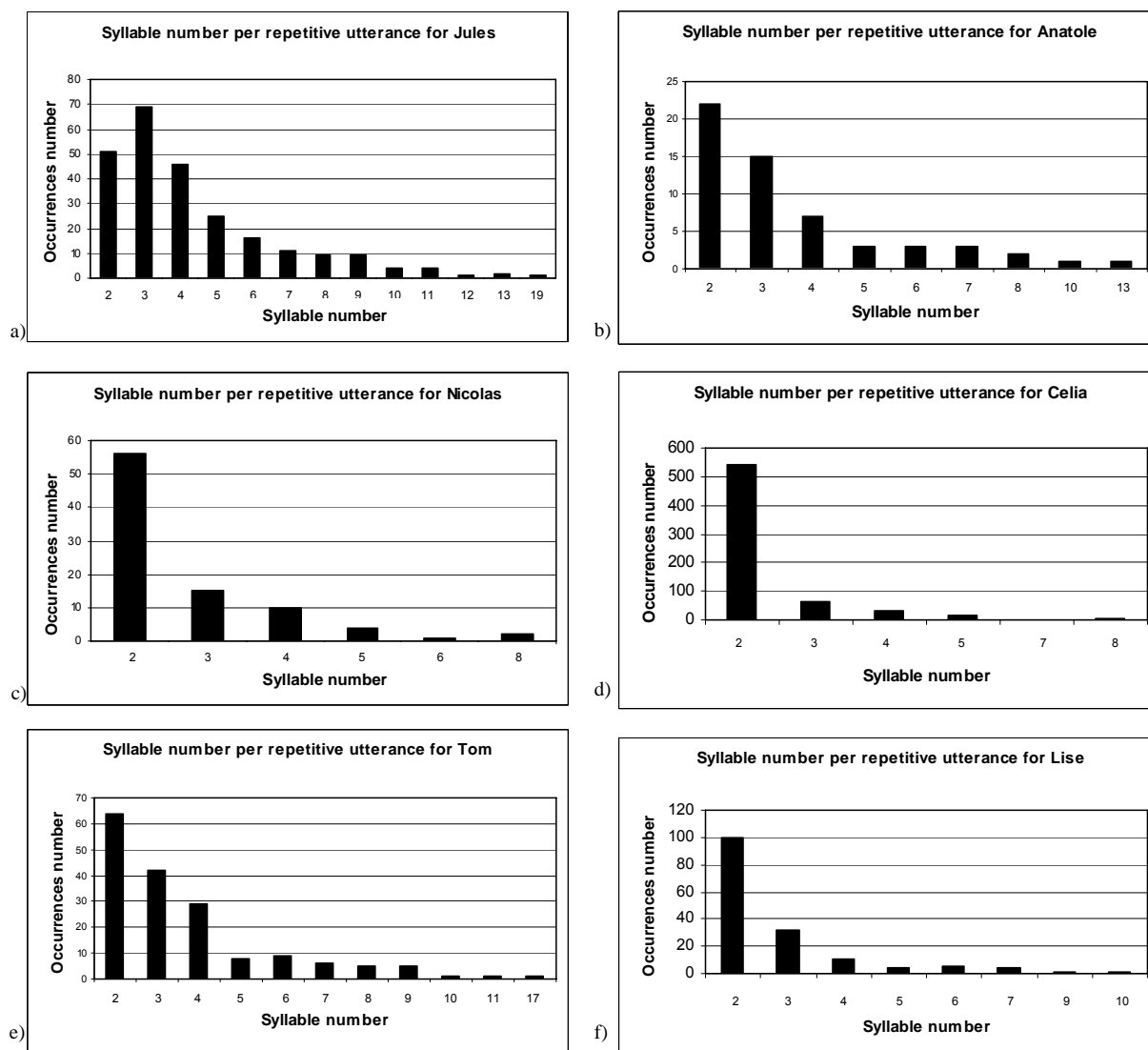


FIGURE III.41: Histogrammes représentant le nombre de syllabes par énoncé répétitif (2 syllabes ou plus) pour chaque enfant français du corpus entre 6 et 18 mois : a) Jules, b) Anatole, c) Nicolas, d) Célia, e) Tom et f) Lise. A la lecture de ces 6 histogrammes, on voit nettement la tendance vers un plus grand nombre d'énoncés à 2 syllabes (excepté pour Jules qui montre une remontée de ses données sur les productions trisyllabiques).

La figure III.42 suivante donne l'histogramme du nombre de syllabes par énoncé répétitif pour les 6 enfants français réunis, mettant clairement en évidence la dominance du patron bisyllabique, les énoncés longs de 2 syllabes représentant 3 ou 4 fois plus d'énoncés que ceux de 3 syllabes.

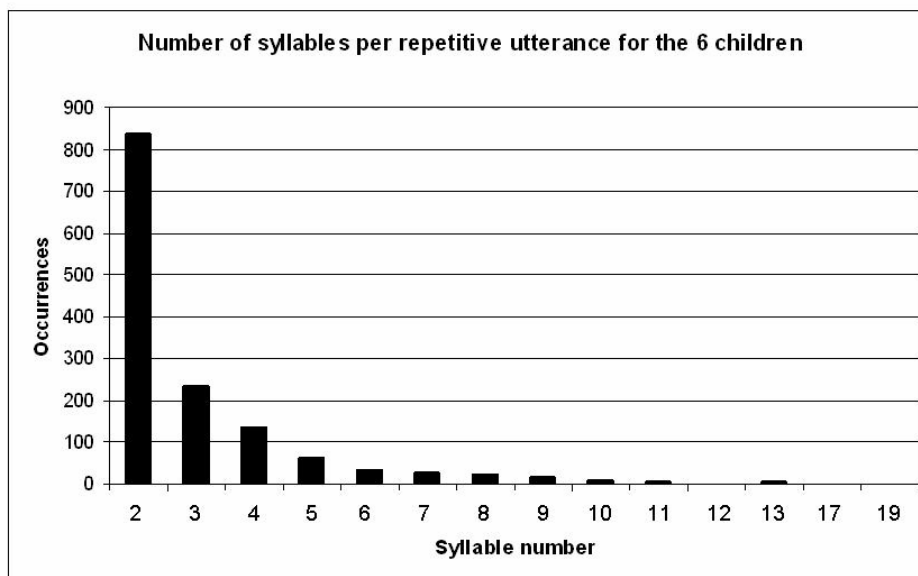


FIGURE III.42 : Histogramme représentant le nombre de syllabes par énoncé répétitif (2 syllabes ou plus) pour les 6 enfants français entre 6 et 18 mois. On remarque un ratio d'environ 3.5 en faveur des productions à 2 syllabes par rapport aux productions à 3 syllabes.

Plus généralement, dans la veine de cette tendance, le ratio harmonique 2:1 trouvé chez les enfants pourrait également correspondre à la fréquence du patron bisyllabique dans les langues du monde. Dans son étude reposant sur un échantillon de 16 langues d'ULSID (très différentes du point de vue typologique), soient 16 lexiques contenant au total 94535 mots et 247252 syllabes, Rousset (2004) a souligné le fait que nous trouvons dans plusieurs langues un pic bisyllabique (voir Figure III.43). Son résultat global ici reproduit indique que les deux tiers des mots sont de 2-3 syllabes. L'intégration de 2 syllabes dans 1 geste de pointer trouvée chez les enfants pourrait alors fournir un patron bisyllabique pour les langues adultes. Notre découverte pourrait ainsi expliquer pourquoi un « chunking » à un gabarit de 2 syllabes s'observe habituellement lorsque les enfants sont exposés à des inputs de *motherese* de plus de 2 syllabes.

## QUAND LE CADRE DE LA PAROLE EST ENCHÂSSÉ DANS LE CADRE DU SIGNE

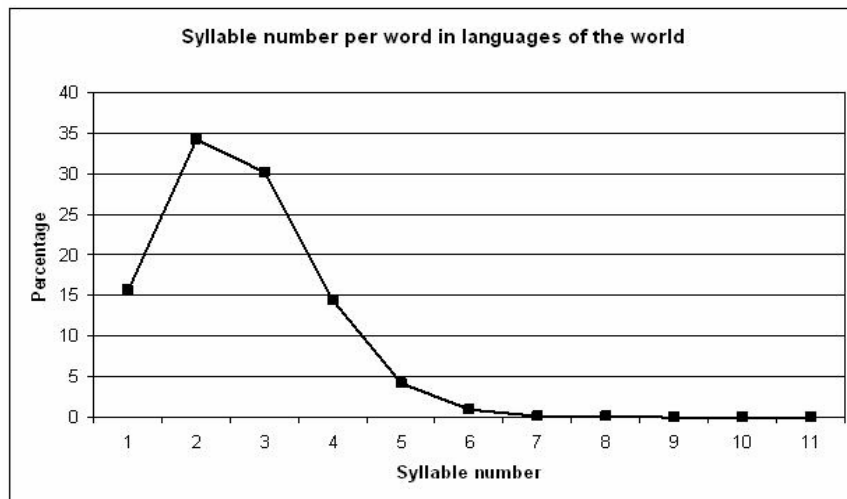


FIGURE III.43 : Un pic à 2 syllabes apparaît dans la distribution (en %) des unités lexicales rangées par nombre de syllabes dans le mot, pour 16 lexiques représentatifs des langues du monde, typologiquement échantillonnées à partir de la base de données ULSID (D'après Rousset, 2004).

CONCLUSION :

**CONCLUSION :**

**PROBLEMES ET PERSPECTIVES**

## Davantage de questions pour une réponse ?

Notre ratio harmonique de 2:1, qui lie de façon flexible les deux cadres, celui de la parole et celui du signe, propose que, de la même façon que la mandibule est la porteuse des lèvres et de la langue dans la cyclicité du babillage-syllabe, l'ensemble bras-main constitue la porteuse de l'index dans les gestes discrets de pointer. Cela ne signifie pas que la syllabe est la mandibule, ni que le pied (*foot*) qu'elle remplit d'un ou deux de ses cycles est le bras. Cela indique fortement que ces contraintes d'articulateurs sont intégrées dans leurs structures de contrôle neurales. Evidemment, le dispositif qui permet au cerveau de l'enfant de devenir capable de tourner le cerveau de la mère vers sa région d'intérêt dès tout petit, seulement par l'œil, ensuite par le pointer et la voix (les enfants plus âgés tournent parfois directement avec leurs mains la tête de leur mère inattentive !), reste une boîte noire mystérieuse qui vient juste d'être ouverte. Appeler cette boîte un mécanisme d'intention partagée parole et signe (*Speech Sign Sharing Intention Mechanism*, S3IM, sigle élaboré à partir du SAM ou *Shared Attention Mechanism* de Baron-Cohen, 1995, au-delà de la proposition SIM, soit *Shared Intention Mechanism* de Tomasello et al., in press) : cela ne restera qu'une simple référence aux hypothèses sur l'intentionnalité inflationnistes/déflationnistes pour la lecture de l'esprit (*mindreading*); à moins que les mécanismes déictiques ne soient mieux contraints (plus par Leslie, 2005 ; que par Pylyshyn, 2000).

Pour notre part, nous continuons à argumenter que ce cadre de la parole et ce cadre du signe sont les deux premières pièces de cette histoire développementale, avant d'aller au-delà du babillage et du pointer : ce sont les racines de la phonologie et de la sémantique. Les deux sont nécessaires pour acquérir plus tard la structure syllabique spécifique, le lexique et la grammaire de la langue ambiante. Invoquer les compétences précoces de découverte de patrons (« *pattern-finding early skills* ») pour les gabarits perceptifs et moteurs —du moment que les bébés sont d'aussi bons extracteurs de patrons (« *pattern-extractors* ») que discriminateurs de sons, et catégoriseurs (voir Tomasello, 2003:28-31)— reste trop peu contraint, ne serait-ce que pour répondre à ces deux questions princeps : (i) Pourquoi des syllabes ? Et (ii) pourquoi une ou deux syllabes dans les premiers mots ? Enfin pourquoi —comme bénéfice gratuit— des premiers énoncés d'un mot ?

Dire que les enfants imitent le mamanais (*motherese*) laisse la question évolutive sans réponse, ne fournissant aucune contrainte pour une approche évo-dévo cohérente, un effort entrepris par MacNeilage, et d'autres précurseurs, pour dériver le langage du non-langage. La découverte des patrons (« *pattern-finding* ») pourra expliquer l'apprentissage des patrons de la langue dans le lexique, dans la phonologie et dans la morpho-syntaxe. Mais avant de pêcher les mots avec leurs constructions par l'oreille (et l'œil), avant d'exprimer même un simple « Qu'est(-ce que) c'est (que) ça? » –prononcé dans le style bébé– on a besoin d'une canne à pêche, à savoir des grognements de voix (*grunts*) et des pointers (œil, bras), comme le rappellent McCune et al. (2003). Nous défendons que cette canne à pêche, ou ce harpon plutôt générique, est un outil pour le gabarit (*template*) phonologique du mot et pour la morphosyntaxe, donnant racine aux démonstratifs (Diessel, 1999b), ceux-ci devenant des relativiseurs et des complémenteurs (de l'anglais *that-that-that* au Nahuatl classique *in-in-in*, etc.: ce que nous appelons les « *that-grammars* » ou « ça-grammaires »), des déterminants du nom (allemand *das Haus*/suédois *huset*, français *l'homme*/roumain *omul*, etc.), des marqueurs verbaux de la personne (latin *ille venit*, français *il vient*, etc.), donnant forme finalement à la morphologie par cliticisation grammaticalisée. Il n'y pas de raison de principe pour ne pas utiliser ces processus dans le développement (ainsi que le défend Vihman, 1999).

La deixis avec des mots et des signes naissants est sans doute enactivée dans les circuits des neurones de la mandibule et du bras. Les bébés ont naturellement besoin des deux pour réussir de façon optimale à capturer conjointement la sémantique et la phonétique disponibles dans leur monde d'interaction. Ainsi accompliront-ils leurs premiers pas avec leurs propres mots en stabilisant leurs syllabes dans un pied.

## **Problèmes d'atypicité ? Les Williams et encore les SMA**

Cette hypothèse que nous avons testée semble très robuste, en tous cas chez les enfants sans pathologie comme nos 6 enfants français. Toutefois, certains cas atypiques pourraient nous poser problème. Qu'en est-il des enfants touchés par le syndrome de Williams ? Les données sont contrastées. Mais Laing et al. (2002 ; pour la comparaison avec le syndrome de Down, cf. Laing et al., 2001) ont trouvé des

enfants atteints de ce syndrome qui ne pointent pas et pourtant apprennent beaucoup de mots, possédant au final un bon vocabulaire. Une solution consisterait à dire qu'il s'agit tout simplement d'un développement atypique et que par conséquent il ne relève pas de la majorité des cas. Toutefois cette explication ne nous satisfait qu'à moitié.

Une première évidence en notre faveur vient des travaux menés par Nazzi (2002, 2005). Celui-ci a observé que les Williams ne sont pas performants pour catégoriser des choses avec des noms. Ils ne sont toujours pas efficaces dans cette tâche à 3-6 ans par rapport aux enfants en développement normal âgés de 20 mois. Mais ils restent en revanche capables de catégoriser visuellement. Ceci dit, le fait qu'ils éprouvent des difficultés à nommer les choses sous-entend l'idée qu'ils ont du mal à ancrer la sémantique de leurs vocalisations. Un bon point pour nous, car s'ils éprouvent ce genre de difficulté, ils peuvent ne pas arriver à notre rendez-vous... Une autre piste pour répondre à cette question viendrait peut-être de la boucle articulatoire. Chez l'adulte, notre boucle s'étend sur 1.6 seconde et permet de retenir environ 10 syllabes (ou un numéro de téléphone par exemple). Chez l'enfant en développement normal cette boucle ne leur permet de retenir que deux syllabes (autant qu'on puisse le savoir dans l'état actuel des mesures). Ainsi, avec un pointer qui permet d'ancrer deux syllabes que l'enfant peut retenir dans sa boucle et ainsi mémoriser à long terme par la suite, nous avons un cadre robuste avec le bras qui permet de gabariser le flux des vocalisations. Mais les Williams ont été caractérisés par un vocabulaire comprenant des mots très longs : il leur faudrait alors une boucle géante pour retenir ces fameux mots rares dont le commun des mortels ignore le sens. Les Williams n'ont donc pas un bon gabarit du bras et ne produisent pas en majorité des mots courts. Ce que nous savons, c'est qu'il existe chez ces enfants plusieurs caractéristiques atypiques : (1) peu de pointers du bras, (2) peu de contact de l'œil (impliquant que les relations triadiques sont très peu pratiquées par ces enfants), (3) pas d'intérêt pour les objets, (4) un défaut de gabarisation du mot, (5) un défaut de référence, (6) pas de couplage signifiant-signifié et (7) une acquisition de mot à la volée (en écho) sans mémorisation. Il apparaît alors que de nombreuses pierres angulaires de notre dispositif de travail (nos deux cadres se rencontrant lors d'un rendez-vous) manquent cruellement au développement de l'enfant Williams. Certes nous n'apportons pas de solution au fait qu'ils acquièrent tout de même un



très grand nombre de mots, dans la quasi-majorité des mots longs et qui plus est des mots rares. Mais ce que nous pouvons relever chez ces enfants, c'est le manque évident de lien entre la sémantique et la phonologie, comme s'ils photographiaient ces mots très facilement mais sans arriver à leur coller une étiquette sémantique. Cela pourrait ainsi constituer une esquisse de réponse aux conséquences du fait qu'ils ne pointent pas.

Les SMA, qui nous ont servi d'entrée à ne donner raison ni à Piaget ni à Chomsky, commencent à être mieux étudiés pour leurs capacités de langage en dynamique interactionnelle. Viodé-Bénony et al. (2002, suite à Bénony, 2000) se sont récemment demandé comment, « la dimension du lien par le langage venant compenser l'absence ou le manque de lien moteur avec l'autre (pointage psychique par les démonstratifs et substitut symbolique de l'action par l'usage accentué de verbes par exemple) », ils parvenaient à des performances remarquables. Notons que ces deux points de leur proposition (italisés par nous), le pointage ou la démonstration, et le focus ou accentuation, sont en droite ligne avec nos préoccupations. L'avenir nous dira si notre « hypothèse étonnante » sur le bras déictique résiste à ces chemins de développement langagier atypique.

## **BIBLIOGRAPHIE**

- ABRY C., STEFANUTO M., VILAIN A., LABOISSIERE R. (2002). What can the utterance “tan,tan” of Broca’s patient Leborgne tell us about the hypothesis of an emergent “babble-syllable” downloaded by SMA? In J. Durand and B. Laks (Eds), *Phonetics, Phonology and Cognition*, Oxford: Oxford University Press, 226-243.
- ABRY C., VILAIN A., SCHWARTZ J.-L. (2004). Introduction: Vocalize to Localize? A call for better crosstalk between auditory and visual communication systems researchers: From meerkats to humans, *Interaction Studies*, 5, 3, 313-325.
- ABRY C., DUCEY V., VILAIN, A., LALEVÉE, C. (*in press*). When the babble-syllable feeds the foot in a point. In Barbara L. Davis and Krisztina Zadjo (Eds.), *Frame/Content Theory and Beyond. A Festschrift for Peter MacNeilage*, London: Erlbaum.
- ACKERMANN H., HERTRICH I. (2003). Cerebellar contributions to speech motor control and auditory verbal imagery: Acoustic / kinematic analyses of ataxic dysarthria and functional magnetic resonance imaging in healthy subjects. In: Solé M.J., Recasens D., Romero J. (Eds). *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, volume 1. Causal Productions, Adelaide, 2003, 163-167.
- ACREDOLO L.P., GOODWYN S.W. (1988). Symbolic gesturing in normal infants, *Child Development*, 59, 450-466.
- AGUIAR A., BAILLARGEON R. (1999). 2.5-month-old infants’ reasoning about when objects should and should not be occluded, *Cognitive Psychology*, 39, 116-157.
- ALIBALI M.W., GOLDIN-MEADOW S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child’s state of mind, *Cognitive Psychology*, 25, 468-523.
- ALIBALI M.W., KITA S., YOUNG A.J. (2000). Gesture and the process of speech production: We think, therefore we gesture, *Language and Cognitive Processes*, 15(6), 593-613.
- ALLEN G.D., HAWKINS S. (1979). Trochaic rhythm in children’s speech. In H. Hollien & P. Hollien (Eds). *Current Issues in the Phonetic Sciences*, Amsterdam: Benjamins, 927-934.
- ALLISON T., PUCE A., MCCARTHY G. (2000). Social perception from visual cues: role of the STS region, *Trends in Cognitive Sciences*, 4, 7, 267-278.
- ANDERSEN E.S., DUNLEA A., KEKELIS L.S. (1984). Blind children’s language: Resolving some differences, *Journal of Child Language*, 11, 645-664.

- ANDERSEN E.S., DUNLEA A., KEKELIS L.S. (1993). The impact of input: Language acquisition in the visually impaired, *First Language*, 13, 23-49.
- ARBIB M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2), 105-124.
- ARZY S., SEECK M., ORTIGUE S., SPINELLI L., BLANKE O. (2002). Induction of an illusory shadow person. Stimulation of a site on the brain's left hemisphere prompts the creepy feeling that somebody is close by, *Nature*, 443, 287.
- ASTAFIEV S., SHULMAN G., STANLEY C., SNYDER A., VAN ESSEN D., CORBETTA M. (2003). Functional organisation of human intraparietal and frontal cortex for attending, looking, and pointing, *The Journal of Neuroscience*, 23:11, 4689-4699.
- ATTINA V., BEAUTEMPS D., CATHIARD M.-A., ODISIO M. (2004). A pilot study of temporal organization in Cued Speech production of French syllables: rules for a Cued Speech synthesizer, *Speech Communication*, 44, 197-214.
- BALLARD D.H. (1997). Deictic codes for the embodiment of cognition, *Behavioral and Brain Sciences*, 20, 723-767.
- BARON-COHEN S. (1995). *Mindblindness*. An essay on autism and theory of mind, MIT Press, Cambridge, Mass.
- BARON-COHEN S., LESLIE A.M., FRITH U. (1985). Does the autistic child have a "theory of mind"?, *Cognition*, 21, 37-46.
- BATES E., BENIGNI L., BRETHERTON I., CAMAIONI L., VOLTERRA V. (1979). Cognition and communication from 9-13 months: correlational findings. In E. Bates (Ed.), *The Emergence of symbols: Cognition and communication in infancy*. New-York: Academic Press.
- BATES E., O'CONNELL B., VAID J., SLEDGE P., OAKES L. (1986). Language and hand preference in early development, *Developmental Neuropsychology*, 2 (1), 1-15.
- BATES E., DICK F. (2002). Language, gesture and the developing brain, *Developmental Psychobiology*, 40(30), 293-310.
- BEKKEN K. (1989). *Is there "motherese" in gesture?*, Unpublished doctoral dissertation, University of Chicago.
- BENONY C. (2000). *Etude psychologique et psycholinguistique de l'acquisition du langage chez des enfants atteints d'une Amyotrophie Spinale Infantile de type II âgés de 25 à 47 mois*. Thèse de Doctorat Nouveau Régime de Psychologie Clinique et de Psychopathologie, Université Paris V-René Descartes.

- BICKERTON D. (1990). *Language and species*, Chicago University Press.
- BICKLEY C., LINDBLOM B., ROUG L. (1986). Acoustic measures of rhythm in infants' babbling, or "All god's children got rhythm", *Proceedings of the 12<sup>th</sup> International Congress on Acoustics*, Volume / Band I A-C, A6-4, Toronto, 24-31 July.
- BLANKE O, ORTIGUE S, LANDIS T, SEECK M. (2002). Stimulating illusory own-body perceptions. *Nature*, 419, 269-270. (Cf. aussi: *Lancet Neurology* (2002), 1; 400, *Trends in Cognitive Neuroscience* (2003), 5; 104-106.)
- BONNOT J-F. (1990). Production de la parole et coarticulation : une analyse critique des principaux modèles, *Travaux de l'Institut de Phonétique de Strasbourg*, volume 20, 172 p.
- BONVILLIAN J.D., ORLANSKY M.D., NOVAK L.L. (1985). Early sign language acquisition and its relationship to cognition and motor development. In J. Kyle & B. Woll (Eds), *Language in sign: An international perspective on sign language*, London: Croom Helm.
- BOVET D., VAUCLAIR J. (1998). Functional categorization of objects and of their pictures in Baboons (*Papio anubis*), *Learning and Motivation*, 29, 309-322).
- BOYSSON-BARDIES B. de, BACRI L., SAGART L., POIZAT M. (1981). Timing in late babbling, *Journal of Child Language*, 8, 525-539.
- BOYSSON-BARDIES B. de. (1996). *Comment la parole vient aux enfants*, Ed. Odile Jacob, 302 p.
- BROOKS R. A. (1986). A Robust Layered Control System For A Mobile Robot, *IEEE Journal of Robotics and Automation*, RA-2, April, 14-23.
- BUCCINO G., BINKOFSKI F., RINK G.R., FADIGA L., FOGASSI L., GALLESE R., SEITZ R.J., ZILLES K., RIZZOLATTI G., FREUND H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study, *European Journal of Neuroscience*, 13, 400-404.
- BUHR R.D. (1980). The emergence of vowels in an infant, *Journal of Speech and Hearing Research*, 23, 73-94.
- BULLOT N., DROULEZ J. (under review). How do objects get predicates? An experimental and conceptual study of the relation between deictic reference, spatial information and object identification.
- BUTCHER C., GOLDIN-MEADOW S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together, In *Language and Gesture*, MacNeill, D. (editor), Cambridge, New York: Cambridge University Press, 235-257.

- BUTTERWORTH B., HADAR U. (1989). Gestures, speech and computational stages: a reply to McNeill, *Psychological Review*, 96-1, 168-174.
- BUTTERWORTH G. (1991). The ontogeny and phylogeny of joint visual attention, in Whiten, A. (ed), *Natural Theories of Mind*, Blackwell.
- BUTTERWORTH G. (1998). What is special about pointing in babies? In *The development of sensory, motor and cognitive capacities in early infancy*, Simion, F., Butterworth, G. (Eds), chapitre 10, 171-190.
- BUTTERWORTH G. (2003). Pointing is the Royal Road to Language for Babies, In *Pointing: Where Language, Culture, and Cognition meet*, edited by Sotaro Kita, 9-33.
- BUTTERWORTH G., GROVER L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy, In *Attention and Performance XIII, Motor Representation and Control*, Edited by M. Jeannerod, Lawrence Erlbaum Publishers, 605-624.
- CAMAIONI L. (1993). The development of intentional communication: A re-analysis. In J. Nadel, & L. Camaioni (Eds.), *New perspectives in early communicative development*. London: Routledge.
- CAMPBELL N. (1992). *Multi-level timing in speech*, Ph.D. Thesis, University of Sussex, U.K.
- CANTALUPO C., HOPKINS W.D. (2001). Asymmetric Broca's area in great apes, *Nature*, 414, 505.
- CAPIRCI O., IVERSON J.M., PIZZUTO E., VOLTERRA V. (1996). Gestures and words during the transition to two-word speech, *Journal of Child Language*, 23, 645-673.
- CAPIRCI O., IVERSON J.M., MONTANARI S., VOLTERRA V. (2002). Gestural, signed and spoken modalities in early language development: The role of linguistic input. *Bilingualism Language and Cognition*, 5 (1), 25-37.
- CAPONE N.C., MCGREGOR K. (2004). Gesture development: a review of clinical and research practices, *Journal of Speech, Language, and Hearing Research*, vol 48, 173-186.
- CAREY S. (2001). Evolutionary and Ontogenetic Foundations of Arithmetic. *Mind and Language*, 16(1), 37-55.
- CAREY S. (2004). Bootstrapping and the origin of concepts, *Dædalus*, 133, 1, MIT Press, 59-68.
- CASELLI M.C., OSSELLA T., VOLTERRA V. (1984). Sign and vocal language acquisition by two Italian deaf children of deaf parents. In F. Loncke, P. Boyes-

Braem & Y. Lebrun (eds.), *Recent Research on European Sign Language*. Lisse: Swets & Zeitlinger, pp.121-128.

- CASELLI, M.C., VOLTERRA V. (1990). From communication to language in hearing and deaf children. In V. Volterra & C. Erting, (eds.) (1990). *From Gesture to Language in Hearing and Deaf Children*. New York: Springer-Verlag. (2nd Edition 1994, Washington, D.C.: Gallaudet University Press), pp. 263-277.
- CASELLI M.C., VOLTERRA V., PIZZUTO E. (1984). *The relationship between vocal and gestural communication from the one-word to the two-word stage*. Paper presented at the International Conference on Infant Studies, New York, NY.
- CASSELL J., PREVOST S. (1996). Distribution of semantic features across speech and gesture by humans and machines. In Messing (ed.), 253-269.
- CASTAIGNE P., LHERMITTE F., SIGNORET J.L., ABELANET R. (1980). Description et étude scannographique du cerveau de Leborgne. La découverte de Broca. *Revue Neurologique*, 136, pp. 563-583
- CASTIELLO U., PAULIGNAN Y., JEANNEROD M. (1991). Temporal Dissociation of Motor Responses and Subjective Awareness, a study in normal subjects, *Brain*, 114, 2639-2655.
- CHAMINADE T., DECETY J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency, *Brain Imaging*, Vol 13, 15, 1975-1978.
- CHEYNEY D.L. et SEYFARTH R.M. (1990). *How monkeys see the world: Inside the mind of another species*, University of Chicago Press.
- CHOMSKY N. (1957). *Syntactic structures*. The Hague, Mouton & co.
- CHOMSKY N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- CHURCH R.B., GOLDIN-MEADOW S. (1986). The mismatch between gesture and speech as an index of transitional knowledge, *Cognition*, 23, 43-71.
- CLARK, R. (1978). The transition from action to gesture. En A. Lock (Ed.) *Action, Gesture and Symbol*. London: Academic Press. (Trad. cast. en A. Perinat. *La comunicación preverbal*. Barcelona: Avesta, 1986).
- CODE C. (2005). Syllables in the brain: Evidence from brain damage. In *Phonological Encoding and Monitoring in Normal and Pathological Speech*, Robert J. Hartsuiker, Roelien Bastiaanse, Albert Postma, Franck Wijnen (Eds), Psychology Press, Hove and New York, 119-136.

- COHEN L.B., CASHON C.H. (2001). Infant object segregation implies information integration, *Journal of Experimental Child Psychology*, 78, 1, 75-83.
- COLLETTA J.-M. (2004). *Le développement de la parole chez l'enfant âgé de 6 à 11 ans : Corps, langage et cognition*, Mardaga, Bruxelles.
- COOK K. (1999). Vervet Monkeys get the « Word » out, *Primate Science Research Highlight, Editorial Intern*, Wisconsin Regional Primate Research Center.
- CORBALLIS M.C. (2002). From hand to mouth. The origins of language, *Gesture*, 5 (1-2), 285-304
- CORBALLIS M.C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness, *Behavioral and Brain Sciences*, 26, 199-260.
- CREISSELS D. (1995). *Eléments de syntaxe générale*, PUF linguistique nouvelle, chapitre 5: la notion de Prédicat, 46-48.
- CUTLER A. (1980). Syllable omission errors and isochrony, In *Temporal variables in speech*, Dechert, H.W. et Raupach, M. (eds), Mouton Publishers, 183-190.
- DAPRATI E., FRANCK N., GEORGIEFF N., PROUST J., PACHERIE E., DALERY J., JEANNEROD M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness of action and self-consciousness in schizophrenic patients, *Cognition*, 65, 71-86.
- DAPRATI E., SIRIGU A. (2006). How we interact with objects: learning from brain lesions, *Trends in Cognitive Sciences*, 10, 6, 265-270.
- DAPRETTO M., BJORK E.L. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth, *Child Development*, 71, 3, 635-648.
- DAVIS B.L., MACNEILAGE P., MATYEAR C. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production, *Phonetica*, 59, 75-107.
- DEAK G.O., FASEL I., MOVELLAN J. (2001). The emergence of shared attention: Using robots to test developmental theories, *First International Workshop on Epigenetic Robotics : Modeling Cognitive Development in Robotic Systems*.
- DECETY J. (2003). L'empathie ou l'émotion partagée, *Pour la Science*, 309, 46-51.
- DEMUTH K. (1996). The prosodic structure of early words, In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, J. Morgan et K. Demuth (eds), Mahwah, N.J.:Lawrence Erlbaum Associates, 171-184.



- DEMUTH K., JOHNSON M. (2003). Truncation to subminimal words in early French, *Canadian Journal of Linguistics*, 48, 211-241.
- DEN OS E.A. (1990). Development of temporal properties in the speech of one child between one and three years of age, *Proceedings of the Institute of Phonetic Sciences*, Amsterdam, 14, 39-52.
- DENNETT D.C. (2000). Making Tools for thinking. In Sperber D. (Ed.), *Metarepresentations : A multidisciplinary perspective*, New York, Oxford University Press, 17-29.
- DE BLESER R., POECK K. (1983). Comments on paper "Neurolinguistic analysis of recurrent utterance in aphasia" by C. Code (*Cortex*, 18, 141-151, 1982), *Cortex*, 19, 259-260.
- DE RUITER J.P. (2000). The production of gesture and speech, In *Language and Gesture*, Edited by McNeill, Cambridge University Press, 284-311.
- DE RUITER J.P, WILKINS D.P (1998). The synchronisation of gesture and speech in Dutch and Arrernte (an Australian Aboriginal language): a cross-cultural comparison, In *Oralité et Gestualité : communication multimodale, Interaction*, G. Santi et alii (eds), Paris : L'Harmattan, 603-607.
- DE WAAL F.B.M., DINDO M., FREEMAN C.A., HALL M.J. (2005). The monkey in the mirror: Hardly a stranger, *Proceedings of the National Academy of Sciences of the United States of America*, August 9, 102, 32, 11140-11147.
- DIAFERIA M.-L., ABRY C. (2005). Du conte au compte: pour une ethnographie du développement de la narration et de la numération. In *DECOLAGE III, Le Développement Conceptuel et Langagier de l'Enfant*, Reims (23 - 24 Juin 2005) <http://www.univ-reims.fr/Labos/Accolade>
- DIESSEL H. (1999a). The morphosyntax of demonstratives in synchrony and diachrony, *Linguistic Typology*, 3, 1-49.
- DIESSEL H. (1999b). *Demonstratives: Form, Function, and Grammaticalization*, John Benjamins.
- DIESSEL H. (2003a). The relationship between demonstratives and interrogatives, *Studies in language*, 27:3, 635-655.
- DIESSEL H. (2003b). Demonstratives in language use and grammar, *San Marino Summer School*, 1-39.
- DIESSEL H., TOMASELLO M. (2000). The development of relative clauses in spontaneous child speech, *Cognitive Linguistics*, 11-1/2, 131-151.

- DREWNOWSKI A., HEALY A.F. (1977). Detection errors on 'the' and 'and' : Evidence for reading units larger than the word, *Memory et Cognition*, 5, 636-647.
- DUBEAU M.C., IACOBONI M., KOSKI L., MARKOVAC J., MAZZIOTTA J.C. (2002). Topography for body-parts motion in the posterior STS region, *From cognitive neuroscience to social science*, Royaumont Abbey, France, 24-26 may, Abstract
- DUCEY-KAUFMANN V., ABRY C., VILAIN C. (in press). When the Speech Frame meets the Sign Frame in a developmental framework, In *Emergence of Language Abilities: Ontogeny and phylogeny* (ELA, Lyon Dec. 2005).
- DUNLEA A. (1989). *Vision and the emergence of meaning*, Cambridge: Cambridge University Press.
- DUNLEA A., ANDERSEN E.S. (1992). The emergence process: Conceptual and linguistic influences on morphological development, *First Language*, 12, 95-115.
- EJIRI K. (1998). Relationship between rythmic behavior and canonical babbling in infant vocal development, *Phonetica*, 55, 226-237.
- EMERY N.J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze, *Neuroscience and Biobehavioral Reviews*, 24, 581-604.
- ERHARD P., KATO T., STRUPP J.P., ANDERSEN P., ADRIANY G., STRICK P.L., UGURBILL K. (1996). Functional mapping of motor in and near Broca's area, *Neuroimage*, 3, S367.
- FASEL I., DEAK G.O., TRIESCH J., MOVELLAN J.R. (2002). Combining embodied models and empirical research for understanding the development of shared attention, In *Proceedings of the 2nd International Conference on Development and Learning*, Cambridge, Massachusetts.
- FERRARI P.F., FOGASSI L., GALLESE V., RIZZOLATTI G. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex, *European Journal of Neuroscience*, 17 (8), 1703-1714.
- FEYEREISEN P. (1997). The competition between gesture and speech production in dual-task paradigms, *Journal of Memory and Language*, 36, 13-33.
- FITCH W.T., HAUSER M.D. (2004). Computational constraints on syntactic processing in a nonhuman primate, *Science*, 16 janvier, 303, 5656, 377-380.
- FOGASSI L., FERRARI P.F. (2004a). Mirror neurons, gestures and language evolution, *Interaction Studies*, 5 (3), 345-363.

- FOGASSI L., FERRARI P.F. (2004b). Neurones miroir, gestes et évolution du langage, *Primatologie*, 6, 263-286.
- FOGEL A., HANNAN T. E. (1985). Manual actions of nine- and fifteen-week-old human infants during face-to-face interaction with their mothers, *Child Development*, 56, 1271-79.
- FOGEL A., THELEN E. (1987). Development of early expressive and communicative action: Reinterpreting the evidence from a dynamic systems perspective, *Developmental Psychology*, 23, 747-761.
- FRANCO F., BUTTERWORTH G.E. (1988). The social origins of pointing in human infancy. Paper presented at *the Annual Conference of the Developmental Psychology Section*, British Psychological Society, Coleg Harlech, Wales.
- GEORGIEFF N., JEANNEROD M. (1998). Beyond consciousness of external reality. A "Who" system for consciousness of action and self-consciousness, *Consciousness and Cognition*, 7(3), 465-477.
- GERKEN L. (1994). A metrical template account of children's weak syllable omission from multisyllabic words, *Journal of Child Language*, 21, 565-584.
- GIL-DA-COSTA R., HAUSER M.D. (In press). Vervet monkeys and humans show brain asymmetries for processing conspecific vocalizations, but with opposite patterns of laterality, In *Proceedings of the Royal Society, Biological Sciences*, 1-21.
- GILLETTE J., GLEITMAN L.R., GLEITMAN H., LEDERER A. (1999). Human simulations of vocabulary learning, *Cognition*, 73, 135-176.
- GIVÓN T. (1998). On the co-evolution of language, mind and brain, *Evolution of Communication*, 2, 1, 45-116.
- GANNON P.J., HOLLOWAY R.L., BROADFIELD D.C., BRAUN A.R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog, *Science*, 279, 220-222.
- GOLDIN-MEADOW S. (2002). Constructing communication by hand, *Cognitive Development*, 17, 1385-1405.
- GOLDIN-MEADOW S. (1999). The role of gesture in communication and thinking, *Trends in Cognitive Science*, vol 3, 11, 419-429.
- GOLDIN-MEADOW S., MORFORD M. (1985). *Gesture in early language: Studies of deaf and hearing children*. Merrill-Palmer Quaterly, 31, 145-176.

- GOLDIN-MEADOW S., BUTCHER C. (2003). Pointing toward two-word Speech in Young Children, In *Pointing: Where Language, Culture and Cognition meet*, S. KITA (ed), Laurence Erlbaum Associates, 85-107.
- GOODALE M.A., MILNER A.D. (1992). Separate visual pathways for perception and action, *Trends in Cognitive Sciences*, 15, 1, 20-25.
- GREEN J.R., MOORE C.A., HIGASHIKAWA M., STEEVE R.W. (2000). The physiologic development of speech motor control: Lip and jaw coordination, *Journal of Speech, Language, and Hearing Research*, 43, 239-255.
- GREEN J.R., MOORE C.A., REILLY K.J. (2002). The sequential development of jaw and lip control for speech, *Journal of Speech, Language, and Hearing Research*, 45, 66-76.
- HAECKEL E. (1896). *Systematische Phylogenie. Zweiter Teil: Systematische Phylogenie der wirbellosen Thiere (Invertebrata)*. Berlin. Verlag von Georg Reimer. 720 p.
- HALLE P., DURAND C., de BOYSSON-BARDIES B. (2005). What role do articles play in infants' recognition of familiar words ? *Conférence Decolage, Développement conceptuel et langagier de l'enfant*, Reims, 23-24 juin 2005, abstract.
- HAMKER F. (2002). How does the ventral pathway contribute to spatial attention and the planning of eye movements?, In *Dynamic Perception*, R.P. Würtz et M. Lappe (Eds), Infix Verlag, St Augustin, 83-88.
- HARRIS L.R., JENKIN M. (2000). *Vision and Attention*, publisher: Springer-Verlag, New York, chap. 1: Vision and Attention.
- HAUSER M.D. (1992). Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustical Society of America*, 363-369.
- HAUSER M.D., CHOMSKY N., FITCH W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve ?, *Science*, 298, 1569-1579.
- HEALY A. F. (1976). Detection errors on the word the: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 235-242.
- HEALY A. F. (1980). Proofreading errors on the word the: New evidence on reading units. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 4557.
- HEALY A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text, *Psychonomic Bulletin et Review*, 1, 333-344.

- HEIDER F., SIMMEL, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.
- HEIM J-L., BOË L-J., ABRY C. (2002). La parole à la portée du conduit vocal de l'homme de Neandertal. Nouvelles recherches, nouvelles perspectives. *Comptes Rendus de l'Académie des Sciences, Palevol*, 1, 129-134.
- HEINE B. et KUTEVA T. (2002). On the evolution of grammatical forms. In Alison Wray, editor, *The transition to Language*. Oxford: Oxford University Press.
- HILAIRE-DEBOVE G., DEMUTH K. (2005). Troncation de mot chez l'enfant francophone, ELA: *Emergence of Language Abilities: Ontogeny and Phylogeny* (Lyon 8-10 dec.) (A).
- HIRSH-PASEK K., GOLINKOFF R.M. (1999). *The origins of grammar: Evidence from early language comprehension*, MIT Press, Cambridge, Mass.
- HOLENDER D. (1980). Interference between a vocal and a manual response to the same stimulus, In *Tutorials in Motor Behavior*, G.E. Stelmach and J. Requin (eds), North-Holland Publishing Company, 421-431.
- HOPKINS W.D., MARINO L., RILLING J.K., MacGREGOR L.A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI), *NeuroReport*, 9 (12), 2913-2918.
- HOUDÉ O. (1998). De la pensée du bébé à l'exemple du nombre, *Sciences Humaines*, 87, 28-31.
- HURFORD J.R. (2003). The Neural Basis of Predicate-Argument Structure, *Behavioral and Brain Sciences*, 26, 3, 261-283.
- IACOBONI M. (in press). Understanding others: imitation, language, empathy, In *Perspectives on imitation: From cognitive neuroscience to social science*, In: Hurley, S., Chater, N. (Eds), Cambridge, MA, MIT Press.
- INDEFREY P., LEVELT W.J.M. (2004). The spatial and temporal signatures of word production components, *Cognition*, 92, 101-144.
- IVERSON J.M., GOLDIN-MEADOW S. (1998). Why people gesture when they speak, *Nature*, 396, 228.
- IVERSON J.M., THELEN E. (1999). Hand, mouth and brain, *Journal of Consciousness Studies*, 6, 19-40.
- IVERSON J.M., TENCER H.L., LANY J., GOLDIN-MEADOW S. (2000). The relation between gesture and speech in congenitally blind and sighted language-learners, *Journal of Nonverbal Behavior*, 24(2), 105-129.

- IVERSON J.M., GOLDIN-MEADOW S. (2005). Gesture paves the way for language development. *Psychological Science*, 16, 367-371.
- IVERSON, J.M., WOZNIAK R.H. (2006). The growth of the growth point, *Trends in Cognitive Sciences*, 10, 6, 241-242.
- JACKENDOFF R. (1983). *Semantics and Cognition*. Cambridge (MA) MIT Press, 283p.
- JACKENDOFF R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*, Oxford University Press, Oxford.
- JACKENDOFF R. (2003). Precis of foundations of Language: Brain, Meaning, Grammar, Evolution, *Behavioral Brain and Sciences*, 26(6), 651-665; discussion 666-707.
- JACKENDOFF R., PINKER S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky), *Cognition*, 97, 211-225.
- JAKOBSON R. (1969). *Langage enfantin et aphasie*, Paris, Editions de Minuit (Traduction de: Kindersprache, Aphasie eine allgemeine Lautgesetze, Uppsala, 1941).
- JOHNSON S.C. (2000). The recognition of mentalistic agents in infancy, *Trends in Cognitive Sciences*, 4, 1, 22-28.
- JONAS S. (1981). The supplementary motor region and speech emission, *Journal of Communication Disorders*, 14, 349-373.
- JUNGBLUTH K. (1999). Two- and three-dimensional deictic systems between speech and writing –Evidences from the use of demonstratives in romance languages, *Proceedings of the Workshop on Deixis, Demonstration and Deictic Belief at ESSLLI XI*, E. André, M. Poesio et H. Rieser (Eds).
- KASSAI I. (1988). Prosodic development: stressing procedures of a Hungarian child. *Budapest: 6<sup>th</sup> International Phonology meeting*, Discussion Papers 1, 32-34. Kassai I. (1991). The emergence of intonation and stress in Hungarian: a case study. *Actes du XI<sup>e</sup> ICPPhS*. Aix-en-Provence, 1, 328-332.
- KATZ S. (2000). Categories of c'est-cleft constructions, *Revue Canadienne de Linguistique*, 45 (3/4), 253-273.
- KENDON A. (1980). Gesticulation and speech, two aspects of the process of utterance. In M.R. Key (Ed), *The relationship of verbal and nonverbal communication*, 207-227.
- KENT R. (1992). *The biology of phonological development*. In Ferguson et al., 65-90.

- KITA S., ÖZYÜREK A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking, *Journal of Memory and Language*, 48, 16-32.
- KONOPCZYNSKI G. (1990). *Le langage émergent : caractéristiques rythmiques*. Hamburg, Buske Verlag.
- KONOPCZYNSKI G. (1993). Le bébé de deux ans a-t-il déjà acquis la structuration rythmique de sa langue maternelle ? Exemples de quelques langues européennes, *Langues*, 3, 184-205.
- KONOPCZYNSKI G. (1998). Interactive Developmental Intonology (IDI) : Theory and Application to French, *Revue Parole*, 7/8, 177-202.
- KOOPMANS-VAN BEINUM F.J. (1992). The role of focus words in natural and in synthetic continuous speech: Acoustic aspects. *Speech Communication*, 11, 439-452.
- KOOPMANS-VAN BEINUM F.J. (1993). Cyclic effects on infant speech perception, early sound production, and maternal speech, *IFA [Institut de Phonétique d'Amsterdam] Proceedings*, 17, 65-78.
- KOOPMANS-VAN BEINUM F.J., VAN DER STELT J. (1986). Early stages in the development of speech movements. In: B. Lindblom & R. Zetterström (Eds), *Precursors of Early Speech*, New York: Stockton, 37-50.
- KOWAL S., O'CONNELL D.C., SABIN E.J. (1975). Development of temporal patterning and vocal hesitation in spontaneous narratives, *Journal of Psycholinguistic Research*, 4, 195-207.
- KOZIMA H. (2002). Infanoid: A babybot that explores the social environment, in K. Dautenhahn, A. H. Bond, L. Canamero, B. Edmonds (eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, Amsterdam: Kluwer Academic Publishers, 157-164.
- KRAUSS R.M. (1998). Why do we gesture when we speak?, *Current Directions in Psychological Science*, 7, 54-59.
- KRAUSS R.M., CHEN Y, GOTTESMAN R.F. (2000). Lexical gestures and lexical access: a process model, In *Language and Gesture*, Edited by McNeill, Cambridge University Press, 261-283.
- KUMASHIRO M., ISHIBASHI H., UCHIYAMA Y., ITAKURA S., MURATA A., IRIKI A. (2003). Natural imitation induced by joint attention in Japanese monkeys, *International Journal of Psychophysiology*, 50, 81-99.
- LAING E., BUTTERWORTH D., ANSARI D., GSÖDL M., LONGHI E., PANAGIOTAKI G., PATERSON S., KARMILOFF-SMITH A. (2002). Atypical

Development of Language and Social Communication in Toddlers with Williams Syndrome, *Developmental Science*, 5, 2, 233-246.

- LAING E., GRANT J., THOMAS M., KARMILOFF-SMITH A. (2001). The influence of phonological and semantic factors on verbal short-term memory in Williams syndrome, Pre-publication manuscript submitted to *Journal of Child Psychiatry and Psychology*, 1-47.
- LALEVEE C. (2003). *Développement du contrôle de la production de parole, du cadre proto-syllabique vers la syllabe : suivi audio-visuel de 2 enfants de 6 à 12 mois*. Mémoire de DEA de Sciences du Langage, Université Stendhal, Institut de la Communication Parlée, Grenoble.
- LALEVEE C., VILAIN A. (2003). Development of speech frame control: a longitudinal study of the oral/nasal control, *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.
- LAMBRECHT K. (1988). Presentational cleft constructions in spoken French. In J. Haiman and S. Thompson (Eds), *Clause combining in grammar and discourse*. Amsterdam: John Benjamins. 135-179.
- LANDAU B., GLEITMAN L.R. (1985). *Language and experience: Evidence from the blind child*, Cambridge, MA: Harvard University Press
- LANGTON S.R.H., WATT R.J., BRUCE V. (2000). Does the eyes have it ? Cues to the Direction of Social Attention, *Trends in Cognitive Sciences*, 4, 2, 50-59.
- LANGTON S.R.H., BRUCE V. (2000). You Must See the Point: Automatic Processing of Cues to the Direction of Social Attention, *Journal of Experimental Psychology: Human Perception and Performance*, 26, 747-757.
- LAUNEY M. (1990). Quand seuls les démonstratifs désignent : prédicats et déictiques en Nahuatl « classique », In *La Deixis, Colloque en Sorbonne*, Morel, M.A., Danon-Boileau, L. (eds), PUF, 221-232.
- LAUNEY M. (1994). *Une Grammaire Omniprédicative, Essai sur la morphosyntaxe du nahuatl classique*, Sciences du Langage, Cnrs éditions.
- LEAVENS D.A. (2004). Manual deixis in apes and humans. In C. Abry, A. Vilain & J-L. Schwartz (Eds), *Special Issue: "Vocalize to localize I". Interaction Studies*, 5 (3), 387-408.
- LEEKAM S.R., HUNNISETT E., MOORE C. (1998). Targets and Cues: Gaze-Following in Children with Autism, *Journal of Child Psychiatry*, 39, 7, 951-962.
- LEGERSTEE M., BARILLAS Y. (2003). Sharing attention and pointing to objects at 12 months: is the intentional stance implied?, *Cognitive Development*, 18, 91-110.



- LESLIE A.M. (1994). ToMM, ToBy, and agency: core architecture and domain specificity, In *Mapping the mind, Domain specificity in cognition and culture*, Hirschfeld L.A. et S.A. Gelman (Eds), Cambridge University Press, 119-148.
- LESLIE A.M, XU F., TREMOULET P., SCHOLL B. (1998). Indexing and the object concept: developing “What” and “Where” systems, *Trends in Cognitive Sciences*, vol 2, 1, 10-18.
- LESLIE A.M. (2005). Developmental parallels in understanding minds and bodies, *Trends in Cognitive Sciences*, 9, 10, 459-462.
- LEVELT W.J.M. (1989). *Speaking : from intention to articulation*. Cambridge, MA, MIT Press.
- LEVELT W.J.M. (1998). The genetic perspective in psycholinguistics or where do spoken words come from?, *Journal of Psycholinguistic Research*, 27, 2, 167-179.
- LEVELT W.J.M., RICHARDSON G., LA HEIJ W. (1985). Pointing and voicing in deictic expressions, *Journal of Memory and Language*, 24, 133-124.
- LIEVEN E., PINE J., BALDWIN G. (1997). Lexically-based learning and early grammatical development, *Journal of Child Language*, 24, 187-219.
- LINDBLOM B. (1991). The Status of Phonetic Gestures. In *Modularity and the Motor Theory of Speech Perception* ed. by Mattingly, Ignatius M. and Michael Studdert-Kennedy, pp. 7-24
- LISZKOWSKI U., CARPENTER M., HENNING A., STRIANO T., TOMASELLO M. (2004). Twelve-month-olds point to share attention and interest, *Developmental Science*, 7(3), 297-307.
- LISZKOWSKI U. (2005). Human twelve-month-olds point cooperatively to share interest with and provide information for a communicative partner, *Gesture*, 5 (1,2), 135-154.
- LOCKE J.L. (1997). A theory of neurolinguistic development, *Brain and Language*, 58, 265-326.
- LÖEVENBRUCK H., BACIU M., SEGEBARTH C., ABRY C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus, *Journal of Neurolinguistics*, 18/3, 237-258.
- LOUWERSE M.M., BANGERTER A. (in press). Focusing attention with deictic gestures and linguistic expressions. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society 2005*.
- MACNAMARA J. (1977). From sign to language. In J. MacNamara (Ed), *Language learning and thought*, New York: Academic Press, 11-35.

- MACNEILAGE P.F., DAVIS B.L. (1990). Acquisition of speech production: frames, then content. In Jeannerod (Ed.), *Attention and Performance XIII: Motor Representation and Control*. Hillsdale: Erlbaum, 452-468.
- MACNEILAGE P.F. (1998). The Frame/Content Theory of Evolution of Speech Production, *Behavioral and Brain Sciences*, 21, 499-546.
- MACNEILAGE P.F., DAVIS B.L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications, *Current Opinion in Neurobiology*, 11, 696-700.
- MAILLARD M. (1987). *Comment ÇA fonctionne*. Thèse de linguistique française pour le doctorat d'état, Université de Paris X-Nanterre.
- MANSER M.B., FLETCHER L.B. (2004). 'Vocalize to localize' - A test on functionally referential alarm calls, *Interaction Studies*, 5, 327-344
- MANSER M.B, SEYFARTH R.M., CHENEY D.L. (2002). Suricate alarm calls signal predator class and urgency, *Trends in Cognitive Sciences*, vol 6, 2, 55-57.
- MAR R.A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation, *Neuropsychologia*, 42, 1414-1434.
- MARESCHAL D. (2000). Object knowledge in infancy: current controversies and approaches, *Trends in Cognitive Sciences*, 4, 11, 408-416.
- MARLER P., EVANS C.S., HAUSER M.D. (1992). Animal signals: motivational, referential, or both? In *Nonverbal vocal communication: comparative and developmental approaches*, Papousek, H., Jürgens U., Papousek M. (Eds), Cambridge University Press, Cambridge, UK, 66-86.
- MARKS I., NESSE R. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders, *Ethology and Sociobiology*, 15, 247-261.
- MASATAKA N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling, *Developmental Science*, 4:2, 158-164.
- MASATAKA N. (2003). From Index-Finger Extension to Index-Finger Pointing: Ontogenesis of Pointing in Preverbal Infants, In *Pointing: Where Language, Culture, and Cognition Meet*, Edited by Sotaro Kita.
- MASUR E.F. (1983). Gestural development, dual-directional signalling, and the transition to words, *Journal of Psycholinguistic Research*, 12, 2, 93-109.

- McCUNE L., VIHMAN M.M. (2001). Early phonetic and lexical development : A productivity approach, *Journal of Speech, Language, and Hearing Research*, 44, 670-684.
- McCUNE L., GREENWOOD A., LENNON E. (2003). *Gestures, grunts and words: The transition to communicative competence*, Tampa, FL:SRCD.
- McNEILL D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago, University of Chicago Press.
- McNEILL D. (2000). Catchments and contexts: non-modular factors in speech and gesture production, In *Language and Gesture*, Edited by McNeill, Cambridge University Press, 313-328.
- McNEILL D., DUNCAN S.D. (2000). Growth points in thinking-for-speaking, In *Language and Gesture*, Edited by McNeill, Cambridge University Press, 141-161.
- MEGUERDITCHIAN A., VAUCLAIR J. (2006). Baboons communicate with their right hand, *Behavioural Brain Research*, 171, 170-174.
- MELIS A.P., HARE B., TOMASELLO M. (2006). Chimpanzees recruit the best collaborators, *Science*, 311, 1297-1299.
- MINEKA S, DAVIDSON M., COOK M., KEIR R. (1984). Observational conditioning of snake fear in rhesus monkey, *Journal of Abnormal Psychology*, 93, 355-372.
- MORFORD M., GOLDIN-MEADOW S. (1992) Comprehension and production of gesture in combination with speech in one-word speakers, *Journal of Child Language*, 19 (3), 559-580.
- MORREL-SAMUELS P., KRAUSS R.M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 3, 615-622
- MUNHALL K., JONES J. (1998). Articulatory evidence for syllabic structure, *Behavioral and Brain Sciences*, 21: 4, 524-525.
- MURPHY K., CORFIELD D.R., GUZ A., FINK G.R., WISE R.J., HARRISON J., ADAMS L. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, 83, 1438-1447.
- NAGEL T. (1974). What is it like to be a bat?, *Philosophical Review*, 83, 435-450.
- NAZZI, T., GOPNIK A, KARMILOFF-SMITH, A. (2005). Asynchrony in the cognitive and lexical development of young children with Williams syndrome. *Journal of Child Language*. 32(2), 427-38.

- NAZZI, T. et KARMILOFF-SMITH, A. (2002). Early categorization abilities in young children with Williams syndrome. *Neuroreport*, 13, 1259-1262
- NEVILLE H.J, BAVELIER D., CORINA D., RAUSCHECKER J., KARNI A., LALWANI A., BRAUN A., CLARK V., JEZZARD P., TURNER R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience, *Proceedings of the National Academy of Sciences of the United States of America*, 9, 922-929.
- OLLER D.K. (1978). Infant vocalization and the development of speech. *Allied Health and Behavioral Science*, 1, 523-549.
- OLLER D.K. (2000). *The emergence of the speech capacity*. London: Erlbaum.
- ÖZÇALIŞKAN S., GOLDIN-MEADOW S. (2005). Gesture is at the cutting edge of early language development, *Cognition*, 96, B101-B113.
- PERNER J., RUFFMAN T. (2005). Infant's insight into the mind: How deep?, *Science*, 308, pp. 214-216.
- PERRETT D.I. (1999). A cellular basis for reading minds from faces and actions. In Hauser M.D. & Konishi, M. (Eds), *The design of animal communication*. Cambridge MA: MIT Press.
- PETERS A.M., MENN L. (1993). False start and filler-syllables: Ways to learn grammatical morphemes, *Language*, 69, 742-777.
- PETITTO L.A., HOLOWKA S., SERGIO L.E., OSTRY D. (2001). Language rhythms in baby hand movement, *Nature*, 413, 35-36.
- PETITTO L.A., HOLOWKA S., SERGIO L.E., LEVY B., OSTRY D.J. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble silently on the hands, *Cognition*, 93, 43-73.
- PHILLIPS M.L., YOUNG A.W., SENIOR C., BRAMMER M., ANDREWS C., CALDER A.J., BULLMORE E.T., PERETT D.I., ROWLAND D., WILLIAMS S.C.R., GRAY J.A., DAVID A.S. (1997). A specific neural substrate for perceiving facial expressions of disgust, *Nature*, 389, 495-498.
- PHILLIPS W., BARON-COHEN S., RUTTER M. (1992). The role of eye-contact in the detection of goals: Evidence from normal toddlers, and children with autism and mental handicap, *Development and Psychopathology*, 4, 375-383.
- PIATTELLI-PALMARINI M. (1979). *Théories du langage, théories de l'apprentissage*. Le débat entre Jean Piaget et Noam Chomsky organisé et recueilli par Massimo Piattelli-Palmarini, Centre Royaumont pour une Science de l'Homme, Editions du Seuil, 533 p.

- PINKER S., JACKENDOFF R. (2005). The faculty of language: What's special about it?, *Cognition*, 95, 201-236.
- PIZZUTO E., CAPOBIANCO M., DEVESCOVI A. (2005). Gestural-vocal deixis and representational skills in early language development, Vocalize to Localize II, *Interaction Studies*, 6, 2, 223–252
- PLACE U. T. (2000). The role of the hand in the evolution of language, *Psychology*, 11, 7.
- POECK K., De BLESER R., GRAF VON KEYSERLINGK D. (1984). Neurolinguistic status and localization of lesion in aphasic patients with exclusively consonant-vowel recurring utterances. *Brain*, 107, pp. 199-217.
- POVINELLI D.J., VONK J. (2003). Chimpanzee minds: suspiciously human?, *Trends in Cognitive Sciences*, 7, 4, 157-160.
- PRABLANC C., PELISSON D. (1990). Gaze saccade orienting and hand pointing are locked to their goal by quick internal loops, In *Attention and Performance XIII, Motor Representation and Control*, M. Jeannerod (Ed.), Lawrence Erlbaum Publishers, 652-676.
- PYLYSHYN Z. (2000). Situating vision in the world, *Trends in Cognitive Sciences*, 4, 5, 197-207.
- RADIN P., *The Trickster : A study in American Indian Mythology* (New York: Schocken books, 1956) 13-14. The telling by Felix White, Sr. is almost identical. See Kathleen Ann Danker, *The Winnebago Narratives of Felix White, Sr.: Style, Structure and Function*, (Ph.D. Thesis, University of Nebraska, Lincoln, May, 1985, 157-161).
- RAO C.S., RAINER G., MILLER E.K. (1997). Integration of What and Where in the Primate Prefrontal Cortex, *Science*, 276, 821-824.
- REID V.M., BELSKY J., JOHNSON M.H. (2005). Infant perception of human action: Toward a developmental cognitive neuroscience of individual differences, *Cognition, Brain, Behavior* (in press).
- REY A. (2002). Les atomes de la lecture, *Fondation Fyssen, Annales*, 17, 83-91.
- RIEMSDIJK H. (1999). *Clitics in the languages of Europe. Empirical approaches to language typology*, In Henk van Riemsdijk (Ed.), Berlin: Mouton de Gruyter.
- RIZZOLATTI G., ARBIB M.A. (1998). Language within our grasp, *Trends in Neurosciences*, 21, 5, 188-194.
- RONDAL J.A. (1997). *L'évaluation du langage*. Hayen, Belgique, Mardaga.

- ROSSATO S., BADIN P., BOUAOUNI F. (2003). Velar movements in French: An articulatory and acoustical analysis of coarticulation, *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, 3-9 August, Barcelona, M.J Solé, D. Recasens, J. Romero (Eds), 3141-3144.
- ROUSSET I. (2004). *Structures syllabiques et lexicales des langues du monde. Données, typologies, tendances universelles et contraintes substantielles*, Thèse de Doctorat en Sciences du Langage, Université Grenoble 3, Institut de la Communication Parlée.
- ROY A., PAULIGNAN Y., FARNÈ A., JOUFFRAIS C., BOUSSAOU D. (2000). Hand Kinematics during reaching and grasping in the macaque monkey, *Behavioural Brain Research*, 117, 75-82.
- SANDLER W. (*in press*). The syllable in sign language: Considering the other natural language modality. In Barbara L. Davis and Krisztina Zadjó (Eds.), *Frame/Content Theory and Beyond. A Festschrift for Peter MacNeilage*, London: Erlbaum.
- SANKOFF G., BROWN P. (1976). The origins of syntax in discourse: a case study of Tok Pisin relatives, *Language*, 52, 631-666.
- SANKOFF G. (1979). The genesis of a language. In Kenneth C. Hill (Ed.), *The genesis of language*, First Michigan Colloquium (Ann Arbor: Karoma), 23-47.
- SATO M., BACIU M., LÖEVENBRUCK H., SCHWARTZ J.-L., CATHIARD M.-A., SEGEBARTH C. ABRY C. (2004). Multistable representation of speech forms: A functional MRI study of verbal transformations. *NeuroImage*, 23, 1143-1151.
- SATO M., SCHWARTZ J.-L., CATHIARD M.-A., ABRY C., LÖEVENBRUCK H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68 (3), 458-474.
- SCASSELLATI B. (2000). Investigating models of social development using a humanoid robot, in *Biorobotics*, Webb, B. et Consi, T. (Eds), Cambridge, MA: MIT Press.
- SCASSELLATI B. (2001). *Foundations for a theory of mind for a humanoid robot*, Doctor of philosophy in Electrical Engineering and Computer Science at the MIT.
- SCHAAL S., STERNAD D., OSU R., KAWATO M. (2004). Rhythmic arm movements are not discrete, *Nature Neuroscience*, 7, 10, 1136-1143.
- SCHAFFER H.R. (1984). *The child's entry into a social world*. New York and London: Academic Press.

- SHIMOJO S., SCHEIER C., NIJHAWAN R., SHAMS L., KAMITANI Y., WATANABE K. (2001). Beyond Perceptual Modality: Auditory Effects on Visual Perception, *Acoustic Science and Technics*, 22, 2, 61-67.
- SIERATZKI J.S., WOLL B. (2002). Toddling into language: precocious language development in motor-impaired children with spinal muscular atrophy, *Lingua*, 112:6, 423-433.
- SHIOMI M, KANDA T., MIRALLES N., MIYASHITA T., FASEL I., MOVELLAN J., ISHIGURO H. (2006). Development of face-to-face communication function for a humanoid robot, *Systems and Computers in Japan*, 37, 14 , 1 -14 (Published Online: 18 Oct. 2006)
- SIMON O., MANGIN J.F., COHEN L., LE BIHAN D., DEHAENE S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe, *Neurosciences*, volume 33, 475-487.
- SIRIGU A., COHEN L., ZALLA T., PRADAT-DIEHL P., VAN EECKOUT P., GRAFMAN J., AGID Y. (1998). Distinct frontal regions for processing sentence syntax and story grammar, *Cortex*, 34, 771-778.
- SLOBIN D.I. (1987). Thinking for speaking. *Proceedings of the Thirteenth Annual Meeting of Berkeley Linguistics Society*, 435-444.
- SLOBIN D.I. (1996). From “thought and language” to “thinking to speaking”. In J.J. Gumperz & S.C. Levinson (Eds), *Rethinking linguistic relativity*, Cambridge: Cambridge University Press, 70-96.
- SLOCOMBE K.E., ZUBERBÜHLER K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role, *Journal of Comparative Psychology*, 119 (1), 67-77.
- SNEDEKER J., GLEITMAN L. (2004). Why it is hard to label our concepts. In Hall & Waxman (Eds.), *Weaving a lexicon*, Cambridge, MA: MIT Press, 255-293.
- SOMMERVILLE A., WOODWARD A.L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy, *Cognition*, 95, 1-30.
- SOMMERVILLE A., WOODWARD A.L., NEEDHAM A. (2005). Action experience alters 3-months-old infants' perception of others' action, *Cognition*, in press, 1-11.
- SOROKIN V.N., GAY T., EWAN W.G. (1980). Some biomechanical correlates of the jaw movements, *Journal of the Acoustical Society of America*, Suppl. 1, 68, S32.
- SPELKE E., PHILIPPS A.T., WOODWARD A.L. (1995). Infants' knowledge of object motion and human action. In Sperber D., Premack D., Premack A.J.

(Eds), *Causal cognition, a multidisciplinary debate*, Oxford: Oxford University Press.

- STARK R.E. (1979). Prespeech segmental feature development. In P. Fletcher and M. Garman (Eds.). *Language Acquisition*, 15-32. New York: Cambridge University Press.
- SUSSMAN H.M., MINIFIE F.D., BUDER E.H., STOEL-GAMMON C., SMITH J. (1996). Consonant-Vowel Interdependencies in Babbling and Early Words: Preliminary Examination of a Locus Equation Approach, *Journal of Speech and Hearing Research*, 39, 424-433.
- SUSSMAN H.M., DUDER C., DALSTON E., CACCIATORE A. (1999). An acoustic analysis of the development of CV coarticulation: a case study, *Journal of Speech, Language, and Hearing Research*, 42, 1080-1096.
- TAGLIALATELA J.P., CANTALUPO C., HOPKINS W.D. (2006). Gesture handedness predicts asymmetry in the chimpanzee inferior frontal gyrus, *NeuroReport*, 17, 9, 923-927.
- THELEN E. (1981). Rhythmical behavior in infancy: an ethological perspective, *Developmental Psychology*, 17, 237-275.
- TOMASELLO M. (1999). *The cultural origins of human cognition*. Cambridge, MA, Harvard University Press.
- TOMASELLO M. (2002). Not waving but speaking, How important were gestures in the evolution of language?, *Nature*, 417, 791-792.
- TOMASELLO M. (2003a). *Constructing a language: A usage-based theory of language acquisition*, Harvard University Press, 408p.
- TOMASELLO M. (2003b). On the different origins of symbols and grammar, In *Language Evolution: The states of the Art*, M.H. Christiansen and S. Kirby (Eds), Oxford University Press, 94-110.
- TOMASELLO M., CALL J., HARE B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent, *Trends in Cognitive Sciences*, 7, 4, 153-156.
- TOMASELLO M., CARPENTER M., CALL J., BEHNE T., MOLL H. (in press). Understanding and sharing intentions: the origins of cultural cognition, *to be published in Behavioral and Brain Sciences*, Cambridge University Press, 1-34.
- TOMASELLO M. (in press). Acquiring linguistic constructions, *Handbook of Child Psychology: Cognitive Development*, R. Siegler and Deanna Kuhn (Eds).
- TOURRETTE C., RECORDONI S., BARBE V., SOARES-BOUCAUD I. (2000). Attention conjointe pré-verbale et théorie de l'esprit à 5 ans : la relation



supposée entre ces deux capacités peut-elle être démontrée ? Etude exploratoire chez des enfants non-autistes. Dans *Autisme : perspectives actuelles*, sous la direction de V. Gerardin-Collet et C. Riboni, L'Harmattan, 61-77.

- TREMOULET P.D., LESLIE A.M., HALL D.G. (2000). Infant individuation and identification of objects, *Cognitive Development*, 15, 499-522.
- UNGERLEIDER L.G., MISHKIN M. (1982). Two cortical visual systems. In *Analysis of visual behavior*, Ingle D.J., Goodale M.A et R.J.W. Mansfield (Eds), The MIT PRESS, 549-586.
- VAN der STELT J. (1976). Une comparaison du rythme de la parole entre des enfants sourds à la naissance et des enfants qui sont devenus sourds, *Prelangage III, Audiophonology*.
- VAN De Walle G.A., CAREY S., PREVOR M. (2000). Bases for object individuation in infancy: Evidence from manual search, *Journal of Cognition and Development*, 1(3), 249-280.
- VAUCLAIR J. (2002). Les chimpanzés et le langage, *Pour la Science*, Hors-Série : La communication animale, janvier-avril, 106-111.
- VAUCLAIR J. (2004). *Développement du jeune enfant : motricité, perception, cognition*. Belin Sup.
- VECERA S.P. (2002). Dissociating 'What' and 'How' in visual form agnosia: a computational investigation, *Neuropsychology*, 40, 187-204.
- VIHMAN M.M., McCUNE L. (1994). When is a word a word ?, *Journal of Child Language*, 21, 517-542.
- VIHMAN M.M., NAKAI S., DE PAOLIS R. (in press). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. To appear in L. Goldstein, D. Whalen et C. Best (eds.), *Papers in Laboratory Phonology 8: Varieties of Phonological Competence*.
- VILAIN A., ABRY C., BADIN P., BROSDA S. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling, *ICPhS99*, San Francisco, 2497-2500.
- VILAIN A., ABRY C., BADIN P. (2000). Coproduction Strategies in French VCVs: Confronting Öhman's Model with Adult and Developmental Articulatory Data, in *Proceedings of the 5<sup>th</sup> Seminar on Speech Production: Models and Data* et *CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, 81-84, Kloster Seeon, Germany.
- VOLTERRA V., CASELLI M.C, CAPIRCI O., PIZZUTO E. (à paraître). Gesture and the emergence and development of language. *To appear in M. Tomasello*

and D. Slobin (Eds), *Elizabeth Bates: a Festschrift*. Mahwah, N.J.: Lawrence Erlbaum Associates.

- VOLTERRA V., CAPIRCI O., CASELLI M.C., VICARI S. (2004). Language in preschool Italian children with Williams and Down syndromes. In S. Bartke & J. Siegmüller, *Williams Syndrome across Languages* (pp. 163-185). Amsterdam: John Benjamins Publishing Company.
- VIODE-BENONY C., LE NORMAND M.-T., GOLSE B. (2002). Etude psychopathologique du langage chez 20 enfants atteints par une Amyotrophie Spinale Infantile de type II âgés de 25 à 47 mois, *La Psychiatrie de l'Enfant*, 65 (1), 133-169.
- VYGOTSKY L.S. (1998). *Imagination and creativity in childhood*. In R.W. Rieber (Ed.), *The collected works of L.S. Vygotsky*, New York, Plenum.
- WERNER H., KAPLAN B. (1963). *Symbol formation*. New-York, Wiley and Sons.
- WILCOX T., BAILLARGEON R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events, *Cognitive Psychology*, 37, 97-155.
- WILKINS D.P. (1999). Spatial deixis in Arrernte speech and gesture: on the analysis of a species of composite signals as used by a central Australian aboriginal group, *Proceedings of the Workshop on Deixis, Demonstration and Deictic Belief at ESSLLI XI*.
- WILKINS D.P. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms), in *Pointing : where language, culture, and cognition meet*, edited by Sotaro Kita. Lawrence Erlbaum, 171–215.
- WILLIAMS J.H.G., WHITEN A., SUDDENDORF T., PERRETT D. (à paraître). Imitation, mirror neurons and autism, To appear in *Neuroscience and Biobehavioral Reviews*.
- WOODWARD A.L. (1998). Infants selectively encode the goal object of an actor's reach, *Cognition*, 69, 1-34.
- WOODWARD A.L., GUAJARDO J.J. (2002). Infants' understanding of the point gesture as an object-directed action, *Cognitive Development*, 17, 1061-1084.
- XU F., CAREY S., QUINT N. (2004). The emergence of kind-based object individuation in infancy, *Cognitive Psychology*, 49, 155-190.
- XU F. (2002). The role of language in acquiring object kind concepts in infancy, *Cognition*, 85, 223-250.

- XU F., CAREY S. (1996). Infants' metaphysics: the case of numerical identity, *Cognitive Psychology*, 30, 111-153.
- ZATORRE R. J., BOUFFARD M., AHAD P., BELIN P. (2002). Where is 'Where' in the human auditory cortex, *Nature Neuroscience*, volume 5, 9, 905-909.
- ZIEGLER W., KILIAN B., DEGER K. (1997). The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage, *Neuropsychologia*, 35, 9, pp. 1197-1208
- ZUBERBÜHLER, K., NOË R., SEYFARTH R.M. (1997). Diana monkey long-distance calls: Messages for conspecifics and predators, *Animal Behaviour*, 53, 589-604.

**LISTE DES FIGURES**

**FIGURE B.1 : Un cadre pour deux cadres.** A environ 1 an, le cadre de la parole va être intégré dans le cadre du signe : une ou deux syllabes dans un gabarit fourni par le pied pour les premiers mots. Pour le cadre de la parole, après l'émergence de la « syllabe » du babillage canonique, il reste deux étapes : le contrôle de la closance pour la « consonne », et la coarticulation (coproduction) pour le contrôle postural de la « voyelle » à l'intérieur de la « consonne ». Pour le cadre du signe, trois flux cérébraux sont en maturation : la voie de détection des événements occipito-pariétale (*When*), la voie dorsale (*Where*) et la voie ventrale (*What*). Leurs produits sont l'objectivité (*Objecthood*) et l'agentivité (*Agentivity*, système *Who*), alors que le système *How* ventro-pariétal fournit l'affordance de la forme, avant la couleur de l'objet via le système *What*. Parmi les « réponses » correspondantes (*Then/There/That*) à ces *Wh*-systèmes, la plus pertinente pour notre posture déictique est la « *That-Path* » fronto-pariétale (mise en évidence par Løevenbruck et al., 2005). Le mécanisme d'attention-intention partagée (SAM-SIM) se développe plus tard que la détection de direction de l'œil (EDD). Avec des cycles de babillage à 3Hz, la prédiction de ce cadre est un ratio Babillage/Pointage de 2:1. (Abry, Ducey-Kaufmann, Vilain, à paraître). ..... 21

**FIGURE I.1 :** Déplacements verticaux des lèvres inférieure et supérieure, au cours de la production répétitive de [bababa] par un bébé de 8 mois et un adulte (Munhall et Jones, 1998 ; p. 525) ..... 37

**FIGURE I.2 :** Distributions de la hauteur du velum pour les voyelles orales, nasales et les consonnes orales et nasales (Rossato et al., 2003, p. 3142) ..... 41

**FIGURE I.3 :** Radiocinématographie d'une séquence [aba] chez un adulte (tracés aux centres des réalisations). On remarquera en (a) qu'à partir d'une posture basse de la langue (dont le dos émerge au-dessus de l'horizontale pointillée, positionnée sur les incisives de la mandibule comme repère), la langue va monter, transportée par la mandibule, laquelle est recrutée par la coordination labio-mandibulaire (flèches en b pour la closion, en c pour la détente). Mais si le dos est plus haut, plus près du plafond de la bouche en (b), il est plus bas par rapport aux incisives qu'en (a) et (c). L'action du muscle hyoglosse a abaissé le corps de la langue (flèche vers le bas en b) pendant que la mandibule montait pour aider à la closion bilabiale. (D'après Vilain, 2000). ..... 45

**FIGURE I.4 :** Extrêmes théoriques des pentes d'équation du locus. Les figures du haut illustrent la représentation de la transition F2 lorsqu'il n'y a aucune coarticulation entre la voyelle et la consonne (le locus, point de convergence des transitions qui n'est pas influencé par le contexte vocalique sur les F2), et la pente zéro qui devrait en résulter. Les figures du bas illustre la coarticulation maximale entre voyelle et consonne avec aucun locus consonantique fixe (celui-ci étant entièrement dépendant du contexte vocalique de F2) et une pente d'équation résultante de 1. (D'après Sussman et al., 1999). ..... 46

**FIGURE I.5 :** Pentes d'équation du locus pour des production contenant [bV] d'une enfant suivie de 7 à 40 mois. Les valeurs des énoncés de babillage (présents jusqu'à 16 mois) sont les carrés noirs, les mots (apparaissant à 12 mois), les carrés blancs, et la norme adulte pour la parole spontanée est donnée par la ligne des petits losanges noirs. (D'après Sussman et al., 1999). ..... 47

**FIGURE II.1 :** Une tête entourée de micros et de caméras (site Medialab du MIT : Projet Oxygen)..... 52

**FIGURE II.2 :** Trois façons différentes par lesquelles le robot peut se représenter son monde (D'après Pylyshyn, 2000). ..... 59

FIGURE II.3 : Les flux multimodaux du système sémantique dorsal <i>Where</i> .....	68
FIGURE II.4 : Modèle pour l'indexation de l'objet (d'après Leslie et al., 1998). Dans le panel A, les objets sont indexés uniquement par la localisation. Dans le panel B, l'indexation de l'objet est réalisée aussi bien par la localisation que par l'information de traits. L'indexation par voie double dans B se développe plus tardivement. ....	69
FIGURE II.5: Expérience de mouvement apparent illustrant la préemption du système <i>Where</i> sur le système <i>What</i> (D'après Carey et Xu, 2000). ....	69
FIGURE II.6 : Les flux multimodaux du système sémantique ventral <i>What</i> . ....	71
FIGURE II.7 : Le système pragmatico-sémantique <i>How</i> . ....	73
FIGURE II.8 : Agnosie visuelle et ataxie optique: (A) explication de la déconnexion ventrale ; (B) explication de l'input épargné (D'après Vecera, 2002). ....	74
FIGURE II.9 : Le système <i>When</i> (étoile MT/V5) à la racine de chaque ségrégation des flux <i>What</i> et <i>Where</i> , pour les informations auditives, visuelles et tactiles. ....	77
FIGURE II.10: Le système <i>When</i> préempte le système <i>Where</i> (D'après l'expérience de Shimojo et al., 2001). Suivi de deux boules, qui partent du haut, (i) se croisent et continuent leurs trajectoires vers le bas; ou (ii) changent de direction après un événement produit à leur rencontre (choc auditif « Poc ! », flash visuel, retour tactile). Par ailleurs, en jouant sur les changements (échanges) de forme (boule qui devient triangle) ou de couleur (boule rouge qui devient verte), on peut constater que le système <i>Where</i> préempte bien le système <i>What</i> (comme dans l'expérience supra de la Figure II.5) : on suit l'objet même si en fin de course on peut constater qu'il a changé de forme et/ou de couleur. ....	78
Que se passe-t-il maintenant si nous comparons ce type de condition répétée à des conditions introduisant un focus sur un élément de la phrase, que ce focus soit intonatif ou syntaxique ? Intonatif : « Buy BOBBY a poppy » ; syntaxique « It's to Bobby that you (have to) buy a poppy ». C'est précisément ce qu'ont testé Løevenbruck et al. (2005) dans leur étude IRMf sur le français. ....	80
FIGURE II.11: a) Condition de deixis intonative « MADELEINE m'amena » ; b) condition de deixis syntaxique « C'est Mad'leine qui m'am'na » ; (c) condition de deixis syntaxique + intonative « C'est MAD'LEINE qui m'am'na » (Løevenbruck et al., 2005).....	82
FIGURE II.12 : Le système déictico-syntaxique <i>That</i> . ....	83
FIGURE II.13 : Le système <i>Then</i> localisé dans le cortex dorsolatéral préfrontal.....	85
FIGURE II.14 : Les 7 grands systèmes cérébraux pragmatico-sémantiques.....	85
FIGURE II.14 : Le système EDD (D'après Emery, 2000). ....	90
FIGURE II.15 : Attention conjointe vs. attention partagée (D'après Emery, 2000). ....	94
FIGURE II.16 : Le mécanisme de la théorie de l'esprit (D'après Emery, 2000). ....	97
FIGURE II.17 : Vue d'ensemble du modèle de Baron-Cohen du développement de l'attention conjointe et de la théorie de l'esprit (D'après Scassellati, 2000). ....	98
FIGURE II.18 : Une décomposition basée sur la tâche en quatre parties des bases de l'attention conjointe (Butterworth 1991, d'après Scassellati, 2000). ....	106
FIGURE II.19 : Progression développementale du suivi du regard (adapté à partir de Butterworth 1991, d'après Scassellati, 2000). ....	112
FIGURE II.20 : Tableau résumant les différents types de traitement chez différents êtres vivants (✓ indique une évidence positive, X indique aucune évidence, ? indique que cela n'a pas été testé ou que l'évidence est controversée) (D'après Emery, 2000). ....	114
FIGURE II.21 : Vue de l'hémisphère gauche du cerveau du macaque, et notamment du couple fronto-pariétal F5-AIP pour la préhension ; avec un aperçu en miroir de la partie supérieure mésiale (D'après Arbib, 2005). ....	119
FIGURE II.22 : L'action intentionnelle humaine. Le but est une boîte ouverte. L'acteur choisit un moyen (plan), représenté par les mains agissantes, qui forme une intention.	

L'action qui en résulte cause un résultat, qui amène à une réaction émotive de l'acteur (d'après Tomasello et al., 2004)..... 124

FIGURE II.23 : Conception de chacun des participants dans une activité collaborative dans laquelle un but partagé et une intention conjointe, avec rôles complémentaires, sont formés. (D'après Tomasello et al., 2004)..... 128

FIGURE II.24 : Représentations somatotopiques des mouvements de bouche, œil, doigts dans le Sulcus Temporal Supérieur. (D'après Dubeau et al., 2002)..... 133

FIGURE II.33 : Le circuit minimal pour la compréhension et l'imitation de l'action. (D'après Dubeau et al., 2002) ..... 134

FIGURE II.26 : Les structures cérébrales fondamentales pour la perception et la cognition sociales chez l'homme ; le cortex orbito-frontal (en gris foncé : on y lit difficilement OFC) est en dessous du cortex préfrontal (PFC). (D'après Allison et al., 2000)..... 139

FIGURE II.26 : La présence d'un visage orienté vers l'apparition d'une cible (petit cercle) diminue irrésistiblement le temps de réaction de 20ms en moyenne, par rapport à une cible dans une autre direction, même si l'on instruit le sujet qu'il doit négliger la direction du visage. (D'après Langton et al., 2000) ..... 140

FIGURE II.27 : figure de gauche, exemple d'habituation et d'événements-tests pour la condition main seule ; figure de droite, exemple d'habituation et d'événements-tests pour la condition main et visage (d'après Woodward et Guajardo, 2002). ..... 148

FIGURE II.28 : Nombre d'occurrences relevées pour les extensions de l'index (carrés), les pointers de l'index (triangles) et les mouvements d'atteinte (ronds) pour 8 enfants japonais suivis longitudinalement de 3 à 16 mois. Nous avons calculé les occurrences additionnées (diamants) des extensions de l'index et des pointers de l'index, pour montrer que les gestes du pointage prennent le relais en continuité des gestes d'extension de l'index, lorsque ceux-ci sont peu à peu remplacés par le pointer (D'après Masataka, 2003). ..... 153

FIGURE II.29 : 4 étapes ontogénétiques dans la représentation lexicale des humains, et la représentation lexicale chez des chimpanzés entraînés (d'après Levelt, 1998). Nous avons ajouté à ce schéma l'élément-clé, qui selon nous permet de lier ces deux systèmes, le système conceptuel et le système syllabique, par les flèches partant du système de « pointing » ou pointer de l'index. .... 162

FIGURE II.30 : Activations cérébrales cartographiées sur l'hémisphère gauche du cerveau pour les tâches : attentionnelle, de préparation de pointer de l'œil (saccade) et de pointer du doigt. Le décours des réponses BOLD (soit la récupération métabolique de la consommation de glucose par les neurones) est donné en-dessus pour les régions du FEF (Left Frontal Eye Field ; l'étoile en A. signale l'absence d'activation au-dessus du seuil choisi) et du sulcus intrapariétal (Left anterior Intra-Parietal Sulcus, actif dans les 3 cas, et tout particulièrement pour le pointer) (D'après Astafiev et al., 2003). ..... 165

FIGURE II.31: Organisation des cris d'alarmes des suricates donnés dans différents contextes de prédateurs selon leurs dimensions établies par une analyse par fonctions discriminantes (DFA) des propriétés acoustiques des cris. T désigne les cris d'alarme destinés aux prédateurs terrestres, A désigne les prédateurs venant des airs et R désigne les cris de recrutement, avec respectivement l, m et h pour les cris de faible, moyenne et grande urgence.(D'après Manser et al., 2002) ..... 173

FIGURE II.32 : le cadre de la parole constitue la phonologie (vocaliser) et le cadre du signe constitue la sémantique (montrer / interroger) dans le cours du développement..... 177

FIGURE II.33 : Les trois phases d'apprentissage du langage. En grisé, les composantes les plus activées en fonction de la phase d'acquisition. (D'après Hirsh-Pasek et Golinkoff, 1999)..... 185

FIGURE II.34 : La deixis syntaxique et la question syntaxique sont possibles sur le nom. Lors d'une extraction syntaxique, la deixis peut être grammaticalisée, mais il n'est pas possible d'extraire syntaxiquement le verbe. Le système intonatif seul permet la deixis sur le verbe (en pointillé sur notre figure), tout comme sur le nom bien entendu. .... 191

FIGURE II.35 : Représentation schématique des facteurs externes et internes à l'organisme reliés à la faculté de langage. FLB inclut le système sensori-moteur, le système conceptuel et intentionnel et d'autres systèmes possibles ; FLN inclut les computations grammaticales noyaux que les auteurs suggèrent limitées à la récursivité (D'après Hauser et al., 2002). ..... 197

FIGURE II.36 : Récursivité dans le groupement visuel (D'après Jackendoff et Pinker, 2005) ..... 200

FIGURE II.37 : Le cube de Necker. Sur la figure de droite, nous avons volontairement biaisé (bloqué) la bistabilité du cube afin de montrer l'un des deux états perceptifs produits par cette figure (le plus courant en perspective post-Giotto), laquelle est en réalité paradoxale (à gauche)..... 204

FIGURE II.38 : La figure présente l'âge auquel chaque enfant a produit ses premières combinaisons à 2 mots en fonction de l'âge auquel l'enfant commençait à produire des combinaisons dans lesquelles geste et parole véhiculaient une information différente (D'après Goldin-Meadow et Butcher, 2003)..... 216

FIGURE II.39 : La figure présente l'âge auquel chaque enfant a produit ses premières combinaisons à 2 mots en fonction de l'âge auquel l'enfant commençait à produire des combinaisons dans lesquelles geste et parole véhiculaient la même information (D'après Goldin-Meadow et Butcher, 2003)..... 216

FIGURE II.40 : a) La figure de gauche montre l'absence de corrélation entre l'âge de début des combinaisons geste + mot complémentaires et l'âge du début des combinaisons à deux mots ; b) La figure de droite présente la corrélation entre l'âge de début des combinaisons geste + mot supplémentaires et l'âge du début des combinaisons à deux mots (D'après Iverson et Goldin-Meadow, 2005). ..... 218

FIGURE II.41 : Nombre d'items lexicaux (en y) dans la parole seule et le geste seul produits par les enfants voyants en fonction de l'âge (en x). (D'après Iverson et al., 2000)..... 232

FIGURE II.42 : Niveaux du seuil d'activation et de l'activation d'entraînement dans le système oro-manuel lors des deux premières années de vie chez l'enfant (D'après Iverson et Thelen, 1999)..... 236

FIGURE II.43 : Nombre de pointers réalisés au total avec la main droite et la main gauche pour l'ensemble des 6 enfants français du corpus (cf. 3<sup>ème</sup> partie). ..... 245

FIGURE II.44 : Patron temporel de la coordination entre son, lèvres, formation de la configuration des doigts de la main et placement de la main, pour la production du Langage Parlé Complété français (les valeurs en italiques sont celles des séquences avec consonnes). (D'après Attina et al., 2004, p. 208) ..... 273

FIGURE II.45 : Comparaison du geste et de la parole affiliée dans le temps en néerlandais et arrernte. (D'après De Ruiter et Wilkins, 1998) ..... 275

FIGURE III.1: Sous chaque production, soit pour chaque pied (foot), on peut mesurer la même longueur de barre, démontrant une tendance à l'isochronie remarquable (D'après Lindblom, 1991)..... 292

FIGURE III.2: Distributions des fréquences cumulées (durée-rang) pour les 4153 syllabes (courbe de tirets-points suivant les données brutes) et 276 pointers (points noirs, avec un fitting approximé par une courbe gaussienne, la plus à droite) produits par 6 enfants français entre 6 et 18 mois. Comme simple indication pour comparaison avec les adultes : les cercles fins correspondent aux 38 valeurs moyennes de temps de mouvement (MT) pour les données de pointer (deixis néerlandaise) de Levelt et al.

(1982) ; et les cercles épais (les plus à gauche) correspondent aux 18 valeurs moyennes de durée de syllabes (analyse d'un passage continu en anglais) mesurées par Campbell (1992). ..... 300

FIGURE III.3: Distributions normalisées des détentes de pointer pour Célia, Anatole et Jules. Leurs données sont plutôt proches de la courbe de distribution générale obtenue pour les 6 enfants (cf. Figure III.2). ..... 301

FIGURE III.4: Distributions normalisées des détentes de pointer pour Lise, Nicolas et Tom. Leurs données sont plutôt plus rapides (Tom, à gauche) ou plus lentes (Lise et Nicolas, à droite) que la tendance générale obtenue à partir des 6 enfants (cf. Figure III.2)..... 302

FIGURE III.5: Evolution des durées de détentes de pointer (diamants) pour les 6 enfants français entre 6 et 18 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis deux premiers points isolés (aucun à 6 mois, et 2 diamants à 7-8 mois), le pointer débute véritablement à 9 mois, et décroît en durée –après une période entre 10-14 mois où il est à environ 800 ms – vers 600 ms à 15 mois. .... 302

FIGURE III.6 : Evolution de la durée (ms) pour les 12 pointers d'Anatole, classés par rang d'apparition. .... 303

FIGURE III.7 : Evolution des durées des détentes de pointer (diamants) pour Anatole entre 7 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis un premier point isolé (1 diamant à 7), le pointer débute véritablement à 12 mois, et décroît en durée –après une période entre 12 et 14 mois où il est d'environ 950 ms– vers 530 ms à 15 mois. .... 304

FIGURE III.8: Evolution de la durée (ms) pour les 31 pointers de Célia, classés par rang d'apparition. .... 305

FIGURE III.9 : Evolution des durées de détentes de pointer (diamants) pour Célia entre 9 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Deux pointers commencent à 9 mois ; on note une majorité de productions entre 10 mois et 12 mois. .... 305

FIGURE III.10: Evolution de la durée (ms) pour les 75 pointers de Jules, classés par rang d'apparition. .... 307

FIGURE III.11 : Evolution des durées des détentes de pointer (diamants) pour Jules entre 10 et 16 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Les occurrences sont assez concentrées dans cette période d'observation. On constate sur la figure une tendance à la baisse des durées des pointers à 14 mois. .... 307

FIGURE III.12 : Evolution de la durée (ms) pour les 21 pointers de Nicolas, classés par rang d'apparition. .... 308

FIGURE III.13 : Evolution des durées des détentes de pointer (diamants) pour Nicolas entre 10 et 16 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis deux premiers points isolés (1 diamant à 10 et 1 diamant à 12 mois), les pointers débutent à 13 mois, la majorité étant produite à 16 mois. .... 308

FIGURE III.14: Evolution de la durée (ms) pour les 37 pointers de Lise, classés par rang d'apparition. .... 309

FIGURE III.15 : Evolution des durées de détentes de pointer (diamants) pour Lise entre 8 et 15 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis un premier point isolé (1 diamant à 8 mois), les pointers débutent véritablement à 9 mois, avec une explosion notable à 10 mois. La durée décroît fortement de 8 à 10 mois, pour se stabiliser autour de 900ms. .... 310



FIGURE III.16 : Evolution de la durée (ms) pour les 100 pointers de Tom, classés par rang d'apparition. .... 311

FIGURE III.17 : Evolution des durées des détentes de pointer (diamants) pour Tom entre 10 et 17 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis quelques premiers points isolés (1 diamant à 10 mois et 2 diamants à 11 mois), les pointers débutent véritablement à 13 mois, avec une explosion notable à ce moment-là. La durée stable jusqu'à 14 mois (autour de 800ms) décroît par la suite de 15 à 17 mois autour de 600ms, soit presque 200ms de gagnées..... 311

FIGURE III.18 : Correspondance entre fréquences du babillage mesurées manuellement et calculées automatiquement. L'abscisse représente les fréquences mesurées à la main à partir de la forme de l'onde ; l'ordonnée les fréquences calculées (cf. texte) par les méthodes d'autocorrélation (AC : cercles) et de Fourier (DFT : triangles). L'ellipse ajoutée par nos soins montre la concentration des données autour de 3 Hz (D'après Bickley et al., 1986). .... 319

FIGURE III.19 : Distribution des fréquences du mouvement de la main dans l'activité rythmique de bébés entendants exposés à la langue des signes (ligne pleine) et pour les bébés exposés à la parole (ligne pointillée), tous âges confondus de 6 à 12 mois. Pour chaque groupe, 200 segments de mouvements ont été aléatoirement sélectionnés. Seuls les bébés exposés aux signes ont une distribution bimodale des fréquences de mouvements : le premier mode (pic de gauche) tombe autour de 1.5 Hz et le second mode (pic de droite) tombe autour de 2.5 Hz. Par contraste, les fréquences de mouvements de main des bébés exposés à la parole sont unimodales, autour de 3 Hz. Une comparaison des deux groupes a révélé que le second mode des fréquences de mouvement produites par les bébés exposés aux signes, vers 2.5 Hz, n'était pas significativement différent du 3Hz des bébés exposés à la parole (D'après Petitto et al., 2001)..... 322

FIGURE III.22 : Evolution des syllabes finales (FS) et des syllabes non-finales (NFS) chez un sujet entre 8 et 24 mois (données replottées d'après Konopczynski 1998) (N.B. : le mois 11 est manquant dans ses données) ..... 326

FIGURE III.23 : Evolution de la durée moyenne des cycles des premières (S1 : diamants) et secondes (S2 : carrés) syllabes dans les performances de Célia entre 6 à 16 mois. Notons l'augmentation dans S2 autour de l'émergence de ses premiers mots à 12 mois. .... 327

FIGURE III.24: Evolution du ratio S2/S1 (seconde syllabe sur la première) pour Célia. Ce ratio met en évidence le changement métrique à la période des premiers mots vers un an (cf.Figure III.23)..... 328

FIGURE III.25: Distribution des fréquences cumulées (durée-rang). Les données brutes correspondent à 4153 syllabes produites par les 6 enfants français entre 6 et 18 mois. 329

FIGURE III.26: Distributions normalisées des cycles de syllabes pour Anatole, Nicolas et Tom. Leurs données sont plutôt proches de la distribution générale observée pour les 6 enfants (ligne pleine)..... 330

FIGURE III.27: Distributions normalisées des cycles de syllabes pour Célia, Lise et Jules. Ils divergent dans leurs plus hautes valeurs de la distribution générale observée pour les 6 enfants (ligne pleine)..... 330

FIGURE III.28 : Evolution de la durée moyenne des cycles syllabiques par mois chez Anatole. .... 332

FIGURE III.29 : Evolution de la durée moyenne des cycles syllabiques par mois chez Célia. .... 333

FIGURE III.30 : Evolution de la durée moyenne des cycles syllabiques par mois chez Jules. .... 334

FIGURE III.31: Evolution de la durée moyenne des cycles syllabiques par mois chez Nicolas. .... 335

FIGURE III.32 : Evolution de la durée moyenne des cycles syllabiques par mois chez Lise. .... 336

FIGURE III.33 : Evolution de la durée moyenne des cycles syllabiques par mois chez Tom. .... 337

FIGURE III.34: Distributions cumulées brutes d'Anatole entre 7 et 15 mois pour ses 210 syllabes (ligne pleine) et ses 12 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées. .... 338

FIGURE III.35 : Distributions cumulées brutes de Célia entre 9 et 15 mois pour ses 1539 syllabes (ligne pleine) et ses 31 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées. .... 339

FIGURE III.36 : Distributions cumulées brutes de Jules entre 10 et 16 mois pour ses 1087 syllabes (ligne pleine) et ses 75 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées. .... 340

FIGURE III.37: Distributions cumulées brutes de Nicolas entre 10 et 16 mois pour ses 235 syllabes (ligne pleine) et ses 21 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées. .... 341

FIGURE III.38 : Distributions cumulées brutes de Lise entre 8 et 15 mois pour ses 441 syllabes (ligne pleine) et ses 37 pointers (étoiles). Les deux courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observées. .... 342

FIGURE III.39: Distributions cumulées brutes de Tom entre 10 et 17 mois pour ses 641 syllabes (ligne pleine) et ses 100 pointers (étoiles). 2 courbes théoriquement prédites sont obtenues en doublant (petits pointillés) et en triplant (tiretés) la distribution des syllabes observée. .... 342

FIGURE III.40 : Les deux distributions cumulées brutes des 6 enfants français entre 6 et 18 mois, pour leur 276 pointers (étoiles) et leurs 4153 syllabes (distribution la plus à gauche). Les deux courbes théoriquement prédites, obtenues en doublant (petits pointillés) et triplant (tiretés) les valeurs de la distribution des syllabes observées, suggèrent que, dans l'ensemble, la durée du *stroke* ou détente de pointer peut être prédite à partir du cycle de babillage observé, comme un cadre (*frame*) ou gabarit (*template*) pour 2 syllabes. .... 344

FIGURE III.41: Histogrammes représentant le nombre de syllabes par énoncé répétitif (2 syllabes ou plus) pour chaque enfant français du corpus entre 6 et 18 mois : a) Jules, b) Anatole, c) Nicolas, d) Célia, e) Tom et f) Lise. A la lecture de ces 6 histogrammes, on voit nettement la tendance vers un plus grand nombre d'énoncés à 2 syllabes (excepté pour Jules qui montre une remontée de ses données sur les productions trisyllabiques). .... 346

FIGURE III.42 : Histogramme représentant le nombre de syllabes par énoncé répétitif (2 syllabes ou plus) pour les 6 enfants français entre 6 et 18 mois. On remarque un ratio d'environ 3.5 en faveur des productions à 2 syllabes par rapport aux productions à 3 syllabes. .... 347

FIGURE III.43 : Un pic à 2 syllabes apparaît dans la distribution (en %) des unités lexicales rangées par nombre de syllabes dans le mot, pour 16 lexiques représentatifs des langues

du monde, typologiquement échantillonnées à partir de la base de données ULSID  
(D'après Rousset, 2004)..... 348

**LISTE DES TABLES**

TABLE III.1 : Nombre de syllabes relevées chez Anatole. Les durées moyennes sont indiquées par mois.....	331
TABLE III.2 : Nombre de syllabes relevées chez Célia. Les durées moyennes sont indiquées par mois. ....	332
TABLE III.3 : Nombre de syllabes relevées chez Jules. Les durées moyennes sont indiquées par mois. ....	333
TABLE III.4 : Nombre de syllabes relevées chez Nicolas. Les durées moyennes sont indiquées par mois.....	334
TABLE III.5 : Nombre de syllabes relevées chez Lise. Les durées moyennes sont indiquées par mois. ....	335
TABLE III.6 : Nombre de syllabes relevées chez Tom. Les durées moyennes sont indiquées par mois. ....	336
TABLE III.7 : Durées moyennes des pointers et des syllabes obtenues pour 6 enfants français suivis par quinzaine entre 6 et 18 mois, avec leurs ratios Pointer/Syllabe. Le nombre d'événements de pointers spontanés avec vocalisation, observables pour tout le corpus est de 276 (de 12 à 100 par enfant). Le nombre de cycles de babillage mesurés par échantillonnage autour de ces événements de pointer est de 4153 au total (de 210 à 1539). Le rapport des moyennes des durées mesurées sur tous les événements (pointers 775.87 : syllabes 355.19), donne un ratio global Pointer/Syllabe de 2.18, soit tout proche de la valeur harmonique prédite de 2:1. On peut donc toujours enchâsser deux cycles de syllabes mandibulaires dans une détente de pointer.....	343