



HAL
open science

La parole, entre objets vocaux et objets du monde

Jean-Luc Schwartz, H el ene Loevenbruck

► **To cite this version:**

Jean-Luc Schwartz, H el ene Loevenbruck. La parole, entre objets vocaux et objets du monde. Michel Denis. La psychologie cognitive, Editions de la Maison des Sciences de l'Homme, pp.165-199, 2012, Cogniprisme, 978-2-7351-1509-9. <hal-00782274>

HAL Id: hal-00782274

<https://hal.science/hal-00782274v1>

Submitted on 29 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



HAL Authorization

La parole, entre objets vocaux et objets du monde

Jean-Luc Schwartz, H el ene L evenbruck
GIPSA-Lab, D epartement Parole et Cognition
UMR 5216, CNRS – Universit e de Grenoble
{Jean-Luc.Schwartz ; Helene.Loevenbruck}@gipsa-lab.grenoble-inp.fr

1. La parole transdisciplinaire

La parole est l'ensemble des m echanismes (de perception, d'action) qui permettent aux humains de communiquer par le langage oral. Elle est la prise en charge directe du langage par le corps¹. Corps perceptivo-moteur (percevoir, produire), corps audio-vocal (une bouche pour parler, une oreille pour entendre), ou plut ot audio-visuo-orofacial (l' eil per oit les gestes du conduit vocal et de la face, autant que l'oreille les entend) et m eme audio-visuo-oro-facio-braccio-manuel : la main accompagne la voix du geste, d'une mani ere qui appara t de plus en plus comme intrins eque et g en erale. Ce corps multiple, dont nous aurons l'occasion de passer en revue les diff erents segments, est aussi un corps cognitif,  evidemment, c'est- a-dire un corps sous contr ole du cerveau pour ses capacit es de coordination motrice et de traitement de l'information perceptive – et r eciproquement, si l'on consid ere que le cerveau est le si ege d'une cognition incarn ee (« embodied ») dans laquelle les actes cognitifs moteurs et perceptifs couplent corps et cerveau de mani ere presque indissociable.

Un corps multiple et cognitif, donc, qu'il s'agit, pour la parole, de brancher sur le langage, c'est- a-dire de mettre en situation de faire sortir le langage du cerveau d'un interlocuteur A pour le faire entrer dans celui d'un interlocuteur B – ou comment A cherche   changer l' etat du cerveau de B,   distance, au moyen de signaux de communication v ehiculant une structure linguistique. La parole est donc l'ensemble des m echanismes cognitifs capables de v ehiculer ces signaux pour faire passer des unit es linguistiques de A   B².

¹ Cette prise en charge est "directe", par contraste avec le couplage  criture/lecture. Ce dernier fournit une prise en charge indirecte, passant par des m echanismes tardifs (dans l'ontog en ese et la phylog en ese) de recodage qui ne sont pas constitutifs du langage propre.

² L a encore, en excluant la boucle  criture/lecture qui n'est qu'un recodage indirect ; et en mettant provisoirement de c ot e les langues sign ees, syst emes de communication visuo-manuelle permettant de produire et percevoir le langage   travers les gestes de la main en l'absence d'audition, syst emes qui sont bien, eux,

Ainsi, notre corps perceptivo-moteur aux segments multiples et sous contrôle cognitif doit-il être branché sur un système de communication sophistiqué, le langage, qui lui impose ses principes de fonctionnement, ses structures propres. La parole est façonnée pour s'interfacer avec ce que Chomsky (1981) nomme « l'organe du langage » ... et peut-être le façonner quelque peu en retour, nous le verrons. La parole est donc, finalement, un système d'articulations par nature, articulation entre gestes (multi-effecteurs, de la voix, de la face et de la main) et percepts (multi-sensoriels, de l'audition, de la vue, voire du toucher) d'une part, et entre signaux et sens d'autre part, c'est-à-dire, dans un premier temps, entre objets sonores (perceptivo-moteurs) et objets du monde (en un sens très général). Ceci fait de la parole un système cognitif par excellence, intégrant en son sein de manière ancienne et profonde des disciplines aussi diverses que :

- la physique, pour étudier les mécanismes de génération acoustique des sons ;
- la (neuro)psycho/physiologie cognitive, pour décrypter les mécanismes de perception et d'action, leur développement, leurs pathologies ;
- les sciences du langage, pour décrire les structures linguistiques, du phonème (phonologie) au mot (lexique) à la grammaire (morphologie, syntaxe) et au sens (sémantique, pragmatique) ;
- le traitement de l'information, pour analyser les contenus informationnels des signaux et les structures formelles des langues, modéliser les processus de perception et d'action, en déduire des algorithmes et des systèmes d'interfaces et de dialogue ;
- la primatologie et l'anthropologie, pour placer ce corps cognitif dans son environnement phylogénétique.

Pour coordonner ces différentes disciplines, on a vu naître dans le monde un certain nombre de laboratoires centrés sur la « communication parlée ». Ces laboratoires regroupent typiquement quelques équipes de chercheurs provenant des disciplines requises, et s'attaquant de manière collaborative à l'étude de la parole. La communication parlée s'est ainsi forgée, dans les cinquante dernières années, une histoire intégratrice, avec ses congrès, ses revues (comme *Speech Communication*, *Journal of Phonetics*, *Computer Speech and Language*, *Language and Speech*, *Journal of Speech Language and Hearing Research* ; et bien sûr des revues plus disciplinaires telles que *IEEE Transactions on Speech and Audio Processing*,

authentiquement du langage, dotés des mêmes capacités fonctionnelles, mais un langage alternatif, non oral et donc non parlé.

Acta Acustica, Journal of the Acoustical Society of America, Hearing Research, Perception and Psychophysics, Brain and Language, Journal of Neurolinguistics, etc.), ses échanges, et même son association internationale, née à Grenoble en 1988 sous le nom de ESCA (*European Speech Communication Association*), devenue maintenant ISCA (*International Speech Communication Association*), et organisant chaque année un congrès, *Interspeech*, réunissant pendant une semaine un millier de chercheurs échangeant sur tous les domaines de la communication parlée. Cette histoire a même été épistémologiquement théorisée par la communauté française, avec les travaux de Boë et Liénard (1988, 1988-1989) et Teston (2004) sur l'émergence des « sciences de la parole », qui se sont constituées en véritable domaine scientifique par une réorganisation de savoirs, partagés au départ par plusieurs disciplines partenaires, à travers un certain nombre de ruptures épistémologiques dont on peut faire l'analyse précise.

Nous allons, dans ce qui suit, organiser la présentation autour de deux questions centrales. Dans un premier temps, nous étudierons de quoi sont constitués les « objets sonores », structurés par le lien perceptivo-moteur, construits dans le développement et probablement façonnés par la phylogenèse. Puis nous plongerons ces formes sonores et gestuelles dans leur relation avec le monde, en analysant comment elles permettent d'attirer l'attention de l'interlocuteur vers les objets du monde auquel réfèrent les objets vocaux, pour construire la relation forme-substance qui définit le langage humain (Saussure, 1916)³.

³ On trouvera abordées plus en détail les différentes questions traitées dans ce chapitre, au sein d'un numéro spécial récent de la *Revue Française de Linguistique Appliquée*, coordonné par L.-J. Boë et al. (2008), auquel nous nous référerons chaque fois que possible pour fournir au lecteur une présentation approfondie sur tel ou tel aspect.

2. Des objets vocaux nés de la matrice perceptivo-motrice

Pour rester dans notre logique résolument pluridisciplinaire, nous allons chercher à décrire les objets vocaux qui servent à communiquer le langage oral, en combinant librement des logiques différentes, celle de la primatologie comparée pour estimer les capacités de communication (alternatives et préexistantes), du développement pour apprécier comment s'acquiert la maîtrise de ces objets, de la psychologie cognitive pour en décrire le fonctionnement mature, et de la phonologie pour savoir comment sont structurés les signaux en systèmes de communication langagière.

Les briques de base d'un système de communication

Pour communiquer, il faut mettre en contact deux agents cognitifs dotés de capacités de production de signaux informatifs, et de perception de ces signaux. Le premier agent doit être capable de contrôler des actionneurs (des membres, de la face, de la voix ...) pour coordonner la production de signaux de communication, informatifs et donc variés. Le second agent doit être capable de percevoir ces signaux, par l'un ou l'autre de ses canaux sensoriels (audition, vision, toucher...) et d'extraire des catégories pertinentes, à travers la variabilité des réalisations observées. Les deux agents doivent enfin pouvoir échanger leur rôle, et connaître ainsi les correspondances entre actions productrices de signaux et catégories perceptives extraites de ces signaux, et ceci en respectant un principe de *parité* qui est que les signaux produits par un agent et perçus par l'autre réfèrent bien au même message.

Les systèmes de communication animaux

Communiquer, c'est disposer de moyens de changer le comportement d'un partenaire, qu'il soit de son espèce ou d'une autre. Cette capacité très générale est présente chez les animaux sous des aspects très génériques : communication sexuée qui englobe l'ensemble des parades amoureuses ; communication de trace ou de territoire, communication agoniste qui implique les comportements de menace ou de marquage du territoire ; communication liée à la nourriture, notamment dans la relation entre parents et progéniture. La communication passe par des signaux, qui peuvent comporter des aspects de variation contrôlée et d'apprenabilité, comme dans la communication vocale par excellence qu'est le chant des oiseaux.

Chez les primates non humains, deux systèmes de communication ont été particulièrement étudiés jusqu'à être proposés comme de possibles précurseurs du langage humain. D'abord,

les gestes communicatifs sont attestés chez de nombreuses espèces de singes : gestes de menace (Meguerditchian et Vauclair, 2006), gestes de désignation ou de « pointage » sur lesquels nous reviendrons longuement dans la section suivante (Hopkins et Leavens, 1998 ; Leavens, 2004). Mais certaines espèces de singes semblent également disposer de systèmes de communication vocale, notamment à travers les cris d'alerte des singes vervets (Cheney et Seyfarth, 1990), distinguant trois types de prédateurs, l'aigle qui vient du ciel, le léopard qui accourt et le serpent caché dans le sol.

Un débat s'est instauré entre primatologues pour estimer lequel de ces deux systèmes de communication, gestuelle ou orale, fournirait le précurseur le plus plausible au langage humain. La communication gestuelle dispose de spécificités qui en font un bon candidat, ce qui était remarqué dès le XVIII^{ème} siècle par le philosophe Condillac. D'une part, elle semble dotée de bonnes propriétés de contrôle intentionnel et de spécialisation fonctionnelle (flexibilité) qui en font par excellence un système de communication sociale puissant (Corballis, 2003 ; Vauclair, 2004 ; Meguerditchian et Vauclair, 2008). D'autre part, ce système est naturellement prédisposé à contenir des capacités référentielles, puisqu'un geste peut renvoyer par ses propriétés intrinsèques à des éléments du monde extérieur (« iconicité »). Le pointage en fournit un point de départ évident, permettant au singe d'orienter l'attention vers l'objet désigné (Leavens, 2004). Arbib (2005) propose que, à partir des capacités d'imitation attestées chez les primates, des gestes plus complexes puissent, par leurs propriétés d'affordance aux objets et aux actions qu'ils impliquent, référer à ces objets et à ces actions (Rizzolatti et Arbib, 1998). De là se mettrait en place un système de « pantomime » précurseur de « protosignes » qui conduiraient au langage humain.

Mais la communication vocale n'est pas sans argument. Contrairement aux visions initiales, elle n'est pas totalement captive d'un système de contrôle émotionnel déclenchant des comportements réflexes. Elle apparaît au contraire de plus en plus comme complexe, flexible et tactique, c'est-à-dire contrôlée jusqu'à un certain point en fonction de l'objectif de communication (voir les travaux de Slocombe et Zuberbühler, 2005, sur les cris associés à la nourriture chez les chimpanzés) jusqu'à englober des capacités de combinatoire syntaxique complexe (avec séquençement contrôlé et informatif de vocalisations « élémentaires » : Arnold et Zuberbühler, 2006). Et elle fournit une continuité de moyens avec la parole, qui semble simplifier le chemin phylogénétique vers le langage oral (MacNeilage, 1998).

De plus en plus, on peut considérer que les liens fonctionnels entre geste et voix, du singe à l'enfant et à l'homme (voir Section 3), en font un système de communication multimodal qui fournit sans doute dans sa globalité un chemin vers le langage. C'est ce qu'Arbib (2005) propose dans sa vision d'une phylogenèse « en hélice », gestes et vocalisations fournissant des passages emboîtés allant progressivement vers la complexité du langage humain. Nous aurons l'occasion de préciser notre vision propre d'un « bootstrap » multimodal vers le langage à travers la deïxis, dans la Section 3.

Les premiers sons de la parole : mettre la voix en mouvement

Pour porter une information par la voix, il faut pouvoir en moduler la couleur acoustique. Pour le chant des oiseaux, cette modulation porte sur la hauteur de la voix, que les humains contrôlent, eux, par leurs cordes vocales. Toutefois pour la parole humaine, la modulation essentielle ne porte pas sur la hauteur, mais sur le timbre, à travers la modulation des résonances du conduit vocal – que l'on nomme les formants. Pour varier les résonances, il faut changer la forme de la bouche – la forme du conduit vocal – en bougeant la mâchoire, la langue, les lèvres, et d'autres « articulateurs ». On peut changer aussi la hauteur du son pour produire la « prosodie », le chant de la parole. Nous y reviendrons.

Prenons alors un moment le point de vue ontogénétique fourni par l'étude du développement de la production du langage chez l'enfant. Le bébé est capable dès la naissance de vocaliser : il crie, il pleure. Mais ses vocalisations sont peu modulées. Il sait pourtant produire du mouvement de ses articulateurs, en bougeant sa mâchoire et sa langue pour se nourrir – téter d'abord, puis commencer à ingérer des aliments. Le miracle survient lorsque vers 7 mois, il couple ces deux capacités, en expérimentant ce que cela fait de vocaliser et bouger la mâchoire : il commence à effectuer des cycles de mâchoire comme il expérimente le mouvement de ses bras et de ses jambes par des cycles de flexion – extension, également vers 7 mois (Thelen, 1981). On observe alors le « babillage canonique » au cours duquel le bébé expérimente les modulations de sa voix en effectuant ces cycles de mâchoire vocalisés, qui produisent des sortes de syllabes – des « protosyllabes » – telles que « ababa » ou « adia », selon que, en fermant sa bouche, le bébé ferme plutôt ses lèvres pour « ba » ou plutôt sa langue ramenée par la mâchoire vers l'avant du palais proche de dents, pour « da » ou « dia » (Davis et al., 2008).

MacNeilage et Davis (2001) ont théorisé cette découverte des sons de la parole à travers la Théorie « Frame-Content » (« cadre et contenu »), selon laquelle la parole, née de ces cycles de mâchoire modulant la voix, débouche ainsi spontanément sur des alternances de « protovoyelles » (lorsque la bouche est ouverte) et de « protoconsonnes » (lorsqu'elle est fermée), qui dessinent le précurseur de la syllabe, présente universellement dans les langues (voir Rousset, 2004). Ils extrapolent ce chemin d'exploration/construction vers la phylogénèse, en faisant de la mastication une étape clé vers la découverte de modulations orofaciales qui fourniraient un précurseur du langage humain (MacNeilage et Davis, 2005).

Des gestes vocaux sélectionnés pour leur efficacité perceptive

Ayant « inventé » la modulation vocale par ce possible « bootstrap » ontogénétique, voire phylogénétique, fourni par les gestes de la mâchoire disponibles dans la mastication, il faut élargir et développer ces capacités de production de signaux vocaux. Ces signaux doivent être informationnels, rappelons-le : donc, contrastifs, susceptibles d'être discriminés et catégorisés efficacement par le partenaire.

Sur ce principe se sont développées des théories de ce que l'on peut baptiser une « morphogénèse des sons du langage », cherchant à déterminer ce qui fonde un « bon geste vocal » et surtout un « bon ensemble de gestes » – ou de sons – pour servir de support à un système linguistique.

Le plus bel exemple est la Théorie de la Dispersion du phonéticien suédois Lindblom, développée depuis le début des années 70. Pour en expliquer le principe, partons des gestes articulatoires permettant de produire des voyelles. Les humains disposent typiquement de trois degrés de liberté pour moduler la forme de leur conduit vocal dans la production des voyelles. En abaissant la mâchoire et la langue, on va vers des voyelles « ouvertes » comme le [a] (de « pas »). En remontant langue et mâchoire, sans fermer complètement le conduit vocal (sous peine de produire des consonnes plutôt que des voyelles), on va vers des voyelles « fermées » que l'on peut produire soit en avançant la langue (comme dans [i], de « pis ») soit en la reculant (comme dans [u] de « pou »). On peut également jouer sur les lèvres, arrondies et presque fermées pour [u] et [y] (de « pus ») ou étirées pour [i]. Ainsi, l'espace des gestes

vocaliques est essentiellement tridimensionnel⁴. Au sein de cet espace, si vous aviez à sélectionner trois voyelles les plus différentes possible, lesquelles choisir ? [i a u] semble un excellent choix : [i] est fermée (langue haute), antérieure (langue en avant) et non arrondie (lèvres étirées), [u] est fermée, postérieure (langue en arrière) et arrondie (lèvres arrondies) et [a] est ouvert (avec en conséquence une langue en position centrale, et des lèvres forcément ouvertes également). Bon choix ! Presque toutes les langues utilisent [i a u], selon la base de données UPSID qui fournit un échantillon représentatif de presque 10% des 6000 ou 7000 langues du monde (Maddieson, 1984 ; Maddieson et Precoda, 1989).

Mais un autre choix semble tout aussi bon : il consiste à échanger configuration de la langue et des lèvres entre [i] et [u], remplaçant [i] par [y], langue antérieure, lèvres arrondies ; et [u] par [ɯ], langue postérieure, lèvres étirées – voyelle présente notamment dans certaines langues asiatiques, et qui ne correspond donc à aucun son prononçable en français (pour la produire, il faut partir d'un « ou » et étirer les lèvres comme pour un « i », sans changer la position de la langue : pas facile ...). Mauvais choix ! Aucune langue au monde ne se contente de ces trois voyelles, et [y], peu présent dans les langues (8% seulement des langues de la base UPSID contiennent cette voyelle, le français comptant parmi ces 8%, au contraire d'autres langues latines comme l'italien ou l'espagnol), n'apparaît *jamais* sans [i]. La raison en est auditive. On caractérise en général une voyelle par ses résonances acoustiques, les formants, et plus précisément par les formants les plus bas, qui sont aussi les plus saillants perceptivement. Dans le plan des deux premiers formants F1 et F2, les voyelles se projettent toutes dans un triangle dont [i a u] sont les sommets, [y] et [ɯ] étant beaucoup plus proches l'une de l'autre que [i] et [u] (voir Figure 1). Ainsi, les langues choisissent les gestes en fonction de leur valeur perceptive : leur « dispersion », dans la terminologie de Lindblom (Liljencrants et Lindblom, 1972 ; Lindblom, 1986, 1990). Si l'on ajoute un second critère perceptif de « bonnes formes acoustiques », stables en mémoire, formes dites « focales » (Boë et Abry, 1986, Schwartz et al., 2005), on obtient la Théorie de la Dispersion-Focalisation (TDF, Schwartz et al., 1997), qui produit d'excellentes prédictions des systèmes vocaliques des langues du monde.

⁴ Si l'on ne tient pas compte de l'ouverture-fermeture du voile du palais, ou velum, permettant de passer des voyelles orales – à velum fermé, sans passage d'air par le nez – aux voyelles nasales à velum ouvert, comme dans le son de “pain” ou “rond”.

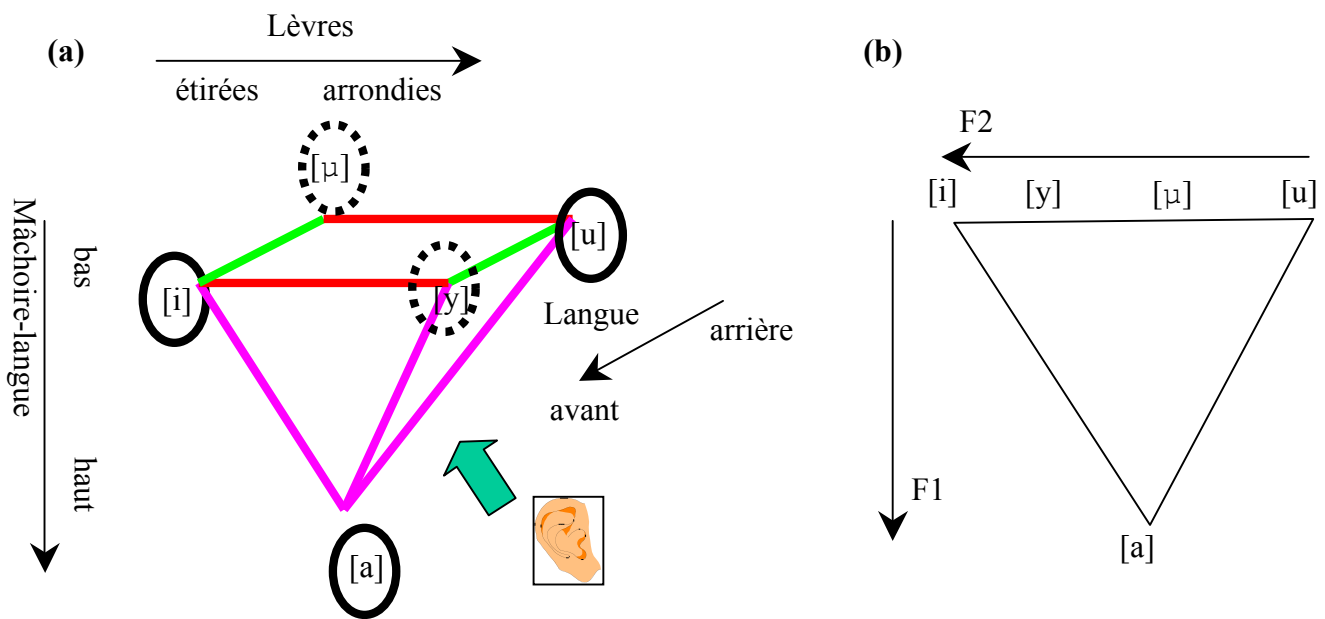


Figure 1. Les trois dimensions articulatoires de l'espace des voyelles orales (a) et leur projection auditive dans l'espace des deux premiers formants (b).

On peut sur le même principe essayer de comprendre pourquoi parmi les fricatives (sons produits avec une obstruction dans le conduit vocal, générant du bruit), les langues sélectionnent plus fréquemment les fricatives « non voisées » comme [s] de « sou » ou [ʃ] de « chou », produites sans action des cordes vocales, plutôt que des fricatives « voisées » comme [z] de « zut » ou [ʒ] de « geai » (Boë et al., 2000); et aussi pourquoi parmi les plosives (sons produits par une fermeture complète du conduit vocal, réalisée par les lèvres ou la langue), [p] (de « pou »), [t] (de « tout ») et [k] (de « cou ») fournissent les meilleurs lieux de fermeture, utilisés prioritairement par les langues (Abry, 2003).

En résumé, les signaux de communication des langues orales sont des gestes modulateurs à forte valeur perceptive, permettant de produire des sons contrastifs perçus efficacement. Sur ce principe, on peut comprendre l'émergence des syllabes (selon la Théorie Frame-Content) puis les consonnes et les voyelles sélectionnées pour leurs qualités acoustiques propres et systémiques (selon la Théorie de la Dispersion-Focalisation).

Des objets sonores et visuels

Ces gestes vocaux sont perçus par l'oreille, mais aussi par l'œil : la perception de la parole est audiovisuelle. Cette propriété a longtemps été conçue comme annexe et essentiellement utile aux sourds via la lecture labiale. On sait maintenant qu'elle est en réalité utile à chacun, particulièrement en situation de perception délicate : dans le bruit (Erber, 1975 ; Benoît et al., 1994) ou en langue étrangère (Davis et Kim, 1998). On commence même à découvrir que voir nous sert à mieux entendre (Schwartz et al., 2004), c'est-à-dire à améliorer ses capacités à extraire le signal dans le bruit, ce qui est générateur de nouvelles techniques de « débruitage du son par l'image » (Girin et al., 2001 ; Sodoyer et al., 2004).

On peut même supposer que les langues ont sélectionné certains phonèmes pour leur visibilité plus que pour leur audibilité : ainsi, le contraste [m]-[n] est bien visible (particulièrement le [m], bilabial et dont la visibilité est fort appréciée des enfants dans le babillage) mais peu audible (Mills, 1987), et pourtant fortement prisé par les langues du monde (Schwartz et al., 2007).

Des objets à contrôler et à catégoriser : éléments d'une théorie perceptivo-motrice

Nous disposons ainsi d'objets vocaux, supports des phonèmes des langues : voyelles et consonnes alternant dans des syllabes universelles, gestes orofaciaux produits par la coordination de la source vocale et des articulateurs-modulateurs (mâchoire, langue, lèvres et autres) pour produire des stimuli perceptifs efficaces. Face à ces objets, les systèmes de perception et de production de la parole s'organisent dans le développement, jusqu'à maturité.

Le système de production de la parole doit gérer les enchaînements, les séquences, au sein de « partitions gestuelles » complexes. Pour ce faire, il utilise des stratégies d'économie articulatoire sophistiquées. Ainsi, pour enchaîner le « p », le « s » et le « eu » de « pseudo », par exemple, le locuteur peut, en même temps qu'il ferme ses lèvres pour préparer le « p », monter sa langue vers son palais derrière ses dents pour préparer le « s » et projeter les lèvres vers l'avant pour le « eu » qui suivra : il ne lui reste alors qu'à lancer presque en même temps les gestes lui permettant de faire sonner le « p » et en dévoilant quasiment instantanément le « s » et le « eu ». C'est ce que l'on nomme la coarticulation, qui suppose de mettre en œuvre des stratégies de contrôle complexe d'un système biomécanique et aéroacoustique sophistiqué (Fuchs et Perrier, 2008).

En retour, la coarticulation se traduit par une superposition articulatoire qui se projette en une coarticulation acoustique complexe, qui fait du son d'un « p » un produit englobant les caractéristiques du « p » mais aussi celles de ses voisins, le « contaminant » en quelque sorte de la coloration, dans notre exemple, du « s » et du « eu ». Les débats entre chercheurs sont vifs pour déterminer comment peut se traiter la résolution perceptive de cette complexité du passage de l'articulatoire à l'acoustique. Le débat oppose tenants d'une solution « auditive » à base d'apprentissage de processus de catégorisation exploitant des connaissances statistiques sur des séquences sonores, et tenants de solutions « motrices » selon lesquelles l'auditeur dispose de moyens de retrouver les gestes articulatoires à l'origine des sons, et, partant, d'en démêler plus aisément l'écheveau.

Nous défendons quant à nous une voie « moyenne » dans laquelle c'est bien sur leur valeur auditive et multisensorielle que les gestes sont catégorisés, mais en intégrant dans le décodage des composantes motrices qui définissent les catégories et co-construisent perception et action (*Perception for Action Control Theory*, Schwartz et al., 2007, 2008). Les unités de la communication, fondant la « parité » entre locuteur et auditeur, sont dans la PACT intrinsèquement perceptivo-motrices.


Ces débats renvoient à une littérature en pleine expansion sur les relations entre représentations perceptives et motrices dans le cortex, au sein d'un arc « dorsal » reliant des aires sensorielles temporales et des aires motrices frontales à travers une connexion pariétale articulant audition, vision et proprioception au sein de représentations multisensorielles phonologiques (voir une revue dans Schwartz et al., 2008).

3. Des objets vocaux aux objets du monde

Les objets vocaux sont les formes sonores signifiantes (Saussure) qui nous permettent de communiquer sur les objets du monde. Comme nous l'avons mentionné dans la Section 2, il a été proposé récemment que le lien entre objets vocaux et objets du monde s'établisse par le geste de pointage. Nous allons essayer de montrer dans cette troisième section que le pointage est beaucoup plus qu'un geste co-verbal, qu'il peut être considéré comme une des racines de la communication parlée car son émergence chez le bébé est liée aux étapes du développement langagier, mais aussi car « montrer du doigt » (et de l'œil) pourrait bien être proche neurophysiologiquement de « montrer de la voix ».

Le pointage, un système puissant de mise en relation multimodale avec les objets du monde

Le pointage est la capacité d'attirer l'attention d'autrui vers un objet du monde pour que cet objet devienne le centre de l'attention partagée. Le pointage présente de nombreuses caractéristiques qui en font un sujet d'étude passionnant pour les sociologues, les anthropologues, les historiens de l'art, les éthologues, les linguistes et les spécialistes du développement de l'enfant.

Une première caractéristique du pointage est qu'il est ubiquitaire. Le geste de pointage canonique, avec l'index et le bras tendus dans la direction de l'objet d'intérêt et les autres doigts repliés sous la paume (appelons-le « pointage avec l'index » ou « pointage digital ») est omniprésent au quotidien, dans la plupart des cultures du monde. On l'observe dans les communications orales aussi bien que signées. Dans certaines langues des signes, le geste de pointage digital est même utilisé linguistiquement, pour l'indexicalité référentielle (les pronoms). Il transparaît également dans notre vocabulaire, dans des mots comme « indiquer », ou « mise à l'index » ainsi que dans nos icônes ou symboles (comme celui-ci: , tellement apprécié de nos collègues phonologues...).

Le pointage présente aussi la caractéristique d'être universel. Il a été affirmé que le pointage digital est une capacité universelle partagée par tous les êtres humains (e.g. Povinelli et Davis, 1994). Dans certaines cultures, cependant, le pointage avec l'index peut être remplacé par un pointage labial (voir Wilkins, 2003). Mais les gestes de pointage, qu'ils soient produits avec l'index, la main tout entière, les lèvres ou le menton, restent un outil de communication crucial chez l'ensemble des humains.

Une autre spécificité du pointage est qu'il a un long passé historique. Bien qu'on ne s'y intéresse scientifiquement que depuis récemment, le pointage digital n'est pas une invention du XXème siècle. L'art européen nous en donne de précieux témoignages. Les peintures de pointages les plus célèbres sont probablement celles de Léonard de Vinci, qui remontent au XVème siècle (La Vierge aux Rochers, 1483-1486, Saint Jean-Baptiste, 1513-1516). Mais on peut trouver des traces de pointage bien plus précoces, notamment dans la Tapisserie de Bayeux, brodée autour de la fin du XIème siècle et qui raconte, à travers un foisonnement de pointages, la conquête de l'Angleterre par les Normands en 1066. De plus, les témoignages de pointage ne se trouvent pas que dans l'art européen, on en trouve aussi en Asie, dans l'art birman du XVIIIème siècle, par exemple.

Le pointage a pour caractéristique d'être une capacité relativement évoluée. Il a longtemps été affirmé que le pointage est une spécificité humaine et que les grands singes, par exemple ne pointent pas (e.g. Povinelli et al., 2003). Mais des travaux plus récents de Leavens et al. (2005), notamment, ont montré que les chimpanzés en captivité sont capables de pointer vers de la nourriture inaccessible. Et, comme les bébés humains, ils le font spontanément, sans apprentissage explicite. Ainsi, bien que le pointage ne semble pas être spécifique aux primates humains, puisque les grands singes (et peut-être d'autres animaux) en sont capables, il semble que ce soit un comportement social sophistiqué, utilisés par des animaux évolués, capables d'imitation et de communication inter-individuelle (Pollick et de Waal, 2007).

Le pointage a aussi la particularité d'être multimodal : il peut être manuel (ou digital), labial, facial, oculaire et... vocal. Comme nous l'avons décrit plus haut, le pointage est le plus souvent digital, c'est-à-dire réalisé avec un index étendu. Il peut être aussi manuel, c'est-à-dire que toute la main est étendue (avec la paume vers le haut ou la main verticale, Kendon, 1996). Il existe aussi un pointage avec les lèvres protruses, observé sur tous les continents (Panama, Australie, Papouasie Nouvelle-Guinée, Ghana, Amérique du Nord ; cf. Wilkins, 2003). En fait, de nombreuses parties du corps sont utilisées pour pointer : la tête, les lèvres, le menton, le coude, le pied, le bras, la main (Kendon, 1996). Le pointage oculaire ou regard déictique est également fréquent. C'est la capacité de regarder alternativement un objet et l'interlocuteur. C'est pour l'interlocuteur une invitation à regarder l'objet qui devient le centre de l'attention partagée. Le pointage vocal correspond à ce que l'on appelle parfois la deixis ou la focalisation. En français, comme dans beaucoup de langues, la deixis vocale peut être

exprimée par l'extraction syntaxique, en utilisant une forme présentative déictique, comme dans l'exemple suivant : « C'est Jean-Luc qui a réparé mon vélo ». Elle peut également être exprimée en utilisant la focalisation prosodique contrastive, c'est-à-dire en appliquant un contour intonatif (une mélodie) spécifique sur le constituant vers lequel on souhaite pointer : « JEAN-LUC_F a réparé mon vélo ». L'effet de ce contour intonatif, qui sera décrit plus loin, est de mettre en relief le constituant pointé, de le contraster, le reste de l'énoncé portant un contour post-focal plat.

Enfin, le pointage est caractérisé par son rôle dans de nombreuses étapes du développement du langage. Le pointage oculaire et, plus tard, le pointage digital sont deux étapes clés du développement cognitif du bébé, liées à des étapes du développement de la parole. Ainsi, on observe d'abord le pointage oculaire, vers 8 ou 9 mois, lors de la mise en place de l'attention conjointe entre l'adulte et le bébé. A cet âge, le bébé devient capable d'inviter l'adulte à regarder un objet en regardant alternativement l'objet et les yeux de l'adulte. Il peut donc alors utiliser le regard pour manipuler l'attention de l'adulte.

Plus tard, vers 9-11 mois, le bébé produit des gestes de pointage, le plus souvent digital. L'émergence de ce geste de pointage est un bon prédicteur de l'apparition des premiers mots (environ 2 mois après) et la production de ce geste est liée au développement du langage entre 9 et 13 mois (Bates et al., 1979; Butcher et Goldin-Meadow, 2000). Le pointage sert non seulement à mettre en valeur un objet mais aussi à établir une connexion entre l'objet du monde et l'objet vocal. Le bébé pour qui l'on pointe vers un chat en disant « un CHAT » ou à qui l'on répond « c'est le CHAT » lorsqu'il désigne un chat, fait le lien entre la forme sonore entendue et l'objet désigné. Le pointage digital est ainsi clairement associé à la construction du lexique.

L'émergence du mot

Ce geste de pointage lié à une production sonore pourrait bien permettre d'amorcer, d'initialiser le découpage en mots du flux sonore continu, un problème central du développement des mécanismes de compréhension de la parole (Milotte, 2008 ; Nazzi, 2008). Même dans la parole adressée aux enfants, les énoncés contiennent généralement plusieurs mots (selon les études de Aslin, 1993, ou Brent et Siskind, 2001, les mots isolés représentent moins de 10 % de l'ensemble des mots prononcés). On sait qu'il existe des indices prosodiques (intonatifs et rythmiques) relativement robustes des débuts et fins de mots

dans toutes les langues. En français par exemple, un groupe accentuel (souvent simplement composé, dans le mamanais – la « langue » utilisée par la maman pour parler à son bébé – d'un mot et son déterminant) en début de phrase est marqué par une montée intonative optionnelle précédée systématiquement d'un creux aligné avec le début du mot (e. g. Welby, 2007). La fin d'un groupe accentuel est marquée (en début de phrase) par une montée, portant approximativement sur la syllabe finale, et un allongement de la syllabe finale (e.g. Di Cristo, 1999 ; Jun et Fougeron 2000 ; Welby et Lævenbruck, 2006). Pour l'adulte, il a été montré que ce type d'indices peut être utilisé pour découper le flux sonore en mots (Christophe et al., 2004 ; Welby, 2007). Mais comment le bébé pourrait-il associer ces indices prosodiques à des frontières de mots, sans connaître les règles d'association, sans savoir au départ où sont les mots ? D'autres indices sont de bons candidats pour permettre le découpage en mots (cf. Nazzi, 2008 pour une revue) : *les indices allophoniques*, c'est-à-dire le fait que la réalisation acoustique de certains phonèmes soit influencée par la position du phonème dans le mot (en français par exemple, /r/ se prononce parfois différemment en attaque de syllabe comme dans « rateau » [ʁ] qu'en coda comme dans « babord » [χ]); *les indices phonotactiques*, c'est-à-dire la connaissance des séquences phonétiques possibles à l'intérieur d'un mot (en français par exemple, /kf/ n'existe pas à l'intérieur d'un mot et signale une frontière entre 2 mots comme « lac familial ») ; *les indices distributionnels* ou les statistiques sur les fréquences d'apparition des mêmes séquences de syllabes, fréquences élevées à l'intérieur d'un mot. Mais tous ces indices nécessitent que l'enfant ait une connaissance, d'une part, de l'existence de mots et, d'autre part, maîtrise déjà quelques mots.

Selon nous, la toute première amorce pourrait bien être le pointage. Le fait que les constructions présentatives (« c'est le CHAT ») soient parmi les plus fréquentes dans le langage adressé à l'enfant (Cameron-Faulkner et al., 2003) est particulièrement intéressant. Dans ce type de construction, le mot désigné est mis en relief par sa position finale dans la phrase et par l'intonation de pointage qui lui est fréquemment associée. L'utilisation d'un geste de pointage digital concomitant pourrait compléter cette mise en relief. Le bébé pour qui l'on pointe avec l'index en utilisant une intonation de pointage pourrait apprendre à faire le lien entre l'objet désigné par la main et l'objet désigné par la voix. Petit à petit, il pourrait ne plus avoir besoin du geste pour comprendre quand la voix montre dans un énoncé et son repérage des mots pourrait ainsi être facilité.

En sens inverse, dans la constitution du mot comme unité de production de la parole, une hypothèse intéressante a été proposée par Ducey et Abry (2004) : le geste de pointage pourrait permettre au bébé de découper son propre flux de babillage en unités lexicales. En effet, l'oscillation mandibulaire a pour fréquence environ 3Hz, ce qui impose que la durée d'une syllabe isolée soit d'environ 330 ms. C'est bien le cas chez l'adulte. Or la durée d'une détente de geste de pointage (c'est-à-dire de l'initiation du geste à l'instant où le bras est complètement tendu) est double, soit environ 660 ms. Le geste de pointage produit de façon concomitante avec des syllabes babillées (mamamamama ou babababa) pourrait ainsi donner le gabarit du mot : deux syllabes, ou un pied, ont le temps d'être prononcées dans un geste de pointage. Cette hypothèse a été vérifiée par Abry et Ducey (2008) qui montrent chez six bébés suivis longitudinalement de 6 à 18 mois, que le rapport entre la durée de la détente du geste de pointage et celle d'une syllabe est de 2,18 en moyenne : le gabarit de découpe du mot est donc construit par le geste de pointage. Ce rapport optimal de deux syllabes consonne-voyelle pour un geste de pointage a également été retrouvé chez l'adulte par Rochet-Capellan et al. (2007).

Du mot à la séquence de mots : l'émergence de la syntaxe

Plus tard, entre 16 et 20 mois, lors de la transition du stade « un mot » au stade « deux mots », on peut observer des combinaisons de mots et de gestes déictiques (comme pointer vers un endroit en prononçant « chat » pour indiquer qu'il y a un chat à cet endroit). Ces combinaisons sont des proto-phrases simples. Le nombre de gestes et de combinaisons gestes-mots produits à 16 mois est corrélé avec la production vocale à 20 mois (Morford et Goldin-Meadow, 1992; Capirci et al., 1996; Goldin-Meadow et Butcher, 2003; Volterra et al., 2005). Le pointage digital semble ainsi clairement associé également à l'émergence de la syntaxe.

Le pointage des enfants peut être aussi vocal, mais peu d'études se sont penchées sur le développement des pointages prosodiques et syntaxiques. On sait que les enfants anglophones maîtrisent la focalisation contrastive prosodique avant l'âge de 4 ans, en l'absence de tout enseignement formel (Hornby et Hass, 1970 ; voir aussi pour le français Ménard et al., 2006). On sait aussi que la capacité de produire des contours intonatifs similaires à ceux des adultes émerge très tôt en français et que l'allongement syllabique est correctement produit dès 2 ans (Konopczynski, 1986). A 18 mois, l'enfant maîtrise les intonations d'interrogations et d'exclamation, notamment. Il a été aussi observé qu'avant la fin de la première année, le bébé maîtrise les montées et descentes intonatives, imitant les questions et assertions des adultes (Karmiloff et Karmiloff-Smith, 2001 ; Snow et Balog, 2002).

Le développement de la syntaxe chez les jeunes enfants a été étudié en détail par Tomasello et ses collègues, notamment (voir aussi Brown, 1971; Sheldon, 1974; Tavakolian, 1981; MacWhinney et Pléh, 1988; McKee et al., 1998; Kidd et Bavin, 2002). Diessel et Tomasello (2000) ont montré en particulier que, chez les enfants anglophones de 2 ans et moins, les clauses syntaxiques les plus précoces et les plus fréquentes utilisées apparaissent dans des constructions présentatives, avec deux clauses, une principale, une relative : comme "Here's a tiger that's gonna scare him" or "That's the sugar that goes in there". Selon ces auteurs, la raison pour laquelle les clauses relatives émergent dans ces constructions présentatives est le caractère préfabriqué de la clause principale. Et, ce qui nous intéresse au plus haut point, la clause principale contient un déictique (« this », « that », « here », « there », « it »). Ces constructions présentatives ont également été observées en français (Vion et Colas, 1987 ; Jisa et Kern, 1998).

Ainsi, le pointage est impliqué lors de nombreuses étapes du développement de la communication humaine. Il semble être le premier outil de communication utilisé par les bébés. Il joue un rôle clef (via le regard) dans le mécanisme d'attention partagée dans l'interaction enfant-adulte. Il semble émerger spontanément, par étapes, et en association avec les productions orales.

Neuroanatomie des mécanismes de pointage

Le rôle crucial du pointage avec l'index dans le développement du langage et l'implication de formes multiples du pointage à différents stades du développement de la communication, d'abord avec l'œil, puis le doigt, puis avec la voix (intonation et syntaxe), suggère que toutes les modalités du pointage pourraient partager un substrat cérébral commun.

Examinons d'abord les corrélats cérébraux des pointages digital et oculaire. Des études de neuroimagerie ont obtenu des activations dans le lobule pariétal inférieur gauche ou droit lors de tâches de pointage digital (cf. Lacquaniti et al., 1997; Kertzman et al., 1997). D'autre part, il a été montré que des patients atteints de lésions pariétales droites et présentant une hémiparésie gauche présentent également des déficits dans des tâches de pointage digital (Edwards et Humphreys, 1999).

Les pointages oculaire et digital ont été examinés en IRM fonctionnelle par Astafiev et al. (2003). Les résultats de l'étude montrent que la préparation d'une saccade oculaire (que l'on peut assimiler au pointage oculaire) active le cortex occipital bilatéral, le cortex frontal bilatéral à la jonction des sillons précentral et frontal supérieur, c'est-à-dire dans le champ oculaire frontal (« frontal eye field » ou FEF, impliqué dans la préparation oculomotrice) et le cortex pariétal postérieur bilatéral, le long du segment horizontal du sillon intrapariétal (IPS). Lors de la préparation d'un pointage digital, on retrouve une activation du FEF bilatéral et de l'IPS bilatéral. On observe en plus, à gauche, des activations du gyrus angulaire, du gyrus supramarginal, du lobule pariétal supérieur, du gyrus précentral dorsal et du sillon temporal supérieur. Il semble donc, d'après ces études, que le cortex pariétal postérieur et le cortex frontal supérieur soient des régions communes qui encoderaient des signaux préparatoires pour les pointages oculaire et digital.

Que sait-on du pointage vocal, sur les plans physiologique et neurofonctionnel ? Sur les plans acoustique et articulatoire, on sait que le pointage vocal et en particulier le pointage prosodique requiert un contrôle très fin de la part du locuteur. Sur le plan acoustique, le pointage prosodique est sophistiqué et demande un contrôle laryngé très précis. En français, il a été montré qu'il correspond à une augmentation de la fréquence fondamentale et de la durée du constituant focalisé (pointé), une diminution de la fréquence fondamentale du constituant pré-focal et une désaccentuation du constituant post-focal (Astésano, 2001 ; Di Cristo, 1998 ; Dohen et Lœvenbruck, 2004 ; Jun et Fougeron, 2000 ; Rossi, 1999), comme dans l'exemple ci-dessous pour l'énoncé « MADELEINE m'amena » (sous-entendu « pas Valérie ») (Figure 2).

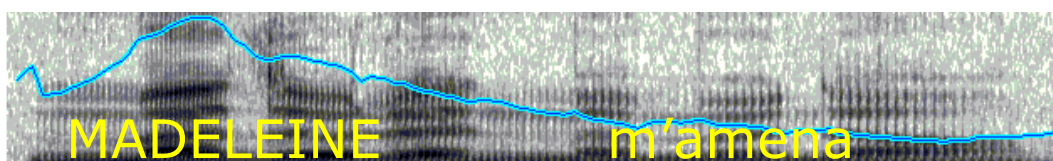


Figure 2. Superposition d'un spectrogramme et du tracé de la fréquence fondamentale (F0) pour l'énoncé « MADELEINE m'amena », avec focalisation prosodique contrastive sur « Madeleine ».

Sur le plan articulatoire, il a été montré par des études des mouvements de la langue (par électromagnéto-métrie) et des lèvres (par suivi vidéo) que le pointage prosodique requiert un

contrôle moteur très précis : il correspond à une articulation plus ample et une durée plus longue du constituant focalisé et est éventuellement accompagné d'une hypo-articulation du constituant post-focal (Løevenbruck, 1999 ; Dohen et al., 2006).

Sur le plan cérébral, nous avons montré que le pointage prosodique en parole intérieure (prononcer mentalement « MADELEINE m'amena ») active un réseau temporo-pariéto-frontal gauche : aire de Wernicke, gyrus supramarginal gauche, région de Broca et insula gauche (Løevenbruck et al., 2005, 2009), alors que le pointage syntaxique (prononcer mentalement « C'est Madeleine qui m'amena ») n'active que la région de Broca. Remarquons d'abord, que lorsque la prosodie joue un rôle *linguistique* (elle peut être la prosodie affective, qui exprime joie ou tristesse), elle active bien les régions classiques du langage, dans l'hémisphère *gauche*, tout comme la syntaxe. Cela conforte ainsi, au passage, son rôle crucial de structure phonologique à part entière, et non pas simplement subordonné à la syntaxe (voir e.g. Beckman, 1996 vs. Dronkers et al., 2000). Ensuite, il est intéressant de noter qu'en plus de recruter le gyrus frontal inférieur gauche, la focalisation prosodique active des régions temporales supérieures et pariétales inférieures. Notre interprétation est que le pointage prosodique nécessiterait que des représentations intégrées (auditives et articulatoires) soient formées via les régions temporales supérieures et pariétales, pour organiser l'articulation et la phonation de façon appropriée. L'absence d'activation temporelle et pariétale pour le pointage syntaxique est cohérente avec le fait que le pointage syntaxique ne correspond pas à des attentes somato-sensorielles ou auditives particulières par rapport à un énoncé « neutre ». En effet, le pointage syntaxique est grammaticalisé, automatisé, il ne nécessite pas un contrôle en ligne, adaptatif, des articulateurs et du larynx. Pour résumer, il semble ainsi que le lobe pariétal gauche, déjà impliqué dans les pointages digital et oculaire, soit également impliqué dans le pointage prosodique (en voix intérieure), mais pas dans le pointage syntaxique.

Si les pointages digital, oculaire et prosodique impliquent tous le lobe pariétal gauche, existe-t-il une somatotopie du pointage dans le cerveau, notamment dans le lobe pariétal ? Pour répondre à cette question, nous avons examiné les activations cérébrales pour plusieurs modalités de pointage (oculaire, digital, prosodique et syntaxique) chez les mêmes sujets (Løevenbruck et al., 2009). Les résultats indiquent que pour le pointage digital, le lobule pariétal supérieur est activé bilatéralement avec une prédominance gauche, ainsi que le lobule pariétal inférieur bilatéralement avec prédominance gauche et le cortex frontal gauche. Pour le pointage oculaire, on observe des activations dans les aires occipitales bilatérales (aires

visuelles), ainsi qu'une activation pariétale gauche et frontale bilatérale. Pour le pointage prosodique, on observe une activation du lobule pariétal supérieur bilatéral, avec prédominance gauche, du gyrus postcentral gauche et des aires périsylviennes bilatérales. Pour le pointage syntaxique, on retrouve l'absence d'activation pariétale gauche, mais une activation des régions périsylviennes droites et du cortex prémoteur bilatéral. En résumé, le lobule pariétal supérieur gauche est activé dans trois modalités de pointage (oculaire, digital et prosodique), mais pas dans le pointage syntaxique. Ainsi, ces trois modalités pourraient bien toutes recruter le lobule pariétal supérieur gauche, le pointage oculaire étant situé dans une zone plus antérieure que le pointage prosodique, lui-même plus antérieur que le pointage digital. Comme nous l'avons expliqué plus haut, les études comportementales (acoustiques et articulatoires) montrent que le pointage prosodique requiert un contrôle très précis du larynx et des articulateurs de la parole, ce qui suggère que les locuteurs utilisent des représentations multisensorielles (auditives et somato-sensorielles) pour produire le pointage prosodique de façon adéquate, tout comme ils le font pour produire un geste digital ou oculaire. Ces représentations seraient formées via l'activation d'aires associatives situées dans le lobule pariétal supérieur. L'absence d'activation pariétale lors du pointage syntaxique pourrait être due à l'absence de construction multisensorielle « en ligne » dans ce type de pointage vocal, par rapport au pointage vocal prosodique. Le pointage syntaxique utilise en effet une construction grammaticalisée, « gelée », et ne nécessiterait pas que des représentations intégrées (auditivo-articulatoires) soient formées.

L'ensemble de ces travaux corrobore notre conjecture que le pointage linguistique en ligne (focalisation prosodique) est inscrit dans le même réseau cérébral que les pointages gestuels (digital et oculaire). Pointages oculaire, digital et vocal seraient ainsi bien liés, autant sur le plan développemental que sur le plan cérébral. Ceci conforte la notion que le pointage, cet outil qui nous fait passer des objets vocaux aux objets du monde, est bien une des racines de la communication parlée.

4. Conclusion : quelques questions de recherche sur la « parole cognitive »

Nous avons tenté dans ce chapitre d'inscrire la parole et le langage dans leur environnement cognitif. Cet éclairage ne permet évidemment pas d'aborder toutes les grandes questions du domaine. Nous allons tenter dans cette dernière partie d'en mentionner quelques-unes, en guise de prospective.

D'abord, il s'agit de comprendre comment fonctionnent les mécanismes de production et de perception des unités de la communication parlée, quelles en sont les représentations. Une question centrale des recherches actuelles, on l'a compris, est celle des couplages perceptivo-moteurs, à la fois pour la production (ou comment la caractérisation perceptive des unités du langage, et les feedbacks sensoriels en ligne, permettent de spécifier les mécanismes de contrôle, les ingrédients des programmes moteurs) et pour la perception (ou comment s'articulent traitements perceptifs et connaissances motrices). Une autre question importante est celle de la multisensorialité, tant vision et audition pour le versant sensoriel, face et main pour le versant moteur, se combinent dans tous les aspects de la communication parlée. En regard de ces questions portant sur les processus et les représentations, la détermination des circuits corticaux par les techniques récentes de neuroanatomie et de neurophysiologie est évidemment d'un grand intérêt.

Ces questions doivent alors être mises dans une perspective développementale. Si les étapes du développement de la communication parlée chez l'enfant commencent à être mieux décrites (vocalisations puis babillage, premiers mots, premiers énoncés, premières phrases) et si quelques amorces (comme le pointage) permettant la découverte des mots chez l'enfant sont maintenant connues, de nombreuses questions restent encore ouvertes. Sur le versant de la production, le rôle de la prosodie et ses liens avec l'émergence du lexique et de la morpho-syntaxe ne sont pas encore bien décrits : il est suggéré que les enfants ne savent pas produire les contours intonatifs de leur langue avant de produire leurs premiers mots, et qu'ils commenceraient à maîtriser un certain nombre de clichés intonatifs lors de leurs premiers énoncés à deux mots. On sait aussi que l'enfant a besoin de connaître quelques mots avant de découvrir la syntaxe. Mais quels sont précisément les liens entre l'émergence du lexique, de la syntaxe et de la prosodie chez l'enfant ? Sur le versant de la perception, là encore les liens entre le niveau « segmental » (le traitement des unités phonémiques) et le niveau « suprasegmental » (le traitement des séquences et de leur organisation prosodique) ne sont

pas complètement expliqués. Sur le versant des interactions, enfin, contrairement à la pauvreté du stimulus revendiquée par Chomsky (1981), il semble bien que les enfants reçoivent une quantité abondante d'informations langagières qui les aident à développer le langage. Mais il est difficile d'évaluer précisément l'« input » que reçoit l'enfant. Il s'agit de déterminer quelles formes lexicales, syntaxiques, prosodiques sont utilisées, et dans quelles quantités et proportions, dans la parole adressée à l'enfant ; et quels types d'interactions adulte-enfant favorisent le développement du langage.

Sur un autre niveau, celui de la phylogénèse, nous avons présenté largement ici les débats sur les chemins vers le langage, et sur les précurseurs à découvrir au sein des mécanismes de la communication animale. Le « puzzle de l'évolution du langage » recèle de toute évidence de nombreuses pièces encore à découvrir. Parmi celles-ci, mentionnons la question des liens avec la théorie de l'esprit ; le rôle des processus d'imitation ; la possible dichotomie entre « faculté de langage au sens large », partagée avec les primates non humains, et « faculté de langage au sens étroit », incluant des mécanismes de récursivité supposés spécifiquement humains (Hauser et al., 2002) ; la compréhension des mécanismes référentiels qui permettent d'espérer « naturaliser » la relation entre le monde et la langue.

Mentionnons pour finir les enjeux applicatifs et sociétaux de ces recherches, et d'abord ceux qui concernent le développement des technologies cognitives de la parole. Mettre la « parole en machines » suppose de définir des systèmes capables à la fois de produire et de reconnaître la parole. Dans le sens de la production, les systèmes de synthèse sont de plus en plus élaborés pour introduire des dimensions expressives et émotionnelles, se compléter d'apparences visibles, inclure des capacités de production de gestes non-verbaux : on se dirige ainsi vers la réalisation d'agents virtuels multimodaux interactifs, ces « agents conversationnels » qui sont au cœur de nombreux workshops en France et à l'étranger (Bailly et al., 2008). Dans le sens de la reconnaissance de la parole, les développements sont nombreux et les résultats parfois spectaculaires, même s'ils s'appuient plus sur les progrès des technologies de calcul et la dimension de corpus d'apprentissage, que sur des percées réelles des théories de traitement statistique de l'information. Enfin, entre synthèse et reconnaissance, les enjeux du dialogue et de l'interaction face à face prennent une importance grandissante (voir Dohen et al., 2010). Dans ces travaux, la dimension cognitive est importante, bien sûr, mais pas toujours déterminante, tant la métaphore de « l'avion qui ne bat pas des ailes » reste signifiante dans ce domaine. Les technologies du traitement de l'information se passent

souvent (hélas ?) fort bien des avancées sur les fonctionnements cognitifs. Cependant, les modèles cognitifs, à la fois théories quantitatives et noyaux d'artefacts et de systèmes de traitement originaux, restent un fort enjeu dans le domaine des technologies de la parole.

D'autres enjeux essentiels résident dans les thématiques de la santé et du handicap. En ce qui concerne les pathologies de l'audition, mentionnons les percées sur les implants cochléaires et plus récemment sur les implants du tronc cérébral ; les recherches sur les technologies pour la langue des signes et le langage parlé complété. En ce qui concerne les troubles de la production de la parole, les recherches théoriques sur la modélisation articulatoire fournissent des assistances pour des systèmes d'apprentissage et de rééducation, et des prédictions pour des actes chirurgicaux de divers types. A un niveau plus central, les recherches en neuro-psycho-linguistique permettent de mieux décrire les réseaux cérébraux qui sous-tendent la perception et la production de la parole, du langage et du geste. Il devient ainsi envisageable d'étudier des pistes de rééducation dans certaines pathologies neurologiques, telles que les lésions cérébrales. En connaissant mieux les régions et les réseaux impliqués dans certaines tâches linguistiques, on peut espérer mieux diagnostiquer ou prévoir les déficits associés à la lésion d'une ou plusieurs régions et travailler de façon plus ciblée et précoce à la rééducation neuropsychologique.

Enfin, les recherches sur le développement des mécanismes de production et de perception de la parole, sur leur chronologie propre, sur leurs possibles difficultés, retards ou accidents de parcours, ont des implications évidentes dans le domaine de l'apprentissage et de la rééducation, notamment dans le secteur de l'orthophonie, des retards de langage, ou en relation avec la lecture et ses pathologies propres, notamment la dyslexie dont on connaît l'importance sociétale croissante.

Ainsi, la parole transdisciplinaire, au carrefour de disciplines multiples et à l'articulation entre langage et cognition, peut aborder questions fondamentales et enjeux applicatifs sur un spectre dont on espère avoir convaincu le lecteur de l'ampleur et de l'intérêt.

Remerciements

Ce chapitre est le fruit de nos travaux, réflexions et discussions avec les personnes suivantes, listées par ordre alphabétique. Nous les remercions toutes sincèrement : M. Baci, M.

Beckman, L.-J. Boë, F. Carota, M.-A. Cathiard, M. Dohen, L. Lamalle, P. Perrier, C. Pichat, M. Sato, C. Savariaux, C. Segebarth, J. Vauclair, A. Vilain, C. Vilain, P. Welby.

Références

- Abry, C. (2003) [b]-[d]-[g] as a universal triangle as acoustically optimal as [i]-[a]-[u].
Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3 – 9 August 2003, 727–730.
- Abry, C. et Ducey, V. (2008). Is pointing the root of the foot? Grounding the “prosodic word” as a pointing word. *The Evolution of Language, Proceedings of the 7th International Conference (EVOLANG7)*, Barcelona, Spain, 12 – 15 March 2008, A. D. M. Smith, K. Smith et R. Ferrer i Cancho (eds.), 3-9.
- Arbib, M. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *The Behavioral and Brain Sciences*, 28, 105-167.
- Arnold, K., et Zuberbühler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature*, 441, 303.
- Aslin, R. N. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*. The Netherlands: Kluwer, 305-315.
- Astafiev, S. V., Shulman, G. L., Stanley, C. M., Snyder, A. Z., Van Essen, D. C. et Corbetta, M. (2003). Functional Organization of Human Intraparietal and Frontal Cortex for Attending, Looking, and Pointing. *Journal of Neuroscience*, 23, 4689-4699.
- Astésano, C. (2001). *Rythme et accentuation en français: invariance et variabilité stylistique*, PhD Thesis, L’Harmattan édition et diffusion.
- Bailly, G., Elisei, F., et Raidt, S. (2008). Boucles de perception-action et interaction face-à-face. In L.-J. Boë, H. Lœvenbruck et A. Vilain (Eds.) *Communiquer par la parole : des processus complexes*. *Revue Française de Linguistique Appliquée*, 2008, XIII (2), 121-131.

- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., et Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.
- Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11(1/2), 17–67.
- Benoît, C., Mohamadi, T., et Kandel, S. (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. *Journal of Speech and Hearing Research*, 37, 1195-1203.
- Boë, L.-J., et Abry, C. (1986). Nomogrammes et systèmes vocaliques. *Actes des 15èmes journées d'étude sur la parole*, Société Française d'Acoustique, 303-306.
- Boë, L.-J., et Liénard, J.S. (1988). La communication parlée est-elle une science ? En doutiez-vous ? Éléments de discussion et de réflexion suivis de repères chronologiques. *Actes des 17èmes journées d'étude sur la parole*, Société Française d'Acoustique, 79-92.
- Boë, L.-J., et Liénard, J.S. (1988-1989). Les sciences de la parole : constitution et développement. Éléments de discussion et de réflexion suivis de repères chronologiques. *Bulletin de l'Institut de phonétique de Grenoble*, vol. 17-18, 1-52.
- Boë, L.-J., Vallée, N., Badin, P., Schwartz, J.-L., et Abry, C. (2000). Tendencies in phonological structures: the influence of substance on form. *Bulletin de la Communication Parlée*, 5, 35–55.
- Boë L.-J., Lœvenbruck H. et Vilain A. (dir.) (2008). Communiquer par la parole : des processus complexes. Numéro spécial de la *Revue Française de Linguistique Appliquée*, Vol XIII, 2.
- Brent, M.R., et Siskind, J.M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33-B44.
- Brown, H. D. (1971). Children's comprehension of relativized English sentences. *Child Development*, 42, 1923–36.

- Butcher, C. et Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeill (Ed.), *Language and gesture*. Cambridge: Cambridge University Press, 235-257.
- Cameron-Faulkner, T., Lieven, E. et Tomasello, M. (2003). A Construction Based Analysis of Child Directed Speech. *Cognitive Science*, 27, 843-873.
- Capirci, O., Iverson, J. M., Pizzuto, E., et Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child language*, 23, 645- 673.
- Cheney, D. L., et Seyfarth, R. M. (1990). *How Monkeys See the World: Inside the mind of Another Species*. Chicago: Chicago University Press.
- Chomsky, N. (1981). *Réflexions sur le langage*. Flammarion.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., et Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51, 523-547.
- Corballis, M. C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *The Behavioral and Brain Sciences*, 26, 199-260.
- Davis, B., Kern, S., Vilain, A., et Lalevée, C. (2008). Des babils à Babel : les premiers pas de la parole. In L.-J. Boë, H. Lœvenbruck et A. Vilain (Eds.) *Communiquer par la parole : des processus complexes*. *Revue Française de Linguistique Appliquée*, 2008, XIII (2), 81-91.
- Davis, C., et Kim, J. (1998). Repeating and remembering foreign language words: does seeing help?. *Proceedings of the AudioVisual Speech Processing conference AVSP'1998*, 121-126.
- Di Cristo, A. (1998). Intonation in French. In D. Hirst et A. Di Cristo (Eds.) *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 195-218.

- Di Cristo A. (1999). Vers une modélisation de l'accentuation du français (première partie). *French Language Studies*, 9, 149-179.
- Diessel, H. et Tomasello M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11, 131-151.
- Dohen, M., Bailly, G., Schwartz, J.-L. (Eds.) (2010). Speech and Face to Face Communication. *Special issue of the Speech Communication journal*, http://www.elsevier.com/authoried_subject_sections/P05/CFP/ctp_speechfacecomm.pdf
- Dohen, M. et Løevenbruck, H. (2004). Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French. *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP 04)*, Jeju island (Korea), 4-8 October 2004, Vol. 1, 785-788.
- Dohen, M., Løevenbruck, H., Hill, H. (2006). Visual correlates of prosodic contrastive focus in French: description and inter-speaker variability. *Proceedings of Speech Prosody 2006*, Dresden, Germany, May 2-5 2006, Vol. I, 221-224.
- Dronkers, N. F., Pinker, S., Damasio, A. (2000). Language and the aphasias. In E. R. Kandel, J. H. Schwartz, et T. M. Jessell (Eds.), *Principles of neural science*, New York: McGraw-Hill, 1169–1187.
- Ducey, V. et Abry, C. (2004). Le cadre de la parole et le cadre du signe : un rendez-vous développemental. *Actes des XXVèmes Journées d'Etude sur la Parole*, Fès, Maroc, 19-22 avril 2004.
- Edwards, M. G. et Humphreys, G. W. (1999). Pointing and grasping in unilateral visual neglect: effect of on-line visual feedback in grasping. *Neuropsychologia*, 37 (8), 959-973.
- Erber, N.P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, 40, 481-492.

- Fuchs, S., et Perrier, P. (2008). Understanding speech production: The PILIOS approach. In L.-J. Boë, H. Lœvenbruck et A. Vilain (Eds.) *Communiquer par la parole : des processus complexes. Revue Française de Linguistique Appliquée*, 2008, XIII (2), 35-44.
- Girin, L., Feng, G., Schwartz, J.-L. (2001). Audiovisual enhancement of speech in noise. *J. Acoust. Soc. Am.*, 109, 3007-3020.
- Goldin-Meadow, S. et Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.) *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Earlbaum Associates, 85-107.
- Hauser, M., Chomsky, N., Fitch, W. T. (2002). The language faculty: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
- Hopkins, W. D., et Leavens, D. A. (1998). Hand use and gestural communication in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 112, 95-99.
- Hornby, P. A. et Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 395-399.
- Jisa, H. et Kern, S. (1998). Relative clauses in French children's narrative texts. *Journal of Child Language*, 25, 623-652.
- Jun, S.-A. et Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology*. Boston: Kluwer, 209-242.
- Karmiloff, K. et Karmiloff-Smith, A. (2001) *Pathways to language: From foetus to adolescent*. Developing Child Series, Harvard University Press.
- Kendon, A. (1996). An Agenda for Gesture Studies. *Semiotic Review of Books*, 7(3), 8-12.
- Kertzman, C., Schwarz, U., Zeffiro, T. A., et Hallett, M. (1997). The role of posterior parietal cortex in visually guided reaching movements in humans. *Exp. Brain Res.*, 114 (1), 170-183.

- Kidd, E. et Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, 31, 599–617.
- Konopczynski, G. (1986). *Du prélangage au langage : acquisition de la structuration prosodique*. Thèse de Doctorat, Université de Strasbourg.
- Lacquaniti, F., Perani, D., Guigon, E., Bettinardi, V., Carrozzo, M., Grassi, F., Rossetti, Y. et Fazio, F. (1997). Visuomotor transformations for reaching to memorized targets: a PET study. *Neuroimage*, 5 (2), 129-146.
- Leavens, D.A. (2004). Manual deixis in apes and humans. *Interaction Studies*, 5, 387-408.
- Leavens, D.A., Hopkins, W.D., et Bard, K.A. (2005). Understanding the Point of Chimpanzee Pointing. Epigenesis and Ecological Validity. *Current Directions in Psychological Science*, 14, 185-189.
- Liljencrants, J., et Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J.J. Ohala and J.J. Jaeger (eds.) *Experimental Phonology*. New–York: Academic Press, 13–44.
- Lindblom, B. (1990). On the notion of possible speech sound. *Journal of Phonetics*, 18, 135–152.
- Lœvenbruck, H. (1999). An investigation of articulatory correlates of the Accentual Phrase in French. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1, 667-670, San Francisco, USA.
- Lœvenbruck, H., Baciú, M., Segebarth, C., et 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, 237-258.

- Løevenbruck, H., Dohen, M., et Vilain, C. (2009). Pointing is ‘special’. In S. Fuchs, H. Løevenbruck, D. Pape et P. Perrier (Eds.) *Some Aspects of Speech and the Brain*. Peter Lang, ISBN 978-3-631-57630-4, 211-258.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-511.
- MacNeilage, P.F. et Davis, B.L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology*, 2001, 11, 696-700.
- MacNeilage, P F. et Davis, B. L., (2005). The Frame/Content Theory of Evolution of Speech: A Comparison with a Gestural-Origins Alternative. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 6, 173-199.
- MacWhinney, B., et Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95–141.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge Univ. Press. 2nd ed. 1986.
- Maddieson, I., et Precoda, K. (1989). Updating UPSID. *UCLA Working Papers in Phonetics* 74, 104–111.
- McKee, C., McDaniel, D., et Snedeker, J. (1998). Relatives children say. *Journal of Psycholinguistic Research*, 27, 573–96.
- Meguerditchian, A., et Vauclair, J. (2006). Baboons communicate with their right hand. *Behavioural Brain Research*, 171, 170–174.
- Meguerditchian, A., et Vauclair, J. (2008). Vocal and gestural communication in non human primates and the question of the origin of language. In L.S. Röska-Hardy et E.M. Neumann-Held (Eds.), *Learning from animals? Examining the Nature of Human Uniqueness*. London: Psychology Press, 61-85.

- Ménard, L., Lœvenbruck, H., et Savariaux, C. (2006). Articulatory and acoustic correlates of contrastive focus in French: a developmental study. In J. Harrington et M. Tabain (Eds.) *Speech Production: Models, Phonetic Processes and Techniques*. New York: Psychology Press, 227-251.
- Millotte, S. (2008). Le jeune enfant à la découverte des mots. In L.-J. Boë, H. Lœvenbruck & A. Vilain (Eds.) *Communiquer par la parole : des processus complexes*. *Revue Française de Linguistique Appliquée*, 2008, XIII (2), 93-102.
- Mills, A.E. (1987). The development of phonology in the blind child. In B. Dodd and R. Campbell (Eds.), *Hearing by eye: the psychology of lipreading*. London: Lawrence Erlbaum Associates, 145-161.
- Morford, M. et 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.
- Nazzi, T. (2008). Segmentation précoce de la parole continue en mots : évaluation interlinguistique de l'hypothèse d'initialisation rythmique. *L'Année Psychologique*, 108, 309-342.
- Pollick, A. S. et de Waal, F. B. M. (2007). Ape gestures and language evolution. *P. N. A. S.*, 104, 19, 8184-8189.
- Povinelli, D. J., Bering, J. M., et Giambrone, S. (2003). Chimpanzee 'pointing': Another error of the argument by analogy? In S. Kita (Ed.) *Pointing: Where language culture and cognition meet*. Lawrence Erlbaum Associates, 35-68.
- Povinelli, D. J. et Davis, D. R. (1994). Differences between chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*) in the resting state of the index finger: Implications for pointing. *Journal of Comparative Psychology*, 108, 134-139.

- Rizzolatti, G., et Arbib, M.A. (1998). Language within our grasp. *Trends in Neurociences*, 21, 188–194.
- Rochet-Capellan, A., Schwartz, J.-L., Laboissière, R., et Galván A. (2007). Two CV syllables for one pointing gesture as an optimal ratio for jaw-arm coordination in a deictic task: A preliminary study. *Proceedings of the European Cognitive Science Conference (EuroCogSci07)*, 608-613, Delphes, Grèce.
- Rossi, M. (1999). La focalisation. In *L'intonation, le système du français: description et modélisation*, Chap. II-6, 116-128. Ophrys.
- Rousset, I., (2004). *Structures syllabiques et lexicales des langues du monde. Données, typologies, tendances universelles et contraintes substantielles*. Thèse en Sciences du Langage, Université Stendhal, Grenoble.
- Saussure, F. de (1972 [1916]). *Cours de linguistique générale*, édition critique préparée par Tullio De Mauro, Paris : Payot.
- Schwartz, J.L., Boë, L.J., Vallée, N., et Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25, 255-286.
- Schwartz, J.-L., Abry, C., Boë, L.J., Ménard, L., et Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory, *Speech Communication*, 45, 425–434.
- Schwartz, J.-L., Berthommier, F., et Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, 93, B69–B78.
- Schwartz, J.-L., Boë, L.J., et Abry, C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M.J. Solé, P. Beddor et M. Ohala (Eds.) *Experimental Approaches to Phonology*. Oxford University Press, 104-124

- Schwartz, J.-L., Sato, M., et Fadiga, L. (2008). The common language of speech perception and action: a neurocognitive perspective. In L.-J. Boë, H. Lœvenbruck et A. Vilain (Eds.) *Communiquer par la parole : des processus complexes. Revue Française de Linguistique Appliquée*, 2008, XIII (2), 9-22.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272–81.
- Slocombe, K.E., et Zuberbühler, K. (2005). Functionally Referential Communication in a Chimpanzee. *Current Biology*, 15, 1779–1784.
- Snow, D. et Balog H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112 (12), 1025-1058.
- Sodoyer, D., Girin, L., Jutten, C., et Schwartz, J.-L. (2004). Further experiments on audio-visual speech source separation. *Speech Communication*, 44, 113-125.
- Tavakolian, S. (1981). The conjoined clause analysis of relative clauses. *Language acquisition and linguistic theory*, S. Tavakolian (ed.), 167–87. Cambridge, MA: MIT Press.
- Teston, B. (2004). L'œuvre d'Etienne-Jules Marey et sa contribution à l'émergence de la phonétique dans les sciences du langage. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, 23, 237-266.
- Thelen, E. (1981) Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology*, 17, 237-257.
- Vauclair, J. (2004). Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 5, 363-384.

- Vion, M. et Colas, A. (1987). Contrôle de la production d'informations nouvelles et anciennes par les enfants de 4 à 11 ans : les constructions présentatives. *Bulletin d'Audiophonologie*, 3, 671-686.
- Volterra, V., Caselli, M. C., Capirci, O., et Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello et D. Slobin (Eds.) *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum Associates, 3-40.
- Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, 49, 28–48.
- Welby, P. et Lævenbruck, H. (2006). Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Rivista di Linguistica/ Italian Journal of Linguistics* 18, 74–124. Current Issues in Tonal Alignment, Mariapaola D'Imperio (éd.).
- Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In Kita S. (Ed.) *Pointing: where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 171-216.