

Les segments du corps en interaction dans la naissance de la parole

Etude longitudinale de 6 enfants français entre 6 et 18 mois

Virginie Ducey-Kaufmann, Christian Abry, Anne Vilain & Claire Lalevée
Institut de la Communication Parlée, UMR CNRS 5009
Université Stendhal, DU, 1180 Av Centrale
BP 25 – 38040 Grenoble Cedex 9
ducey@icp.inpg.fr

Résumé

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 6 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.

Mots-clés : *babillage, syllabe, pointer, pied, premiers mots*

1. Deux contrôles proximaux des segments du corps pour la naissance de la parole : la mandibule et le bras

Le problème général que rencontre l'enfant dans différents domaines de son développement est celui du contrôle des segments du corps, contrôles proximaux dans l'espace proprioceptif (p. ex. celui des joints des segments du bras) et contrôles distaux des effecteurs finaux atteignant des buts extéroceptifs (p. ex. celui visuel et tactile de toucher une cible du doigt). Selon la théorie *Frame/Content* de MacNeilage (1998), il en va de même pour la parole, qui est caractérisée par le contrôle cyclique auditori-moteur de la mandibule, contrôle proximal, qui se manifeste dans le *babillage canonique*, apparaissant autour de six-sept mois, bien avant l'apparition du contrôle indépendant et coordonné des articulateurs distaux, effecteurs finaux portés par la mandibule, essentiellement les lèvres et la langue. En d'autres termes, bien avant que le contenu segmental (*content*), soient les *contacts* « consonantiques » et les *postures* « vocaliques », ne soit maîtrisé (Vilain & Abry, 2000), le contrôle rythmique « syllabique » était déjà né (*frame*). Cette différence entre contrôle proximal et distal a été d'abord illustrée par un enregistrement de Munhall & Jones (1998), qui ont comparé les patrons de mouvement

des lèvres supérieure et inférieure chez un bébé de huit mois babillant [bababa], par rapport à un adulte prononçant [bababa]. Sur l'enregistrement de l'enfant de huit ans, on voit parfaitement que la lèvre supérieure (compliante) se laisse purement et simplement soulever par la poussée de la lèvre inférieure portée par la mandibule. Cette contribution principale de la mandibule a été mise en évidence par Green et al. (2000 ; 2002 tout spécialement : leur figure 4), la part active des lèvres débutant plus tard. Ils ont ainsi apporté les premières données longitudinales en faveur de l'évolution vers un contrôle autonome de la coordination de ces deux articulateurs distaux, les deux lèvres, pour leur mise en contact dans la closure bilabiale [b].

Mais nous pensons que, s'il est nécessaire de replacer le développement de la parole dans le développement des contrôles moteurs — ainsi que l'avait commencé la regrettée Esther Thelen (1981), en montrant que le babillage canonique apparaît dans une période d'activité rythmique des bras et des jambes— cela ne saurait être suffisant. Les coordinations et les couplages reconnus entre systèmes, du type œil-tête-main, dans le développement cognitivo-moteur de l'enfant, qui sont reconnus classiquement comme étant à l'œuvre dans les interactions dyadiques et triadiques, ne doivent pas faire oublier l'interaction majeure entre le contrôle oro-laryngé et le contrôle céphalo-brachial, dans la co-occurrence complémentaire du geste et de la parole (Goldin-Meadow, 2003), et ceci tout particulièrement aux stades cruciaux du babillage canonique et du pointer de l'index. Dans ce sens, contrôle rythmique de la mandibule et détente du bras sont, selon nous, deux contrôles proximaux qui doivent naturellement se coordonner pour le développement de la parole et du langage.

2. Une hypothèse de travail pour trois questions princeps sur le mot, le pied et la syllabe.

Nous devons à Edy Veneziano (d'une brève mais dense conversation avec C. Abry, le 10 décembre 2005 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 prétendrons hardiment proposer du même coup) ?

Notre hypothèse de travail (*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*) vient d'être 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 cet article, 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 & Davis (Davis et al., 2002), le *schème vocal moteur* pour Vihman (McCune & Vihman, 2001), sont des propositions soutenues par des données en faveur du filtrage articulatoire des premiers mots (1). 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 (2). 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*) (3) est prosodiquement contrôlé, accentué, intonné, focalisé, etc. 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 (p. ex. en anglais, *muse, music, musical, musically*). Ainsi, nous proposons que le geste de pointer déictique-sémiotique –un capteur sémantique (*semantic grabber*) pour le célèbre *fast mapping* lexical, et un précurseur syntaxique pour les connecteurs de type *that*– soit considéré comme l'unité de contrôle principale formant le pied, en quelque sorte un « pointer-pied ».

Cette proposition de « babillage-syllabe » et « pointer-pied » 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, il 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 600-700ms. 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 600-700ms.

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 un bébé qui est dans le profil de pointer général, mais qui expose 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.

3. Le mode de la distribution des durées de la détente du pointer

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. Le 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 ! 276 événements de pointer *et* de vocalisation ont pu être mesurés, ce qui correspond approximativement aux 207 cris de conflits collectés à partir de 37 chimpanzés lors d'une enquête de terrain de 5 mois (Slocombe & Zuberbühler, 2005).

La présence de gestes de pointer a été observée de 7,20 mois à 17,18 mois. Les premières occurrences étaient observées 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 (40ms) ; idem pour la fin du temps de mouvement (*Movement Time* ou MT) (4). La figure 1 donne une distribution cumulative brute, obtenue avec MT ou la durée des détentes (*strokes*) en x et le rang de ces valeurs de durée en y (rang normalisé, pour les comparaisons qui vont suivre).

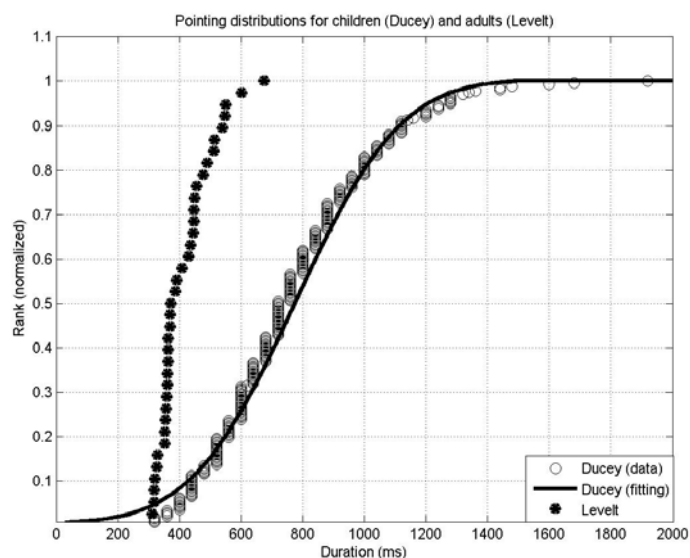


Fig.1 : Courbe des fréquences cumulées (durée-rang) pour les 276 détentés (cercles) des 6 enfants français entre 6 et 18 mois (mesurés par Ducey), approximée par une gaussienne. Les durées moyennes du pointer des adultes avec la main droite (alors qu'ils prononçaient des déictiques néerlandais) mesurées par Levelt et al. (1985), disponibles pour 38 conditions (points étoilés à gauche), sont tracées pour comparaison (les deux courbes sont normalisées en rang).

Les valeurs s'étendent de 280ms à 1920ms (un premier *stroke* de pointer très lent, voir Fig.4). La durée moyenne était de 776ms avec ± 267 ms comme écart-type ; l'approximation gaussienne est réalisée sur les données brutes. Nous avons tracé une comparaison avec les adultes en utilisant les moyennes des 38 expériences faites avec la main droite dans Levelt et al. (1985), pour lesquelles le temps de mouvement était disponible. Dépendant du nombre de choix dans les déictiques (*this/that lamp*, en néerlandais *dit/dat lampje*), de la direction et de la distance des LEDs-cibles (champ gauche/droit ; proche/éloigné), les valeurs moyennes de temps de mouvement étaient rangées de 313ms (un déictique, pour la direction proche droite parmi 2 LEDs, réponse « *on-line* » instantanée) à 675ms (deux déictiques, pour une direction gauche éloignée parmi les 4 LEDs, et une réponse « *offline* », i.e. après que la lumière se soit allumée, le sujet attendait la question de l'expérimentateur « *Which light ?* », avant de commencer). Dans l'ensemble et comme attendu, les pointers spontanés des enfants sont plus lents que ceux des adultes, mettant en évidence un facteur maturationnel. Les observations, pour chaque enfant (normalisées en rang, dès lors que le nombre des observations collectées pour chacun était différent), situées par rapport à la courbe cumulée générale, montrent que trois d'entre eux (Fig.2) sont plus proches de ce comportement général (Anatole : 763ms \pm 255 ; Jules : 769ms \pm 222 ; et Célia : 826ms \pm 261), tandis que les autres (Fig.3) sont soit plus rapides (Tom : 681ms \pm 239) ou plus lents (Nicolas : 926ms \pm 291 ; et Lise : 927ms \pm 313).

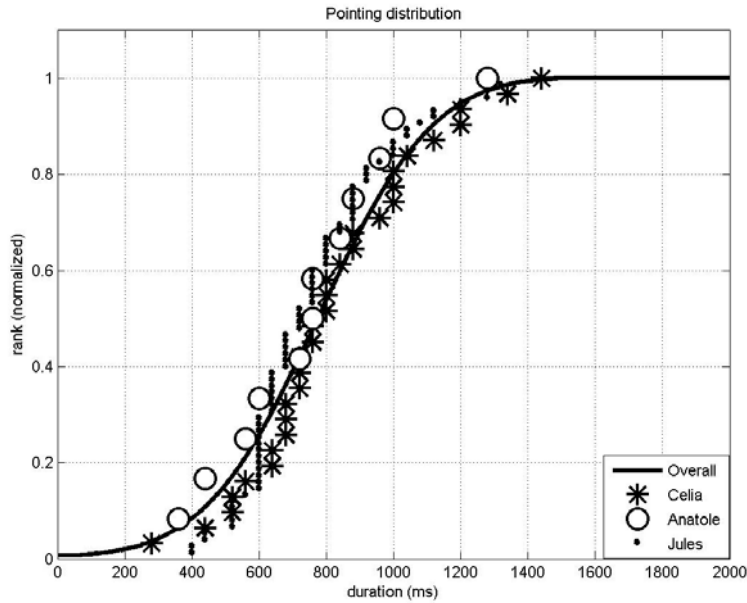


Fig.2 : Distributions normalisées des détentes de pointer pour Célia, Anatole et Jules. Leurs données sont plutôt proches de la courbe de distribution générale obtenue pour les 6 enfants (cf. Fig.1).

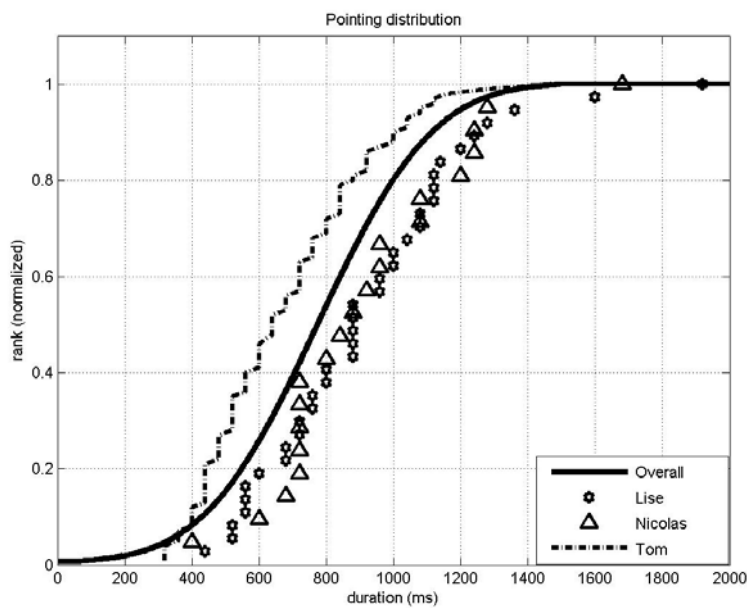


Fig.3 : Distributions normalisées des détentes de pointer pour Lise, Nicolas et Tom. Leurs données sont plutôt plus rapides (Tom) ou plus lentes (Lise et Nicolas) que la tendance générale obtenue à partir des 6 enfants (cf. Fig.1).

De 9 à 17 mois (laissions de côté les deux rares occurrences à 7-8 mois), on observe une tendance décroissante claire dans les moyennes et les distributions, d'environ 1s à environ 600ms à 15-17 mois, via une chute vers 800ms de 10 à 14 mois (Fig.4). A nouveau, la maturation correspond à des gestes plus rapides.

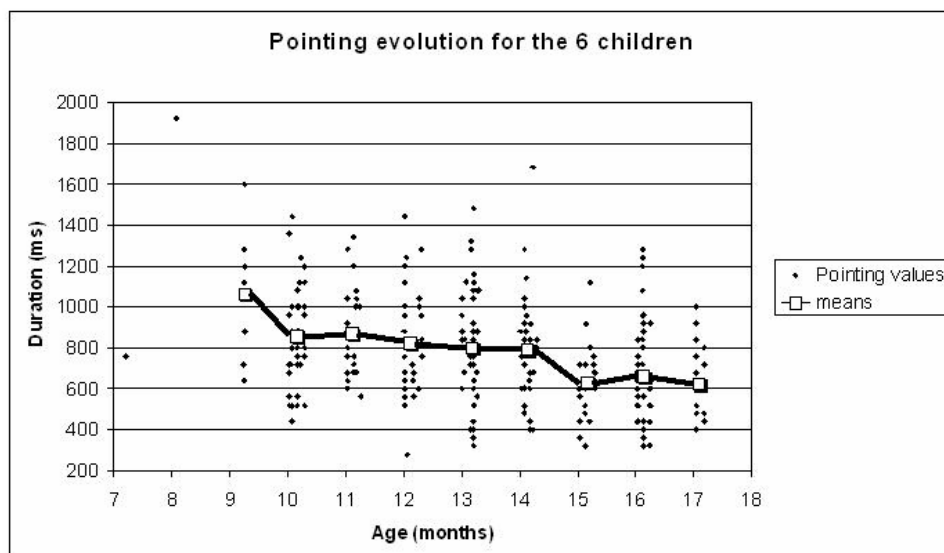


Fig.4 : Evolution des durées de détentes de pointer (diamants) pour les 6 enfants français entre 6 et 18 mois (les données sont tracées avec une échelle décimale <mois,jour>, pour obtenir un groupement visuel par mois). Hormis deux premiers points isolés (aucun à 6 mois, et 2 diamants à 7-8 mois), le pointer débute véritablement à 9 mois, et décroît en durée –après une période d’environ 800ms entre 10-14 mois– vers 600ms à 15 mois.

4. Le mode de la distribution des cycles du babillage et le ratio pointer/babillage

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 (voir pour l’influence claire des patrons linguistiques de l’anglais américain, du français et du gallois, Vihman et al., 2004). 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 (1Hz) 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 (comme discuté dans Petitto et al., 2004). 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. 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 non-controversées des activités rythmiques de la main obtenues dans la même gamme de fréquence pour les bébés signant ou non. Ce 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 différent. Par conséquent, et par rapport à la proposition d’une étape rythmique générale de Thelen (1981), en précisant : (i) une proposition quantifiée de contrôle d’isofréquence maturationnelle pour la main et la bouche, nous pouvons prédire que, (ii) avec cette clause conditionnelle que l’activité rythmique (mandibulaire) de la mâchoire dans le

babillage canonique, ne soit pas différente de l'activité rythmique de l'ensemble bras-main, les deux vont fonctionner autour de 3Hz, et finalement, (iii) pour pouvoir inclure 2 cycles de mandibule, un geste de pointer discret va offrir une détente (*stroke*) d'environ 1.5Hz.

Mais quel sont les soutiens empiriques pour cette proposition d'un cycle de babillage à environ 3Hz, comme le sont ces mains de six bébés ? Nous avons utilisé l'approche source/filtre de Koopmans-van Beinum et Van der Stelt (1993) dans AMSTIVOC pour étiqueter notre corpus. Elles notent que :

« 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 424ms, 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 & Van der Stelt, 1993, p. 76).

Pris ensemble, ces résultats montrent clairement une tendance maturationnelle vers des cycles de « babillage/syllabe » plus rapides. Nous pouvons ajouter à ces mesures de fréquence de babillage canonique plutôt parsemées dans la littérature, les données de Bickley et al. (1986) sur une enfant. Ils ont testé la DFT (transformée de Fourier numérique) comparée à AC (fonction d'auto-corrélation) obtenues à partir de l'enveloppe, soit la courbe d'intensité, du signal audio du babillage. Les deux méthodes ont donné une dispersion des fréquences autour de 3Hz (leur Fig.3). Ils ont montré que la mise en évidence par ces mesures d'une activité rythmique de babillage canonique pouvait être obtenue 2 mois plus tôt (à environ 6 mois) que par la simple transcription auditive (8 mois). 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.

Si l'on revient à la durée moyenne de la détente de pointer de 776ms, cela donne bien un ratio de 2:1 (2.18), soit le ratio harmonique prédit de *2 syllabes pour 1 pointer*. Ce résultat démontre que les deux systèmes peuvent fonctionner en harmonie. Pour obtenir les capacités de ces deux systèmes, l'approche globale que nous avons préalablement utilisée était neutre par rapport à la sémantique, comme la présence supposée de mots. Nous n'avons pas non plus considéré la relation pointer-syllabe intra-événement. Une approche plus détaillée est nécessaire pour montrer que la réalisation de cette capacité globale est aussi individualisée et longitudinalisée que possible.

5. 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 (voir Fig.2). 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 & 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 ?

Sur la Fig. 5, les premières (S1) et secondes (S2) syllabes, débutent toutes deux en isochronie autour de 400ms (2.5Hz) à 6-8 mois ; tombent à 200ms (S1) et 300ms (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 600ms et oscille entre 500 et 600-700ms pour les 13-15 mois restant du suivi (qui finit à 2.5Hz à nouveau). Ce profil est quelque peu lissé quand le ratio S2/S1 à l'intérieur de chaque item mesuré est tracé (Fig.6), mettant à nouveau en évidence le saut à 12 mois, où la seconde syllabe devient 2,5 fois plus longue que la première, suivi par une décroissance de ce ratio à 13-15 mois.

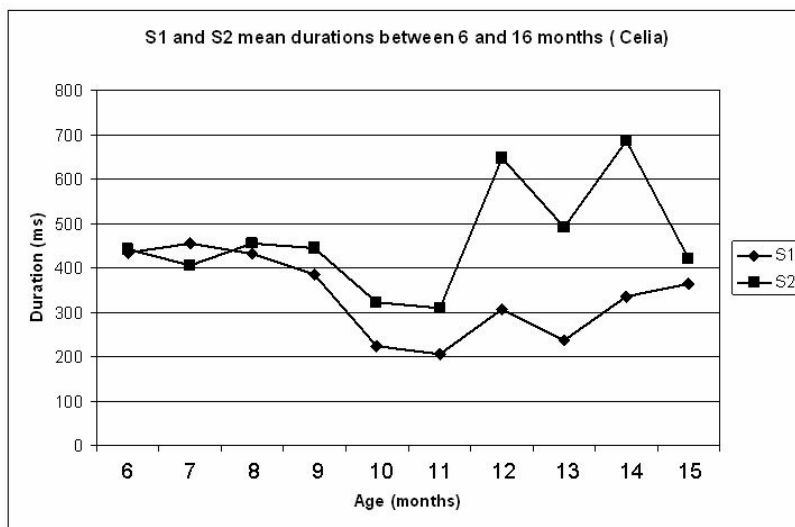


Fig.5 : Evolution de la durée moyenne des cycles des premières (S1 : diamants) et secondes (S2 : carrés) syllabes dans les performances de Célia entre 6 à 16 mois (N.B. : un regroupement légèrement différent des résultats de 15 mois et plus, donne les mêmes évolutions d'ensemble pour ces Figs 5 & 6, in Abry et al., à paraître). Notons l'augmentation dans S2 autour de l'émergence de ses premiers mots à 12 mois.

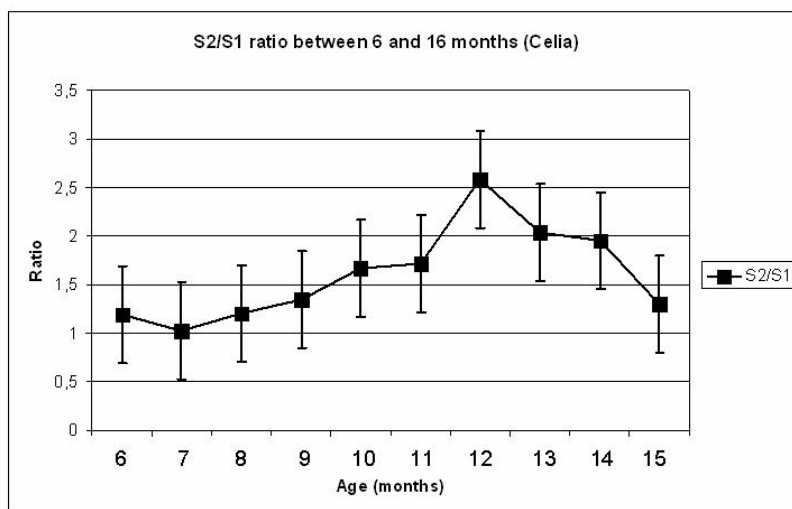


Fig.6 : Evolution du ratio S2/S1 (durée de la seconde syllabe par rapport à la première) pour Célia. Ce ratio met en évidence le changement de métrique à la période des premiers mots, à un an (cf. Fig.5).

Ce comportement est aussi illustré par un autre bébé français (Konopczynski, 1998, p. 186, sa Fig.2). Ce sujet présentait les mêmes changements dans le patron des syllabes. Les premières et secondes syllabes commençaient en isochronie à 8 mois (à environ 4Hz) et augmentaient en durée autour des premiers mots, la seconde syllabe atteignant un maximum à 12 mois, suivi par une chute à 15 mois. Elles oscillaient ensuite en parallèle, gardant leurs valeurs à part, leur durée moyenne globale étant d'environ 300ms (3.3Hz) à 2 ans. Prises ensemble, ces observations pourraient constituer une preuve que le patron iambique français est acquis (dans le babillage) autour des premiers mots. Nous avons remarqué par ailleurs que Célia présentait un pic remarquable de dissyllabes CVCV (Fig.7), par rapport à toutes ses autres productions (V, VCV, CVCVCV... ; aucune de celles-ci n'atteignant ces quelques 40%), et ceci à 11 mois, soit juste avant le démarrage des premiers mots.

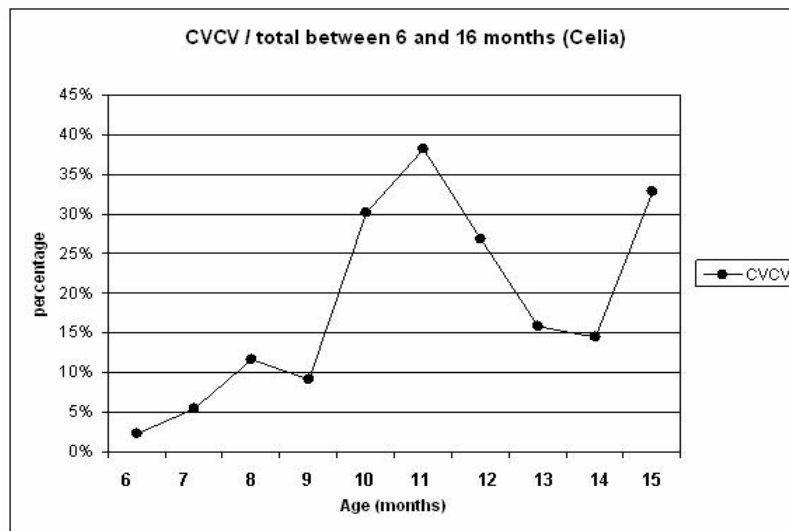


Fig.7 : Evolution du pourcentage de dissyllabes pour Célia. On notera un pic de près de 40% à 11 mois, juste avant l'apparition des premiers mots (un pourcentage que n'atteignent aucun des autres patrons observés, V, VCV, CVCVCV...).

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 900ms ne pourrait pas être contenue dans son pied moyen de 540ms... Mais cela ne se passe pas ainsi. Les mots CVCV de Célia correspondant à ces gestes de pointer durent entre 450-650ms (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 680ms. 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 fait en coupant dans le flux de babillage les première et seconde syllabes... Ainsi, il y a ici une production autonome des mots parallèle à cette restructuration dans le babillage. Ce résultat ne signifie pas que de tels changements dans la métrique (prosodique) soient indépendants.

6. 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 avec 1 ou 2 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. (2005)) : cela ne restera qu'une simple référence aux hypothèses sur l'intentionnalité inflationnistes/déflationnistes pour la lecture de l'esprit (*mind-reading*) ; à moins que les mécanismes déictiques ne soient mieux contraints (plus par Leslie et al., 1998, 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, p. 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 MacNeillage, et d'autres précurseurs, pour *dérivée 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 et des pointers (œil, bras), comme le rappellent McCune et al. (2003). Nous défendons que cette canne à pêche est un outil pour le gabarit phonologique du mot et pour la morphosyntaxe, donnant racine aux démonstratifs (Diessel, 1999), 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.), et finalement donnant forme 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 l'illustre Vihman, 1999).

La deixis avec des mots et des signes naissants est sans doute enactivée dans les circuits des nerfs et 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 accomplissent-ils leurs premiers pas en langue, avec leurs propres *mots* d'un *pied*, réussissant à stabiliser leurs *syllabes* dans ce *pied*.

Remerciements : A Stefanie Brosda pour l'héritage de son corpus ; à Romain Trollat pour les mesures du babillage de Célia ; à Coriandre Vilain pour son aide dans les graphiques et statistiques ; et à Alain Arnal et Christophe Savariaux pour la gestion des fichiers vidéos. Les recherches pour cette contribution ont été soutenues par un projet européen ESF au sein du programme *Origin of Man, Language, and Languages* (« COG-Speech ») et par l'ACI sur les *Systèmes Complexes dans les Sciences Sociales* du programme Cognitique français (« Pati papa ? »).

Notes

(1) 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 & Van der Stelt, 1986, Locke, 1997, Oller, 2000), avec ou sans emphase sur les gestes... et les grognements (voir McCune et al., 2003). Nous les prenons comme des acquis nécessaire et précieux : qui dirait que les premiers pas des jeunes enfants commencent à partir de compétences en locomotion zéro ? Depuis la page dans *Nature* d'Iverson & Goldin-Meadow (1998), la gestuelle dans le développement du langage a été illustrée par ces auteurs et bien d'autres, parmi lesquels Butterworth (2003), Bates et al. (voir Bates & 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).

(2) 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 pour une tâche typique de dénomination d'image (*picture naming*), ne traitent jamais de la question : pourquoi un total de TR de 600ms à partir de l'exposition 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 ; sans que la production de parole imaginée devienne plus rapide que la parole à haute voix (phénomène mesuré en premier par Landauer, 1962), les deux étant « cérébralement incarnées ».

(3) Sa formulation la plus fameuse dans une approche développementaliste, publiée par Allen & 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 & Demuth (2005), qui obtiennent des réponses de *réduplication dissyllabique* aux expériences de

truncation de mots avec leurs plus jeunes sujets (environ 2 ans). En fait Demuth & 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 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 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, qu'il s'agit probablement d'une unité de contrôle en *production* (discussion privée avec C. Abry, *InterSpeech-ICSLP*, Jeju, Corée, 2004).

(4) Nous avons 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 ne parlerons pas ici ; voir récemment Pizzuto et al, 2005). Ce problème sera seulement résolu 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 (*Movement Time*, 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 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 obtennent 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.

(5) A l'exception de l'exemple cité d'entrée, illustratif du patron de mouvement des deux lèvres d'une petite fille de huit mois, exposée à l'anglais, mis en évidence par Munhall & Jones (1998), avec une fréquence de 2.6Hz sur la lèvre inférieure portée par la mandibule (un rôle de porteuse corroboré plus tard, à 1 an, par Green et al., 2000 ; 2002).

(6) Les cycles de babillage n'ont pas été enregistrés par Petitto et al. (2004). En ce qui concerne les bébés entendant exposés aux signes, ils étaient « immergés dans un monde de sourds signant hautement exclusif » et « ils ne produisaient pas de babillage vocal syllabique systématique » (p. 51, note 2 ; une affirmation qui aurait à elle seule mérité tout un article).

Références bibliographiques

Abry, C., Ducey, V., Vilain, A., & Lalevée, C. (à paraître). When the babble-syllable feeds the foot in a point. In *Festschrift in Honour to Peter MacNeilage*. London: Erlbaum.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

- Bates, E., & Dick, F. (2002). Language, gesture and the developing brain. In B. J. Casey & Y. Munakata (Eds.), *Special issue: Converging method approach to the study of developmental science. Developmental Psychobiology*, 40(3), 293-310.
- Bickley, C., Lindblom, B., & Roug, L. (1986). Acoustic measures of rhythm in infants' babbling, or "All God's children got rhythm". *Proceedings of the 12th International Congress on Acoustics*, Volume / Band I A-C, A6-4, Toronto, 24-31 July.
- 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.
- Boysson-Bardies, B., De Bacri, N., Sagart, L., & Poizat, M. (1981). Timing in late babbling. *Journal of Child Language*, 8, 525-539.
- Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing. When language, culture, and cognition meet* (p. 9-33). London: Erlbaum.
- 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.
- 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, p. 39-52.
- Diessel, H. (1999). *Demonstratives: Form, function, and grammaticalization*. Amsterdam: Benjamins.
- Gerken, L. (1994). A metrical template account of children's weak syllable omission from multisyllabic words. *Journal of Child Language*, 21, 565-584.
- Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where Language, Culture and Cognition meet* (p. 85-107). Mahwah, NJ: Laurence Erlbaum Associates.
- 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(1), 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(2), 66-79.
- Hilaire-Debove, G., & Demuth, K. (2005). Troncation de mot chez l'enfant francophone. Actes du colloque ELA: *Emergence of Language Abilities: Ontogeny and Phylogeny (Lyon 8 -10 dec.)* (Actes à paraître).

- 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.
- Konopczynski, G. (1998). Interactive developmental intonology (IDI): Theory and application to French. *Revue Parole*, 7-8, 177-201.
- 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 of infant speech: Perception, early sound production, and maternal speech. *Proceedings of the Institute of Phonetic Sciences of Amsterdam*, 14, 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* (p. 37-50). New York: Stockton.
- Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646.
- 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.
- Leslie, A.M., Xu, F., Tremoulet, P., & Scholl, B. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, 2(1), 10-18.
- Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24, 133-164.
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language*, 58, 265-326.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-546.
- McCune, L., Greenwood, A., & Lennon, E. (2003). *Gestures, grunts, and words: The transition to communicative competence*. Tampa, FL: SRCD.
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44, 670-684.
- Munhall, K. G., & Jones, J. A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences*, 21(4), 524-525.
- Oller, D. K. (2000). *The emergence of the speech capacity*. London: Erlbaum.

- 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.
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2005). Gestural-vocal deixis and representational skills in early language development. In C. Abry, A. Vilain & J.-L. Schwartz (Eds.), *Special issue: "Vocalize to Localize II"*. *Interaction Studies*, 6(2), 223-252.
- Pylyshyn, Z. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, 4(5), 197-207.
- 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.
- 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.
- 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.
- Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology*, 17, 237-257.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675-735.
- Vihman, M. M. (1999). The transition to grammar in a bilingual child: Positional patterns, model learning, and relational words. *Journal of Bilingualism*, 3(2), 267-301.
- Vihman, M., DePaolis, R., Nakai, S., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336-353.
- Vihman, M. M. & McCune, L. (1994). When a word is a word?. *Journal of Child Language*, 21, 517-542.
- Volterra, V., Caselli, M.C., Capirci, O., & Pizzuto, E. (2004). Gesture and the emergence and development of language. In M. Tomasello & D. Slobin (Eds.), *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates* (p. 3-40). London: Erlbaum.